The Dissertation Committee for Jori Barash
certifies that this is the approved version of the following dissertation:

Essays on Selection Markets and Contract Design

Committee:

Robert Town, Supervisor

Victoria Marone

Michael Geruso

Jeffrey Clemens

Essays on Selection Markets and Contract Design

by

Jori Barash

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2025

Abstract

Essays on Selection Markets and Contract Design

Jori Barash, PhD
The University of Texas at Austin, 2025

SUPERVISOR: Robert Town

Contract flexibility can improve cost-effectiveness and fairness in labor markets by targeting agents with the largest social benefits. This dissertation extends insights and empirical methods from health insurance markets to three new domains. I demonstrate how without increasing public spending or harming decision-makers, policymakers can target new patients to high-capacity physicians, undertreating physicians to strong incentives, and high-ability students to selective universities.

Understanding the determinants of physicians' treatment decisions helps inform policies that incentivize appropriate amounts of treatment and mitigate health disparities. In Norway, large administrative datasets and national regulations allow me to measure how primary care physicians heterogeneously respond to shocks and investigate mechanisms. I use this population of physicians for two papers that study shocks to the number of patients and the reimbursement rate. In Chapter 1, I document that physicians spend less time treating existing patients ("crowd-out") after a quasi-random shock to the total number of patients. I combine heterogeneity analysis with a theoretical model of decision-making to compare explanations and policy responses. Increasing the number of physicians may more effectively reduce crowd-out than incentives for greater treatment per physician. Even with a fixed supply of physicians, targeting assignment of patients to low-crowd-out physicians

could eliminate most of the documented effect.

In Chapter 2, I expand on the decision-making model to investigate the social efficiency of contract choice. Typically, physicians all face the same uniform incentives. With a menu of payment contracts, each physician chooses a combination of a piece-rate and fixed fee. Since a regulator does not perfectly observe physicians' differences or the underlying treatment needs of patients, a menu can help introduce variation in incentives to prevent undertreatment without large increases in public spending or creating perverse incentives for physicians to leave the public system. My empirical framework derives the socially optimal menu and measures its benefits in settings with heterogeneous altruistic agents and unobserved differences in clients' needs. Generally, the impacts of menu design are theoretically ambiguous; in Norway, my estimates of physician heterogeneity imply that a menu would lead to large welfare gains.

In Chapter 3, I study how more flexible financial aid contracts can improve how students select across universities. Financial aid is often a simple function of income. I estimate the impact of allowing financial aid to vary with a richer set of student characteristics and a future outcome: on-time graduation. I leverage students' selection on unobserved ability into a historical loan forgiveness program to predict their response to counterfactual financial aid. More flexible contracts lower the price of selective universities for students who are more likely to graduate, increasing statewide graduation. In this context, those students benefit the most from attendance and tend to come from underrepresented demographics.

Table of Contents

List of Tables

List of Figures

Chapter 1: Do New Patients Displace Existing Patients' Treatment?

Abstract

This paper estimates the effect of a physician's number of registered patients ("enrollment") on short-run treatment intensity in the context of Norwegian primary care. I instrument for enrollment with quasi-random administrative patient assignments. The estimated effect of enrollment is negative but small for several measures of treatment intensity. For example, with one new patient registration, the average physician spends 3 fewer minutes per month across incumbent patients. Descriptive evidence suggests that crowd-out exacerbates under-treatment. Crowd-out is larger among physicians who reach their stated capacity or initially work part-time. Drawing on a model of physician decision-making, this heterogeneity implies that capacity constraints dominate income effects in explaining crowd-out. With capacity constraints, increasing the number of physicians may more effectively reduce crowd-out than incentives for greater treatment per physician. Fixing physician supply, an alternative patient assignment rule could reduce crowd-out from administrative assignment by 86 percent.

1.1 Introduction

Healthcare policy interventions often target the supply of physicians, underpinned by the assumption that physicians who take on too many patients may deliver an inadequate amount ("intensity") of treatment. For example, regulatory bodies impose patient limits on physicians, and governments incentivize entry into underserved and less profitable areas.[1] The assumption aligns with economic theory, which predicts that an increase in a physician's workload might lead to a decrease in patients' treatment intensity due to time constraints or stronger incentives for leisure. Suggestive evidence reinforces the theoretical prediction: health outcomes tend to be worse in areas with more patients per physician (Macinko et al., 2007; Chang et al., 2016; Falcettoni, 2018; Vallejo-Torres and Morris, 2018; Basu et al., 2019).[2]

Causal evidence is limited on whether increasing the number of patients per physician impacts treatment or damages health. Related correlations may be biased because unobserved factors likely influence the equilibrium number of patients per physician. For example, patients with worse health may live in areas that are less attractive to physicians. Within an area, patients may choose to receive treatment in greater numbers from relatively high-quality physicians.

In this paper, I estimate the effect of a primary care physician's number of registered patients ("enrollment") on short-run treatment intensity. I find that enrollment lowers several measures of treatment intensity. Although this crowd-out effect is economically small, enrollment also increases avoidable hospitalizations. I present a model of physician decision-making to illustrate competing explanations for crowd-out: capacity constraints and income effects. Building on the model, heterogeneity across physicians is most consistent with capacity constraints. Consequently, increasing the number of physicians may reduce crowd-out more effectively than incentives for existing physicians to increase treatment. I contextualize findings by bounding the impact of a counterfactual registration policy on treatment intensity and introducing suggestive evidence that crowd-out exacerbates undertreatment.

[1]Limits on the number of patients exist in at least Norway, Denmark, and China.

[2]Likewise, primary care physicians per capita is a strong correlate of the causal effects of commuting zone on life expectancy, estimated using patient moves (Finkelstein et al., 2021).

Norway's institutional features are ideal for isolating exogenous variation in enrollment from potential confounding factors. First, Norwegian municipalities often automatically reassign the patients of physicians exiting the market. Auto-reassignment to nearby physicians is quasi-random, so the resulting changes to enrollment are plausibly independent of patient health and physician quality. The mechanical increase in enrollment from auto-reassigned patients is compounded by the fact that other patients in the same region suddenly must choose among fewer nearby physicians. Both mechanisms result in a persistent increase in nearby physicians' enrollment. Second, Norway's registration system encourages long-term patient-physician relationships, which allows me to create a large balanced panel of patients. With a balanced panel, I can control for persistent heterogeneity and make relatively weak assumptions about patient composition and unobserved determinants of health. With an unbalanced panel, average treatment intensity may vary due to the idiosyncratic needs of others joining or leaving the physician, leading to biased estimates. Third, national agreements dictate an identical reimbursement schedule for physicians, ruling out confounding changes to treatment intensity from financial incentives. In other contexts, physicians with greater enrollment may be able to negotiate higher reimbursement rates, which in turn incentivize greater treatment intensity and attenuate estimates. Fourth, virtually all of Norway's primary care physicians and residents are part of the country's universal public healthcare system, so the design of the registration system can have far-reaching consequences for health and public expenditure. Moreover, the scale of administrative data provides the statistical power to estimate even small effects, particularly for subsamples to test model predictions, e.g., part-time physicians or chronically ill patients.

I estimate the effect of enrollment on short-run treatment intensity using two-stage least squares, instrumenting for enrollment with the number of auto-reassigned patients. Each observation is a physician's aggregate monthly treatment intensity among incumbent patients. I find suggestive evidence supporting the exclusion assumption that auto-reassignments' timing and size are independent of factors determining changes to existing patients' health that are specific to a physician. First, the composition of nearby physicians' patients does not predict the number of auto-reassigned patients, conditional on availability, municipality, and at least one auto-reassigned patient. Second, prior to auto-reassignment, trends in incumbents' treatment intensity are uncorrelated with the number of auto-reassigned patients. I show this using a complementary event study design, comparing physicians with large and small auto-reassignments. Third, consistent with monotonicity and relevance assumptions, auto-reassignment shifts the entire distribution of enrollment increases relative to other months. To minimize threats to identification, I include fixed effects for each physician and event-month.

I find that an increase in enrollment results in no more than a small decrease in several measures of primary care treatment intensity among incumbent patients in the short run. For example, one new patient registration results in 3 fewer minutes of encounters among all incumbent patients, equivalently 0.05 percent of the mean. Crowd-out is consistently small among particular types of treatment, alternative specifications, and subsets of patients and physicians. When including both incumbents and new patients, I do not find evidence of a net increase in physicians' aggregate treatment intensity, consistent with capacity constraints. Physicians meet the needs of new patients with a commensurate decrease in treatment intensity among incumbent patients. Finally, I find a precise increase in incumbents' avoidable hospitalizations, suggesting that crowd-out in treatment intensity harms patients.

Heterogeneity in the effect of enrollment on physician hours helps distinguish between the mechanisms of income effects and capacity constraints. First, crowd-out effects appear concentrated among physicians who reach their stated capacity. Second, I do not find strong evidence of heterogeneity between physicians with different financial incentives. Third, crowd-out effects appear concentrated among part-time physicians. Given other findings, this heterogeneity suggests that capacity constraints are idiosyncratic and physicians with lower capacity select into part-time work. Descriptive evidence is consistent with idiosyncratic constraints among a fraction of physicians: weekly hours bunch near a physician-specific maximum for a small number of physicians over ten years.[3]

Capacity constraints imply that increasing the supply of physicians would most reduce crowd-out. However, with large fixed costs of physician entry and small crowd-out effects, such a policy may not be cost-effective. Also, Norway's current policy of limiting maximum enrollment may not have meaningful impacts on treatment intensity, because crowd-out effects are largest among low-workload physicians. Using a fixed supply of physicians, I explore the implications of a counterfactual registration policy. I simulate a new auto-reassignment rule for patients of departing physicians designed to minimize crowd-out. This exercise incorporates estimated heterogeneity in predicted crowd for combinations of patient and physician characteristics. Relative to status quo random assignment, targeted assignment can eliminate 86 percent of the crowd-out from status-quo auto-reassignment.

Crowd-out only harms incumbent patients if the resulting treatment intensity is inefficiently low. In absence of data on the efficient level of treatment, I show that crowd-out increases avoidable hospitalizations. I also contextualize findings with three stylized facts. First, incumbent patients in the analysis sample have lower initial treatment intensity and higher avoidable hospitalizations than observably similar patients in the full population (both 5 percent of the mean). Crowd-out effects increase that gap. Second, prior survey evidence suggests that both patients and physicians consider treatment intensity to be too low. Third, national statistics show that treatment intensity is low in Norway relative to most OECD countries.

To the best of my knowledge, this paper is the first to estimate the effect of the number of registered patients per physician on treatment intensity. As a design parameter within public healthcare systems, physician enrollment is understudied relative to patient cost-sharing (see Kiil and Houlberg, 2013, for a review) and physician

[3]By contrast, the distribution of physician labor supply does not bunch near a shared limit, e.g., 60 hours per week.

payments (Clemens and Gottlieb, 2014; Brekke et al., 2017; Einav et al., 2018; Eliason et al., 2018; Cabral et al., 2021). This paper also complements studies of other healthcare disruption effects. First, I extend well-documented disruption effects in emergency care (Jena et al., 2017; Chan, 2018; Hsia and Shen, 2019; Hoe, 2022) to the less-studied context of primary care. Representing a third of physicians in Norway, primary care physicians deliver the bulk of diagnoses, basic procedures, and chronic disease management, with large impacts on patient health, specialist referrals, and total healthcare spending. Unlike emergency physicians, primary care physicians do not choose their patients, they are responsible for those patients for long periods of time, they do choose their working hours, and they have financial incentives on the intensive margin of care.

Second, I estimate the effect of a persistent shock to workload, rather than a temporary disruption (Shurtz et al., 2018; Harris et al., 2020; Freedman et al., 2021; Kovacs and Lagarde, 2022). Variation from persistent shocks may have stronger external validity for answering the central policy question of whether health systems have a sufficient supply of physicians. Other persistent shocks are relatively difficult to study due to simultaneous changes to incentives, e.g., insurance expansions (Garthwaite, 2012; Carey et al., 2020).

Third, I measure spillover effects of physician exit: I study the patients of nearby physicians while prior work focuses on the patients of departing physicians (Kwok, 2018; Fadlon and Van Parys, 2020; Bischof and Kaiser, 2021; Simonsen et al., 2021; Zhang, 2022; Sabety et al., 2021; Sabety, 2023).[4] Physician workload shocks disrupt care differently than when patients switch physicians. For example, new physicians may lack soft information from long-term relationships. I also show suggestive evidence of a limitation of this literature: patients' choice of a new physician

[4]This paper extends a related analysis in Sabety (2023), Table A19. First, I decompose imprecise clinic-level effects to show crowd-out among incumbents. Second, I use quasi-random variation in enrollment. After physician exit, patients' health and remaining physicians' quality might influence the choice to switch establishments.

may be endogenous even if the timing of physician exit is exogenous. The few patients who make active choices experience a different trajectory of treatment intensity than those who are quasi-randomly reassigned.

The remainder of the paper proceeds as follows. Section 1.2 describes the empirical setting including identifying variation and data. Section 1.3 presents a model of physician-making and uses its comparative statics to guide the empirical strategy. Section 1.4 describes the baseline estimates of crowd-out, heterogeneity for subsamples of physicians and patients, and robustness. Section 1.5 discusses policy implications with suggestive evidence of status quo undertreatment and a counterfactual assignment rule.

1.2 Background

1.2.1 Institutional Setting

Norway's public healthcare system guarantees all residents access to a primary care physician. Norwegians frequently visit their physician – on average 3 times per year – for chronic disease management, consultations about symptoms, diagnostic tests, and general medical procedures, but also sick leave certification and referrals to specialists and non-urgent hospital care (Norway, 2023). Patients must be referred by physicians to specialists and non-emergency hospital visits. Norwegians may elect to purchase private health insurance to lower wait times for private specialists, but primary care nearly exclusively occurs within the public system.

To facilitate access, Norwegian primary care is carefully regulated. First, the number of registered patients per physician must generally remain between 500 and 2500 (Lovdata, 2017). Second, national guidance states that patients should not wait more than five days for a consultation in most circumstances. Within these limits, physicians should prioritize the treatment of patients with the greatest health need. Third, Norway regulates the supply of primary care physicians through a fixed number of contracts in each municipality. Physicians may move across municipalities or retire, but the total number of contracts and the locations of practices are slow to change. Fourth, patients register with physicians through a centralized system. Patients can request a new physician twice each year, but switches are rare (Iversen and Lurås, 2011). Physicians must accept all new patients until reaching a maximum number of patients. Physicians and municipalities must agree on the upper bound and weekly hours of operation to provide sufficient capacity for all nearby residents. As a result, these contract details rarely change. Fifth, prices for healthcare services are generally fixed across physicians. Some prices change once per year when renegotiated between the regulator and the physicians' union (Legeforening, 2022). Physicians also receive a fixed fee for each registered patient, representing approximately 30 percent of revenue. In heterogeneity analyses, I exploit one exception to fixed prices: physicians with additional training receive supplementary reimbursement per patient visit. Sixth, when registered physicians are unavailable, patients can receive treatment at primary care emergency centers, which are comparable to stand-alone urgent care centers in the United States. Patients can also seek second opinions from other primary care physicians.[5]

The identifying variation in this paper comes from administrative assignment, which occurs when an exiting physician ends his contract and no new physician accepts ownership of the patient list. Contracts typically end because the physician is retiring or moving to a different municipality. Exiting physicians provide the municipality with six months to find a replacement. In 30 percent of exits, no new physician accepts ownership of the list and the municipality notifies patients on the list that they might be reassigned. Approximately 8 percent of patients actively choose a new physician and the rest are administratively assigned. Patient reassignment is random among nearby physicians with enrollment lower than their upper bounds.[6]

[5] On average, 80 percent of primary care visits are with the registered physician.

[6] Typically, only a few new patients are auto-reassigned to each nearby physician. Sometimes, administrative assignments coincide with an increase in physicians' upper bound on enrollment. This increase may reflect renegotiation so that a municipality can satisfy requirements for capacity. Table 1.4 evaluates robustness by excluding these cases.

Municipalities do not have access to patient characteristics when determining reassignment. Appendix Table A.2 verifies that patient composition is conditionally uncorrelated with auto-reassignments. Appendix Figure A.5 shows the distribution of auto-reassignments and Appendix Figure A.4 illustrates the corresponding distribution of enrollment increases. After being auto-reassigned, 83 percent of patients choose to stay at their assigned physician for at least six months. If auto-reassigned patients do not request a preferred physician, they may be auto-reassigned again to expand the patient list of entering physicians.

1.2.2 Data

The estimation sample is a balanced panel of physician-months in the six months before and after auto-reassignment. Outcomes reflect aggregated treatment intensity among incumbent patients who are registered to physicians both before and after auto-reassignment. The panel structure helps attribute short-term changes in treatment intensity to sudden changes in enrollment rather than the composition of patients' underlying treatment need. I construct the sample using restricted administrative records on registration, individual demographics, and healthcare reimbursement, which are maintained by the Norwegian Directorate of Health and Statistics Norway.[7] These records contain nearly all of Norway's residents and primary care physicians from 2008 to 2017.

The estimation sample excludes potentially confounding variation. Appendix Table A.1 shows the number of physicians impacted by each sample condition. Each physician must practice in a single location during the entire period and each patient must be registered for the entire period. I also exclude physicians with fewer than 500 patients or with overlapping changes in reimbursement rates due to additional training. Both the physician and patient must have identification numbers to attribute treatment intensity to a particular physician of interest, which excludes

[7] See Appendix A.1 for additional details.

recent migrants. I also exclude temporary replacement physicians and shared patient lists.

To assess pre-trends in treatment intensity before auto-reassignment, physicians must not receive auto-reassignments from exiting physicians in the six months before a focal auto-reassignment. Physicians may receive subsequent auto-reassignments during the following six months. Moreover, I restrict auto-reassignments to those coming from exiting physicians with a dissolved patient list of at least 20 auto-reassigned patients. I exclude physicians who immediately lower their enrollment upper bound. In these cases, incumbent patients might be auto-reassigned. This restriction reinforces that physicians in the sample experience a persistent increase in enrollment from auto-reassignment.

Table 1.1: Summary Statistics: Treatment Intensity per Physician

Variable	Mean	Std. Dev.	10th Percentile	Median	90th Percentile
Enrollment	1,262.00	354.23	811.00	1,237.50	1,710.00
Visits	372.52	210.86	51.00	368.00	641.00
Spending	10,858.77	27,152.92	1,056.89	10,111.69	17,832.71
Hours	103.25	42.84	45.29	107.94	153.00
Avoidable Hosp.	3.61	2.93	1.00	3.00	7.00
Follow-up Visits	110.29	74.31	21.00	102.00	204.00
Bill Lines	1,140.84	700.84	322.00	1,060.00	2,019.00
Diagnostics	72.59	88.72	8.00	51.50	150.00
Procedures	274.05	215.50	57.00	228.00	541.00
per Visit	2.94	0.84	2.04	2.81	4.01
Physician-Spells	2,722				

Notes: Summary statistics reflect monthly totals six months prior to auto-reassignment, across physicians. A physician may appear more than once if they experience multiple auto-reassignment spells. All treatment intensity measures are restricted to primary care at the assigned physician. Spending (reimbursement) is measured in USD. I classify Avoidable Hospitalizations based on diagnosis codes (see Appendix A.1.

The final analysis sample reflects 2,065 unique physicians and 2,335,982 patients, representing nearly one-third of the population. Appendix Figure A.4 shows that sample restrictions result in only negligible changes to the distribution of enrollment changes. The vast majority of incumbents experience a small number of auto-reassigned patients joining their physician. In 64 percent of physician-spells, only 1 patient is initially auto-reassigned. Table 1.1 shows the distribution of aggregate treatment intensity six months before new patients are auto-reassigned to the physician. Treatment intensity varies widely across physicians, and most physicians spend less than 40 hours per week with patients.

1.3 Research Design

1.3.1 Conceptual Framework

To guide intuition about the treatment intensity effects of increased enrollment, I extend the model of physician decision making in Ellis and McGuire (1986). Distinguishing between mechanisms is important because disruption in treatment intensity due to income effects can be mitigated through reimbursement incentives with a fixed supply of physicians, while avoiding disruption due to capacity constraints may require increasing the supply of physicians.[8]

Consider a physician who simultaneously chooses treatment intensity m_i for each registered patient $i \in 1, ..., N$, given that patient's initial illness severity. The physician has additive preferences over net income, leisure, and health production. Each patient draws a random severity from a distribution $\lambda_i \sim F$. Importantly, F is fixed for the incumbent patients of a physician when enrollment N changes. Health production $h(m_i, \lambda)$ is a function that measures the value of health given each patient's initial severity and treatment intensity. Net income $\sum_i \pi(m_i)$ and workload $\sum_i m_i$ aggregate across patients, so the choice of one patient's m_i depends on other patients' severity $\lambda_{i'}$. The physician's relative weight on health production is

[8] Perhaps a regulator can incentivize investments in increased capacity, but healthcare is labor-intensive with respect to high-skilled labor, technology investments are unlikely to change the fundamental tradeoffs, especially in the short and medium run.

altruism α. Functions u, π, h are strictly increasing and weakly concave, while distaste for workload l is strictly increasing and strictly convex. Finally, the physician has a capacity constraint on workload, $\sum_{i'} m_{i'} \leq \bar{M}$, with shadow cost μ. The problem takes the form,

$$\max_{m_i \geq 0, \forall i \in 1,...,N} u\left(\sum_i \pi(m_i)\right) - l\left(\sum_i m_i\right) + \alpha \sum_i h(m_i, \lambda)$$
$$\text{s.t.} \sum_i m_i \leq \bar{M} \quad [\mu]$$

with the following optimality conditions $\forall i' \in 1, ..., N$,

$$u'\left(\sum_i \pi(m_i)\right)\pi'(m_{i'}) - l'\left(\sum_i m_i\right) + \alpha h_m(m_{i'}, \lambda) + \mu = 0$$
$$\sum_i m_i \leq \bar{M}\,,$$

resulting in the optimal choice of treatment intensity $m_i^*(\lambda_1, ..., \lambda_N)$.

If enrollment N exogenously increases and capacity is slack ($\mu = 0$), then expected treatment intensity decreases through income effects. In expectation, new patients require some treatment, resulting in higher aggregate treatment intensity, which lowers leisure, weakly increasing the marginal utility of leisure and decreasing the marginal utility of net income. The physician responds by setting the loss in marginal health production equal to the loss in marginal utilities of net income and leisure. For each realization of severity λ, these changes unambiguously lower optimal treatment intensity for an incumbent patient. Enrollment only lowers treatment intensity in expectation because realized treatment intensity may increase due to idiosyncratic draws of λ.

Curvature of preferences results in larger decreases in treatment intensity, e.g., if the physician's initial workload is large. A physician with a relatively large initial workload has a high marginal distaste for incremental treatment intensity relative to leisure. When enrollment increases, any new treatment intensity from new patients has to be compensated by relatively large decreases in treatment intensity for incumbent patients. On the other hand, there is an ambiguous relationship between crowd-out and observed correlates of severity λ. If sicker patients benefit more from marginal treatment ($\frac{d^2h}{dmd\lambda} > 0$), then their crowd-out effects should be smaller.

The capacity constraint also drives incumbents' treatment intensity to decrease in enrollment. When the constraint binds, the total workload stays the same, so incremental treatment for new patients must correspond to lower treatment for incumbents. The distaste for the total workload is unaffected. A binding capacity constraint implies that no policy can fully counteract the short-run crowd-out effect of increased enrollment. By contrast, with slack capacity, income effects imply that well-targeted reimbursement incentives ($\frac{d\pi}{dm}$) can mitigate crowd-out effects.

The intuition behind these comparative statics also holds with several extensions to the model. First, F can vary arbitrarily across patients and physicians. Second, u, π, l, α, h, and $\bar{M}$ can vary arbitrarily across physicians as long as they are unaffected when N exogenously changes. Third, physicians may choose treatment intensity sequentially in expectation over other patients' severity.

1.3.2 Empirical Strategy

To estimate the effect of enrollment on treatment intensity, I use two-stage least squares:

$$Y_{jt} = \beta_1 \widehat{\text{Enroll}}_{jt} + \beta_j + \beta_t + \epsilon_{jt} \tag{1.1}$$

$$\text{Enroll}_{jt} = \gamma_1 \text{Auto}_{jt} + \gamma_j + \gamma_t + \varepsilon_{jt}\,. \tag{1.2}$$

Here, Y_{jt} is the outcome of interest, e.g., physician hours with incumbent patients, where subscript j indexes physicians and t indexes months relative to auto-reassignment ($t = -6, -5, -4, ..., 4, 5, 6$). Enroll_{jt} is physician j's total enrollment in month t, including incumbents and newly joined patients. Auto_{jt} reflects the cumulative number of patients auto-reassigned.[9] β_1 is the coefficient of interest, representing the causal effect of increasing enrollment by one patient. γ_1 represents the effect of one recently auto-reassigned patient on enrollment. β_j is a fixed effect for physician-spell j, controlling for unobserved heterogeneity in average treatment intensity across physicians.[10] β_t residualizes trends in treatment intensity relative to auto-reassignment, absorbing general disruption effects of auto-reassignment. ϵ_{jt} and ε_{jt} are random idiosyncratic errors. All regressions cluster standard errors by physician. The empirical strategy requires three identifying assumptions.

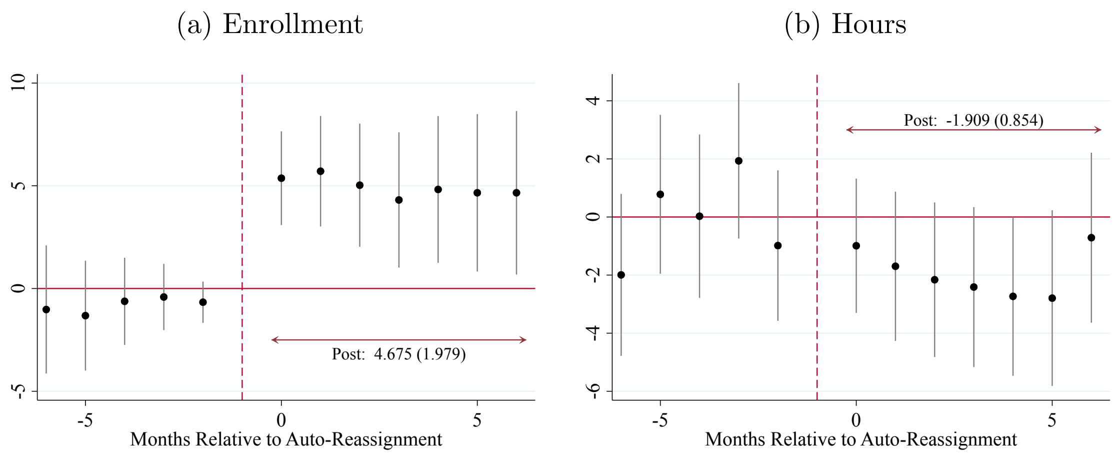
EXCLUSION. The exclusion assumption is that the number of auto-reassigned patients only affects incumbent patients' treatment intensity through the change in enrollment. In the econometric model, physician-spell fixed effects absorb time-invariant characteristics, and event-month fixed effects absorb physician-invariant trends from receiving auto-reassigned patients, so the timing and size of auto-reassignment must be independent of factors determining *changes* to incumbent illness severity that are specific to a physician, e.g., missed preventative care or a local viral outbreak.

Ginja et al. (2022) report that administrative assignment is conditionally random and show that assigned physicians' value-added does not correlate with reassigned patient characteristics. In particular, auto-reassignment should be random conditional on availability and municipality, which are absorbed by physician-spell fixed effects. Appendix Table A.2 verifies that the number of auto-reassigned patients does not correlate with destination physician characteristics conditional on availability, a shared municipality, and the exiting physician. Fixed effects absorb any level differences in treatment intensity across physicians.

[9] Auto_{jt} is similar to the interaction of post-treatment and treatment status in a difference-in-differences design, where $\text{Auto}_{jt} = 0\,\forall t < 0$. Most auto-reassigned patients register in event-month 0. Auto_{jt} is cumulative to reflect that additional patients are sometimes auto-reassigned after event-month 0.

[10] A physician-spell includes the six months before and after auto-reassignment for physician receiving auto-registrations.

Figure 1.1: Trends Among Incumbent Patients: Large versus Small Auto-Reassignments



Notes: This figure shows estimates and 95% confidence intervals of the difference between physicians with large and small auto-reassignments in each month relative to auto-reassignment. The omitted month is -1. I define large auto-reassignments as those with two or more reassigned patients (969 of 2,722 physician-spells). Each panel represents a separate regression with the dependent variable as indicated. The physician-month sample reflects aggregate treatment intensity among a fixed set of incumbent patients. Standard errors are clustered at the physician level. Estimates correspond to β_{1t} in the following regression: $Y_{jt} = \beta_{0t} + \beta_{1t}\text{AutoHigh}_j + \beta_j + \epsilon_{jt}$. "Post" indicates the average difference in outcomes after auto-reassignment between physicians with large and small auto-reassignments. Standard errors are in parentheses, and the coefficient is β_1 in the following regression: $Y_{jt} = \beta_{0t} + \beta_1\text{AutoHigh}_j \times 1[t \geq 0] + \beta_j + \epsilon_{jt}$. See Appendix Figures A.8 and A.9 show event-study estimates for other outcomes.

Likewise, physicians with relatively large or small auto-reassignments have similar trends in outcomes prior to auto-reassignment. Figure 1.1 uses an alternative event-study design to show that differences in the trend before auto-reassignment are not statistically distinguishable from zero for enrollment and hours. Section 1.4.3 discusses the second-stage effect of enrollment on hours implied by event-study estimates. Appendix Figures A.8 and A.9 show event-study estimates for other outcomes and Appendix Figure A.2 shows the trends in mean outcomes.

Relevance and Monotonicity. Auto-reassignments must strictly increase enrollment. The relevance assumption is directly testable through the first-stage F-statistic ($F > 117$ in the baseline specification). Consistent with a mechanical increase in enrollment, in Column (1) of Table 1.2 shows that the first stage coefficient is precise and indistinguishable from 1. The coefficient could have been larger than 1 because auto-reassignment coincides with physician exit. As shown in Appendix Figure A.3, for remaining nearby physicians with availability, enrollment increases due to both auto-reassigned patients and voluntarily switching patients. Voluntarily switching patients have to choose among fewer physicians than before the physician exit.

Appendix Figure A.6 shows that the change in enrollment is approximately monotonic in the number of auto-reassigned patients. Likewise, Appendix Figure A.4 shows that auto-reassignments shift the full distribution of enrollment changes rightward. In months with auto-reassignments, large enrollment increases become more likely and small increases (or decreases) become less likely. Figure 1.1 shows that physicians with large auto-reassignments experience a sudden precise increase in enrollment that is persistent. The point estimate implies that enrollment increases by five more patients for physicians with large auto-reassignments relative to physicians with a single auto-reassigned patient.

1.4 Enrollment and Treatment Intensity

1.4.1 Baseline Results

Table 1.2 shows that enrollment lowers physician treatment intensity, but the effects are small. One additional patient lowers hours spent in incumbent patient encounters by 0.048, equal to 0.05 percent of the mean. Importantly, this crowd-out effect is shared among all incumbents. The average incumbent loses just 0.14 seconds of treatment per month.[11] Bill lines, which reflect the count of distinct services, decrease by 0.342. Estimates for total reimbursement and patient visits are imprecise, but the confidence intervals rule out large decreases. Counterintuitively, the point estimate on spending is positive. Without a proportional decrease in spending, enrollment affects the qualitative composition of treatment, e.g., replacing relatively low-reimbursement treatment with high-reimbursement alternatives.[12] Columns (6)-(9) show that enrollment decreases intensity for several subsets of treatment and these effects are of similar size relative to the mean: physicians perform fewer services per visit, and crowd-out is similar among high-value services like procedures and diagnostics.

Appendix Table A.3 fails to find evidence that enrollment increases physician's total labor supply. In this table, outcomes reflect the treatment intensity among all registered patients rather than only incumbent patients.[13] Point estimates are generally small and imprecise, with confidence intervals ruling out large changes. No effect on total treatment intensity would be consistent with the theoretical framework: physicians meet the needs of new patients by reducing the treatment intensity of existing patients. Although enrollment does not precisely affect physicians' total labor

---

[11] I divide the effect size by the average of 1262 incumbent patients per physician.

[12] To see this, suppose there are two services with fixed reimbursement rates: m_L reimbursed at p_L and m_H reimbursed at $p_H > p_L$. If Spending $p_L m_L + p_H m_H$ is unchanged while Bill Lines $m_1 + m_2$ decreases, then either m_H or m_L (but not both) must increase. The physician has a greater incentive for m_H to increase.

[13] The sample includes incumbent patients, auto-reassigned patients, and other currently registered patients who do not have identification numbers or switch physicians during the 13-month spell.

supply, point estimates are generally positive. Likewise, the effects on incumbents' treatment intensity in Table 1.2 represent 20-38 percent of average treatment intensity among auto-reassigned patients.[14]

Consistent with crowd-out in utilization being harmful to patients, Column (10) of Table 1.2 shows that enrollment increases incumbents' avoidable hospitalizations. The estimate is precise but small: enrollment would need to increase by 82 patients for a single incumbent patient to experience an additional avoidable hospitalization in the following year.[15] New patients have relatively high avoidable hospitalizations, and accordingly, Appendix Table A.3 shows a larger point estimate from a regression using all registered patients. Prior work (e.g., Sabety, 2023) suggests that new patients may be harmed by switching physicians.

1.4.2 Heterogeneity and Mechanisms

Drawing on the comparative statics of the conceptual framework, physician heterogeneity helps distinguish income effects and capacity constraints as explanations for crowd-out. In this section, I focus on heterogeneity in the effect of enrollment on physicians' total hours spent treating incumbent patients. Appendix Tables A.5, A.6, A.7, and A.8 show heterogeneity for the first-stage and other measures of treatment intensity. Differences in point estimates across subsamples are at best suggestive because of the lack of precision.

First, columns (1) and (2) suggest that crowd-out is concentrated among physicians who have plausibly binding capacity. To proxy capacity, I divide the sample based on whether physicians reach 99 percent of their contracted upper bound on enrollment. For physicians near capacity, crowd-out is precise and similar in magnitude to the full sample. For physicians with slack capacity, the point estimate is imprecise. The subsample with binding capacity has a greater average workload, so

[14] This range excludes imprecise point estimates and Lines Per Visit.
[15] The calculation is $\frac{1}{1.017\times10^{-3}} \times \frac{1}{12} = 82$.

income effects could also explain the heterogeneity.

Second, columns (3) and (4) show similar noisy estimates of crowd-out for low- and high-fee physicians. Over half of primary care physicians are certified as general practitioners, and certification entitles these physicians to supplementary reimbursement for each patient visit. All else equal, with slack capacity constraints and strong income effects, low-fee physicians would have more crowd-out. By contrast, if capacity constraints bind for both sets of physicians, then the difference in incentive would have no impact. Without a large difference in estimates across subsamples, additional counterfactual reimbursement may be unlikely to eliminate crowd-out.

Third, columns (5) and (6) show that crowd-out is concentrated among part-time physicians, which is inconsistent with symmetric income effects. I classify physicians as full-time if they treat patients for at least six hours per weekday before auto-reassignment. Part-time physicians have a lower initial workload, so greater crowd-out is surprising regardless of whether income effects or capacity constraints dominate. All else equal, with income effects, part-time physicians should have a lower marginal taste for leisure and smaller crowd-out. Likewise, with a symmetric capacity constraint, e.g., 60 hours per week, capacity should be slack for part-time physicians, and crowd-out should be smaller or non-existent. To further distinguish between explanations for crowd-out, Appendix Figure A.7 plots the distribution of weekly hours. Although hours do not bunch near a global maximum (Panel A), hours sometimes bunch near a physician-specific maximum (Panel B). Taken together, this evidence suggests that, at least in the short run, physicians have idiosyncratic capacity constraints. Idiosyncratic capacity constraints also rationalize the precise crowd-out among physicians that reach their stated capacity, which can be as low as 500 patients.

Finally, columns (7)-(12) show estimates of crowd-out among six subsets of patients. In each column, outcomes reflect sums of treatment intensity among the indicated subset of incumbent patients. The sample still includes all physician-months in the sample. Point estimates are relatively large among patients who are young, not chronically ill, and female. To the extent that such characteristics proxy for low severity, this heterogeneity is consistent with marginal crowd-out being less harmful for low-severity patients.

1.4.3 Robustness

Table 1.4 shows that the effect of enrollment on hours is small, similar in magnitude, and generally precise across alternative specifications and samples.

First, I add physician time-varying controls for stated capacity and patient composition, e.g., the share age 65 or older and mean income. The point estimate is stable which further allays concerns about instrument validity. Second, I restrict the sample to physicians who receive the most auto-reassigned patients, producing a nearly identical point estimate. The estimate might change if auto-reassignment directly affected incumbents' treatment intensity, i.e., not through enrollment. For example, the potential salience of large auto-reassignments does not appear to produce a nonlinear relationship between auto-reassignments and treatment intensity. Third, I remove fixed effects for the month relative to auto-reassignment. Fourth, I add fixed effects for calendar months to control for seasonality. Fifth, I exclude physicians who ever have less than eight hours of treatment per month to address the possibility of intertemporal substitution. Physicians might respond to a temporary constraint in one month by increasing treatment intensity in another month. Sixth, for the same reason, I exclude the month of auto-reassignment and surrounding months. Seventh, I exclude physicians who increase their contracted upper bound on enrollment during the spell. One might be concerned that auto-reassignment is not quasi-random among these physicians. Eighth, I subset to physicians that have an avoidable hospitalization six months before auto-reassignment. Intuitively, incumbent crowd-out should be limited for patients with a salient health shock. Instead, the point estimate is larger but still small. Ninth, I use an alternative specification for the first stage which interacts total auto-reassignments with the number of months relative to auto-reassignment. Tenth, I weight observations by the number of patients because underlying treatment decisions are at the patient level and effects might correlate with baseline enrollment.

The event study design illustrated in Figures 1.1, A.8, and A.9 is also consistent with the main results: an increase in enrollment leads to a small decrease in treatment intensity. For example, dividing point estimates, 1 additional patient leads to $\frac{1.909}{4.675} = 0.408$ (standard error = 0.239) fewer hours spent with incumbent patients. This point estimate is equivalent to less than 2 seconds per month per patient. The 0.408 estimate is larger than, but not statistically distinguishable from, the baseline estimate of -0.048 (0.026) in Table 1.2.

This paper focuses on the baseline instrumental variable design rather than the event study because the former equally weights all auto-reassignment events to produce a policy-relevant treatment effect. The event study design limits identifying variation to an arbitrary margin like 1 versus 2+ auto-reassignments. A related caveat of the event study is that the comparison group is still treated. Physicians with a single auto-reassigned patient experience a relatively small shock, but enrollment still grows on average. Alternative comparison groups also have issues. Physicians without availability are not eligible to receive auto-reassignments, and full patient lists might suggest that these physicians differ systematically. Other physicians might be affected even without auto-reassignments because all patients have fewer choices once a physician exits.

Table 1.2: Effect of Enrollment Among Incumbent Patients

	Enrollment (1)	Hours (2)	Bill Lines (3)	Spending (4)	Visits (5)
Cuml. Auto-Joins	1.055				
	(0.055)				
	[<0.001]				
Enrollment		−0.048	−0.345	2.230	0.004
		(0.026)	(0.055)	(4.597)	(0.152)
		[0.070]	[<0.001]	[0.628]	[0.980]
Dep. Var. Mean	1274.917	101.349	1126.464	10299.286	370.160
New Pat. Mean		0.124	1.313	131.205	0.400
F-Statistic	117.129	20.101	30.679	9.458	28.624
Observations	35,386	35,386	35,386	35,386	35,386
	Lines Per Visit (6)	Procedures (7)	Diagnostics (8)	Follow-ups (9)	Avoidable Hosp. (10)
Enrollment	−0.002	−0.038	−0.058	−0.023	0.001
	(0.000)	(0.019)	(0.015)	(0.021)	(0.000)
	[<0.001]	[0.045]	[<0.001]	[0.268]	[0.024]
Dep. Var. Mean	2.940	70.908	270.428	135.084	3.619
New Pat. Mean	3.244	0.096	0.293	0.165	0.004
F-Statistic	36.423	46.320	43.653	26.550	6.869
Observations	34,578	35,386	35,386	35,386	35,386

Notes: This table displays estimates of coefficients from regressions of Equations 1.1 and 1.2. Each column represents a separate regression with the dependent variable as indicated in the table. The physician-month sample reflects aggregate treatment intensity among a fixed set of incumbent patients. Standard errors clustered at the physician level are reported in parentheses, and p-values are reported in brackets. "Cuml. Auto-Joins" indicates the number of patients auto-reassigned to a physician since the start of the spell. New Pat. Mean is the average of the dependent variable among auto-reassigned patients, averaged across physicians in the six months after auto-reassignment. Regressions also include fixed effects for spell and event-month.

Table 1.3: Heterogeneity in the Effect of Enrollment on Hours

	Capacity		Fee Level		Schedule	
	Slack (1)	Binds (2)	Low (3)	High (4)	Part-Time (5)	Full-Time (6)
Enrollment	−0.025	−0.051	−0.045	−0.039	−0.083	−0.005
	(0.023)	(0.027)	(0.027)	(0.030)	(0.004)	(0.010)
	[0.281]	[0.057]	[0.095]	[0.197]	[<0.001]	[0.609]
Dep. Var. Mean	94.515	107.514	84.576	113.236	78.104	113.770
1st Stage F-Stat.	33.577	134.240	83.519	46.769	100.088	66.967
Observations	16,783	18,603	14,677	20,709	12,324	23,062
	Age		Diagnoses		Gender	
	Under 65 (7)	Over 65 (8)	Healthy (9)	Chronic (10)	Male (11)	Female (12)
Enrollment	−0.042	−0.007	−0.037	−0.012	−0.013	−0.036
	(0.020)	(0.006)	(0.020)	(0.006)	(0.007)	(0.019)
	[0.035]	[0.248]	[0.061]	[0.045]	[0.068]	[0.053]
Dep. Var. Mean	72.288	29.061	58.032	43.316	42.905	58.444
1st Stage F-Stat.	112.037	112.037	112.037	112.037	112.037	112.037
Observations	35,386	35,386	35,386	35,386	35,386	35,386

Notes: This table displays estimates of coefficients from regressions of Equation 1.2. Each column represents a separate regression among a subsample with the dependent variable as indicated in the table. Columns (1-6) reflect aggregate treatment intensity among all incumbent patients for subsets of physicians. Columns (7-12) reflect aggregate treatment intensity among subsets of incumbent patients. For example, Column (1) includes physicians with enrollment consistently less than 99 percent of initial stated capacity, and Column (2) includes all other physicians. Standard errors clustered at the physician level are reported in parentheses, and p-values are reported in brackets. High-fee physicians receive supplementary reimbursement for each visit due to a certificate from additional training. Full-time physicians treat patients for at least six hours per weekday, on average, during the six months prior to auto-reassignment. Regressions also include fixed effects for spell and event-month.

Table 1.4: Robustness of the Effect of Enrollment on Hours

		Estimate			Mean	F-Stat.	Obs.
(1)	Add Controls	−0.044	(0.030)	[0.149]	101.349	88.295	35,386
(2)	Top 5%	−0.045	(0.028)	[0.109]	96.568	77.400	2,158
(3)	Drop Event-Month	−0.045	(0.024)	[0.056]	101.349	100.906	35,386
(4)	Calendar Month	−0.061	(0.023)	[0.007]	101.349	61.752	35,386
(5)	Hours Always 8+	−0.034	(0.020)	[0.095]	110.662	69.265	30,472
(6)	Drop Middle Months	−0.056	(0.030)	[0.064]	101.028	61.412	27,220
(7)	Constant Ceiling	−0.077	(0.057)	[0.178]	102.791	72.963	29,328
(8)	Avoidable Hosp.	−0.081	(0.005)	[<0.001]	103.043	88.793	32,097
(9)	Alt. 1st Stage	−0.050	(0.026)	[0.053]	101.349	3509.565	35,386
(10)	Weighted	−0.057	(0.022)	[0.011]	101.349	98.327	35,386

Notes: This table displays estimates of coefficients from regressions that vary the specification in Equation 1.2. Standard errors clustered at the physician level are reported in parentheses, and p-values are reported in brackets. The other columns are the mean of the dependent variable, the first-stage F-statistic, and the number of observations (physician-months). Variations to the specification are not cumulative. Specification (1) adds controls, including stated capacity, system entry and exit, and average patient characteristics: age 65+, chronic illness, female, income, and disability benefit receipt. (2) subsets to spells with auto-reassignments per incumbent above the 95th percentile. (3) drops fixed effects for months relative to auto-reassignment. (4) includes fixed effects for calendar months, e.g., January. (5) subsets to spells during which physician total hours are always above eight per month. (6) drops observations within 1 month of auto-reassignment. (7) subsets to spells during which the contracted upper bound on enrollments is non-increasing. (8) subsets to spells that begin with at least one avoidable hospitalization. (9) Alt. 1st Stage uses an alternative specification: enrollment is a function of event-months interacted with the number of total auto-reassignments per incumbent during the 13-month spell. (10) weights observations by the number of incumbents.

1.5 Policy Discussion

The previous sections show that registration of new patients lowers the treatment intensity of existing patients. The effect is small, and idiosyncratic capacity constraints offer the best explanation. Going forward, I consider the policy implications of these findings. Section 1.5.1 explores whether crowd-out should be avoided. Several stylized facts suggest that baseline treatment intensity is low in Norway, so crowd-out may exacerbate undertreatment. To eliminate crowd-out, capacity constraints imply that incentivizing the entry of physicians would be most effective. However, Norwegian municipalities already struggle to recruit physicians (of Health and Welfare, 2023), so such policies may be prohibitively expensive. Section 1.5.2 bounds the effects of a relatively costless intervention. Instead of randomly assigning patients of exiting physicians, Norway can target new patient-physician assignments with low expected crowd-out and eliminate 86 percent of the effect.

1.5.1 Broader Evidence of Undertreatment

Three pieces of evidence suggest that auto-reassigned patients inefficiently lower the treatment intensity of incumbent patients. First, I use data on the universe of Norwegian primary care utilization to benchmark the treatment intensity and health outcomes of incumbent patients in the analysis sample. Appendix Table A.4 shows this comparison, using all Norwegian patient-months in 2015. The coefficient of interest is the mean difference for in-sample patient-months after flexibly controlling for month, age, gender, tenure, and chronic illnesses. Relative to a population mean of 0.122 hours per month, in-sample patients receive 0.006 fewer hours of treatment (p-value < 0.001). Fewer hours do not necessarily imply undertreatment, e.g., if out-of-sample patients need less treatment. Comparing health outcomes within Norway helps alleviate this concern: relative to a mean of 0.0034 avoidable hospitalizations, in-sample patients have 0.0002 more avoidable hospitalizations (p-value < 0.01). These associations highlight the potential for policy interventions to mitigate crowd-out of primary care treatment intensity and perhaps improve health outcomes.

Second, as shown in prior surveys, Norwegian residents are concerned about primary care undertreatment. Over 25 percent express dissatisfaction with the duration of consultations (Kjøllesdal et al., 2020). This number is quite large relative to treatment intensity. On average, only 20 percent of residents visit their registered primary care physician each month. Likewise, despite national guidance requiring near-immediate access to physicians, 40 to 50 percent of patients are not satisfied with wait times for appointments (Bjertnæs et al., 2023; Kjøllesdal et al., 2020). Long wait times can lead to adverse health outcomes like avoidable hospitalizations when symptoms are not promptly addressed. With short appointments and long wait times, patients may receive inadequate specialist referrals: 25 percent of patients do not express satisfaction with referrals (Kjøllesdal et al., 2020).

Likewise, Norwegian physicians express dissatisfaction about time spent with patients. Rosta et al. (2019) document declining overall job satisfaction, largely driven by increasing dissatisfaction with work hours, amount of responsibility given, and freedom to choose treatment methods. In focus groups, physicians express concern that their workloads are growing and large enough to cause issues for patient safety and physician motivation (Svedahl et al., 2019).

Third, as of 2019, average physician visits are much lower in Norway than OECD countries overall: 4.5 vs. 6.8 (Tikkanen and Abrams, 2020). If on average, patients in OECD countries receive appropriate treatment, then lower visits in Norway would be consistent with undertreatment. Cross-country evidence is more mixed when considering specific primary care services. Relative to population, fewer elderly residents are immunized but more women are screened for breast cancer. Avoidable hospitalizations from diabetes and hypertension are relatively low. Finally, although capacity constraints best explain crowd-out, it may be difficult for Norway to increase capacity by hiring additional physicians. Norway has among the highest physicians per capita: 4.8 per 100,000 vs. an average of 3.5.

1.5.2 Policy Counterfactual

Since current auto-reassignments crowd out incumbent patients' treatment intensity, and baseline treatment is relatively low, Norwegian policymakers might consider replacing random assignment with targeted assignment.[16] I show that targeted assignment can lower crowd-out by 86 percent.

First, I estimate crowd-out effects for four subsamples including combinations of both physician heterogeneity and patient heterogeneity. High-type physicians either work part-time at baseline or reach 99 percent of stated capacity with reassigned patients. High-type patients are younger than 65. Second, I use point estimates to predict counterfactual crowd-out.[17] For each auto-reassignment event, I only consider assignments to physicians in the same municipality. For each physician, crowd-out is a sum of effects among young and old patients. If physicians in the same municipality do not have enough spots, I assume remaining reassignments to other municipalities have the worst-case crowd-out.

For the counterfactual, I first ration auto-reassigned patients to the physicians with the lowest total crowd-out, until updated enrollment reaches 99 percent of stated capacity. When enrollment approaches stated capacity, crowd-out is higher. I compare aggregate predicted hours of crowd-out under this counterfactual to random assignment across available slots within a municipality.[18] With targeting, predicted crowd-out over ten years drops from 83,877 to 11,483 hours. This corresponds to 1534 fewer avoidable hospitalizations if I apply the point estimates from Table 1.2.[19]

[16]Changes in treatment intensity may not fully reflect changes in welfare from this counterfactual, e.g., if patients have low taste for physicians with low crowd-out.

[17]I estimate imprecise positive point estimates for subsets with low crowd-out physicians. To be conservative, I treat these as zero reduction in hours per auto-reassigned patient.

[18]As before, I assume reassignments to other municipalities have the worst-case crowd-out.

[19]The calculation is $(83,877 - 11,483) \div 0.048 \times 0.001017$.

1.6 Conclusion

This paper provides causal evidence that new patients minimally crowd out the short-run primary care treatment intensity of existing patients. I leverage administrative assignment of new patients following physician exit for exogenous variation in nearby enrollment. This finding implies that physicians can shift along their labor supply curve without large frictions. In Norway, relaxing the regulated bounds on enrollment might allow highly demanded physicians to serve more patients without excessive declines in existing patients' treatment intensity. For example, Denmark allows physicians to choose patients after enrollment reaches an upper bound (Forde et al., 2016). Likewise, existing subsidies for rural physicians through supplementary base payments may be justified.[20] To minimize crowd-out, policymakers might consider requiring subsidy recipients to schedule full-time opening hours. If the finding holds in other institutional settings, policies like insurance expansions that increase the number of covered patients can increase healthcare access without resulting in large negative effects for the already-insured. However, estimates may not be informative in settings where patients' treatment needs already meet or exceed physicians' latent capacity.

Guided by a theoretical framework, heterogeneity in treatment effects suggests that idiosyncratic capacity constraints best explain observed crowd-out. To contextualize estimated crowd-out effects, evidence from out-of-sample comparisons, patient and physician surveys, and cross-country comparisons suggest that baseline treatment intensity is low. As a result, policymakers may seek to increase treatment intensity, especially for incumbent patients of physicians with rising enrollment. Since idiosyncratic capacity constraints better explain crowd-out than income effects, expanding capacity by subsidizing physician entry may better reduce crowd-out than incentivizing existing physicians to work more hours. However, both policies may be quite costly relative to the small level of crowd-out. Simulations suggest that Norway

[20] Norway has an equalization grant for physicians in municipalities with fewer than 5000 residents.

can eliminate 86 percent of crowd-out by replacing random auto-reassignment with targeted re-assignment.

Chapter 2: Should Physicians Choose Their Reimbursement Rate? Menu Design for Physician Payment Contracts

Abstract

Experts can leverage asymmetric information to induce demand for their services, complicating the design of payment contracts. In healthcare, physicians are widely believed to induce excessive treatment under a piece rate contract ("fee-for-service") and inadequate treatment under a flat-fee contract ("capitation"). A single contract that mixes fee-for-service and capitation payments may balance these forces for an average physician, but heterogeneous physicians plausibly have different socially optimal contracts. I study whether offering physicians a menu of contracts can improve welfare relative to a single contract. I first develop a model of treatment decisions, showing that welfare impacts are theoretically ambiguous and depend on the correlation between physicians' altruism, cost of effort, and patient needs. I then estimate the model using administrative data on Norwegian primary care physicians and their patients. In this population, the status quo single contract is inefficient. Physicians prefer a menu and respond by spending more time treating patients without increasing aggregate expenditure. The increase in patient health is equivalent to 5 percent of expenditure, with the largest gains for older, chronically ill, and rural patients.

A central challenge in healthcare is that physicians almost invariably have private information about the appropriate amount of treatment to provide to a patient. Compared to physicians, patients lack medical expertise, and medical records do not fully convey a physician's information to third-party payers. As a result, common reimbursement arrangements may not fully align physicians' incentives. Most often, payers reimburse physicians for each unit of treatment ("fee-for-service"). This arrangement can incentivize wasteful spending because a physician may only be willing to accept a marginal payment that exceeds the effective marginal cost. An alternative is to pay physicians an upfront flat fee based on a patient's expected costs. Removing the financial incentive to spend more time with patients may result in inadequate treatment. Moreover, insurance programs typically use a uniform fee schedule, which sets the same incentives for all physicians.[1] However, a physician's socially efficient incentive structure may vary with unobserved characteristics like an idiosyncratic cost of effort.

I present the first empirical evidence that replacing a single fee schedule with a shared menu of contracts can improve patient health without increasing spending. A menu allows each physician to choose a combination of a fee-for-service rate and a flat fee. I show how, for some distributions of unobserved physician heterogeneity, a menu can efficiently separate physicians across multiple fee-for-service rates. I estimate one such distribution in the context of Norwegian primary care, evaluate the social cost of information asymmetry, and derive a budget-neutral and voluntary menu of linear contracts that maximizes patient health. The welfare increase from offering multiple contracts is equivalent to 5 percent of initial expenditure.

I present a model of physician decision-making to quantify the expenditure and health impacts of counterfactual reimbursement schemes. In the model, physicians

[1]Fixed administrative fee-for-service rate schedules are employed by public insurers in Australia, Canada, China, Denmark, France, Germany, Japan, Norway, Singapore, Sweden, Switzerland, and Taiwan. These schedules generally cover primary care and sometimes also cover specialist and hospital services. In the United States, 44 percent of healthcare spending is paid by public insurance programs according to a fee schedule and private insurers increasingly negotiate physician reimbursement rates as a multiple of Medicare or Medicaid rates (Gottlieb et al., 2020).

choose a reimbursement contract and then treatment hours.[2] For each patient on a fixed list, a physician chooses treatment hours to maximize a weighted sum of private net income and patient health production (e.g., as in Ellis and McGuire, 1986).[3] Drawing on novel evidence from plausibly causal reduced-form research designs, I supplement this model with three types of physician heterogeneity: altruism is the weight on patient health relative to private profit, cost of effort represents the difficulty of spending time with patients, and productivity makes treatment more effective in improving health. Compared to a regulator, physicians have private information about both their characteristics and patients' initial illness severity. Patients' health returns to treatment are decreasing in effort and increasing in illness severity.

This model provides intuition for why a budget-neutral menu of contracts can sometimes increase patient health relative to a default uniform contract. A menu can increase the effort of physicians who spend relatively little time with patients. To see this, consider a simplified example with two physicians and a two-contract menu, which consists of a low-rate contract ($45 per hour and $b per patient) and a high-rate contract ($50 per hour and $0 per patient). For each physician, there is a social gain and a private gain from increasing the fee-for-service rate from $45 to $50. The social gain – the health benefit of increased physician effort – is larger for the physician who works fewer hours at $45 per hour.[4] Offered a menu, the physician with the larger private gain (which exceeds $b) will choose the higher $50 rate. The menu outperforms a uniform contract when this physician also has the larger social gain.

Depending on the correlation structure of physician heterogeneity, private and

[2]Each contract consists of a base payment per patient-month and a reimbursement rate per unit of treatment, e.g., an hour of patient interactions. In the United States, Medicare reimburses physicians based on the relative time and difficulty associated with furnishing a Medicare physician fee schedule service, measured as "relative value units."

[3]In the model, physicians do not exclude some patients to spend more time with others.

[4]This simplified example assumes that the patients of these physicians have similar needs, and locally, patient health increases in treatment hours, but the health benefit may not outweigh the cost of effort.

social gains may be aligned so that a menu of contracts can strictly increase welfare. Continuing the example, suppose that the two physicians vary in cost of effort and altruism. First, an older physician might work fewer hours at $45 due to trouble hearing patients (high cost), which implies a larger social gain of switching to $50. Second, a rural physician might have long-term relationships with patients (high altruism), leading to a larger private gain from switching to $50, because he especially values the incremental health. Combining these two forces, a menu can efficiently allocate an older rural physician to the high fee-for-service rate and a younger urban physician to the low rate. On the other hand, a menu is unlikely to be efficient if one physician has relatively high cost and low altruism, e.g., from being older and urban. High cost and low altruism both lower treatment hours, so in this case, the private gain is small while the social gain is large. For the same reason, a menu is unlikely to be efficient if physicians only vary along one dimension.

For a menu to improve a uniform contract, physicians must be sufficiently differentiated. I find novel reduced-form evidence that Norwegian primary care physicians vary along multiple dimensions. Consistent with heterogeneity in cost of effort, treatment hours vary widely across observably similar patients, and persistent physician heterogeneity explains a large share of this variation.[5] Consistent with heterogeneity in productivity, some physicians cause worse health outcomes among quasi-randomly assigned patients.[6] Consistent with heterogeneity in altruism, treatment hours change heterogeneously across physicians in response to increased reimbursement rates.[7] With this multi-dimensional heterogeneity, physicians' efficient reimbursement rates may be dispersed enough that a menu of contracts can meaning-

[5]Figure B.6 illustrates the identification intuition. On average, a physician with a high cost of effort treats all types of patients less than an otherwise similar physician with a low cost of effort.

[6]Productivity augments treatment hours in health production, so treatment hours are less dispersed among patients of a high-productivity physician than among patients of an otherwise identical low-productivity physician.

[7]Relatively altruistic physicians are less responsive to a reimbursement rate increase because they have less scope to vary treatment hours. At any reimbursement rate, these physicians sacrifice profit to further improve patient health.

fully increase efficiency. Still, to simulate the effects of counterfactual reimbursement schemes, I need to estimate the joint distribution of physician heterogeneity including its correlation structure.

Norway's institutional setting and data are particularly well-suited for estimating each physician's cost of effort, altruism, and productivity. I exploit a large and sudden increase in the reimbursement rate to identify altruism. Local regulations rule out several sources of potentially confounding variation. For example, payment rates are otherwise uniform across physicians and physicians do not choose their patients. Moreover, the restricted administrative data reflect the public healthcare utilization of nearly all Norwegian residents. I observe individual procedures, detailed demographics, medical histories, and adverse outcomes like avoidable hospitalizations and mortality. To estimate parameters of the structural model, I construct a balanced sample of registered patients and maximize the likelihood of observed treatment hours.[8] With patient records outside of the estimation sample, I can relax and test assumptions that may be necessary in other settings.[9]

I estimate considerable heterogeneity in physicians' marginal cost, altruism, and patients' treatment needs, implying large social costs of information asymmetry. Parameter estimates accurately predict treatment hours across physicians and across time for each physician, even for physicians outside of the estimation sample. With perfect information, the regulator would offer a different contract to each physician. Physician heterogeneity corresponds to widely dispersed full-information fee-for-service rates which incentivize greater treatment hours, increasing welfare by

[8]I focus on total treatment hours rather than subsets of care like procedures or diagnostics, which represent a small share of reimbursement and time. In Norway, primary care physicians screen for illness, manage chronic conditions, approve paid sick leave, and refer patients to specialist and non-emergency hospital services.

[9]For example, I test whether physicians' hours bunch at capacity constraints, whether patients systematically sort toward physicians with high health production, whether physicians with reimbursement rate increases are selected on unobserved characteristics, and whether in-sample patients and physicians are nationally representative.

$8.39 per patient-month or approximately 70 percent of baseline spending.[10] Welfare reflects patients' benefits from incremental treatment hours, as perceived by physicians. The underlying assumption is that physicians all perceive the same production function that maps treatment hours and patient need to a socially relevant measure of patient health. Alternative measures have limitations: for example, inadequate primary care may not have measurable effects on adverse health outcomes until several years later, in part because outcomes like mortality are rare and highly random. Likewise, due to asymmetric information, patient satisfaction may have a weak relationship with objective treatment quality.

With imperfect information about physician heterogeneity, the optimal menu of contracts still meaningfully increases welfare over the status quo. The difference amounts to $33 million per year across the Norwegian population. For comparison, the best uniform contract improves welfare by $22 million. The menu consists of seven traded contracts that mostly exchange higher fee-for-service rates for lower base payments. Higher rates imply greater perceived health for all patients. Gains are largest among patients with high need and low initial treatment, and these patients tend to have physicians with high cost of effort and low altruism. All else equal, these physicians have low private gains from high fee-for-service rates, but an efficient separation of physicians across contracts is possible because of variation in productivity and patient characteristics that shift severity.

Relative to the status quo, the gains from a menu of linear contracts are striking because menus are rarely featured in physician contract design.[11] Relative to full-information contracts, the menu's impact is somewhat modest, highlighting the significance of information asymmetry and the potential for further flexibility

[10] All welfare comparisons are measured relative to the status quo before observed reimbursement rates increase.

[11] I study a budget-neutral menu that maximizes patient health. Historically, menu initiatives prioritized lower spending but had limited impact, e.g., Quebec's 1999 reform studied in Fortin et al. (2021) and Medicare's Comprehensive Primary Care model.

in contracting. For example, I find evidence for large regional health disparities because rural communities tend to have both high-severity patients and low-treatment physicians. A national menu of linear contracts helps narrow these disparities, but there is room for further improvement from complementary policies. For example, the regulator could incentivize high-treatment physicians to establish practices in high-need communities.

Several robustness analyses suggest that welfare improvements are not driven by an idiosyncrasy of the empirical approach or setting. For example, counterfactual outcomes are similar when I incorporate more flexible specifications like preferences for leisure or large perturbations to the estimated joint distribution of physicians' cost, altruism, and productivity. Shifting from a uniform contract to a menu of contracts might therefore improve outcomes in settings other than Norwegian primary care.

In this paper I synthesize a large theoretical literature on physician contracting into an empirical framework for menu design. In both this paper and the stylized settings featured in prior work, the distribution of physician heterogeneity determines which types of contracts are efficient (Jack, 2005; Choné and Ma, 2011; Naegelen and Mougeot, 2011; Allard et al., 2014; Barham and Milliken, 2014; Wu et al., 2017; Wu, 2020; Ji, 2021). I characterize the optimal menu of contracts in terms of parameters that can be estimated with panel variation in reimbursement. I derive the menu for Norwegian primary care physicians to provide the first empirical evidence that any uniform contract is strictly less efficient. I also extend the empirical literature on socially optimal menu design with multi-dimensional consumer heterogeneity from insurance to a new selection market (physician labor supply) while incorporating unique dimensions of heterogeneity (Fang and Wu, 2018; Marone and Sabety, 2022; Ho and Lee, 2023). Similar to the study of health insurance menus in Marone and Sabety (2022), I estimate a joint distribution of agent types and characterize the relative efficiency of a uniform contract. In a parallel exercise, I use the graphical framework from Einav et al. (2010) to provide intuition for how a two-contract menu can increase efficiency for only some distributions of physicians. Outside of healthcare, there are few studies that empirically evaluate menus of contracts in selection markets, where multi-dimensional heterogeneity is first-order (Bellemare and Shearer, 2013; D'Haultfœuille and Février, 2020; Taburet et al., 2024).

I simultaneously estimate three key correlated dimensions of heterogeneity, which extends the literature that separately documents variation in physicians' altruism (Hennig-Schmidt et al., 2011; Godager and Wiesen, 2013; Douven et al., 2017; Galizzi et al., 2015) and practice style (Epstein and Nicholson, 2009; Chan and Chen, 2022; Doyle et al., 2010; Gowrisankaran et al., 2017). Policies that assume physicians vary along only one dimension may have unintended consequences.[12] Consistent with prior work, I show that physician treatment decisions respond to financial incentives (Brekke et al., 2017; Einav et al., 2018; Eliason et al., 2018; Clemens and Gottlieb, 2014; Cabral et al., 2021; Xiang, 2021). I find heterogeneity in this response, and decompose physician heterogeneity into structural physician types and variation in patient treatment need. This decomposition enables welfare analysis in contexts where selection affects both expenditure and healthcare quality.[13]

My framework emphasizes unobserved patient severity and a menu of linear contracts rather than a non-linear uniform contract. In primary care, dermatology, and dentistry – but also non-healthcare settings like indigent criminal defense – the regulator cannot observe the socially efficient level of effort and instead must rely on altruistic agents to exercise discretion in allocating effort across clients. In such settings, aligning incentives through differentiated contracts can improve welfare relative to targeting a fixed level of effort for each combination of patient and physician. In related work, Gaynor et al. (2023) estimate distributions of cost and altruism of dialysis clinics and derive the optimal non-linear uniform contract for an anti-anemia

[12]For example, if an insurer believed that physicians only vary in productivity, they might end contracts for physicians with low treatment. However, reimbursing these physicians at higher rates might be more cost-effective.

[13]In a setting with no effects on healthcare quality, Einav et al. (2021) study hospitals' selection into bundled contracts and subsequent changes in spending.

drug. I extend the framework from that paper to include unobserved patient severity and heterogeneity in productivity. Although a nonlinear uniform contract can achieve greater patient health than a menu of linear contracts, the uniform contract is not cost-effective without making many physicians worse off.

In Section 2.1 I present the theoretical model and provide intuition about the importance of correlated physician heterogeneity. In Section 2.2 I describe the empirical setting and present novel reduced-form evidence consistent with multi-dimensional physician heterogeneity. In Section 2.3 I discuss the parameterization and identification to recover the estimates, which are summarized in Section 2.4. In Section 2.5 I demonstrate the efficiency of a counterfactual menu of contracts, evaluate robustness, and discuss extensions.

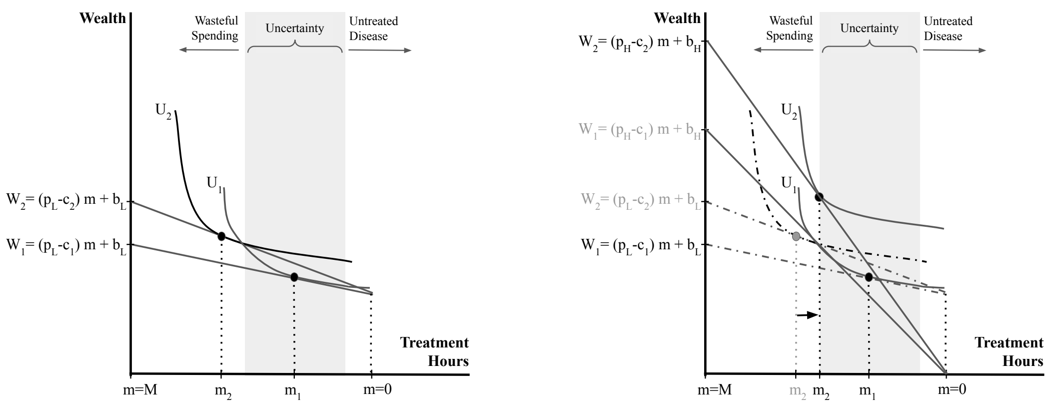
2.1 Theoretical Framework

2.1.1 Graphical Intuition for a Menu of Contracts

A uniform contract may be inefficient when physicians are heterogeneous. I first show this using a stylized graphical example. Consider the canonical model in which a worker chooses the number of hours to work $m \in [0, M]$ given a wage contract (p, b), where p is the reimbursement rate and b is the base payment. With this contract and private marginal cost of effort c, the worker earns wealth $W(m) = (p - c)\, m + b$. Privately optimal labor supply is where the indifference curve is tangent to the contract budget constraint. The budget constraint is steeper for smaller values of marginal cost.

Figure 2.1 plots wealth W against treatment hours m for two physicians, each with their own marginal cost and indifference curve. Typically, a competitive labor market implies that the reimbursement rate p should be the marginal product of labor. In many healthcare markets, the regulator does not observe the underlying treatment need, so the efficient level of labor supply is also unobserved. Labor supply that is too high may correspond to wasteful spending. Labor supply that is too low may lead

Figure 2.1: Two Contracts May Be More Efficient Than One



(a) Initial Uniform Contract (b) Menu of Contracts

Notes: This figure shows a stylized example with two physicians, in which a two-contract menu is more efficient than a uniform contract. The x-axis plots treatment hours $m \in [0, M]$ from right to left. Each panel shows the indifference curves of these physicians and the budget constraint(s) implied by simple reimbursement contract(s) with a base payment b and an hourly wage p. The shaded region includes the efficient level of labor supply which is unobserved to the regulator. In the left panel, the single status quo contract is efficient only for Physician 1. In the right panel, the regulator optimally offers a menu with two contracts to lower the labor supply of Physician 2.

to untreated disease. The shaded region reflects the regulator's uncertainty about patient severity and, consequently, the efficient level of labor supply. The figure is drawn in Panel A so that the initial uniform contract (p_L, b_L) is likely efficient for Physician 1, but the labor supply of Physician 2 is inefficiently high. Panel B introduces a second contract with a higher reimbursement rate p_H and a lower base payment b_H. Physician 2 chooses the new contract and lowers labor supply while increasing wealth. Labor supply is unchanged for Physician 1, who is indifferent between the two contracts.

The introduction of a second contract increased expenditure and moved labor supply closer to the efficient level. Whether this is efficient depends on the costs and preferences of physicians, as well as the social tradeoff between expenditure and patient health. Figure B.1 shows a counterexample where a uniform contract is efficient. If the physicians are nearly identical, then the differences between their choices of labor supply under a uniform contract may be negligible. Likewise, a uniform contract with a sufficiently large reimbursement rate p and small base payment b can induce any two physicians with quasi-concave preferences into the shaded region, but improvements in patient health may not justify the corresponding increase in expenditure. Below, with multi-dimensional heterogeneity for a continuum of physicians, the relative efficiency of a uniform contract still depends on the distribution of physician types and the social tradeoff between health and expenditure.

2.1.2 Model

I develop a model of physician decision-making to quantify expenditure and health outcomes under counterfactual menus. In the model, each physician has private information about her multi-dimensional type and patients' illness severity. A regulator designs a menu from which each physician chooses a contract. Next, each patient's severity of illness is drawn from a known distribution. Based on the severity and contract, the physician chooses the treatment intensity for each ill patient. Treatment intensity, physician productivity, and illness severity jointly determine a patient's health benefits.

Reimbursement Contracts. A contract maps treatment intensity m into a physician's revenue $x(m)$. Motivated by the empirical setting, I focus on contracts with a linear form, also called a two-part tariff: $x(m) = p\,m + b$. For example, the average physician in my sample receives $p = \$43$ per hour of patient interactions and $b = \$4$ per registered patient per month.[14]

The Physician. A physician determines treatment intensity m for each registered patient on a panel. Ex-ante, patients are characterized by a distribution of illness severity, $F(\lambda)$. Ex-post, realizations of severity λ are only observed by the physician. The physician also has private information about her type $\theta = \{c, \alpha, \gamma\}$,

[14] Sections 2.2.2 and B.2.2 describe how I calculate reimbursement per hour using data on higher-resolution services, e.g., visits, procedures, and diagnostics.

which is distributed in the population according to $G(\theta)$. Private cost of effort c is an opportunity cost of providing treatment. Altruism α is the weight on utility derived from patient health production relative to utility derived from net income. Both intrinsic and extrinsic forces may motivate physicians to value patient health, e.g., prosociality and reputation. Productivity γ^{-1} measures physician skill in terms of how efficiently treatment intensity translates into patient health benefits. A high-productivity physician needs relatively low effort to produce a certain amount of patient health. This notion of productivity is distinct from heterogeneous diagnostic skill (e.g., Abaluck et al., 2016). Here, a low-skill physician always requires more time to fully treat patients, rather than sometimes under-diagnosing them.

Before observing realized patient severity, the physician chooses the contract with the highest expected indirect utility: $p^*_\theta = \arg\max E\left[V(p;\lambda,\theta) \mid \lambda \sim F\right]$. Following the literature on physician-induced demand, e.g., Ellis and McGuire (1986), indirect utility V is a weighted average of private net income $(p-c)\,m + b(p)$ and patient health production $h(m,\gamma\lambda)$:

$$V(p;\lambda,\theta) \equiv \max_{m\geq 0}\,(p-c)\,m + b(p) + \alpha h(m,\gamma\lambda)\,. \tag{2.1}$$

After selecting a contract, the physician observes each patient's severity and chooses a corresponding quantity of treatment: $m^*(p) = \arg\max V(p;\lambda,\theta)$. Incremental treatment will earn additional revenue and influence patient health, but the value does not necessarily outweigh the additional cost of effort.[15]

The Regulator. The regulator observes the distributions of physician types θ and patient severity λ but not the realizations. The regulator chooses the menu of contracts $b(p)$ to maximize expected patient health production subject to a global budget constraint and each physician's participation constraint.[16] Expenditure, i.e.,

[15] Appendix B.1.3 relaxes and tests the assumption of linear cost of effort with a taste for leisure and a constraint on aggregate treatment intensity.

[16] Equivalently, the regulator maximizes a weighted sum of expectations over health production, expenditure, and physician indirect utility. In reality, it may be politically difficult to increase

total payments to physicians, cannot exceed the budget threshold, which incorporates the government's opportunity cost of healthcare spending. Non-health goods and services are also valued and taxation may distort behavior. Participation in the public system is optional, so the expected indirect utility of the physician must stay above a threshold. In the long run, physicians may choose an alternative medical specialty, practice location, or non-healthcare occupation. Physician exit is undesirable because a small number of physicians cannot realistically treat all patients.

The regulator's objective is:

$$\begin{aligned}
\max_{b(p)} \quad & \int_\theta E[h(m^*(p^*_\theta;\theta), \gamma\lambda;\theta) \mid \lambda \sim F]\, dG(\theta) && (2.2)\\
\text{s.t.} \quad & \int_\theta E[p^*_\theta m^*(p^*_\theta;\theta) + b(p^*_\theta) \mid \lambda \sim F]\, dG(\theta) \leq \bar{B} && [\mu_B, \text{ Budget}]\\
& E[V(p^*_\theta;\theta) \mid \lambda \sim F] \geq \bar{v}(\theta),\ \forall\theta && [\mu_{P,\theta}, \text{ Participation}]
\end{aligned}$$

where μ_B and $\mu_{P,\theta}$ are the shadow costs of expenditure and participation.[17] The social objective partially coincides with the physician objective because of altruism and the participation constraints, but otherwise differs because the regulator is budget-constrained, limiting physician payments. The optimal menu of contracts ("second best") satisfies the constraints as well as the first-order condition: in expectation, marginal health production equals marginal reimbursement minus marginal indirect utility, weighted by shadow costs:

$$\int_\theta E\left[h_m(m^*(p^*_\theta;\theta), \gamma\lambda) - \mu_B p^*_\theta m^* + \mu_{P,\theta} V_m(p^*_\theta, \theta) \mid \lambda \sim F\right] dG(\theta) = 0\,.$$

The first-order condition provides intuition about how physician quality is context-dependent, so physicians are not necessarily vertically differentiated. The degree to which a physician contributes to the social objective depends on both the type θ and

healthcare budgets even with a positive aggregate net impact, so I focus on strict constraints for exposition and counterfactual analysis.

[17]Privately optimal treatment intensity also depends on patient severity λ which is omitted for readability.

menu $b(p)$: $h(m^*(x;\theta),\gamma\lambda) - \mu_B p^*_\theta m^* + \mu_{P,\theta} V(x,\theta)$. Likewise, persistent variation in treatment intensity across physicians does not necessarily convey quality.

To benchmark social efficiency, consider the regulator's problem without information asymmetry about physician types θ. In this case, the regulator sets a personalized contract for each physician. The base payment b_θ^{FB} is just high enough for each participation constraint to bind, and the reimbursement rate p_θ^{FB} induces the efficient level of treatment intensity, $m^*(p_\theta^{FB};\theta)$ ("first-best"). Now, a stricter condition can hold for every physician:

$$E\left[h_m(m^*(p_\theta^{FB};\theta),\gamma\lambda) - \mu_B p_\theta^{FB} m^* + \mu_P V_m(p_\theta^{FB},\theta) \mid \lambda \sim F\right] = 0\,.$$

This first-order condition implies that the efficient reimbursement rate increases with physicians' marginal cost and decreases with altruism (See Appendix B.3.1). As the budget constraint relaxes, this level converges to private marginal cost.

2.1.3 Conditions for Efficient Self-Selection

The principal question of this paper is whether introducing a choice among contracts ("self-selection") is socially efficient. With the stylized example in Figure 2.1, a menu of two contracts may be more efficient than a uniform contract, but this depends on the distribution of types and the social tradeoff between health production and expenditure. This subsection extends that intuition to the full model: when starting from a reference contract, under what conditions will introducing a second contract strictly increase social welfare? I present a sufficiency condition and illustrate how efficient self-selection is facilitated by a dispersed and correlated distribution of cost, altruism, and productivity. From comparative statics, physicians who choose the high reimbursement rate will have relatively low cost, high altruism, and high productivity because they have the largest private benefit, all else equal. Increases in the reimbursement rate also lead to relatively large increases in public expenditure among these physicians, potentially outweighing the gains in health production.

Suppose that the regulator starts with a reference contract (p_L) and adds a higher fee-for-service contract to the menu (p_H). This two-contract menu increases efficiency if expected health production increases among the set of physicians who prefer the higher reimbursement rate, without increasing average expenditure. Let $\Delta z(p) \equiv z(p_H) - z(p_L)$, then

$$E[\Delta h(m(p), \gamma\lambda) \mid \Delta E\, V(p) \geq 0, \Delta E[pm(p) + b(p)] \leq 0] \geq 0\,. \tag{2.3}$$

All physicians who choose p_H will increase treatment intensity relative to p_L. If h is locally monotonic and concave in m, then an increase in treatment intensity necessarily increases health production. As a result, the problem simplifies to a question of feasibility: are any physicians willing to choose the high contract when the reduction in base payments offsets expected increases in fee-for-service reimbursement? Necessarily, physicians choosing the high fee-for-service contract must value incremental health production more than incremental costs on average. Importantly, physician contract choice is a selection market – expenditure on the high-fee-for-service contract depends on the set of physicians who choose it. A decrease in expenditure on base payments must offset both the mechanical ($m(p_L)\Delta p$) and behavioral ($p_H \Delta m(p)$) increases in fee-for-service expenditure among physicians who choose the high-fee-for-service contract:

$$E\left[\Delta\left(pm(p,\lambda) + b\right) \middle| \, \Delta E[V(p,b,\lambda)] \geq 0\right] \leq 0 \tag{2.4}$$

Comparing the partial derivatives of indirect utility and expenditure highlights the roles of correlation and dispersion.[18] Physicians are more likely to choose the high-fee-for-service contract if they have low cost, high altruism, high productivity, or high patient severity $E\lambda$.[19] If physicians only vary along one of these four dimensions,

[18]See Appendix B.3.1 for derivations and a similar discussion with weaker assumptions.

[19]As an aside, these statics may also be informative about the characteristics of physicians who choose to accept long-term positions with fee-for-service rather than salary reimbursement, e.g., private practice vs. HMO employment in the United States.

self-selection leads to more positive incremental expenditure, potentially violating the budget constraint. In direct contrast, physicians are most likely to decrease expected fee-for-service expenditure if they have high cost, low altruism, low productivity, or low patient severity, all else equal. With correlation among physician types, partial derivatives do not necessarily imply that physicians who most prefer higher rates will most increase expenditure, e.g., those with both high cost and high altruism.

The sufficiency condition for efficiently adding a high-fee-for-service contract requires additional assumptions to generalize to the broader question of menu design with any number of contracts. For example, if the fee-for-service rate of the reference contract is lower than the optimal uniform contract, it may be efficient to add a higher fee-for-service contract that attracts all physicians. A separating equilibrium in which more than one contract is traded also requires that some physician types prefer the low-fee-for-service contract: $\exists \theta : \Delta V(p, \theta) < 0$. With menus of three or more contracts, it may be efficient to offer a contract that decreases health production among some physicians if that lowers expenditure enough to efficiently subsidize higher fee-for-service rates and health production for other physicians. As a starting point, the intuition from the sufficiency condition is helpful for re-framing the problem as a sequence of two-contract menus that span a large set of reimbursement rates.[20]

2.2 Empirical Setting

The theoretical framework establishes that for some distributions of physician types, a menu of contracts can increase welfare relative to a uniform contract. Here I extend the framework to estimate such a distribution, derive the optimal menu, and

[20] In the closely related context of health insurance contracts, Chade et al. (2022) "decouple" a similar menu design problem. This requires quasiconcave household utility with respect to insurance coverage level. In the empirical application, I find that the optimal menu meets a related condition: each physician's expected indirect utility is quasiconcave with respect to reimbursement rate among traded contracts.

measure its impacts. I first explore several necessary assumptions in the setting of Norwegian primary care. I present institutional details in Section 2.2.1 which support the assumption that the focal variation in treatment intensity is driven by physician heterogeneity and contracts rather than patient composition. In Section 2.2.2 I detail the construction of a balanced estimation sample of patients that further removes potentially confounding variation. In Section 2.2.3 I introduce reduced-form evidence consistent with physician heterogeneity in cost, altruism, and productivity, which suggests that the status-quo uniform contract may be inefficient.

2.2.1 Institutional Setting

In Norway, each practicing primary care physician can increase their reimbursement by becoming certified as a general practitioner. In 2023, physicians without the certificate received $33 for a basic consultation and certified physicians received $44. As a result, with no changes to treatment intensity, a newly certified physician would suddenly earn 24 percent more fee-for-service revenue.[21] Crucially for causal inference, certification does not formally change a physician's patient pool, treatment options, or responsibilities. Physicians become eligible for the certificate by completing two years of additional part-time training and also having four years of full-time practice experience. Training includes both coursework and small-group meetings with other physicians, guided by national learning objectives.[22] Once the training is completed, physicians can apply for the certificate, which they typically receive after several months. Supplementary payments begin around that time and continue for five years. Before 2017, 80 percent of physicians received this certificate during their careers.[23]

[21] 24 percent reflects an average within the estimation sample, including reimbursement for other services provided during consultations.

[22] In 2019, physicians needed to meet 88 learning objectives. For example, Objective #18 is to understand challenges with over- and under-treatment.

[23] In March 2017, it became mandatory for most primary care physicians to start training toward certification. In March 2019, municipalities became responsible for facilitating supervised hours requirements and subsidizing part of the costs.

Apart from certification, physicians face nationally uniform reimbursement incentives. On average, physicians receive 70 percent of their revenue from fee-for-service payments, at rates listed in a national administrative schedule.[24] For example, in 2021, physicians received $17 for an E-consultation, made up of $16 from national health insurance and $1 from a patient copay (Legeforening, 2022).[25] In 2023, the schedule included 189 reimbursement codes, covering broad categories of physician services. The most commonly billed codes cover unspecified time spent with patients, rather than a specific procedure or diagnostic, highlighting the importance of physicians' discretion in choosing treatment intensity (See Table B.1).[26] The other 30 percent of revenue comes from base payments of approximately $4 per registered patient per month. Both fee-for-service rates and base payments are negotiated annually between the regulator and the physicians' union. If prices were instead negotiated individually between physicians and payers, as is common in the United States, it would be difficult to attribute variation in treatment intensity to reimbursement rates rather than physician skill or patient composition.

Within the scope of these national reimbursement agreements, physicians contract directly with municipalities. Among other details, these contracts stipulate the maximum number of registered patients and opening hours. Each physician agrees to meet the primary care treatment needs of between 500 and 2500 registered patients. National guidance states that physicians must be accessible to registered patients within contracted opening hours and that patients should not wait more than five days for a consultation in most circumstances (Lovdata, 2017). If physicians are unavailable, registered patients may seek treatment from stand-alone urgent care centers. Physicians provide consultations about symptoms, diagnostic tests, and general

[24]As of 2016, over 95 percent of physicians face this mixed contract. The remainder are fixed-salary employees of municipalities with no fee-for-service reimbursement.

[25]Once a patient reaches an annual individual cap on copayments, the public insurer funds the entire $17.

[26]In the United States, most claims for primary care consultations also include one of a small number of procedure codes.

medical procedures to registered patients. They also sign off on sick leave and refer patients to all specialist and non-emergency hospital services.

Patients often choose to remain with their registered physician for years at a time. One contributing factor is the centralized registration system, which allows patients to request a new physician twice per year. Patients can choose any physician who has fewer patients than the contracted maximum. The choice set changes infrequently due to the national licensing system, which fixes the total number of local physicians in the short term. I am able to construct a representative balanced panel for the estimation sample because physicians and patients tend to have long-term relationships.

2.2.2 Data

The estimation sample is a balanced panel of patients who are registered to certified physicians in the six months before and after certification (a "spell").[27] I focus on short-term variation and fix the composition of patients to attribute any sudden change in treatment intensity to the sudden change in marginal reimbursement. I construct the sample using restricted administrative records on registration, individual demographics, and healthcare reimbursement, which are maintained by Statistics Norway and the Directorate of Health.[28] These records nearly span the universe of Norway's residents and primary care physicians from 2008 to 2017.

The estimation sample excludes potentially confounding variation. First, each physician must only practice in one location during the entire period and each patient must be registered for the entire period. Second, both the physician and patient must have identification numbers to attribute treatment intensity to a particular physician of interest, which excludes recent migrants. I separately consider primary care from

[27] I classify the first month a physician is certified based on when they first receive a supplementary payment, including reimbursement codes 2dd, 2dk, 6ad, 11dd, 11min, and 14d, which is generally consistent with the certification date.

[28] See Appendix B.2.1 for additional details on data sources.

urgent care centers or second opinions. Third, each physician must provide some treatment during every month of the spell to exclude irregular variation that arises from the physician's absence, e.g., an anticipatory effect or temporary replacement physician. Table B.2 provides more detail on sample selection. In robustness analyses, I compare the estimation sample to a similarly defined control sample that includes patients whose physicians do not experience sudden changes in reimbursement.[29]

I construct measures of treatment intensity and marginal reimbursement rates that aggregate over the particular types of services provided. Treatment intensity m equals patient-month fee-for-service revenue divided by marginal reimbursement. This measure of intensity roughly corresponds to hours of treatment per patient-month ("simulated hours"). Marginal reimbursement p_{kt} is a "simulated wage" equal to the reimbursement per hour a physician would receive for providing the average bundle of services to a patient of type k in month t. I group patients with similar characteristics into ten types, and for each type, I use all Norwegian patients to calculate the average bundle of services received and the average hours required to provide that bundle.[30] I inflate all money-metric variables by Norway's monthly all-goods-and-services CPI to January 2023 USD.

The final estimation sample is approximately representative, and it includes 619 unique physicians and 643,363 patient-spells (13 months each).[31] Table 2.1 describes the distribution of selected characteristics and outcomes six months before

[29]To accommodate computer memory constraints, I use a 10-percent random subsample of physicians who never receive the certification supplement during the study period. I randomly select a 13-month spell that meets the same conditions as the main estimation sample, except for certification. Spells prior to certification are also safe comparisons, but I exclude these from the control sample to be conservative when analyzing selection into certification.

[30]See Appendix B.2.2 for additional details on constructing measures. For example, hours reflect time spent in encounters with registered patients and not work like administrative tasks. Table B.3 shows average characteristics and sample share separately for each patient type, including the simulated wage.

[31]When estimating the structural model, I split this sample into three parts and use the best-fitting set of estimates.

Table 2.1: Registered Patient Summary Statistics

	Control Sample	Estimation Sample					
	Mean	Mean	Std. Dev.	% > 0	10th	50th	90th
Patient Characteristics							
Reimbursement	8.59	8.33	25.49	20.74	0.00	0.00	30.92
Simulated Hourly Rate	43.82	43.76	6.86	100.00	32.38	45.49	50.95
Simulated Hours	0.19	0.18	0.56	20.74	0.00	0.00	0.68
Base Payment	4.03	4.01	0.11	100.00	3.84	4.02	4.13
Age	40.54	37.57	22.78	100.00	6.67	36.58	69.00
Chronic Illness	0.23	0.21	0.41	21.03	0.00	0.00	1.00
Months Registered	43.89	40.93	32.32	98.99	6.00	36.00	84.00
Physician Characteristics							
Max Enrollment	1268.60	1273.48	293.21	100.00	900.00	1220.00	1600.00
Physician Hours/Week	28.36	26.56	9.44	100.00	13.13	27.33	37.27
Physician Age	42.87	40.23	5.92	100.00	34.08	38.83	48.67
Patients Age 60+	0.23	0.19	0.10	100.00	0.07	0.18	0.32
Patients with Chronic Illness	0.23	0.21	0.06	100.00	0.14	0.20	0.29
Patients	131800	643363					
Physicians	136	619					

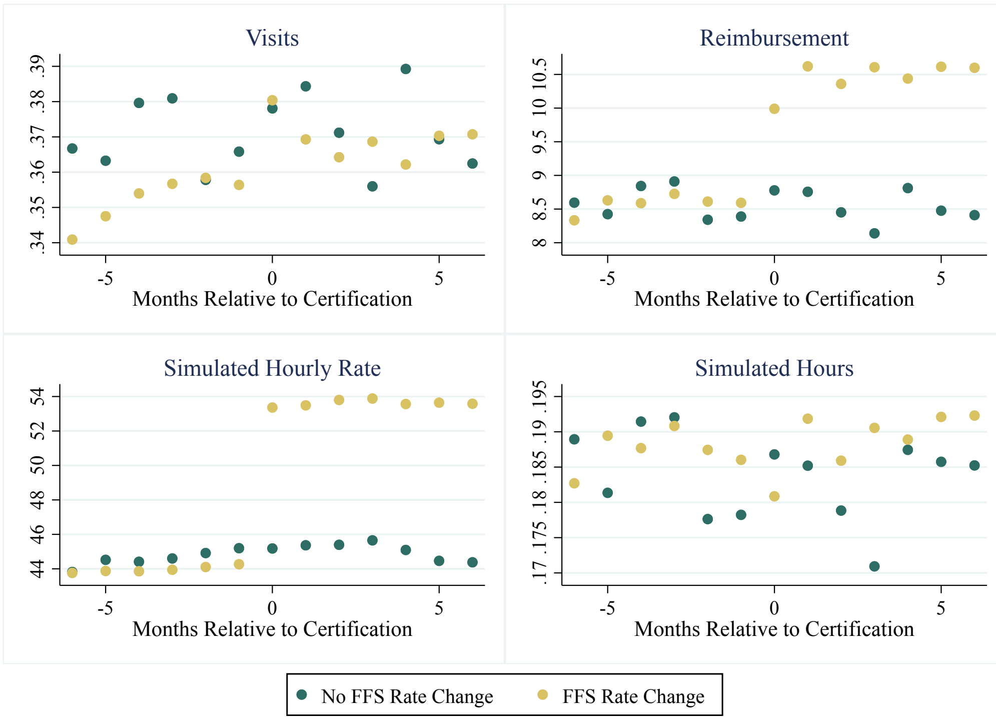
Notes: Summary statistics reflect registered patients' monthly totals six months before certification (or a randomly selected month for patients in the control sample). % > 0 indicates the share of patients with a strictly positive measure (row). Other columns reflect the mean, standard deviation, and 10^{th}, 50^{th}, and 90^{th} percentiles. Monetary measures are in USD. Physician Characteristics are also averaged across patients. The last two Physician Characteristics reflect shares of registered patients.

certification, and three facts stand out.[32] First, most patients do not visit their physician during a typical month. Second, the average physician spends 28 hours per week with registered patients (90^{th} percentile = 37) suggesting that with sufficient reimbursement, physicians can increase treatment intensity. Third, there is meaningful heterogeneity across physicians for proxies of mean patient severity like average age and chronic illness.

Trends suggest that treatment intensity varies systematically with marginal reimbursement and short-run changes are persistent. Figure 2.2 plots the trend in raw means, showing that visits, total reimbursement, and simulated hours all increase sud-

[32]See Table B.4 for the distributions of additional variables. See Table B.5 for comparisons to the Norwegian population: patients have similar characteristics, and certifying physicians are more often young and female.

Figure 2.2: Raw Means of Treatment Intensity Relative to Certification



Notes: These plots show averages of treatment intensity outcomes across patient-months in the estimation and control samples in each month relative to certification. Each sample is a balanced panel of patients, and in the estimation sample, Month 0 is the first month in which the registered physician received a certification supplement. I include registered patients' Visits and fee-for-service Reimbursement for the focal certified physician. Simulated hours equals monthly reimbursement divided by the Simulated Hourly Rate, an aggregation of service-level reimbursement rates that varies with patient characteristics, described in Appendix B.2.2.

denly after certification in the estimation sample but not the control sample. Unlike treatment intensity, trends in the number and composition of registered patients do not change with certification (See Figure B.2). These plots and most subsequent analyses reflect short-run variation around the sudden change in incentives which usually occurs months after physicians complete the prerequisite training. Short-run variation might obscure differences in long-run trends between certified and non-certified physicians that limit validity. For example, physicians who pursue certification might also make cost-reducing investments, or training might have delayed effects. Mitigating these concerns, Figure B.3 shows that even over five years, certification corresponds to a sudden and persistent increase in related measures of treatment intensity. Raw means suggest that the effects of certification might be overstated if using a longer time horizon, because treatment intensity dips during the middle of training.

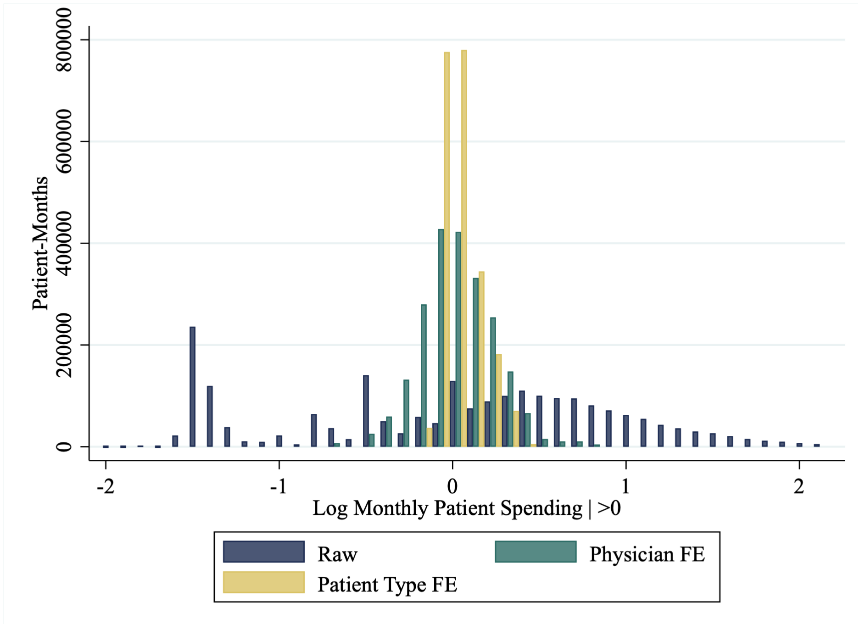
2.2.3 Stylized Facts

A necessary condition for physician self-selection is variation in physician types. I show novel reduced-form evidence consistent with heterogeneity in physicians' cost, altruism, and productivity. First, I show descriptively that observably similar patient receive more treatment at some physicians than at others, driving a large share of variation in treatment intensity. Second, I exploit quasi-random patient assignment to estimate the heterogeneous causal effects of physicians on treatment and adverse outcomes, consistent with variation in cost and productivity. Third, with a stacked difference-in-differences model, I show that treatment intensity increases in marginal reimbursement across a range of measures, highlighting the role of altruism. Fourth, I show heterogeneity in this effect, which suggests dispersion in altruism.

Cost and Productivity. Figure 2.3 shows the persistent variation across physicians in how intensively they treat observably similar patients. To make this comparison, I regress log reimbursement on fixed effects for each physician and 108 bins of patients with similar observed characteristics, as well as other controls.[33] Reimbursement per patient-month is approximately log-normally distributed with significant dispersion, while variation across patients with different observed characteristics (e.g., age, gender, chronic diagnoses) is relatively small. The limited dispersion across patients' observed characteristics implies that the regulator can only weakly predict patients' underlying treatment need and must generally defer to physicians' judgment about the appropriate level of treatment intensity. Physician fixed effects are more

[33] I regress log reimbursement on an indicator for post-utilization, physician fixed effects, high-resolution patient observed-type fixed effects (combinations of age bins, primary diagnosis, gender, and an indicator for lagged hospitalization), a time trend, and a quadratic function of patient age, among patient-months with positive reimbursement.

Figure 2.3: Decomposition of Treatment Intensity



Notes: This histogram shows the plot of log reimbursement for patient-months in the estimation sample with any utilization (Raw), as well as fixed effects from a regression of that outcome on an indicator for post-certification, physician fixed effects, high-resolution fixed effects for patients with similar observed characteristics (combinations of age bins, primary diagnosis, gender, and an indicator for lagged hospitalization), and a quadratic function of patient age.

dispersed, highlighting the large role of physicians in treatment intensity, similar to recent work such as that of Badinski et al. (2023).

These physician fixed effects should not be interpreted causally if, for example, patients with high unobserved severity systematically register with certain physicians. Fortunately, conditionally random patient assignment in Norway allows me to recover plausibly causal estimates of assignment to each physician ("assignment effects") on subsequent log treatment intensity, following the approach in Ginja et al. (2022).[34]

[34]When one physician exits, the municipality reassigns remaining patients to nearby available physicians, and the assignment is conditionally random. This variation exists for a subset of physicians. The research design compares patients of the same exiting physician who are assigned to different nearby physicians to recover those nearby physicians' assignment effects, controlling for the exiting physician, year, and nearby physician's municipality and availability. I shrink all physician assignment effects using Empirical Bayes.

As shown in Figure B.5, there is substantial dispersion in these physician effects even after shrinking effects to account for estimation error, reinforcing the importance of persistent physician heterogeneity. Limited patient selection is consistent with evidence from Norway that a patient's choice of physician is uncorrelated with the physician's effect on mortality (Ginja et al., 2022). In Norway and other settings, patients tend to respond to public measures of quality like star ratings (Bensnes and Huitfeldt, 2021; Vatter, 2022; Brown et al., 2023; Chartock, 2023). By contrast, treatment intensity does not appear to drive patient switching (Iversen and Lurås, 2011).

Continuing to use random patient assignment, I estimate effects of individual physicians on related outcomes to distinguish cost and productivity as drivers of persistent physician heterogeneity. In the model, low-productivity physicians treat patients multiplicatively more – leading to variation in assignment effects on log reimbursement – while low-cost physicians treat patients additively more – leading to variation in levels of reimbursement. Figure B.5 shows significant variation among both sets of assignment effects. For example, moving from the 10^{th} to 90^{th} percentile of physician treatment intensity corresponds to 1.19 additional visits each month over a patient mean of 0.34. I also estimate dispersion in assignment effects on avoidable hospitalization, which is largely uncorrelated with assignment effects for treatment intensity. This pattern suggests that due to physicians, health can vary even among patients with identical treatment hours and illness severity. Other natural experiments show dispersion across physicians in measures of productivity like resource use and skill, e.g., avoiding hospital readmissions (Doyle et al., 2010; Gowrisankaran et al., 2017; Chan et al., 2022; Chan and Chen, 2022; Kwon, 2023).

Altruism. Altruism is identified by how physicians' choice of treatment intensity responds to the reimbursement rate. Intuitively, relatively altruistic physicians have less scope to change treatment intensity when the reimbursement rate changes. At any reimbursement rate, these physicians sacrifice profit to provide greater health production.[35] To evaluate the effect of higher reimbursement from certification on treatment intensity, I estimate the following stacked difference-in-differences regression:

$$Y_{ijt} = \beta_1 Post_{jt} \times Certified_j + \boldsymbol{\beta}_x \boldsymbol{X}_{jt} + \gamma_i + \gamma_{y(t)} + \gamma_{m(t)} + \epsilon_{ijt} \qquad (2.5)$$

where Y_{ijt} is the outcome of interest for patient i of physician j in month t. $Post_{jt}$ is an indicator for months in which physicians receive certification supplements, $Certified_j$ indicates the main estimation sample of certified physicians rather than randomly selected non-certified physicians. β_1 is the coefficient of interest, X_{jt} is a vector of practice characteristics following Brekke et al. (2017), and $\gamma_i, \gamma_{y(t)}, \gamma_{m(t)}$ are fixed effects for patient, year, and calendar month.

A threat to identification would require that patients of certified physicians systematically need more treatment in the six months after certification than in the six months before for reasons other than certification, beyond the variation captured by time-invariant differences between patients and shared time shocks. Such variation is unlikely. First, physicians are not suddenly eligible to provide more expensive services. Second, as shown in Figure 2.3, future treatment need is difficult to anticipate, so physicians likely have little scope or incentive to strategically time their application for certification after completing the training. Alternative explanations are generally incompatible with Figure 2.2, which shows that average reimbursement does not trend differently for certified versus non-certified physicians in the months before certification.[36]

Table 2.2 shows that higher reimbursement rates result in greater treatment intensity. I observe precise increases both in visits, which are directly incentivized, and other measures of treatment intensity, which suggests complementarity between

[35]For any health production function with fixed concavity, the responsiveness of treatment intensity to marginal reimbursement, $\frac{dm}{dp}$, is proportional to inverse altruism, $\frac{1}{\alpha}$, among patients with positive treatment intensity.

[36]See Section 2.2.2 for discussion of the long-run variation shown in Figures B.3 and B.13.

Table 2.2: Main Effects of Certification on Treatment Intensity

	Post × Certified		Mean (Pre)	R^2	Obs.
Visits	0.015***	(0.001)	0.355	0.401	9,301,956
Reimbursement	2.093***	(0.106)	8.581	0.213	9,301,956
Simulated Hours	0.006**	(0.002)	0.187	0.186	9,301,956
Procedures	−0.001	(0.001)	0.071	0.237	9,301,956
Diagnostics	0.009***	(0.002)	0.229	0.266	9,301,956
Extra Time Codes	0.002***	(0.001)	0.086	0.230	9,301,956
Other Reimbursement	−0.303***	(0.076)	2.486	0.099	9,301,956
Specialist Reimbursement	0.245	(0.310)	19.702	0.190	9,301,956
Acute Hospitalizations	−0.000	(0.000)	0.019	0.153	9,301,956

Notes: This table estimates equation 2.5 using the pooled estimation and control samples, showing the coefficient on the interaction of indicators for the main (certified) estimation sample and post-certification. The unit of analysis is a patient-month and the sample includes the six months before and after a physician becomes certified for registered patients, among complete spells. Unless otherwise indicated, all outcomes are specific to a physician-patient pair with registration numbers, and zeroes are included. Visits includes any in-person encounter. Reimbursement indicates fee-for-service revenue. Simulated Hours is reimbursement divided by a price index as described in Section 2.2.2. Procedures, Diagnostics, and Extra Time Codes are counts of reimbursement codes grouped by the chapter of the reimbursement code. These categories are mutually exclusive but not exhaustive. Other Reimbursement includes treatment by any primary care physician other than the registered one, e.g., at community health clinics. Specialist Reimbursement includes all non-primary physician care eligible for public reimbursement. Acute Hospitalizations are unscheduled with admission within six hours. Mean (Pre) is an average of patient-months in the six months before certification, excluding the control sample. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level.

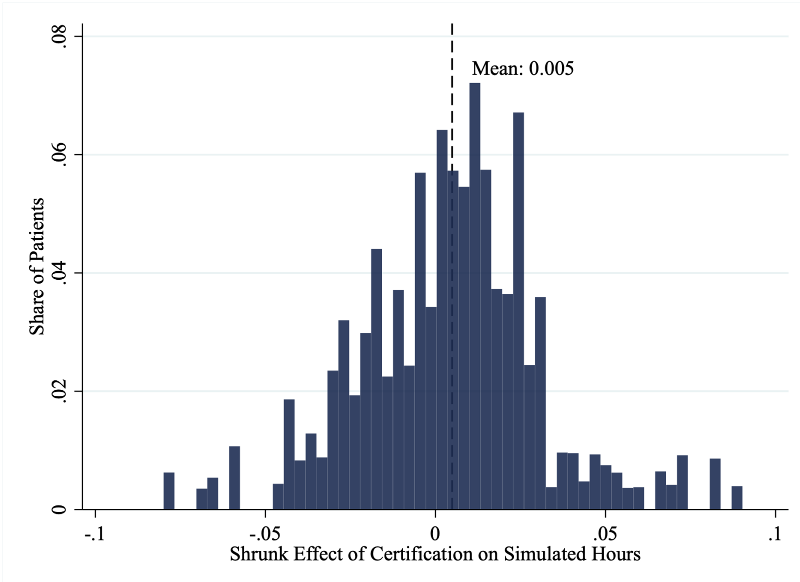
visits and in-visit services.[37] Simulated hours, which combines all categories of treatment, increase by approximately 3 percent of the pre-certification mean. Relative increases are similar for sub-categories of reimbursement codes like diagnostics and extra time per visit. Notably, increased treatment intensity provided by the registered physician coincides with small decreases in primary care from other physicians. The counterfactuals below focus on the treatment intensity of registered physicians and might overstate incremental expenditure from higher marginal reimbursement rates relative to this substitution effect.[38] I do not find evidence that certification immediately affects specialist treatment or acute hospitalizations.

Consistent with dispersion in physicians' altruism, I find heterogeneity in the effect of certification on treatment intensity. I extend the difference-in-differences analysis to include a post-certification indicator for each physician. Figure 2.4 is a plot of the physician-specific estimates after adjusting for error. Although the average physician increases treatment intensity post-certification, there is meaningful heterogeneity including precise negative estimates, motivating the test for income effects in Section 2.5.2. Estimates do not correlate precisely with physicians' observed characteristics like employment history or the maximum number of patients. Dispersion in altruism is consistent with experimental evidence of heterogeneity (Godager and Wiesen, 2013; Hennig-Schmidt et al., 2011). To interpret estimated elasticities exclusively as altruism, physicians must not vary in their ability to increase treatment intensity. In Section 2.5.2 I discuss several tests of this assumption. For example, in Figure B.14 I present descriptive evidence that capacity constraints do not bind in this setting and in Figure B.15 I show that high-altruism and low-altruism physicians respond similarly after observed shocks to patient health.

[37]Brekke et al. (2017) perform a similar analysis, finding a comparable effect on visits but no evidence of effects on treatment intensity per visit. The difference might be due to lower power from the narrower sample, confounding effects of changing patient composition from the underlying unbalanced patient panel, or confounding time-specific shocks from the lack of a comparison group like non-certified physicians.

[38]Changes to health production might be understated if registered physicians do not fully internalize substitution with other providers.

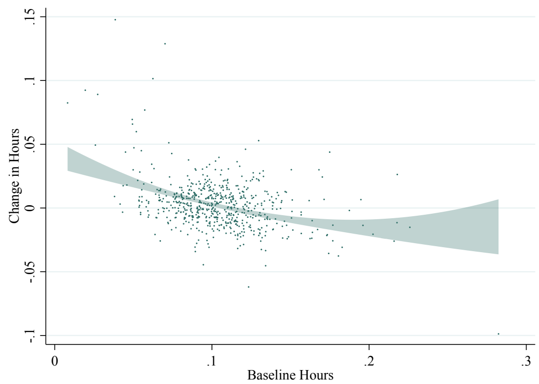
Figure 2.4: Distribution of Physician-Level Effect of Certification on Simulated Hours



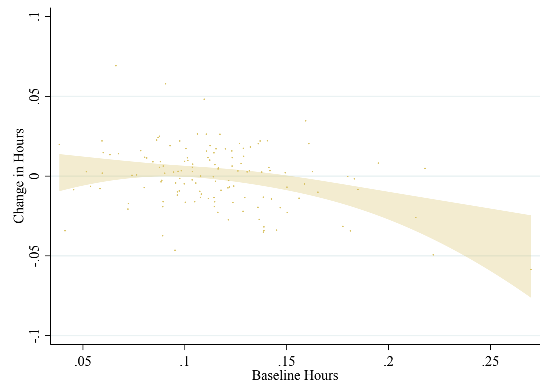
Notes: This histogram shows estimates of β_{1j} from equation 2.5 where the effect of certification is allowed to vary by certified physician. I shrink estimates to the mean using Empirical Bayes. Frequencies are weighted by the number of patients. Estimates are based on a subsample of spells starting 2010-2012.

Correlation. Dispersion in physicians' cost of effort, productivity, and altruism satisfy a necessary condition for physician self-selection. However, to separate physicians across contracts, these dimensions of heterogeneity should also be correlated. In Section 2.1.3, I illustrate how physicians with high efficient reimbursement rates must have relatively high willingness-to-pay for higher rates. Before estimating the correlation structure for physician types in the next section, I check for a consistent pattern in the raw data. Figure 2.5 shows that when the reimbursement rate increases, physicians with large increases in treatment hours (e.g., from low altruism) tend to initially provide low treatment intensity (e.g., from high cost). This pattern is likely not only regression to the mean because it does not clearly hold among physicians without reimbursement rate variation.

Figure 2.5: Raw Data Consistent with Correlated Cost and Altruism



(a) Certified Physicians



(b) Non-Certified Physicians

Notes: These plots shows the correlation between pre-certification treatment intensity and the change in treatment intensity (post-certification relative to pre). Each point is a physician. I calculate the average hours of treatment per registered patient in the six months before certification and the six months after certification. The placebo certification date is randomly selected for the control sample of non-certified physicians. The shaded region indicates a 95 percent confidence interval for a quadratic prediction.

2.3 Empirical Model

I estimate the joint distribution of physician heterogeneity to predict behavior under counterfactual menus and determine whether introducing a menu would increase efficiency relative to a uniform contract. In this section I review additional assumptions to support estimation as well as the intuition for which patterns in the data help to recover each parameter.

2.3.1 Parameterization

I estimate the distributions of physician heterogeneity and patient illness severity by maximizing the likelihood of observed treatment intensity. Privately optimal treatment intensity sets marginal net income equal to marginal health production scaled by altruism. The key assumption supporting empirical analysis is that conditional on observed characteristics, patient severity λ is independent of the reimbursement rate p and physician type θ.[39] To generate a likelihood, I make two parametric assumptions that I later relax in Section 2.5.2. First, since economies of scale are unlikely in this setting, I continue to assume that costs increase linearly in treatment intensity: $c(m) = c\,m$.[40] Second, health production is quadratic in the distance between treatment intensity and patient severity scaled by productivity: $h(m, \lambda; \gamma) = H - \frac{1}{2}(m - \gamma\lambda)^2$. Quadratic functional forms are common in the insurance literature to model households' valuation of treatment intensity, e.g., Cardon and Hendel (2001), Einav et al. (2013), and Marone and Sabety (2022). Given these assumptions, privately optimal treatment intensity takes the form:

$$m^*(p, \lambda, F) = \max\{0, \frac{p - c}{\alpha} + \gamma\lambda\}\,. \tag{2.6}$$

Gaynor et al. (2023) use a special case of this parameterization where γ is constant across physicians and λ is a deterministic function of patient characteristics.

The final step is to solve for the model residual, the unobserved component of patient severity. I parameterize the distribution of severity as a two-stage process. Conditional on being positive, severity is distributed log-normal, where the mean varies with observed characteristics: $(\ln \lambda \mid \lambda > 0) \sim N(\beta_\lambda X_\lambda, \sigma_\lambda)$.[41] I parameterize the probability that severity is positive as $Pr(\lambda > 0) = \frac{\exp d_0 + d_1 \beta_\lambda X_\lambda}{1 + \exp d_0 + d_1 \beta_\lambda X_\lambda}$. This step helps rationalize why patients often have zero treatment intensity, similar to Ho and Lee (2023). Appendix B.3.2 presents the full expression of the conditional likelihood.

[39]In Sections 2.2.3 and 2.5.2 I discuss evidence supporting this assumption.

[40]For example, the regulator dissuades a large number of patients per physician by approving the entry of each new practice. Similarly, the maximum number of patients per physician can be up to 2500 but most physicians choose a much lower maximum. I exclude the small number of physicians who share a workload with other physicians.

[41]These characteristics include fixed effects for each of the 10 observed patient types, fixed effects for calendar months, normalized lagged treatment intensity, an indicator for zero lagged treatment intensity, indicators for cancer, diabetes, COPD, Asthma, and CVD, indicators for 1 or 2+ of these chronic illnesses, indicators for female and disability receipt, percentile of income as of 2016, indicators for 1 or 2+ acute hospital visit in the last 6 months, and indicator for registering with the current physician in the last 6 months and a scaled time trend.

2.3.2 Identification Intuition

An altruistic physician places high weight on patient health production relative to private net income. When reimbursement rates increase the altruistic physician's treatment intensity is relatively unresponsive, despite the incentive of higher marginal revenue. For any concave health production function, responsiveness $\frac{dm}{dp}$ is proportional to inverse altruism $\frac{1}{\alpha}$. Next, consider the distribution of treatment intensity across patients of one physician at a time. If two physicians and their patients are otherwise identical – the same altruism, productivity, and mean patient severity – then a high-cost physician will have the entire distribution of treatment intensity shifted to the left of a low-cost physician. Likewise, all else equal, a low-productivity physician will have a more dispersed distribution than a high-productivity physician. In figure B.6, I show stylized visual examples of these patterns. Conditioning on physician heterogeneity, the remaining correlation between treatment intensity and patients' observed characteristics identifies the conditional means of the distribution of patient severity. Variance in residual treatment intensity reflects the variance of unobserved patient severity.

The key assumption supporting identification is that the data include within-physician variation which separately shifts marginal utility from net income (in this case, $p-c$) and marginal health production ($h_m(m, \gamma\lambda)$). Increased marginal revenue p shifts marginal net income, and patient characteristics X shift marginal health production via expected severity $E[\lambda|X]$. For example, older patients likely need more care on average so there are different returns to health from treatment. Besides the additive separability of net income and health production, an implicit assumption is that these terms have different second derivatives with respect to treatment intensity.

2.3.3 Estimation

To recover parameters of the model, I maximize the likelihood of observed treatment intensity for patients of certified physicians in the six months before and after a change in marginal reimbursement from certification: $l(m \mid \theta_i, p, F)$.[42] Parameters include the conditional means and variance of patient severity $F(\lambda)$, and each certified physician's marginal cost c, altruism α, and productivity γ^{-1}. Estimated parameters are sometimes simple transformations of model parameters.[43] The full distributions of productivity and patient severity are not separately identified, so I fix the intercept of log severity $\beta_{\lambda,0}$ at zero.[44] To accommodate computer resource constraints, I separately estimate parameters for three subsamples: 2008-2010, 2011-2013, and 2014-2016. I use the 2011-2013 subsample for counterfactuals because parameter estimates best predict treatment hours.

2.4 Estimates

Parameter estimates are sensible and fit the data, accurately predicting treatment intensity both in- and out-of-sample. To assess the model fit, I first plot observed treatment intensity against predicted values. Figure 2.6 shows a correlation of nearly 1 for both the estimation sample and a control sample of never-certified physicians.[45] Estimates predict treatment intensity well both across physicians and over time for particular physicians. Table B.11 shows corresponding regressions: the coefficient on predicted treatment intensity is approximately 1, even when including physician fixed effects in columns (3) and (5). Column (5) shows that, conditional on estimates, patient covariates explain little remaining variation in treatment inten-

[42]I use L-BFGS-B with the Python module JAX to calculate the analytic gradients of the log-likelihood objective. The box constraints are that cost, altruism and productivity are strictly positive and that marginal cost is no more than ten times as large as marginal revenue.

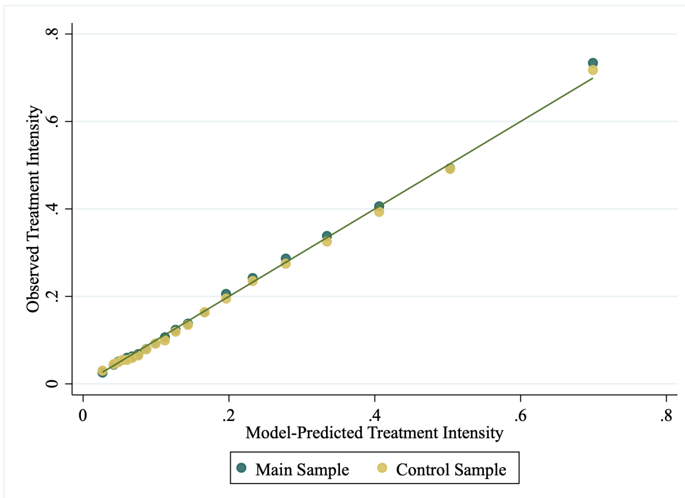
[43]c is a multiple of the fee-for-service rate six months before certification, α is scaled by 1000, and σ_λ is exponentiated. The transformation of c implies that marginal cost varies across patients of the same physician.

[44]With this normalization, I assume that a young long-term low-income male patient with no major diagnoses or lagged utilization in the first month of the sample has diminishing returns to treatment after 1 hour per month.

[45]The control sample is a nearly identical balanced panel of patients for randomly selected spells of other physicians with no reimbursement variation from certification (See Section 2.2.2).

sity.[46] In counterfactual analysis, estimates also rationalize the choice of physicians to become certified even though that choice is not used to estimate the model. All physicians in the estimation sample have higher expected indirect utility EV after certification, with an average of \$1.80 per patient-month. Figure B.8 shows the distribution of this change in EV across physicians.

Figure 2.6: Model Fit: Ventiles of Predicted Treatment Intensity



Notes: This plot shows ventiles of predicted patient-month treatment intensity on the x-axis against means of actual treatment intensity on the y-axis. The 45-degree line is also plotted.

The correlation between estimated cost, altruism, and productivity reinforces the potential for efficient self-selection. Figure B.7 shows the joint density of physician heterogeneity. High-cost physicians tend to have low altruism, productivity, an mean patient illness severity. The upper panel of Table 2.3 shows that observed characteristics explain some of this variation.[47] For example, productive physicians tend to be younger and born outside of Norway. They hold larger lists of patients, make greater use of diagnostics relative to procedures, and historically worked under fee-for-service contracts. The bottom panel shows residual variation in physician types

[46] Adding patient covariates does not increase in R^2 and slightly increases the coefficient on predicted intensity.

[47] All standard errors are adjusted for noise in parameter estimates.

Table 2.3: Correlates of Physician Heterogeneity

	$\ln c$	$\ln \alpha$	$\ln \gamma$
Constant	0.902***	8.418***	−0.348***
	(0.168)	(0.280)	(0.011)
Age	0.031	0.024	0.035***
	(0.028)	(0.048)	(0.002)
Max Enrollment	−0.011	0.019	−0.015***
	(0.032)	(0.052)	(0.002)
Pr(Diagnostic)	−0.057*	0.023	−0.085***
	(0.030)	(0.049)	(0.002)
Ever Fixed-Salary	0.113	−0.050	0.113***
	(0.184)	(0.297)	(0.010)
Female	0.018	−0.049	−0.003
	(0.060)	(0.101)	(0.004)
Migrant	−0.104*	−0.022	−0.021***
	(0.063)	(0.110)	(0.004)
Rural Municipality	0.099	−0.091	0.003
	(0.077)	(0.127)	(0.004)
Trend	0.121	−0.639	−0.138***
	(0.304)	(0.517)	(0.018)
S.D. Residual	0.227***	0.318***	0.145***
	(0.031)	(0.029)	(0.002)
$\rho(\ln c, \ln \alpha)$	−0.269*		
	(0.139)		
$\rho(\ln c, \ln \gamma)$	0.561***		
	(0.101)		
$\rho(\ln \alpha, \ln \gamma)$	−0.295**		
	(0.137)		

Notes: This table regresses log physician-level estimates of cost c, altruism α, and inverse productivity γ on observable characteristics. Standard errors come from the delta method using the approximate Hessian of parameter estimates. Continuous covariates are normalized by mean and standard deviation relative to the full population of physicians. Max Enrollment is the largest number of patients a physician agrees to have on their registered list. Pr(Diagnostic) is the share of reimbursement lines that are diagnostic relative to procedures. Ever Fixed-Salary is an indicator for physicians ever working as employees, rather than contractors, of municipalities with no marginal reimbursement. S.D. Residual is the standard deviation of the residual of log estimates after regressing on covariates. ρ indicates the correlation between residuals.
*, **, and *** denote statistical significance at the 10%, 5%, and 1% level.

is also widely dispersed and correlated.[48]

Patient observable characteristics explain a moderate share of the variation in treatment intensity by shifting illness severity (See Table B.6). Seasonality and particular chronic illnesses are major determinants of patients' treatment needs. For example, utilization is much lower in August than in January, and diabetes patients are more likely to visit a primary care physician than cancer patients. Other coefficients are precise but unexpectedly low in magnitude relative to raw correlations with treatment intensity, e.g., lagged treatment intensity and gender. Increasing lagged treatment by one standard deviation would only increase the health shock about as much as the average difference between January and April. This small coefficient reinforces the assumption that the distribution of health shocks is conditionally independent across months within each patient. Finally, conditioning on the full set of patient covariates, patient severity is highly dispersed and difficult to predict.

2.5 Counterfactual Menus of Contracts

2.5.1 Baseline Counterfactuals

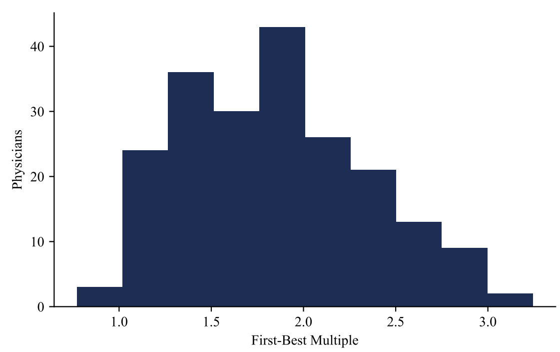
Using estimates, I simulate physicians' choices under counterfactual menus to illustrate the welfare effects of self-selection. First, I quantify the cost of information asymmetry by solving for the personalized contracts offered by the regulator with perfect information. Second, I benchmark to the status quo and find that the existing reimbursement supplement is nearly optimal if the regulator can only offer a single (uniform) contract. Third, I demonstrate that even an arbitrary two-contract menu can increase welfare relative to a uniform contract because the distribution of physician heterogeneity satisfies key properties of dispersion and correlation. Fourth,

[48] Rather than introduce a menu, a regulator could condition the reimbursement rate on observed physician characteristics. The substantial unobserved heterogeneity suggests that targeting observed characteristics may be ineffective. Likewise, targeting may be infeasible given, e.g., legal protections for age and physicians' collective bargaining.

I derive the menu of linear contracts that maximizes welfare given imperfect information. I conclude by assessing the equity implications of the optimal menu.

To scale health production into dollars, I assume that the regulator values incremental health production from certification as much as incremental expenditure. This assumption implies that the regulator is 3.1 times as altruistic as the median certified physician.[49] Table 2.4 compares aggregate health production across counterfactual menus, relative to the pre-certification status quo. Columns for expenditure and physician indirect utility reflect the budget and participation constraints. To focus on the role of reimbursement in treatment intensity, I fix other sources of variation at values six months before certification: enrollment, the share of patient types for each physician, and pre-certification fee-for-service rates. I provide additional detail on how I measure counterfactual outcomes and search for counterfactual menus in Appendix B.2.3.

Figure 2.7: Dispersion in Efficient Reimbursement Rates



Notes: The y-axis is the count of physicians in each bin. The x-axis is a multiple of pre-certification fee-for-service that maximizes scaled health production subject to strict physician-level participation constraints and a global budget constraint (average expenditure must be less than status quo post-certification). The grid of fee-for-service multiples includes 200 points between 0 and 2.5. The base payment is the lowest level for each physician to satisfy the participation constraint for each physician.

[49]For comparison, Gaynor et al. (2023) calibrate a comparable parameter at 52.6 times the median altruism among providers based on the value of a statistical life-year. My approach does not internalize the regulator's valuation of certification training beyond immediate changes to health production.

Table 2.4: Annual Counterfactual Outcomes for Norwegian Population ($M)

	Health Production	Share of Max	Expenditure	$E[V]$
Pre-Certification	0.0	0.000	0.0	0.0
Post-Certification	139.0	0.264	138.9	113.6
	(0.4)	(0.001)	(0.4)	(0.4)
Efficient Contracts	525.8	1.000	137.2	0.0
	(3.0)	(0.000)	(0.6)	(0.0)
Optimal Uniform Contract	153.7	0.292	132.5	103.6
	(2.1)	(0.003)	(0.5)	(0.5)
Optimal Menu of Contracts	176.5	0.336	144.9	109.1
	(1.9)	(0.003)	(0.4)	(0.6)

Notes: This table shows key outcomes from realized and counterfactual contract menus, scaled annually to the Norwegian population (5.24M). All outcomes are based on ex-ante expectations over patient-months using estimated distributions of G and F, weighted across physicians by enrollment. Enrollment, the share of patient-types, pre-certification fee-for-service rates, and base payments are fixed at values six months before certification. Post-certification fee-for-service rates are fixed at values in the month after certification. Counterfactuals vary fee-for-service rates and base payments, enforcing participation and budget constraints. Health production is scaled such that the regulator is indifferent between incremental expenditure and incremental expenditure from certification. Share of Max divides the first column by its maximum from efficient contracts. Expenditure includes both fee-for-service and base payments. $E[V]$ is the expected indirect utility per patient-month of private physicians. Standard errors, shown in parentheses, are calculated across 25 bootstrap estimation samples, with randomly selected patient-months within physician and re-solved counterfactual menus. Figures B.11 and B.12 further illustrate the distribution of counterfactual contracts across bootstrap samples.

With perfect information about physician heterogeneity, personalized contracts would increase expected health production by $525 million per year nationally. In this first-best allocation, efficient contracts achieve nearly four times the gain in health production of the observed reimbursement rate increase at a lower cost while satisfying strict participation and budget constraints. I identify efficient contracts by selecting the fee-for-service rate for each physician from a grid that maximizes $E[\alpha_R h(m^*, \lambda) - p\, m^*]$. I set base payments so that in expectation, each physician is indifferent between the efficient contract and the status quo. Figure 2.7 shows substantial heterogeneity in the efficient reimbursement rates.[50] On average, efficient rates are 84 percent above the initial status quo rate with substantial variation (SD = 49 percent). Efficient rates are far above the status quo because a large share of physicians have high cost of effort or low altruism. In the status quo, these physicians spend relatively little time with patients, so the health benefits of incremental treatment are large.

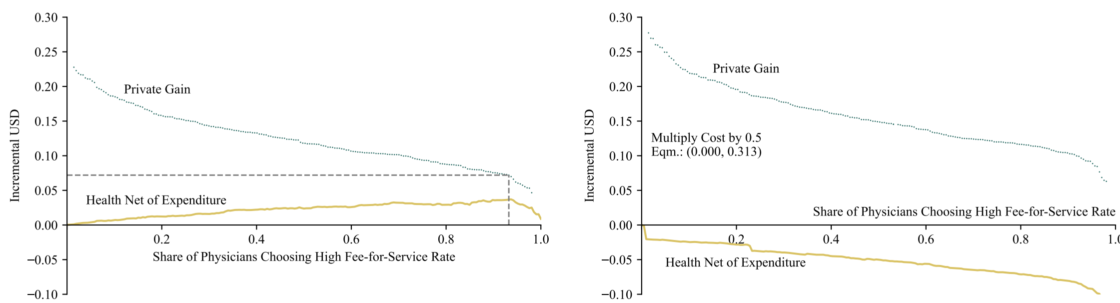
In the status quo, the reimbursement rate increases by 24 percent which improves health production by approximately one-fourth as much as efficient rates. Part of the difference is because the new status quo rate is too high for some physicians. For example, the most altruistic physicians do not change treatment intensity enough to justify the mechanical increase in expenditure. Since most physicians have even higher efficient rates, the regulator could still improve health production at lower cost with an ever higher uniform reimbursement rate and lower base payments. On average, physician surplus would be lower, but all physicians weakly prefer this contract to the initial status quo.

Even a two-contract menu achieves meaningful efficiency gains relative to the best uniform contract. Reinforcing the intuition from Section 2.1.3, this intermediate

[50] Throughout this section, I discuss multiples of counterfactual reimbursement rates. For example, 1.2 indicates 120 percent of the initial fee-for-service rate. This approach preserves variation in fee-for-service rates across patients while allowing simple graphical comparisons across counterfactuals. In a robustness check below, I consider a unique reimbursement rate for each type of patient with similar observed characteristics.

exercise shows how the self-selection may not increase welfare for some distributions of physicians. I adapt the graphical framework for selection markets introduced in Einav et al. (2010) and extended by Marone and Sabety (2022). I start with the optimal uniform contract (p_L) and add a contract (p_H) to the menu with a marginally higher reimbursement rate. If p_H requires accepting a relatively low base payment, then only a fraction of physicians with large private benefits ($EV(p_H, 0) - EV(p_L, 0)$) will choose it. Physicians with large private benefits have relatively low cost, high altruism, and high productivity (See Appendix B.3.1). However, these characteristics also predict relatively large increases in expenditure which might outweigh the corresponding increase in health production, especially if cost is low relative to altruism.

Figure 2.8: Two-Contract Menus: Setting Incremental Base Payments



(a) Estimated Distribution

(b) Counterfactual Distribution: $\frac{1}{2} \times c$

Notes: This figure shows outcomes under a menu that includes the best uniform fee-for-service rate and a fee-for-service rate that is incrementally higher while varying the difference in the base payment between these contracts. The x-axis orders a continuum of physicians according to their decreasing private gains from an increased reimbursement rate. The green line is incremental social surplus for each percentile of private gain: expected (scaled) health production minus expenditure among all patients (and all physicians). Grey dashed lines indicate the optimal share of physicians choosing the high-fee-for-service contract and the corresponding difference in base payments between the two contracts. Panel A shows the estimated distribution of physician heterogeneity. Panel B multiplies estimated marginal cost by 0.5.

Figure 2.8a shows the tradeoff between increased health production and increased expenditure across physicians, ordering physicians in decreasing order by their private benefit from an increased fee-for-service rate ("WTP"). The WTP curve is like a demand curve, indicating participation in the high-fee-for-service contract for various prices Δb, i.e., lower base payments. I also summarize welfare as incremental social surplus: expected health production minus expected expenditure, relative to the low-fee-for-service contract, where expenditure reflects both fee-for-service and base payment changes in equilibrium.[51] For each share of physicians choosing high-fee-for-service, I show the average incremental surplus across all patients. The regulator sets incremental base payments to maximize expected social surplus: 93 percent of physicians choose the high-fee-for-service contract with a \$0.07 lower base payment. With smaller differences in the base payment, more physicians would choose the high-fee-for-service contract and expenditure would outweigh incremental health production. Figure B.10 shows that variation in social surplus is best explained by cost. Variation in WTP is best explained by mean patient severity rather than cost, altruism, or productivity.

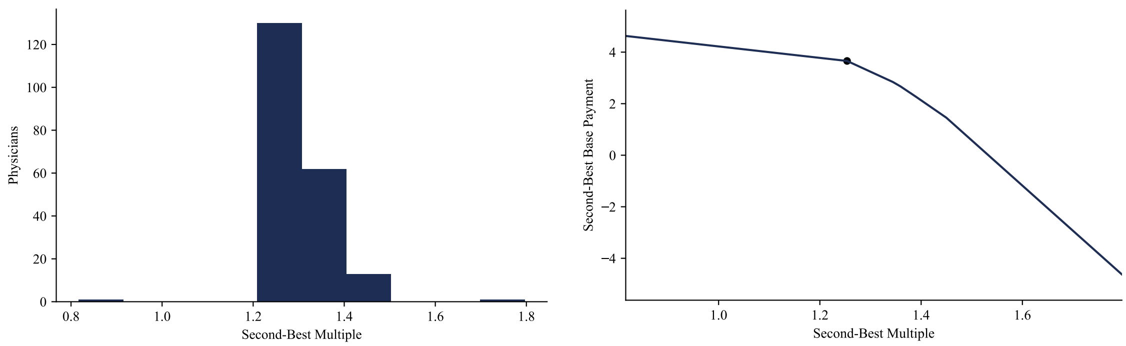
Figure 2.8b illustrates a counterexample where the two-contract menu is not more efficient than the uniform contract. This panel repeats the previous exercise with a counterfactual distribution of physician heterogeneity. When marginal costs are half as large, WTP is greater and efficient rates are lower. WTP and social surplus are not sufficiently correlated for both contracts to be traded. Incremental expenditure would always exceed incremental health production. The regulator sets the incremental base payment high enough for a corner solution where all physicians choose the low-fee-for-service contract. Lower cost of effort makes efficient rates more affordable and more similar, so efficient rates are all relatively close to a feasible uniform contract.

The optimal 7-contract menu achieves large efficiency gains by separating some physicians into high-fee-for-service contracts: \$33 million per year more than the status quo or 34 percent of first-best. To search for this menu, I adapt the line-search algorithm from Marone and Sabety (2022) and Azevedo and Gottlieb (2017).

[51] At virtually any quantile of WTP, some physicians will be inefficiently selected into the high-fee-for-service contract and some will be inefficiently selected into the low-fee-for-service contract, relative to full information with the same restricted menu.

Most physicians choose just one of three contracts (Figure 2.9a) and the optimal base payment decreases concavely in the fee-for-service rate (Figure 2.9b).[52] Perhaps a smaller menu would involve lower implementation costs: Figure B.9 shows that while increasing the number of contracts per menu generally improves welfare, most efficiency gains can be achieved with a small number of contracts.

Figure 2.9: Optimal Menu of Contracts



(a) Physicians' Reimbursement Rates (b) Base Payment Schedule

Notes: In Panel A, the y-axis is the count of physicians in each bin. The x-axis is a multiple of pre-certification fee-for-service that maximizes scaled health production subject to strict physician-level participation constraints and a global budget constraint (average expenditure must be less than status quo post-certification). Panel B plots base payments versus multiples of status-quo fee-for-service rates for the optimal menu. The point indicates the optimal uniform contract.

Redistribution across patients drives some of the gains in average welfare from efficient contracts and the optimal menu. To explore redistribution, I disaggregate counterfactual outcomes across physician types. In Table 2.5 I categorize physicians into 16 groups based on whether each of cost, altruism, productivity, and expected patient severity are above or below the median. For both efficient contracts and the optimal menu, health production increases most among the 40 percent of physicians with high cost and low altruism. Changes in expenditure are relatively small. All else equal, these physicians tend to have low private gain from increased rates. The

[52]Moreover, Figure B.12 shows that across bootstrap samples, the optimal menu consistently lies on approximately the same curve of Base Payment versus fee-for-service Multiple.

menu can separate some of these physicians into high fee-for-service rates because productivity and average patient severity are highly dispersed.

Table 2.5: Counterfactual Outcomes by Physician Type

Physicians		Efficient Contracts		Menu of Contracts		
Type	Share	$\Delta E[h(m)]$	$\Delta E[p\,m+b]$	$\Delta E[h(m)]$	$\Delta E[p\,m+b]$	$\Delta E[V(p)]$
$c_L, \alpha_H, \gamma_L, F_L$	0.171	1.311	0.554	0.855	1.505	1.177
$c_H, \alpha_L, \gamma_H, F_L$	0.160	14.165	3.420	2.984	1.545	1.161
$c_H, \alpha_L, \gamma_H, F_H$	0.155	20.003	4.486	5.793	3.140	2.303
$c_L, \alpha_H, \gamma_L, F_H$	0.154	1.441	0.622	1.155	2.549	2.012
$c_L, \alpha_L, \gamma_H, F_L$	0.049	3.860	1.367	1.749	1.963	1.527
$c_H, \alpha_L, \gamma_L, F_H$	0.047	21.554	4.820	4.905	2.010	1.412
$c_L, \alpha_H, \gamma_H, F_H$	0.045	3.201	1.251	2.336	4.003	3.155
$c_H, \alpha_H, \gamma_L, F_L$	0.037	5.670	1.977	1.722	1.567	1.206
$c_L, \alpha_H, \gamma_H, F_L$	0.033	2.172	0.847	1.218	2.031	1.630
$c_H, \alpha_H, \gamma_L, F_H$	0.031	6.620	2.372	2.468	2.529	1.953
$c_H, \alpha_H, \gamma_H, F_L$	0.025	5.997	2.052	1.942	2.064	1.654
$c_H, \alpha_L, \gamma_L, F_L$	0.022	7.306	2.315	2.157	1.366	0.977
$c_H, \alpha_H, \gamma_H, F_H$	0.019	6.240	2.288	3.405	3.974	3.067
$c_L, \alpha_L, \gamma_L, F_L$	0.017	5.755	1.285	6.489	2.202	1.129
$c_L, \alpha_L, \gamma_L, F_H$	0.017	6.694	1.690	10.557	4.990	2.875
$c_L, \alpha_L, \gamma_H, F_H$	0.017	4.359	1.639	3.106	3.880	2.872

Notes: This table shows average outcomes for efficient (personalized) contracts and the optimal menu of contracts, disaggregated across groups of physicians (rows). For physician types, the subscript "H" indicates above-median, and "L" indicates below median. Physician type is a combination of physicians' cost c, altruism α, productivity γ^{-1}, and expected patient severity F. $\Delta E[h(m)]$ represents the change in health production relative to the status quo, for efficient contracts and the optimal menu of contracts. Likewise, $\Delta E[p\,m+b]$ represents incremental expected expenditure and ΔEV represents incremental expected indirect utility. Outcomes are averages across patients within each group, measured in USD.

2.5.2 Robustness

Relaxing restrictions on model assumptions and sample construction suggests that the efficiency of self-selection does not rely on an idiosyncrasy of the empirical approach or setting. First, I find evidence for external validity within Norway: including out-of-sample physicians in counterfactuals does not change the main finding.

The model predicts that physicians would select into the certification sample based on observed and unobserved characteristics. Table 2.1 shows that non-certified physicians have slightly higher treatment intensity which may be explained by relatively old and chronically ill patients. Non-certified physicians are also more likely to be old, born outside of Norway, and use diagnostics. To explore unobserved differences for non-certified physicians, I estimate the distribution of unobserved heterogeneity based on the relatively weak assumption that non-certified physicians have the average log altruism among certified physicians with identical observed characteristics. This assumption is necessary because the identification of altruism requires observing the same physician with different fee-for-service rates. Physicians in the main estimation sample can still be selected on observed heterogeneity in altruism and both observed and unobserved heterogeneity in cost and productivity. Reinforcing this assumption, unobserved heterogeneity in altruism is precise and small relative to the mean (Table 2.3). Likewise, Table B.11 shows that estimates fit observed treatment intensity well for both samples. If non-certified physicians were meaningfully selected on unobserved heterogeneity in altruism, predicted treatment intensity would be a weak predictor of observed treatment intensity. Finally, repeating the counterfactual analysis for the combined population of certified and non-certified physicians results in similar outcomes.[53]

Second, I find evidence for external validity outside of Norway: even large perturbations of estimates rarely change the main finding. In Table B.9, I first perturb cost c, altruism α, and productivity γ^{-1}. A menu increases efficiency when halving or doubling estimates, removing unobserved heterogeneity by replacing estimates with the sample mean for one or two dimensions at a time, or limiting dispersion in estimates by halving variance or dropping outliers. When doubling the variance of cost or altruism, I cannot find efficiency gains from a menu. Although this perturbation

[53]See Table B.9. This specification should be interpreted with caution because estimates are pooled across subsamples. Results are also similar when excluding rural physicians. Anecdotally, rural physicians may face unusual circumstances.

increases the variance of efficient rates, it weakens the correlation between incremental health and physicians' private gain from increased rates and makes efficient rates relatively unaffordable. The variance of severity σ_λ and regulator altruism α_R have large impacts on the levels of counterfactual outcomes. With sufficiently low regulator altruism, a uniform contract is optimal.

Robustness to scaling variance and dropping outliers of physician heterogeneity also suggests that estimation error does not drive the main findings. With overestimated heterogeneity in cost, altruism, and productivity, the gains from self-selection might appear artificially large. Likewise, bootstrapped standard errors are small across aggregate counterfactuals outcomes (Table 2.4), physician-specific contracts (Figure B.11), and the relationship between the reimbursement rate and base payment in the optimal menu (Figure B.12).

Third, descriptive evidence reinforces the exclusion assumption that high-severity patients do not systematically choose particular physicians. The assumption simplifies the analysis by avoiding dynamic considerations. In practice, patients can freely switch between physicians whose enrollment is below the contracted maximum, up to twice per year. As a result, physicians might perceive a link between current treatment intensity decisions and future enrollment, e.g., through reputation effects, which would increase future revenue. Likewise, patients with higher unobserved severity might systematically sort towards physicians with higher expected health production (low cost, high altruism, high productivity). Descriptive evidence suggests that these are not first-order concerns. Figure B.2 shows that enrollment and the share of enrolled patients that are over 60 or chronically ill do not systematically vary with certification, unlike treatment intensity and health production.[54] Enrollment and the share of patients with higher treatment need should increase if patients are sorting towards physicians with greater health production after cer-

[54]Figure B.3 shows that certified and non-certified physicians experience similar trends in enrollment for at least two years after certification.

tification due to increased treatment intensity. Likewise, as shown in Section 2.2.3, physicians' fixed effects in treatment intensity are highly dispersed whether estimated among all patients or only quasi-randomly assigned patients. To test for medium-run sorting, I regress an indicator for switching physicians in the next six months on model-predicted health production, patient covariates, and fixed effects for year and calendar month.[55] Column (2) of Table B.10 shows that the correlation between health and switching to a new physician is imprecise, with point estimates that are small in magnitude. By contrast, expected health production is predictive of (lower) future avoidable hospitalizations and mortality. In Figure B.13, raw means also suggest a decline in avoidable hospitalization after three years. Likewise, cumulative mortality is 36 percent lower than among patients of non-certified physicians.

Fourth, motivated by Ellis and McGuire (1986) and McGuire and Pauly (1991), I test for income effects – nonlinear cost of effort – with a likelihood ratio and cannot reject the baseline model. Income effects can also rationalize why some physicians reduce treatment intensity by a small amount in response to newly registered patients (Barash, 2024) or an increase in reimbursement rates (Figure 2.4). To estimate physicians' marginal disutility of expected workload, I extend the theoretical framework and estimation strategy with additional assumptions, detailed in Appendix B.1.3. If income effects do exist, they seem too small relative to unobserved variation in patient severity to be economically meaningful. Figure B.14 tests the related assumption that physicians do not face binding capacity constraints. Over ten years, the distribution of physicians' monthly treatment intensity varies smoothly near each physician's maximum. Monthly treatment intensity should bunch at high values if some physicians occasionally reach capacity, e.g., due to idiosyncratic variation over time in the number of patients or realized severity. Next, Figure B.15 shows that the treatment intensity of high-altruism and low-altruism physicians is similarly

[55] I use model estimates to calculate expected health production for each patient in the main estimation sample during the six months post-certification. I measure switching 7-12 months after certification.

responsive to the shock of a first avoidable hospitalization. This suggests that estimates of high altruism are not biased by an unobserved constraint. Likewise, the across-time variance of pre-certification workload is similar for low- and high-altruism physicians.[56] I do not find evidence that patients of high-altruism physicians are more likely to seek treatment elsewhere.[57] Finally, as shown in Table B.9, the optimal menu of contracts leads to similar welfare gains over a uniform contract when I impose a capacity constraint and repeat counterfactuals.[58] The constraint limits large expenditure increases on high-severity patients when the health production curve is relatively flat, while high rates still permit large gains for less severe patients. As a caveat, physicians might respond to counterfactual contracts by spending less time on other work, e.g., at nursing homes or universities, that is socially valuable but unmeasured.[59]

Finally, counterfactual outcomes are nearly identical with an alternate health production parameterization from the insurance literature (Cardon and Hendel, 2001; Einav et al., 2013; Marone and Sabety, 2022). Those papers use a quadratic function with a linear term which results in a convenient expression for treatment intensity: $h_0 + h_1(m - \gamma\lambda) - \frac{h_2}{2}(m - \gamma\lambda)^2$. In the baseline approach, I assume $h_1 = 0$ because it is not separately identified from the mean of private marginal cost apart from functional form.[60] To test the alternate parameterization, I re-estimate the model with $h_1 \geq 0$. I focus on non-negative values because previous studies estimate a parameter close to

[56] I aggregate hours for each physician in each month before certification and then calculate the across-month variance. This physician-specific variance does not correlate precisely with estimated altruism. If some physicians are less responsive to certification because of capacity, then low altruism should correlate with low variance. Such physicians would work a similar amount each month (at capacity).

[57] Patients registered with high-altruism physicians receive relatively little primary care from secondary opinions and urgent care centers. If the registered physician was capacity-constrained, patients might seek more treatment from other physicians.

[58] In this case, I bound workload (total simulated hours per physician-month) below the 99th percentile reached after certification.

[59] Table B.9 shows similar gains to a menu when excluding physicians that initially work part-time, i.e., those who spend fewer than 25 hours per week with registered patients.

[60] h_0 is also not identified but does not affect choices. h_2 is absorbed in altruism.

1, and health production should initially increase in treatment. I estimate $h_1 = 0.073$.

2.5.3 Extensions

Even when considering more flexible contract structures, a menu of linear contracts tends to dominate a uniform contract, and the gap between a menu and personalized contracts under full information remains large. First, instead of requiring all physicians to choose from a single menu of contracts ("Baseline"), I derive multiple menus of contracts that incorporate observed heterogeneity. Table B.8 shows that offering ten menus – one for each bin of patients with similar characteristics – does not lead to larger welfare gains.[61] One explanation is that in the baseline counterfactual, each contract includes a multiple of the status quo fee-for-service rate. This approach preserves variation in fee-for-service rates across types of patients who on average consume different bundles of services. Further expanding variation in fee-for-service rates across patients for a given physician may have limited benefits.

Second, Table B.9 shows that counterfactual outcomes are similar to Baseline when separating the analysis between urban and rural patients.[62] The limited benefit of regional contracts is surprising because in the status quo, regional variation is one of the few exceptions to nationally uniform reimbursement.[63] Baseline (national) contracts might perform relatively well because within-region physician heterogeneity is larger than across-region heterogeneity. Consistent with prior literature, Table B.7 shows that health disparities among rural patients remain a pressing concern. Relative to the status quo, eliminating information asymmetry about physicians improves patient health by \$13 for the most rural patients and \$6 for most urban patients. A

[61] I use the same procedure as before, except that each counterfactual fee-for-service rate is a level rather than a multiple of the status quo.

[62] I use the same procedure as Baseline, but separately for rural and urban physicians. Approximately one-fourth of physicians are classified as rural because they practice in a low-centrality municipality.

[63] For example, in Norway, physicians in very small municipalities receive additional payments per registered patient. In the United States, Medicare reimbursement adjusts for rural status, the share of low-income patients, and a local wage index.

national menu of contracts helps narrow the gap in health, but only by a fraction. Similarly, regional menus achieve less than one-third of first-best welfare gains.

Third, relaxing the linear structure of contracts does not increase the welfare achievable with a uniform contract. Although such contracts are rare in healthcare settings and perhaps difficult to implement, larger welfare gains may be possible when revenue is a flexible function of treatment intensity. For example, after a large amount of treatment, the marginal return to health may be small, and low marginal reimbursement would limit relatively inefficient spending. I find that the optimal nonlinear uniform contract substantially improves patient health relative to a menu of contracts. The gain is around half as large as from efficient linear contracts. However, the participation constraint requires large increases in expenditure on base payments, so the net gains to welfare are small: 23 percent of first-best. Without base payments, 56 percent of physicians are worse off, and these losses represent up to 5 percent of status quo revenue. Appendix B.1.5 provides details on deriving the nonlinear contract (extending Gaynor et al., 2023), and compares distributional outcomes across counterfactuals. For each segment along a grid of treatment intensity, the optimal marginal payment maximizes incremental health production net of incremental private costs, among patients with marginal treatment intensity. I find that the optimal nonlinear contract is approximately linear beyond low levels of treatment intensity (Figure B.16), redistributing away from patients with low severity to most other patients with relatively high severity (Figure B.17).

Institutional differences may explain the different impacts of a non-linear uniform contract in this setting relative to Gaynor et al. (2023). With primary care and the large estimated dispersion in unobserved patient severity, there does not seem to be a narrow range of medically appropriate treatment intensity for a non-linear contract to target. Moreover, my estimates imply that marginal health production is nearly universally positive, so decreasing treatment intensity is not generally efficient. In Gaynor et al. (2023), more than half of observed treatment intensity was high enough to damage health because medication dosage exceeded a known cutoff.[64]

In addition to contract flexibility, the regulator might further improve patient health through policies that complement contracts by shifting the allocation of patients across physicians. For example, in Table 2.5, the decomposition across physician and patient types suggests that perhaps high-severity patients should not be registered with high-cost low-altruism physicians. The combination of relatively high severity, high cost, and low altruism partly explains the larger impacts of counterfactual contracts in rural areas (see Table B.7). In reality, physicians decide where to establish a practice and most will move at least once during their career. At these times, contract heterogeneity could incentivize different location choices. For example, it might induce better match quality to increase the base payment for a high-fee-for-service contract in areas where nearby patients have relatively high observed severity. Alternatively, the regulator could incentivize patients to switch to under-subscribed high-quality physicians. While this question is beyond the scope of the current work, it may be a fruitful path for future research: combining efficient reimbursement rates with optimal patient switches can increase incremental social surplus by 17 percent relative to efficient reimbursement rates alone.[65]

So far, while fixing the distribution of physicians, information asymmetry remains costly even after adding contract flexibility – perhaps the regulator could further improve patient health through complementary long-run investments that alter the distribution of physician heterogeneity. At reasonable reimbursement rates, the regulator prefers a physician with low cost of effort, high altruism, and high productivity. Public subsidies for support staff or telehealth might lower cost of effort;

[64]These characterizations mostly refer to Figure 3 in that paper, which is based on a patient with median observed severity.

[65]This exercise involves a stylized example of two vertically differentiated physicians at the 10^{th} and 90^{th} percentile of (initial) efficient fee-for-service rates. I begin by counterfactually assigning both physicians the average patient distribution, corresponding fee-for-service rates, and average enrollment. I alternate between searching for first-best contracts and looping through the maximally profitable patient switch for a given set of contracts. This method maintains the initial number of patients per physician and converges after 53 percent of patients have switched.

performance benchmarks might increase altruism, and promoting long-term patient-physician relationships might increase productivity via soft knowledge. For example, performance benchmarks can increase information and facilitate learning, particularly about past patients that have since left the list. Physicians currently do not observe all the long-term impacts of treatment, like utilization and avoidable hospitalizations.

2.6 Conclusion

In this paper I present a framework for deriving the optimal menu of physician reimbursement contracts. The framework incorporates unobserved patient illness severity and physicians' endogenous choices of contract and treatment intensity. I characterize the conditions on multidimensional physician heterogeneity under which self-selection among a menu of contracts is more efficient than a uniform reimbursement contract. These conditions are met in the empirical example of Norwegian primary care physicians. I estimate the distributions of physician and patient heterogeneity, exploiting the sudden large variation in marginal reimbursement when physicians become certified as general practitioners. I find large efficiency gains from introducing self-selection, and that finding is robust to several model enrichments, estimate perturbations, and alternative samples.

The most direct policy implication is that the Norwegian National Insurance Scheme could cost-effectively improve access to primary care by offering a menu of 2-7 linear contracts. These contracts are easy to understand because they have the same linear structure as status quo reimbursement. The difference is that each contract exchanges a higher multiple on service-level reimbursement for lower revenue per registered patient-month. The regulator could use its existing data and infrastructure to administrate the policy counterfactual as an occasional settlement payment. Moreover, the menu of contracts is efficient even as a voluntary reform: physicians can still choose the status quo contract, which might make it acceptable to the association that negotiates reimbursement on behalf of physicians. I also find evidence that is consistent with reductions in hospitalization and mortality. By contrast, economic theory and empirical evidence alike predict that Norway's recent initiative to increase base payments for relatively ill patients will not immediately affect treatment intensity, because marginal incentives are unchanged.[66]

Beyond Norway, my framework for evaluating the efficiency of self-selection is broadly applicable to settings featuring heterogeneous altruistic agents that experience panel variation in marginal reimbursement. In healthcare, this includes systems in which many physicians derive most revenue from contracts with a single payer, e.g., several countries' health agencies or Kaiser Permanente in the United States. External validity might be limited in settings where prices are negotiated or patients frequently switch physicians based on reputations for treatment intensity. Outside of healthcare, menu design may be an effective tool in the markets for indigent defense attorneys, K-12 educators, and social workers. These agents are likely altruistic – sacrificing some profit to improve outcomes for their clients and students – and also heterogeneous in marginal cost and productivity. The frequent lack of compensation for incremental effort may also contribute to capacity constraints and disparities in outcomes. My framework uses reimbursement variation, which often exists in these settings, but even cross-sectional data can be sufficient with additional assumptions.[67]

Why are uniform contracts ubiquitous if the potential gains from self-selection are large? First, variation in incentives across physicians may conflict with norms concerning uniformity. Moreover, before considering multidimensional unobserved physician heterogeneity, policymakers may not find it intuitive that increasing dispersion in patients' treatment could be efficient. Second, there may be fixed costs of

[66]On the other hand, such a reform may effectively deter exit in the long term. With sufficient exit, capacity constraints may bind and reduce treatment intensity.

[67]For example, a simulation-based estimator could recover a parametric distribution of altruism with cross-section variation in reimbursement under a stronger exclusion assumption. Client severity must be conditionally independent of agent type and reimbursement, which is unlikely if, e.g., high-quality agents receive higher reimbursement. See Lee (2021), Biasi (2021), or Hanushek et al. (2023) for reimbursement variation among attorneys and teachers.

introducing counterfactual menus, e.g., costly experiments in reimbursement variation to derive the optimal menu or incremental costs of negotiation with a physicians' union.

I also explore related applications of the model that may be productive directions for future research. First, several studies decompose dispersion in healthcare utilization between broadly supply-side or demand-side factors. I begin to further decompose supply-side factors by simulating dispersion in treatment intensity with counterfactual distributions and characteristics. Consistent with existing evidence that patients imperfectly perceive physician quality, I find evidence that patients may not be optimally allocated across physicians to maximize cost-effective health production. Future work might consider self-selection in the context of physician entry, incorporating reimbursement contracts as well as the number and composition of nearby patients. I also do not find evidence of income effects or capacity constraints in Norway, but these features may add nuance to contracting in related settings.

Chapter 3: Why Don't Graduation Incentives Work? Match Quality and Financial Aid Design

Abstract

Recent policy changes limit the scope for university admission decisions to equitably ration spots. I investigate whether selective universities can instead use graduation-contingent loan forgiveness to allocate spots to the students who most benefit from attendance. Identifying variation comes from an existing loan forgiveness program that incentivized greater on-time graduation. Participants' relatively high graduation rates appear to be driven by selection on ability into loan take-up rather than program effects. Exploiting a discontinuity from Pell Grant eligibility, I find no detectable effect of loan take-up on course load, course completion, part-time work, on-time graduation, or earnings. I also document how selective universities increase graduation more for some students than others. I incorporate selection on unobserved ability into a structural model of students' college choice, loan forgiveness take-up, and graduation. Using model estimates, I show how a counterfactual loan forgiveness program could shift college choice, leading to greater welfare, statewide graduation, and demographic equity.

3.1 Introduction

Some universities are better than others at improving students' long-run outcomes (Dale and Krueger, 2002, 2014; Cunha et al., 2017; Chetty et al., 2022). Limited capacity at these high-quality universities presents an allocation problem. In settings with decentralized admissions like the United States, this challenge is compounded by private information. Some students may benefit more than others from attending a particular university ("high match quality"), but the university cannot fully observe students' preferences and ability. Universities might use demographics as a signal of match quality, but recent policy changes restrict this practice.[1] When screened out of high-quality institutions, students take on debt to attend universities where they may struggle to graduate and experience a meaningful boost in earnings (Looney and Yannelis, 2024). The costs of such inefficiency may be large: much of the $36.5 billion of annual federal spending subsidizes the enrollment of students who do not ultimately graduate (Board, 2019).

Universities screen prospective students through both admissions and financial aid, but financial aid is far less studied as a tool to improve equity and efficiency. This paper sheds light on how financial aid can improve match quality between students and colleges. I exploit variation from a novel program in Texas that incentivized on-time graduation. Conditional on a rich set of observed characteristics, participants in this program graduated on time at much higher rates. However, this apparent effect is not causal if observed characteristics do not fully capture graduation chances. With a difference-in-discontinuity design, I find no evidence that the program improved on-time graduation or proxies for short-term effort. As a result, program participation helps identify the distribution of students' expectations about graduation. To explore improved design of graduation incentives, I estimate a model of students' college choice, incentive take-up, and graduation outcomes. I find that counterfactual graduation incentives could substantially improve welfare and equity by influencing students' choice of university.[2]

[1] For example, the Supreme Court case *Students for Fair Admissions v. Harvard* prohibited universities from considering race when determining admission.

[2] By contrast, the incentive program as implemented had no measurable impact on ex-ante welfare. I use a social objective which includes students' welfare, universities' priority over student demographics, universities' profit, and a social externality from student graduation.

I use data from Texas public universities, a setting particularly well-suited to studying unobserved variation in graduation likelihood. First, the B-On-Time ("BOT") program attracted students who expected to graduate. The program offered large interest-free loans and forgave debt if students completed their degree in four years with a 3.0 GPA. Second, I can measure how students with diverse backgrounds and abilities respond to financial aid. All Texas students in the top ranks of their high school by GPA are guaranteed admission to public universities. Third, comprehensive administrative records allow me to precisely measure student ability and the returns to selective universities. For the universe of Texas students, from kindergarten through early careers, I observe demographics, educational achievement, employment, and earnings.

The key identifying variation comes from BOT loans, which lowered the cost of attending college for participating students. BOT was established through an act of the state legislature in 2003 to increase timely degree attainment. Funded by a statewide tuition set-aside, the program offered loans of up to $8000 per year to students at 4-year public universities. Virtually all Texas residents were eligible, regardless of ability, need, or program of study. BOT loans were converted to grants for students who graduated on time. Students who did not graduate on time could still benefit by using the program's zero-interest loans instead of high-interest private debt. Despite large program incentives, only 8 percent of eligible students opted to participate. Overall, BOT participants were twice as likely to graduate on time than non-participants. Conditional on observed characteristics like demographics, financial need, and proxies of ability, BOT recipients were 9 percentage points more likely to graduate on time, over a mean of 25 percent.

Higher graduation among BOT recipients appears to entirely reflect selection rather than reduced moral hazard. BOT participants have an incentive to increase effort towards academic progress so their loans will be forgiven.[3] To identify the

[3]Some students may not have positive returns to effort.

causal effect of BOT, I leverage the discontinuity in Pell Grant eligibility. Pell Grants are the largest federal aid program and discontinuously phase out with family income. Several other institutional and private financial aid programs use the same thresholds, so relative to marginally lower-income students, Pell-ineligible students typically have a $2000 greater net cost of attendance. Since net cost varies at the cutoff with or without BOT, I apply a difference-in-RD design, comparing students just above versus just below the cutoff, in cohorts with access to BOT versus cohorts without access. Combined with BOT availability, the Pell cutoff doubles the probability of BOT participation. Despite BOT participants' high graduation rates, I do not find evidence that BOT causally affects 4-year graduation or proxies for effort like course load, part-time work, and persistence. The null effect for on-time graduation is similar across several alternative specifications and dimensions of heterogeneity. The lack of evidence suggests that, despite greater incentives to graduate on time, the marginal BOT participant cannot exert more effort towards academic progress. Instead, I find suggestive evidence that BOT distorts choices across fields of study, leading to less difficult majors, i.e., those with higher mean graduation rates. Although BOT increases eventual graduation (within six years), it fails to increase early-career earnings.

The null effects of BOT participation suggest that students at the same university have different graduation chances, but do graduation chances also vary across universities for a given student? I document that enrollment at the two most selective public universities (the "flagships") increases 4-year graduation by 7 percentage points, 6-year graduation by 10 percentage points, and early career annual earnings by $2510. Moreover, point estimates vary across subsamples by family education, family income, GPA, and high school quality. By controlling for the set of universities that admit a student, I capture the admission offices' collective information about student ability (as in Dale and Krueger, 2002; Mountjoy and Hickman, 2021). This heterogeneity builds on prior evidence that the returns to selective universities vary within university by students' characteristics like income, race, gender, and high school (for a recent review, see Lovenheim and Smith, 2023).

BOT loans may have failed to increase graduation because the program targeted students' effort rather than students' choice of university. For example, BOT was voluntary and available at all universities, so it did not shift the *difference* in net cost between universities. One alternative is to implement mandatory graduation-contingent loan forgiveness, but only at selective universities. This alternative could help selective universities attract students with high chances of graduation by lowering these students' effective net cost of attendance. Replacing low-graduation students at selective universities with high-graduation students has ambiguous welfare effects, depending on the joint distribution of students' tastes, graduation at selective universities, and graduation at other universities. For example, a student with high graduation at a selective university might have similarly high graduation at every university.

To explore how counterfactual graduation incentives can improve the selection of students across universities, I present a model of student's college choice, incentive take-up, and graduation. Students first choose a university, trading off the net cost of attendance, graduation likelihood, and distance.[4] Next, students apply for loan forgiveness if the expected savings outweigh an idiosyncratic application cost. Finally, each student graduates according to a stochastic process. Anticipating student choices, each university designs a net cost schedule. At this stage, universities only observe students' financial need and maximize a weighted sum of profit and idiosyncratic preferences over student composition. A regulator designs subsidies and loan forgiveness to maximize expected consumer surplus plus the net social value of graduation, subject to budget, capacity, and incentive compatibility constraints. Match quality is the contribution of a student-university match to this social objective. Students may not fully value the benefits of graduation, e.g., additional tax revenue on

[4] Students at the same university face different net costs of attendance, graduation likelihoods, and distances from their high school.

higher earnings.

Using estimates of the model parameters, I document substantial variation in student's match quality with flagship universities, so full-information financial aid offers could lead to large welfare gains. Match quality varies due to both student preferences and idiosyncratic effects of flagship enrollment on graduation ("value-added").[5] Both types of heterogeneity are only partially explained by the student characteristics that universities observe. With full information, a regulator could offer each student a personalized financial aid offer that incorporates both observed characteristics and graduation chances at each university.[6] These first-best financial aid offers increase welfare by $27k per student without lowering revenue or increasing flagship enrollment.

The large welfare gain from full information is driven by three factors. First, universities' private objectives do not necessarily coincide with the social objective. Universities deviate from profit maximization most for middle-income students whose graduation chances do not depend as much on attending a selective university. Second, due to status quo institutional constraints – namely, staffing and state legislation – universities only consider financial need when allocating most financial aid. By considering additional observed characteristics, a regulator could increase welfare by 80 percent of the first-best schedule, even without graduation incentives. Third, in the status quo, BOT offers identical graduation incentives at every university.

For comparison, I find that BOT did not affect student-university match quality. Most students perceived BOT application costs as too high for loan forgiveness

[5] Relative to regression estimates in prior studies of heterogeneous college quality, these graduation effects rely on weaker assumptions about how students choose colleges (see, e.g., Dale and Krueger, 2002; Andrews et al., 2016; Mountjoy and Hickman, 2021).

[6] Counterfactuals vary the cost of attending a flagship university, holding other prices fixed. For every counterfactual, universities' producer surplus increases by $400-800 per student, consistent with institutional frictions and incentive compatibility. Full information does not include information about student's idiosyncratic preferences.

to impact college choice.[7] Without effects on effort or college choice, BOT could only impact welfare as an ex-post transfer to students with relatively high ex-ante graduation. BOT's funding might be deployed more efficiently by lowering the net cost of flagship attendance for a broad set of students. Its limited efficacy might explain why the program was discontinued in 2016.

On the other hand, counterfactual loan forgiveness can improve the selection of students into flagship universities by increasing the price *difference* between flagships and alternatives for targeted students. I first combine screening on observed characteristics with a universal loan forgiveness policy comparable in size to BOT. This combined policy increases the welfare gain from 80 percent to 92 percent of first-best financial aid offers. Counterintuitively, screening financial aid based on the full set of observed characteristics decreases graduation by 3.5 percentage points while increasing welfare. Increasing state-wide graduation requires allocating capacity at a selective university to students with the largest graduation value-added. However, student characteristics that increase value-added often also increase price sensitivity or weaken tastes for graduation. In these cases, it may be efficient to offer a low net cost at a less productive university. If the regulator cannot target financial aid based on student characteristics, universal graduation incentives alone increase welfare by 25 percent of the first-best schedule.

Counterfactual financial aid schedules also tend to increase flagship enrollment for students from historically underrepresented backgrounds. Screening with graduation incentives increases the flagships' share of students of color, female students, first-generation students, and students from high schools with historically low college graduation. Consistent with cross-subsidization, average family income becomes nearly identical between flagship and non-flagship universities.

This study contributes most to the literature on price discrimination in higher

[7] In the model, when students choose a university, BOT lowers net cost if the expected forgiven amount outweighs the expected application cost.

education (Waldfogel, 2015; Epple et al., 2017, 2019, 2021). These studies focus primarily on competition between colleges and inferring the colleges' objective functions from financial aid offers with third-degree price discrimination. Empirical results typically suggest that colleges target student composition, offering relative discounts to high-ability and under-represented minority students. This paper emphasizes the opposite question: how can second and third-degree price discrimination shift the allocation of students across colleges to maximize a social objective? This market design emphasis is most closely related to Fillmore (2022), which considers information disclosure through FAFSA and its role in price discrimination.

A literature in public finance derives optimal income-based financial aid schedules, focusing on the extensive margin of college enrollment (Lans Bovenberg and Jacobs, 2005; Lawson, 2017; Colas et al., 2021). I show how the intensive margin of college match quality can also generate fiscal externalities and affect the optimal schedule.[8] To do this, I adapt intuition from the study of insurance. In selection markets, it may be efficient to price-discriminate on both ex-ante (e.g., risk scores or test scores) and ex-post characteristics (e.g., filing a claim or graduation). I find that the effect of screening on both ex-ante and ex-post characteristics is less than the sum of its parts.

The null reduced-form effects of status quo graduation incentives are somewhat unusual relative to prior literature.[9] Several papers use variation from eligibility discontinuities and typically find that grant aid affects student outcomes (van der Klaauw, 2002; Evans and Nguyen, 2019; Denning et al., 2019).[10] Related graduation incentives often improve outcomes, e.g., universal state grants in Norway (Gunnes

[8]By characterizing an optimal aid schedule, I also build on related studies that decompose the welfare and composition effects of targeted financial aid (e.g., Dobbin et al., 2022).

[9]There are some exceptions to the positive effects of grant aid. Rubin (2011) finds no effect on initial enrollment at the Pell eligibility threshold. Likewise, Rattini (2023) finds that aid generosity in Italy lowers attempted credit hours and increases time-to-graduation.

[10]See Nguyen et al. (2019) for a review. Not all studies use discontinuity evidence. For example, Murphy and Wyness (2022) finds that unexpected means-tested aid increases persistence, test scores, and graduation with honors.

et al., 2013), need-based privately-funded grants in Wisconsin (Goldrick-Rab et al., 2016), penalties in Germany (Mathias et al., 2006), and short-term incentives in the Netherlands (Leuven et al., 2010), but not loan tax rebates in Finland (Hämäläinen et al., 2016) or relaxing exam requirements (Malacrino et al., 2024). Overall, this literature suggests that increasing the cost of late college completion sometimes incentivizes more effort during the early years of a program.

Going forward, Section 3.2 summarizes the process of admissions and financial aid. Section 3.3 uses complementary reduced-form designs to evaluate the effect of BOT loans. Section 3.4 presents a theoretical model of students' college choice, loan take-up, and graduation, using it to interpret the effects of BOT. Section 3.5 describes the dataset and empirical model used to estimate structural parameters. This section also tests the underlying assumption that flagship enrollment benefits some students more than others Section 3.6 characterizes parameter estimates. Section 3.7 shows how counterfactual financial aid offers affect welfare, graduation, and equity effects. Section 3.8 concludes.

3.2 Background

In the U.S., university admissions and financial aid are largely decentralized. Before students apply, each university sets the tuition and fees that compose the list price. Universities may also determine capacity, the targeted number of enrolled students in the subsequent year. At public universities, list price is lower for students from the same state. List price may vary across majors within a university, but variation in list price across universities is relatively large. Students can use online calculators to estimate financial aid for each program before applying. Applications for admission generally involve submitting essays, an intended major, and measures of academic performance like high school transcripts or test scores. Universities' admission decisions balance academic preparedness, student interest, and in a broad sense, diversity of student composition. Anticipating that not all new students will ultimately enroll, universities admit more students than capacity.

After admitting students, universities determine financial aid largely based on financial need. Historically, financial need is measured relative to the expected family contribution ("EFC"), a function of family income and wealth. To determine their EFC, students complete the centralized Free Application for Federal Student Aid. Students may also immediately qualify for merit-based aid, e.g., the National Merit Scholarship or Texas Grant, but need-based aid tends to dominate. Students' initial net cost of enrollment for each program is the difference between a list price and financial aid from federal, state, and institution sources. The final net cost can vary if students apply to special programs like private scholarships. After students choose which program to enroll in, they pay their final net cost through a combination of family resources and loans.

In Texas, state regulations constrain universities' admission and financial aid decisions. Public universities must automatically admit students in the top 10 percent of their high school cohort. Since 2018, the University of Texas at Austin must admit 75 percent of students through this rule, so the effective threshold lowered to approximately 6 percent. Since 2003, Texas public universities could raise tuition above a previously enforced upper bound. However, a fixed share of incremental tuition revenue must subsidize the tuition of low-income students.

Also in 2003, Texas introduced a state loan program that provides identifying variation. The program, named B-On-Time ("BOT"), aimed to increase on-time graduation through financial incentives. BOT consisted of zero-interest loans, and the state forgave debt if students graduated within four years of first enrollment. As part of this requirement, students needed to primarily enroll in courses required for their major. All in-state students attending a public university could apply for BOT to supplement their initial financial aid offer. Unlike most state programs, BOT did not require high financial need.

In other settings, graduation incentive programs often consist of a transfer after graduation. The effects of such programs may differ with liquidity constraints, hyperbolic discounting, or loss aversion. Also, over four years, BOT loans can add up to as much as \$32,000, which may be relatively salient. BOT is also quite different than recent federal policies forgiving loans of students long after graduation. Students' choices of university enrollment might change when they anticipate loan forgiveness – both the initial enrollment and how long to work towards graduation.

3.3 Effects of B-On-Time Loans

B-On-Time Loans increased incentives for on-time graduation, but few students participated. This section explores whether such incentives affected student behavior. Participants were much more likely to graduate on time than non-participants with identical characteristics. However, evidence from a discontinuity in financial need suggests that the program did not increase on-time graduation for marginal participants. Together, these facts imply that students selected into the program based on unobserved heterogeneity in predicted graduation.

3.3.1 Econometric Models

I apply two complementary research designs to disentangle the treatment effect of graduation incentives from Roy-style selection. The first estimating equation assumes selection on observables:

$$Y_{ijt} = \beta_1 \mathrm{BOT}_{ijt} + \beta_2 \mathbf{X_i} + \gamma_{jt} + \epsilon_{ijt}\,, \tag{3.1}$$

where Y_{ijt} is an outcome of interest for student i, with a financial aid application at school j in fiscal year t. β_1 measures the effect of BOT take-up on Y under a strict assumption: conditional on students covariates $\mathbf{X_i}$, BOT take-up is uncorrelated with unobserved determinants of the outcome. This assumption may be violated, e.g., in a regression with on-time graduation as the outcome. Students with high unobserved ability may be more likely to graduate on-time, while implies greater expected loan forgiveness, leading to greater selection into BOT. In this case, the coefficient of interest would be biased upward.

The second estimating equation applies a difference-in-discontinuity design:

$$\begin{aligned} Y_{ijt} = {} & \beta_1 \text{Above}_i + \beta_2 \text{Avail}_{jt} \text{Above}_i + \beta_3 f(\text{EFC}_i) + \beta_4 \text{Above}_i f(\text{EFC}_i) \\ & + \beta_5 \text{Avail}_{jt} f(\text{EFC}_i) + \beta_6 \text{Avail}_{jt} f(\text{EFC}_i) \text{Above}_i + \epsilon_{ijt} \,, \end{aligned} \tag{3.2}$$

where Y_{ijt} is an outcome of interest for student i, with a financial aid application at school j in fiscal year t. EFC_i is expected family contribution relative to the cutoff, the running variable. Above_i indicates Pell ineligibility. Avail_{jt} indicates years when BOT was available. Some alternative specifications include student-level covariates $\mathbf{X_i}$ and fixed effects γ_{jt}.

The coefficient β_1 represents the discontinuity in the outcome in years when BOT was unavailable. It should not necessarily be interpreted as the effect of Pell eligibility, because potential outcomes may not be continuous at the Pell eligibility cutoff. Students in the sample have already applied and been accepted by universities. If Pell-ineligible students face systematically different costs of attendance or choice sets ex-ante, those in the sample may have greater unobserved tastes or resources for attending college. For example, if students with greater unobserved resources are more likely to graduate, then β_1 would be biased upwards.

If the discontinuity in potential outcomes is time-invariant, then the coefficient of interest β_2 is an unbiased estimate of the incremental discontinuity in the outcome in years during which BOT was available ("BOT years") relative to when it was unavailable. Equivalently, other determinants of the outcome must not systematically vary at the cutoff in BOT years. I consider two potential violations of this assumption. First, BOT may directly crowd out other sources of financial aid. If students' enrollment choices depend on BOT, universities could offer less grant aid and achieve similar enrollment patterns. In this case, if non-BOT aid improves outcomes, β_2 would be biased down. Second, the Pell cutoff increases over time, and

BOT years are on average later than non-BOT years, so if potential outcomes vary in the level of EFC, then β_2 may reflect that difference. I cluster standard errors at the level of university-year.

Comparing estimates across research designs is suggestive of selection into BOT on unobserved graduation likelihood. For on-time graduation, the most credible estimate of BOT's effect is the local average treatment effect from Equation (3.2), calculated by dividing the reduced-form estimate by the first-stage estimate. In the reduced form, on-time graduation is the outcome, and in the first stage, BOT take-up is the outcome. The average treatment on the treated from Equation (3.1) should be biased upwards relative to the LATE if there is substantial selection into BOT on unobserved graduation likelihood. To address the possibility that potential outcomes vary with EFC, I also estimate Equation (3.1) among students within a narrow bandwidth of the Pell eligibility threshold.

3.3.2 Data

To estimate the effect of B-On-Time Loans on student outcomes, I focus on Texas high school graduates who submitted a financial aid application to a Texas public university between 2001 and 2017. Before 2017, I can measure 4-year graduation, the primary outcome of interest. Universities must be present in all years and each have at least one student with BOT loans during the sample period. The discontinuity design uses family resources as the running variable, so I remove the mass point of students who have an EFC of $0. The Pell eligibility threshold is different for students with unusually high gross cost of attendance, so I remove these students too. To reduce measurement error, students in the sample must eventually enroll in a public 4-year university.

Figure 3.1 plots the discontinuities in raw mean outcomes at the Pell eligibility threshold. In both sets of years, federal aid drops off discontinuously at the threshold and generates unmet financial need (Panel A).[11] Pell-ineligible students are approximately twice as likely as Pell-eligible students to take up BOT loans when BOT is available (Panel B). BOT appears to substitute for traditional loans, which vary smoothly at the cutoff when BOT is available, but increase discontinuously in other years (Panel C). Unlike other outcomes, on-time graduation varies smoothly at the cutoff for all years (Panel D). Figures C.0.1 and C.0.2 plot additional outcomes.

3.3.3 Results

Graduation incentives are associated with substantially higher graduation rates. First, Panel A of Table 3.1 shows estimates for on-time graduation. Regardless of specification or sample, BOT participants have 12 percentage points higher on-time graduation. Column (1) shows a raw correlation among students that apply for financial aid. Relative to this sample's mean graduation of 26 percent, the estimate of 12 percent is almost implausibly large. To ease comparisons across research designs, Columns (2)-(3) focus on students near the Pell eligibility threshold at bandwidths of \$5000 and \$1000, for which mean graduation is 41 percent. These specifications also add controls for students' initial covariates and fixed effects for year, major, high school, and admission set. The coefficient on BOT participation is nearly identical across specifications. If unobserved selection into BOT varied by financial need, then these point estimates would likely also vary.

Estimates from the difference-in-discontinuity design suggest that BOT loans did not impact on-time graduation for marginal students. If the selection-on-observables estimate reflected a treatment effect, then on-time graduation should increase discontinuously at the Pell ineligibility threshold for BOT-eligible students relative to BOT-ineligible students. Columns (4)-(6) of Table 3.1 show point estimates are nearly zero. Although imprecise, confidence intervals exclude large reduced-form increases in graduation. The local average treatment effect of BOT – which nor-

[11] Federal aid mostly consists of Pell grants.

malizes these estimates by the first-stage estimates – is orders of magnitude smaller than the selection-on-observables estimate. The difference is consistent with substantial selection on unobserved gains into BOT loans. Importantly, BOT take-up is a relevant instrument. The first-stage estimate is large and precise: Pell ineligibility increases BOT take-up by 3 percentage points over a mean of 10 percent (Panel C of Table 3.1). This estimate is similar across specifications that add control variables or exclude observations near the cutoff.

Estimates for 6-year graduation are precise, closing the gap generated by Pell ineligibility in years when BOT is unavailable. Anticipating loan forgiveness, students may make incremental degree progress during the first four years of a program. The type of student making such progress may not be on the margin of on-time graduation. After four years of study, students may perceive a lower cost of finishing their degrees despite needing to repay BOT loans.

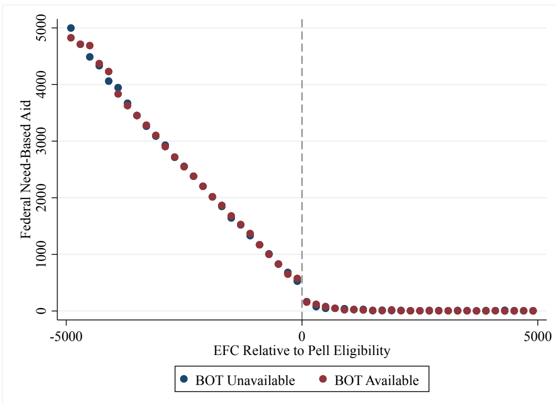
Similarly, Table 3.2 shows that although BOT impacts financial aid, there is limited evidence of reduced-form effects on students' program choice, effort, or long-run outcomes. If flagship universities increase graduation rates, then graduation incentives lower the difference in expected net cost between flagships and alternatives. Price-sensitive students might therefore respond to BOT by increasing flagship enrollment, but I do not detect an effect. Once at a university, BOT participants might choose less effort-intensive fields of study to that they can graduate on time. I find that BOT leads to choices of major with higher mean graduation. The point estimate for majors' mean earnings is positive but imprecise. Once committed to a program, BOT participants might exert more effort toward degree progress if such effort is effective. Students might register for summer courses or refocus time towards studying, which in turn should contribute to degree progress. I do not find evidence that BOT increases attempted credit hours, decreases part-time work, or increases degree progress. Other studies find that students respond to grant aid by reducing part-time work. The lack of a precise effect here might suggest that the reduction in loans and net cost was not large enough to relax students' budget constraints. Since

BOT does not seem to increase effort or on-time graduation, it is surprising that BOT increases 6-year graduation. However, there is no effect on early-career earnings. Typically, college graduation carries some signaling value that corresponds to an earnings increase. Perhaps, the 6-year effect is an artifact of choosing less difficult majors while not exerting incremental effort.

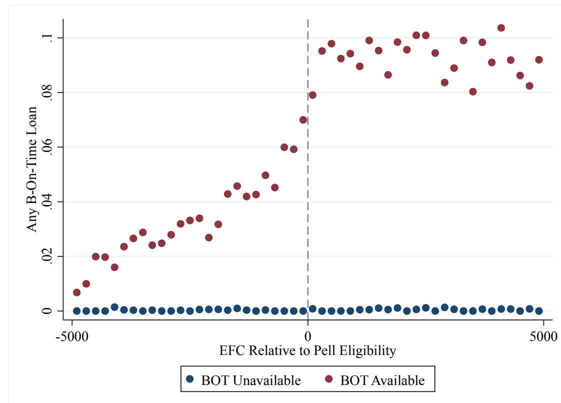
Robustness specifications validate the interpretation of the DRD coefficients: BOT does not affect graduation. First, Table 3.1, Column (5) shows that estimates are similar without controls for students covariates. This helps address concerns about sample selection at the Pell eligiblity threshold. Second, Column (6) shows that estimates are similar when excluding observations near the threshold. Third, estimates are nearly identical to Column (5) when I do not control for students' net cost of attendance. The baseline specification controls for net cost because Table 3.2 shows a \$8900 lower net cost for marginal students. If net cost directly affects graduation, that effect appears to be uncorrelated with the direct effect of BOT.

Fourth, I find limited evidence that BOT directly crowds out other sources of financial aid. Universities would need to anticipate BOT receipt based on student characteristics because students apply for BOT after receiving an initial financial aid offer. Table 3.2 shows no evidence for incrementally higher federal aid from Pell ineligibility in BOT years. The point estimate is approximately zero and the confidence interval rules out large changes. Similarly, point estimates are small and imprecise for merit and non-discretionary aid. Small increases in the gross cost of attendance and discretionary aid are precise, but point estimates add to approximately zero, so these margins do not explain the effect on net cost. Fifth, I do not find evidence that the effect is confounded by increases in the Pell eligibility threshold over time. In alternative specifications, interactions with the cutoff level or share of students receiving BOT are imprecise.

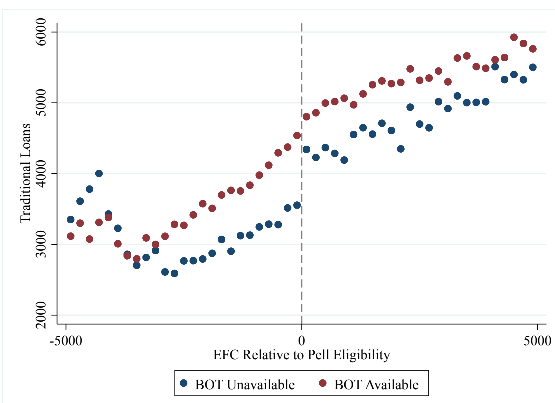
Figure 3.1: Pell Eligibility and BOT Loans



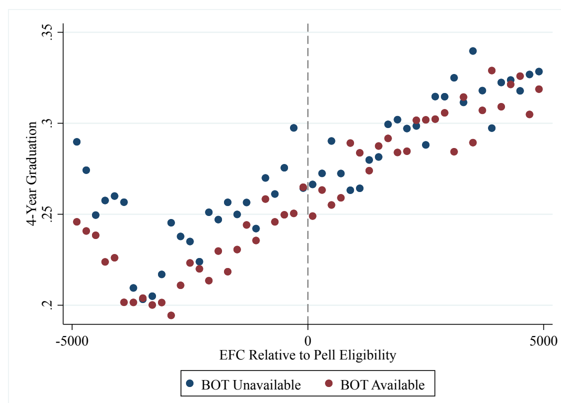
(a) Federal Need-Based Aid



(b) Any B-On-Time Loans



(c) Traditional Loans



(d) Graduation Within 4 Years

Notes: The x-axis shows expected family contribution (EFC) minus the maximum EFC for Pell Grant eligibility, which varies by year. Points represent averages of students within $500 bins of EFC. BOT Available includes students with initial financial aid offers in years when BOT loans were widely available.

Table 3.1: B-On-Time Loans Take-Up and Graduation

	(1)	(2)	(3)	(4)	(5)	(6)
			(a) Four-Year Graduation			
Any B-On-Time Loans	0.120	0.119	0.112			
	(0.005)	(0.010)	(0.023)			
	[<0.001]	[<0.001]	[<0.001]			
Pell Ineligible × BOT Available				−0.000	−0.001	0.010
				(0.011)	(0.011)	(0.012)
				[0.996]	[0.932]	[0.439]
Outcome Mean	0.255	0.408	0.407	0.243	0.255	0.254
R^2	0.002	0.176	0.304	0.089	0.074	0.075
Observations	284,039	54,247	9,792	256,979	284,039	272,689
			(b) Six-Year Graduation			
Any B-On-Time Loans	0.135	0.047	0.039			
	(0.005)	(0.008)	(0.017)			
	[<0.001]	[<0.001]	[0.023]			
Pell Ineligible × BOT Available				0.032	0.027	0.035
				(0.014)	(0.013)	(0.016)
				[0.027]	[0.048]	[0.034]
Outcome Mean	0.515	0.828	0.821	0.513	0.515	0.514
R^2	0.003	0.157	0.304	0.119	0.094	0.095
Observations	253,267	46,894	8,629	227,552	253,267	242,962
			(c) Any B-On-Time Loans			
Pell Ineligible × BOT Available				0.030	0.027	0.036
				(0.009)	(0.008)	(0.010)
				[<0.001]	[<0.001]	[<0.001]
Outcome Mean				0.036	0.033	0.033
R^2				0.102	0.097	0.097
Observations	242,962	242,962	242,962	256,984	284,046	272,695

Notes: This table shows regression estimates relating B-On-Time Loans and 4-year graduation from the Selection On Observables design – Columns (1)-(3) from Equation (3.1) – and the Difference-in-Discontinuity Design – Columns (4)-(6) from Equation (3.2). Column (1) is a raw correlation. Column (2) controls for students' initial covariates and fixed effects for year, major, high school, and admission set, among students with an EFC within \$5000 of the Pell eligibility threshold. Column (3) restricts to a bandwidth of \$1000. Column (4) is a local quadratic difference-in-discontinuity with controls for students' initial covariates, final net cost of attendance, and fixed effects for the financial aid year, among students with an EFC within \$5000 of the Pell eligibility threshold. Column (5) only controls for the net cost and year. Column (6) is identical to (4) but excludes observations within \$200 of the cutoff. Panel names refer to the outcome variable. Panel C represents the first stage effect, so dividing by this coefficient produces the local average treatment effect of Any B-On-Time Loans on Graduation. Although BOT only rewards 4-year graduation, it also predicts overall degree completion. Panel B shows results for 6-year graduation. With controls, the raw correlation of 13 percentage points attenuates to 5 percentage points, over a mean of 83 percent.

Table 3.2: Difference-in-Discontinuity: Pell Ineligible × BOT Available

	Estimate	Std. Err.	P-Value	Outcome Mean	R^2	Obs.
Financial Aid						
Federal Aid	0.006	(0.021)	[0.789]	1.661	0.837	256,984
Any B-On-Time Loans	0.030	(0.009)	[<0.001]	0.036	0.102	256,984
Traditional Loans	−0.418	(0.123)	[<0.001]	3.987	0.231	256,984
Net Cost of Attendance	−0.889	(0.196)	[<0.001]	12.115	0.609	256,984
Program Choice						
Enrolled at a Flagship	−0.001	(0.002)	[0.617]	0.169	0.933	256,984
E[Graduation | Major]	0.011	(0.005)	[0.053]	0.288	0.104	256,984
Effect of Major on Earnings	0.029	(0.144)	[0.839]	−0.248	0.043	256,981
Effort Proxies						
Credit Hours Attempted	1.052	(0.974)	[0.281]	68.077	0.587	256,984
Contemporary Employment	0.003	(0.010)	[0.770]	0.759	0.039	256,984
Max Class Rank	0.031	(0.025)	[0.227]	2.301	0.671	256,796
Outcomes						
Graduation Within 4 Years	−0.000	(0.011)	[0.996]	0.243	0.089	256,979
Graduation Within 6 Years	0.032	(0.014)	[0.027]	0.513	0.119	227,552
Earnings After 8-10 Years	1.119	(0.950)	[0.240]	73.061	0.365	256,984
Balance						
Top 10% Within HS	−0.013	(0.009)	[0.159]	0.213	0.210	269,037
Top 25% Within HS	−0.023	(0.011)	[0.038]	0.402	0.140	269,037
Advanced HS Courses	0.009	(0.069)	[0.897]	3.906	0.252	284,046
White	−0.027	(0.010)	[0.009]	0.421	0.221	284,046
Free-Lunch	0.019	(0.010)	[0.054]	0.244	0.127	284,046
Discipline Days	−0.014	(0.161)	[0.931]	1.834	0.027	284,046

Notes: This table shows estimates of the difference in discontinuities at the Pell eligibility threshold in years with B-On-Time loans, i.e., β_2 from Equation (3.2). The specification corresponds to Column (4) in Table 3.1: a local quadratic difference-in-discontinuity with controls for students' initial covariates, final net cost of attendance, and fixed effects for the financial aid year, among students with an EFC within \$5000 of the Pell eligibility threshold. All monetary measures are scaled down by 1000.

3.4 Theoretical Framework

3.4.1 Model

I model the choices of students, universities, and a regulator to rationalize the effects of BOT loans and quantify welfare and graduation under counterfactual schedules. The regulator first sets a policy, then universities set financial aid schedules, then students choose a university, and then students exert costly effort toward degree completion and human capital accumulation. I describe these components in reverse order.

The student. Student i is characterized by characteristics at the time of college application, $\mathcal{X} \equiv \{X, h_0\}$, and preferences over university characteristics θ_i. X consists of characteristics observed by the econometrician such as demographics and measures of ability. h_0 is the stock of human capital.

After college, students' earnings increase in their human capital stock and graduation status, which is used as a signal by employers. During college, anticipating future earnings, students repeatedly choose how much costly effort to exert studying. Effort increases human capital accumulation and the probability and speed of graduation. Both outcomes may vary across universities based on educational quality and accessibility. Effort below a threshold is equivalent to dropping out and forgoing graduation to enter the labor market. Effort and part-time paid work are substitutes and bounded above by a time constraint. Students value current consumption and leisure as well as the present discounted value of future earnings. Each semester, idiosyncratic shocks to income and demands for time can influence students' choices. The sequence of shocks induce an ex-ante probability of graduation g_{ij}. For example, small changes to family income can discontinuously change financial aid, making attendance unaffordable. Alternatively, the combination of current human capital and local labor market conditions creates variation in the value of full-time work without degree completion relative to current college enrollment.

At the start of college, after choosing a program, students can pay a cost to accept graduation incentives. The random idiosyncratic portion of this cost is realized after choosing a program.[12] To help smooth future consumption, students accept graduation incentives if the incentives lower the expected net cost of attendance. At this stage, students anticipate the probability of graduation. I assume that students are risk-neutral.

Students choose the university program that maximizes expected indirect utility. Indirect utility is a linear function of the program's mean quality, net cost, distance from family, the continuation value of enrollment, and idiosyncratic taste shocks. Preferences over program characteristics can vary across students. Students maximize indirect utility in expectation because of uncertainty about future choices. First, net cost depends on a student's anticipated choice of BOT take-up and future graduation. At this point, the idiosyncratic BOT application cost is not yet realized. Second, the continuation value of enrollment depends on a student's anticipated choices of costly effort. Counterfactuals may understate changes to indirect utility if students value other time-varying program characteristics, e.g., the demographics of other enrolled students.

THE UNIVERSITY. University j is characterized by its preferences over student characteristics, θ_j, capacity Q_j, per-student budget R_j, and two production functions, human capital $f_{h,j}$ and degree completion $f_{g,j}$. At the time of financial aid determination, the university only observes a subset of characteristics $X_0 \in \mathcal{X}$. Anticipating students' enrollment choices, the university chooses a financial aid schedule to maximize the expected sum of revenue and valuations of enrolled students' characteristics, subject to budget and capacity constraints. Revenue equals the student-paid net cost plus federal and state subsidies. Marginal valuations may be higher for students with underrepresented demographics if universities seek to maintain diversity. Diversity might be valued due to its downstream effects on reputation and educational quality.

[12]With BOT loans, students choose to accept graduation incentives. In contrast, counterfactual loan forgiveness may be universal, so the upfront cost is zero.

Universities cannot increase enrollment beyond capacity without substantial fixed costs, so financial aid helps ration spots among admitted students.[13]

Generally, a financial aid schedule P_j maps student characteristics and outcomes to a student-specific net cost of attendance for each university j. In the status quo, financial aid schedules depend only on university-observed X_0, i.e. expected family contribution. Students can apply for small additional scholarships, but at the time of university choice, they anticipate paying the net cost from the schedule.

The Regulator. The regulator values a weighted sum of students' consumer surplus, universities' producer surplus, and externalities of higher education, in expectation. Subject to a statewide budget constraint, regulation can affect students' net cost through restrictions on list price, in-kind tuition subsidies, or direct transfers to students. Through such regulation, a financial aid schedule might condition on a broad set of student characteristics – i.e., those in administrative records – as well as graduation outcomes.

Generally, an equilibrium is characterized by a set of financial aid schedules that clears the market subject to subsidies, regulation, constraints, and students' privately optimal choices of enrollment. To illustrate the effects of loan forgiveness, I simplify the problem: the regulator directly sets financial aid schedules, valuing consumer surplus and externalities, subject to universities' constraints and an added participation constraint. In expectation, each university must weakly prefer a counterfactual financial aid schedule to the status quo. The regulator may be able to improve producer surplus because institutional constraints limit universities' information and which financial aid schedules are feasible.

[13]In Texas, flagship university capacity is regulated.

3.4.2 Discussion

The model helps interpret estimates of the effects of BOT loans. If students' initial choice of effort is an interior solution, then BOT increases effort by increasing the marginal benefit of on-time graduation and relaxing the budget constraint. BOT also relaxes the budget constraint: with a low net cost, students require less income from part-time employment, which in turn frees up time for effort and leisure. Among marginal BOT recipients, productive effort must already be at its upper bound, because BOT does not measurably affect proxies for effort. Moreover, graduation incentives may be unlikely to increase effort among other students whose mean observed effort is similar to marginal BOT recipients'.[14] I interpret the upper bound of effort as structural in the context of marginal changes to financial aid. However, other policies outside the scope of this paper may improve the initial stock of human capital or mitigate external constraints on student resources and result in greater effort and graduation.

In the model, all else equal, average graduation at flagship universities is weakly increasing in graduation incentives. Even if students do not directly value graduation, incentives lower the net cost for high-graduation students and increase the net cost of low-graduation students. High-graduation students replace others if all students are sensitive to net cost. However, incentives may not increase welfare relative to fully screening on observables. Here, the benchmark is designing a financial aid schedule that depends on all student characteristics and maximizes social welfare. In the status quo, financial aid schedules are privately rather than socially optimal and they depend on a subset of student characteristics. If graduation value-added is positively correlated or uncorrelated with flagship graduation likelihood, then greater flagship graduation means lower non-flagship graduation.[15] Likewise, if a marginally

[14]Here, I also assume that these sets of students do not vary systematically in their inputs to the costly effort problem, e.g., the upper bound on effort and the cost of effort.

[15]Graduation value-added is a student's difference in graduation chances between flagships and other public universities.

selected student is equally likely to graduate at all universities (i.e., zero value-added), then he would not derive additional surplus from flagship enrollment. Other students who prefer flagships are made worse off. These students either pay a higher net cost to enroll at the flagship or they are forced to attend their second-choice university. Lower welfare among these students could be offset with a large social weight on graduation relative to student consumer surplus, and a sufficiently positive correlation between flagship graduation chances and value-added.

3.5 Estimation

3.5.1 Data

I focus on students admitted to at least one flagship university and one other public university. Students in the estimation sample both graduated from a Texas high school in the top 10 percent of their class and submitted a FAFSA between 2001 and 2017. Relative to the sample used in Section 3.3, this estimation sample includes greater variation in family income and relatively high-achievement students. I observe one financial aid offer per student per year, typically for the university at which they first enroll. I use the first observed offer and calculate the net cost as the cost of attendance minus all forms of grant aid, scholarship aid, and work-study funding. Cost of attendance is typically greater than tuition due to estimated costs of living like housing. Cost of attendance may vary across students within a school-year due to state rules governing tuition discounts which are separate from aid.[16] I predict net cost for all students by regressing observed net cost on interactions of EFC bins and continuous EFC percentile, separately for each year.[17] I cap net cost for students in the top bin of EFC. I measure miles between each student's K-12 district and each university. I measure college graduation for public universities and compare the

[16]I impute cost of attendance using the school-year median for some students, bound net cost between \$0 and \$100,000, and impute missing values of net cost as \$0.

[17]I group students into 31 bins; the lowest bin includes those with an EFC of \$0 and the remainder are binned by the within-year EFC percentile.

college graduation year to the high school graduation year.

3.5.2 Heterogeneity in Value-Added

To motivate key features of the structural model, I test whether the returns to selective universities vary across students. Consistent with prior studies, average returns to flagship enrollment are large for graduation and earnings. Effects vary substantially across subsamples of student demographics. For example, flagship enrollment most increases 4-year graduation for first-generation students, free-lunch-eligible students, and students of color.

I estimate the value-added of flagship universities as the difference in outcomes among students with similar ability and choice sets, following Mountjoy and Hickman (2021) and Dale and Krueger (2002):

$$Y_{ij} = \beta_1 1[\text{Enrolled Flagship}] + \gamma_{a(i)} + \beta_2 \mathbf{X_i} + \epsilon_{ij}\,, \tag{3.3}$$

where Y_{ij} is an outcome of interest for student i with initial enrollment at school j. The key assumption is that, conditional on observed characteristics, students symmetrically sort on unobserved characteristics into a flagship university or an alternative. In that case, β_1 represents the effect of initial flagship enrollment on the outcome. $\gamma_{a(i)}$ is a fixed effect for the set of universities that admitted student i, which may capture both students' and universities' private information about unobserved ability through application and admission decisions. I estimate this equation on the full sample before comparing estimates across subsamples.

First, Table 3.3 shows a large average treatment effect of flagship enrollment on 4-year graduation (Panel A), 6-year graduation (Panel B), and early-career earnings. Column (1) only controls for basic student characteristics that are observed by the university such as financial need and test scores. Columns (2)-(4) add covariates that are not directly observed by the university, which attenuates the coefficient of interest by almost half. This finding is consistent with prior work suggesting students select a university based on their unobserved ability. Estimates are still large:

flagship enrollment increases 4-year graduation by 6.4 percentage points over a mean of 41.4 percent; this drives three-quarters of the increase in 6-year graduation; and earnings increase by $2,621 over a mean of $54,587. Universities do not observe students' other admissions which proxy ability through selection. Column (3) takes the conventional approach of an admission set fixed effect, while Column (2) instead controls for mean graduation rate by admission set. Both approaches result in similar estimates, which helps reduce the parameter set for estimating the structural model below. In prior studies, estimates are similar when using high-dimensional fixed effects or a low-dimensional approximation (Dale and Krueger, 2002; Mountjoy and Hickman, 2021). Column (4) adds distance as a control, producing similar estimates and nearly identical R^2, easing concerns about remaining unobserved variation. Across several settings, distance shifts college enrollment and studies frequently assume that distance is independent of ability.

Second, Tables 3.4 and 3.5 show that returns to flagship enrollment vary across student characteristics, particularly relative to mean outcomes. For example, I compare students by the average 4-year college graduations rates of their high school. Students from schools with below-median graduation have a larger point estimate of the effect of flagship enrollment on their own 4-year graduation: 7.6 percent versus 6.3 percent. As a share of mean outcomes, the estimate for high-graduation high schools is nearly twice as large: 23 percent versus 12 percent. The relative returns are similar for 6-year graduation and much smaller for earnings (3.5 percent versus 5.3 percent). For other subsamples too, larger relative returns on 4-year graduation do not necessarily translate to larger returns for 6-year graduation or earnings. First-generation students, free-lunch-eligible students, and students of color all have relatively large returns to 4-year graduation. On the other hand, I do not find evidence of heterogeneity by distance or high school rank. Heterogeneity in returns to education is broadly consistent with the literature. For example, Andrews et al. (2016) use quantile regression to estimate the returns to Texas flagship enrollment. They find a dispersed distribution of earnings effects across students with heterogeneity by race and ethnicity.[18]

3.5.3 Empirical Model

I estimate preferences and graduation value-added by maximizing the joint likelihood of the college choice, BOT take-up, and the probability of graduation:

$$l_i(Y_i, B_i, G_i, X_i) = \int_{g_{f,i}} Pr(G_i \mid X_i, Y_i) Pr(B_i|X_i, Y_i, g_{f,i}) Pr(Y_i|X_i, g_{f,i}) dF(g_{f,i} \mid X_I)$$

where G_i indicates observed graduation within 4 years for student i, g is the latent ex-ante probability of on-time graduation, X_i includes observed characteristics (e.g., demographics, financial need, and the admission set), Y_i indicates the choice of university, and B_i indicates take-up of BOT loans. Students can choose UT Austin, Texas A&M, or an outside option.[19]

The key timing assumption is that students do not observe the exact cost of applying for BOT until after university enrollment. The university choice incorporates expectations over the distribution of BOT application cost. Since effort is largely unobserved, I also simplify the broader theoretical framework by parameterizing the continuation value of enrollment as linear in ex-ante graduation probability $g_{f,i}$. As a slight abuse of notation, I refer to $g_{f,i}$ as a student-university characteristic. Graduation probabilities vary flexibly depending on whether students enroll at a flagship university. This flexibility reflects that the underlying production functions may vary with university quality. I parameterize $g_{f,i}$ as a logistic function of student characteristics and a normal random variable $\epsilon^g_{f,i}$.

$$g_{f(j),i} = \Phi\left(\beta^g_x X_i + \epsilon^g_f\right)$$
$$Pr(G_i \mid X_i, f_i) = G_i \, g_{f,i} + (1 - G_i)\,(1 - g_{f,i})$$

[18]At UT Austin, the earnings premium ranges from 2.7 percent at the 9$^{\text{th}}$ percentile to 31.7 percent at the 97$^{\text{th}}$ percentile.

[19]Some students are not admitted to both UT Austin and Texas A&M. $f \in \{0, 1\}$ indicates enrollment at a flagship university.

Students take up BOT loans if the expected reduction in net cost of attendance outweighs an idiosyncratic application cost, which is a function of student characteristics and a normally distributed error. Likewise, the expected net cost is reduced by expected loan forgiveness.

$$Pr(B_{ijkt} = 1) = Pr(g_{f(j),i}\bar{B} - a_{ijkt} > 0),\ a_{ijkt} \sim N(a_{jkt}, \sigma_B)$$
$$E[P^B_{ijkt}|g] = \max\{0, a_{jkt} - g_{f(j),i}\bar{B}\}$$

Students choose a college to maximize indirect utility, a function of net cost of attendance, graduation likelihood, residual mean quality, and a taste shock:

$$u_{ij} = -\alpha_i P_{ij} + \beta^u_{g,i} g_{ij} + \beta^u_{d,i} d_{ij} + \delta_{ij} + \epsilon^u_{ij}$$
$$u_{i0} = -\alpha_i \bar{P}_{i0} + \beta^u_{g,i} g_{i0} + f(|\ \mathcal{J}_0\ |) + \epsilon^u_{i0}$$
$$P_{ij} = P^0_{ij} - E[P^B_{ijkt}|g]$$
$$\theta_i = f_\theta\left(\Pi_\theta X_i + \epsilon_\theta\right),\ \theta \in \{\alpha, \beta_g, \beta_d, \delta\}$$

I normalize the outside option intercept δ_{i0} to 0. All else equal, students who are admitted to more non-flagship universities may have a stronger preference for the outside option. To address this, I include a quadratic function of the count of other admissions. $\beta_{g,i}$ and α_i are log transformations of the linear index because indirect utility should not increase in net cost or decrease in graduation. I allow for correlation between ϵ^g_0 and ϵ^g_1. The preference shocks are distributed Type-I extreme value. I integrate over ϵ^g using quadrature to calculate the likelihood.

Based on reduced-form findings, I assume that BOT has no direct impact on graduation. As a result, observed selection into BOT – among students with identical observed characteristics – identifies the variance in ex-ante graduation likelihood. Conditioning on BOT take-up and university choice, residual variation in graduation rates identifies $E[g_{f(j)}]$. Residual correlation between BOT take-up and observed characteristics identifies mean BOT application cost. The correlation between selection into flagship universities and the corresponding ex-ante graduation probability g_1

separately identifies $\beta^u_{g,ik}$ and δ_{jt} as long as g_1 and δ_{jt} have different functional forms. I assume financial need is exogenous so formulaic financial aid is also exogenous of preference shocks. Then the correlation between residual enrollment and financial aid identifies α_{ik}.

Table 3.3: Mean Estimates of Value-Added

	(1)	(2)	(3)	(4)
	(a) Four-Year Graduation			
Enroll Flagship	0.130	0.064	0.070	0.070
	(0.003)	(0.003)	(0.007)	(0.007)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]
R^2	0.048	0.155	0.157	0.157
Outcome Mean	0.414	0.414	0.414	0.414
Observations	94,839	94,839	94,839	94,839
	(b) Six-Year Graduation			
Enroll Flagship	0.174	0.097	0.101	0.102
	(0.003)	(0.003)	(0.005)	(0.005)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]
R^2	0.056	0.197	0.217	0.217
Outcome Mean	0.703	0.703	0.704	0.704
Observations	77,705	77,705	77,705	77,705
	(c) Annual Earnings 8-10 Years Post-HS (Thousands)			
Enroll Flagship	2.191	2.621	2.454	2.510
	(0.314)	(0.319)	(0.486)	(0.484)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]
R^2	0.025	0.033	0.066	0.067
Outcome Mean	54.587	54.587	54.680	54.680
Observations	56,833	56,833	56,833	56,833

Notes: This table shows the effects of flagship enrollment on medium-run outcomes, with estimates from Equations (3.3). Column (1) includes basic initial student characteristics observed by every university and a fixed effect for year of enrollment. Column (2) adds average graduation rates conditional on each of high school, major, and admission portfolio. Column (3) adds an admission portfolio fixed effect, which absorbs the corresponding mean graduation rate, and clusters standard errors by admission portfolio. Column (4) adds a control for distance to the nearest flagship.

Table 3.4: Heterogeneity in Value-Added

	HS $\geq$ 60 miles		High-Graduation HS		First-Generation	
	(N)	(Y)	(N)	(Y)	(N)	(Y)
(a) Four-Year Graduation						
Enroll Flagship	0.071	0.070	0.076	0.063	0.062	0.078
	(0.009)	(0.006)	(0.007)	(0.008)	(0.007)	(0.008)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Outcome Mean	0.419	0.410	0.337	0.492	0.441	0.387
R^2	0.152	0.172	0.141	0.143	0.160	0.162
Observations	48,269	45,085	46,621	46,695	48,303	44,887
(b) Six-Year Graduation						
Enroll Flagship	0.111	0.094	0.092	0.112	0.113	0.091
	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Outcome Mean	0.709	0.698	0.648	0.761	0.738	0.669
R^2	0.225	0.219	0.197	0.233	0.242	0.203
Observations	39,265	36,975	38,857	37,383	39,008	37,094
(c) Annual Earnings 8-10 Years Post-HS (Thousands)						
Enroll Flagship	3.867	1.135	1.798	3.122	2.971	1.896
	(0.530)	(0.601)	(0.477)	(0.818)	(0.818)	(0.442)
	[<0.001]	[0.059]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Outcome Mean	55.239	54.159	51.240	58.556	57.081	52.439
R^2	0.074	0.077	0.074	0.057	0.066	0.082
Observations	28,818	26,680	29,176	26,297	27,459	27,877

Notes: This table shows heterogeneity in the effects of flagship enrollment on medium-run outcomes, with estimates from Equations (3.3). The specification corresponds to Column (3) from Table 3.3. It includes controls for basic initial characteristics, fixed effects for the admission portfolio and year of enrollment, and average graduation by high school and major. Each column represents a separate regression on a subsample of students.

Table 3.5: Heterogeneity in Value-Added: Additional Characteristics

	Top 10%		Free-Lunch		White	
	(N)	(Y)	(N)	(Y)	(N)	(Y)
	(a) Four-Year Graduation					
Enroll Flagship	0.073	0.070	0.067	0.078	0.070	0.068
	(0.008)	(0.009)	(0.007)	(0.008)	(0.009)	(0.007)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Outcome Mean	0.414	0.416	0.437	0.350	0.390	0.441
R^2	0.149	0.183	0.151	0.183	0.173	0.148
Observations	46,211	47,019	69,872	23,306	48,069	45,364
	(b) Six-Year Graduation					
Enroll Flagship	0.111	0.093	0.107	0.085	0.088	0.113
	(0.007)	(0.006)	(0.006)	(0.008)	(0.007)	(0.006)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Outcome Mean	0.701	0.707	0.725	0.639	0.675	0.732
R^2	0.200	0.258	0.222	0.217	0.218	0.225
Observations	39,350	36,787	57,912	18,196	37,579	38,743
	(c) Annual Earnings 8-10 Years Post-HS (Thousands)					
Enroll Flagship	2.541	2.631	2.571	2.233	1.799	2.997
	(0.488)	(0.659)	(0.583)	(0.735)	(0.665)	(0.797)
	[<0.001]	[<0.001]	[<0.001]	[0.002]	[0.007]	[<0.001]
Outcome Mean	54.861	54.602	56.607	48.755	51.210	57.909
R^2	0.068	0.084	0.060	0.105	0.075	0.066
Observations	29,681	25,705	42,343	13,018	26,388	29,137

Notes: This table shows heterogeneity in the effects of flagship enrollment on medium-run outcomes, with estimates from Equations (3.3). The specification corresponds to Column (3) from Table 3.3. It includes controls for basic initial characteristics, fixed effects for the admission portfolio and year of enrollment, and average graduation by high school and major. Each column represents a separate regression on a subsample of students.

3.6 Estimates

This section provides context for how parameter estimates shape outcomes across counterfactual financial aid schedules. Counterfactual financial aid can increase welfare through selection by improving student-university match. First, I describe the heterogeneous effects of flagship enrollment on enrollment. Graduation value-added correlates sensibly with observed characteristics. Conditional on characteristics, there is substantial variation in graduation chances driving selection into flagship enrollment and BOT take-up. Second, I describe heterogeneity in preferences for flagship universities, even among students with identical observed characteristics. Graduation value-added and preferences are correlated due to demographic variation. Third, model estimates imply heterogeneity in universities' marginal costs across students. I interpret this as a relative valuation for middle-income students. In the status quo, graduation is relatively low because these students face large discounts despite their low value-added.

Parameter estimates in Table 3.6 show how ex-ante graduation correlates with student characteristics. Relative to regression estimates, these correlations correct for bias from asymmetric selection. Graduation at any university is greater for female, white, high family education, Advanced Placement, and Gifted program students. Average graduation rates by admission portfolio, high school, and college major are highly predictive of graduation. Notably, family income is not particularly predictive of graduation chances after controlling for other characteristics and unobserved selection. The one exception is that at non-flagship universities, students with the lowest family income, i.e. zero expected family contribution, have precisely greater graduation chances than students with the highest family income. In estimation, I make the relatively strong assumption that financial aid does not directly impact graduation. This assumption may be reasonable because financial aid is a function of income, and generally, income is not precisely correlated with graduation. Estimates vary for flagship and non-flagship graduation chances so observed characteristics also drive variation in graduation value-added, the increase in graduation chances from enrolling at a flagship. These correlations may not be causal if characteristics correlate with the unobserved component of graduation.

The bottom rows show substantial variation in the unobserved component of graduation. For flagships, a one-standard-deviation increase is roughly equivalent to either attending a high school with 25 percentage points greater college graduation or the combined impact of being female, white, having two college-educated parents, and attending a high school gifted program. Second, residual graduation likelihoods are highly correlated between flagship and non-flagship universities: the confidence interval includes 1. Counterfactual financial aid that targets selection on ex-ante observed characteristics may be almost as efficient as a policy that targets selection on unobserved characteristics via graduation incentives.

Although family income is not precisely correlated with graduation, higher-income students place more weight on graduation when choosing a university. Several other demographics that correlate with graduation and value-added – female, white, family education – correspond to lower marginal utility from graduation. Advanced courses and expected college graduation by admission portfolio, high school, and major are positively correlated.

Consistent with prior research, net cost plays a large role in determining college enrollment. Net cost sensitivity is higher for students with more positive point estimates: lower income, white, lower family education, fewer AP courses, and no gifted program. Estimates contrast prior literature in suggesting that preferences for nearby universities are small and generally uncorrelated with student characteristics. The exception is that students with high family education are less deterred by distance.

The intercept term of university indirect utility absorbs remaining differences in the perceived quality of flagship universities relative to alternatives. Generally, demographics that correlate with high graduation correspond to lower mean quality.

Table 3.6: Demand Model Estimates

	Intercept	Net Cost	Miles	Value-Added	BOT Cost	$P(Grad \mid f=0)$	$P(Grad \mid f=1)$
Intercept	9.388	1.499	−0.344	0.811	2.807	−7.677	−6.641
	(0.309)	(0.022)	(0.042)	(0.131)	(1.048)	(0.124)	(0.092)
EFC Percentile	−3.552	−0.343	0.168	0.710	−0.144	−0.054	0.322
	(0.919)	(0.092)	(0.237)	(0.028)	(0.332)	(0.202)	(0.172)
Female	−0.527	0.013	−0.010	−0.183	0.039	0.179	0.417
	(0.107)	(0.010)	(0.027)	(0.013)	(0.041)	(0.025)	(0.022)
White	0.841	0.115	−0.039	−0.068	0.162	0.114	0.233
	(0.129)	(0.012)	(0.033)	(0.013)	(0.063)	(0.027)	(0.025)
Family Education	−0.480	−0.053	0.070	−0.012	−0.001	0.059	0.085
	(0.078)	(0.007)	(0.019)	(0.007)	(0.027)	(0.016)	(0.014)
Advanced Courses	−0.198	−0.007	0.005	0.123	−0.019	0.047	0.019
	(0.018)	(0.002)	(0.005)	(0.004)	(0.009)	(0.003)	(0.003)
Gifted Program	−0.243	−0.016	0.033	−0.021	0.044	0.156	0.130
	(0.109)	(0.011)	(0.028)	(0.011)	(0.041)	(0.024)	(0.021)
E[Graduation | Admissions]				0.592	0.387	4.347	3.708
				(0.047)	(0.179)	(0.115)	(0.097)
E[Graduation | HS]				0.551	0.219	4.811	3.927
				(0.078)	(0.225)	(0.125)	(0.102)
E[Graduation | Major]				1.916	1.046	6.238	5.863
				(0.104)	(0.175)	(0.137)	(0.114)
HS Residual Family Education				0.190	−0.095	0.150	0.679
				(0.063)	(0.245)	(0.125)	(0.113)
EFC = 0	3.618	0.349	−0.024	−0.619	−1.023	0.150	0.064
	(0.291)	(0.030)	(0.070)	(0.039)	(0.351)	(0.066)	(0.055)
EFC Percentile 1-20	4.887	0.328	−0.034	−0.769	−0.964	0.141	0.085
	(0.479)	(0.048)	(0.114)	(0.036)	(0.357)	(0.101)	(0.085)
EFC Percentile 21-40	4.637	0.293	−0.103	−0.209	−0.357	0.135	0.083
	(0.678)	(0.065)	(0.165)	(0.019)	(0.249)	(0.138)	(0.117)
EFC Percentile 41-60	4.253	0.291	−0.172	0.292	0.181	0.133	−0.043
	(0.878)	(0.084)	(0.219)	(0.018)	(0.289)	(0.175)	(0.151)
EFC Percentile 61-80	2.885	0.189	−0.107		1.790	0.157	0.154
	(0.965)	(0.095)	(0.238)		(0.717)	(0.196)	(0.167)
UT Austin	11.504						
	(0.319)						
Other 4-Year Net Cost	−0.008						
	(0.172)						
Other Admissions	−0.325						
	(0.093)						
Other Admissions^2	−0.049						
	(0.015)						
Standard Deviation					1.338	0.000	0.977
					(0.438)	(0.580)	(0.027)
Correlation P(Grad | f=1)						0.973	
						(0.004)	

Notes: This table shows model estimates for the Top 10 percent sample with asymptotic standard errors calculated using the approximate Hessian. The coefficient on Net Cost in indirect utility is the negative exponential of a linear combination of student characteristics. The coefficient on Value-Added in a positive exponential of a linear combination of student characteristics. $P(Grad \mid f = 0)$ and $P(Grad \mid f = 1)$ are logistic transformations of a linear combination of student characteristics. Standard Deviation refers to a normally distributed random term of each linear combination. The random components of $P(Grad \mid f = 0)$ and $P(Grad \mid f = 1)$ are correlated. An indicator for expected family contribution ("EFC") above \$150k is omitted. Indicators for each financial aid year are not shown.

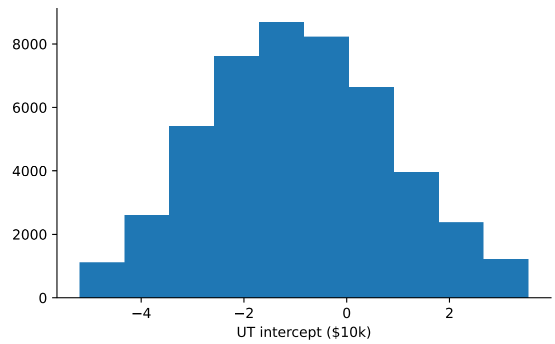
This pattern rationalizes relatively low enrollment at the high-graduation alternative. A few additional terms discipline the outside option. As expected, students with more non-flagship admissions have a greater outside option value. On the other hand, I do not find evidence of correlation with average net cost of students' non-flagship alternatives, which mitigates concern that students' preferences for non-flagship universities are incompletely modeled.

Figure 3.2 plots the distributions of students' valuations of university characteristics by normalizing marginal utility for each characteristic by the marginal utility of net cost. Valuations for increased graduation are large and approximately log-normally distributed, consistent with the log-normal earnings potential of college graduates. By contrast, the distaste for distance explains a small fraction of the variation in flagship enrollment. As expected, most students prefer a nearby school, all else equal.

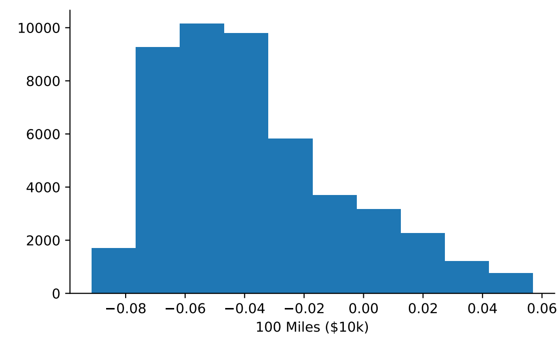
Large and dispersed application costs rationalize why so few students apply for BOT loans. These costs are perhaps too large to be plausible but might be rationalized by information frictions outside the model. Due to federal regulations, universities could not advertise BOT loans, so a large share of students may have been unaware of the program. Application costs might be much smaller among the set of students with knowledge of the program. Also, application costs rationalize some students' distaste for the coursework requirements of BOT loans. To graduate in four years, students could not take many courses outside their majors, and could not choose certain majors that are unlikely to be completed in four years. Students' idiosyncratic preferences for such courses and majors are outside of the model except for unobserved variation in the BOT application cost.

Large application costs for BOT loans imply that BOT had virtually no impact on ex-ante welfare. When choosing a university, the expected gross reduction in net cost of attendance from BOT is negligible relative to expected application costs. If application costs were much lower, some students might have chosen a different uni-

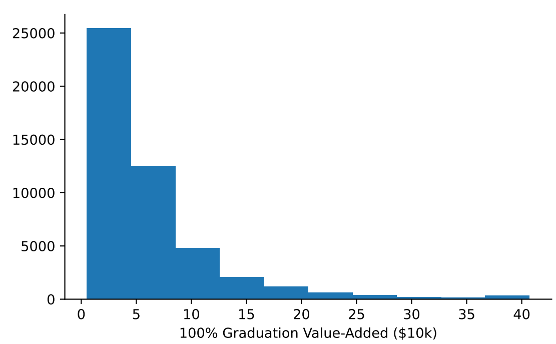
Figure 3.2: Distribution of Student Valuations of University Characteristics



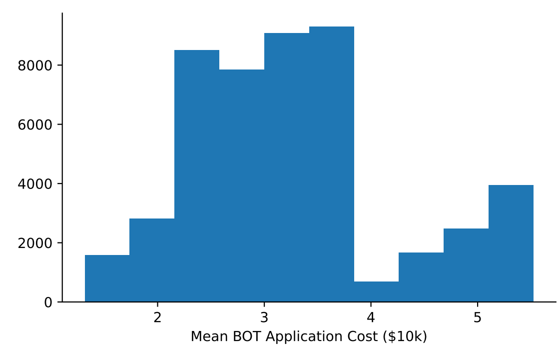
(a) UT Austin Intercept



(b) Distance



(c) Value-Added in Graduation



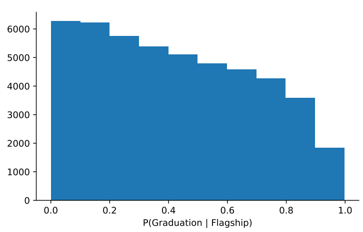
(d) B-On-Time Loans Application Cost

Notes: This figure shows distributions across students in valuations of university characteristics, converted to units of $10,000 by dividing marginal utility estimates by the marginal (dis)utility of net cost of attendance.

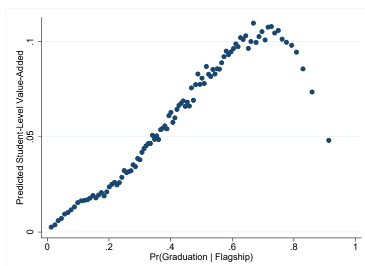
versity with higher expected graduation and lower net cost via BOT. After enrolling, students realize their idiosyncratic application costs, which are sometimes low enough to apply. Some of these students graduate and enjoy the ex-post transfer of forgiven BOT loans. However, with a budget constraint, this small ex-post benefit comes at the relatively large cost of greater prices for most students who do not receive BOT loans or do not graduate on time. I rely on the timing assumption that students do not know their exact application cost until after enrolling at a university.

Dispersion in graduation value-added helps counterfactual financial aid to increase statewide graduation. Figure 3.3, Panel B shows that higher flagship graduation generally corresponds to higher graduation value-added. Equivalently, graduation increases in ability at all universities, but the slope is larger at flagships. This pattern is consequential for student equity and counterintuitive given prior work studying large compositional changes of selective universities (e.g., Black et al., 2020; Bleemer, 2021). Generally, in those studies, underrepresented students reap relatively large benefits from attending selective universities, while other students are similarly successful regardless of university. In this setting, targeting enrollment towards students with high graduation value-added (and improving statewide graduation rates) may also increase disparities across demographics and universities.

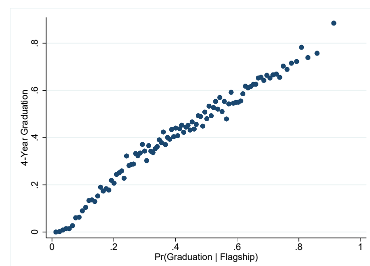
Figure 3.3: Predicted Ex-Ante Graduation Likelihood



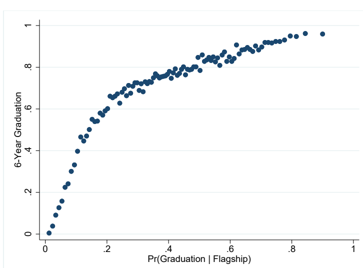
(a) Distribution



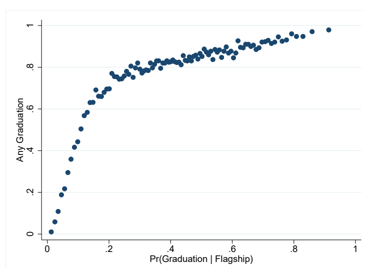
(b) Predicted Flagship Value-Added



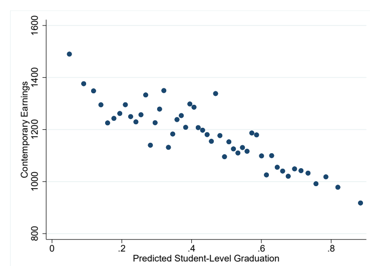
(c) Observed 4-Year Graduation



(d) Observed 6-Year Graduation



(e) Observed Any Graduation



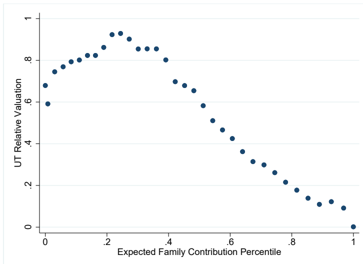
(f) Contemporary Earnings

Notes: This figure shows the distribution and correlates of students' ex-ante likelihood of on-time graduation. "Predicted" indicates that a measure is calculated using estimates of the structural model. Figures (b)-(f) are scatter plots showing the means for each percentile of graduation likelihood.

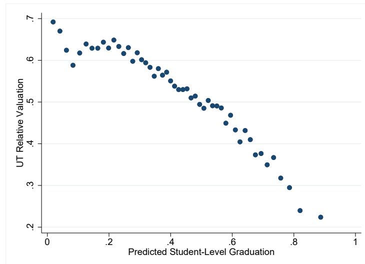
Panel A highlights that predicted ex-ante graduation likelihood at flagship universities is highly dispersed, most students have low graduation chances, and even the highest-ability students have meaningful uncertainty about graduation. Estimates fit the data well for both targeted and untargeted outcomes. Predicted 4-year flagship graduation is highly correlated with observed graduation over several time windows. In Panel C, the relationship is highly linear reflecting strong fit. The slope is less than 1 because idiosyncratic preferences drive some students to enroll at non-flagship universities and graduate at lower rates. The difference is largest for students with high flagship graduation chances. At 6 or more years after first enrollment, higher predicted graduation corresponds to higher realized graduation. The relationship is not linear because students with intermediate ex-ante chances graduate eventually but in more than 4 years. Finally, Panel F shows that students with high graduation chances exert more effort towards studying, proxied by less part-time work.

Estimates imply that flagship universities have strong non-pecuniary incentives to enroll middle-income students over low-income students and low-income students over high-income students. I first calculate the effective marginal cost for each student and university by inverting the first-order condition for net cost, given a student's net cost elasticity and federal financial aid. True marginal costs should be approximately equal across students, so I subtract the estimated marginal cost of the wealthiest students to calculate the universities' relative valuation of each student. Figure 3.4 shows the UT Austin's implied relative valuation of student expected family contribution ("EFC"). Relative valuations increase from minimum EFC to the 25^{th} percentile and then decrease. The peak implies that enrolling a middle-income student is valued approximately \$9500 more than enrolling a high-income student. Since low-income students have lower expected graduation, value-added, and SAT scores, the first-order condition also implies that universities devalue these characteristics. Net cost schedules are a function of EFC, so these secondary correlations do not necessarily represent how universities would target composition with full information. For comparison, relative valuations are uncorrelated with distance and highly correlated between both flagship universities.

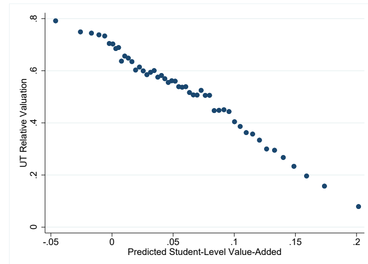
Figure 3.4: Flagship Universities' Relative Valuations of Student Characteristics



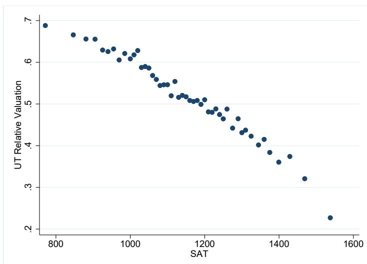
(a) Expected Family Contribution



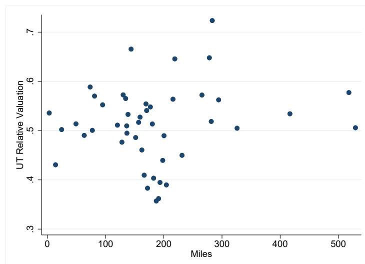
(b) Graduation Likelihood



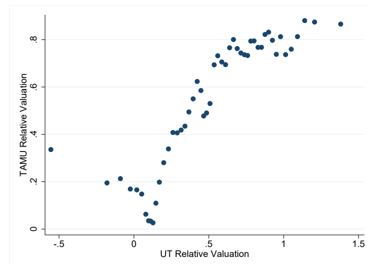
(c) Graduation Value-Added



(d) SAT Score



(e) Miles from High School



(f) Texas A&M

Notes: This figure shows the correlates of UT Austin's valuation of students (in units of $10,000 per student) and their characteristics. Relative valuations are the difference in implied marginal cost between students with EFC above $150,000 and a given student. Marginal costs are implied by student own-price elasticities and inverting the university's first-order condition. Figures (b)-(f) are scatter plots showing the means for each bin of the x-axis with an equal number of students. By construction, financial aid offers are purely functions of EFC, so any other correlation with relative valuation is driven by its correlation with EFC.

3.7 Counterfactual Financial Aid

Using the estimates from Section 3.6, I simulate students' choices under counterfactual financial aid schedules to illustrate the welfare effects of graduation incentives. First, I quantify the total cost of information frictions by deriving students' first-best financial aid offers. In the first best, the regulator has perfect information about all observed characteristics and unobserved graduation likelihood. Second, I decompose information frictions by setting financial aid without graduation chances, instead using the full set of ex-ante characteristics. Third, I add graduation incentives which largely close the gap between these two scenarios. Fourth, I show that graduation incentives alone can improve welfare, fixing the status quo average net cost. I conclude by assessing the equity implications of optimal financial aid schedules.

Table 3.7: Average Counterfactual Outcomes Per Admitted Student

	Social Surplus	Consumer Surplus	Graduation (%)	Enrollment (%)	Revenue	Producer Surplus
(1) First-Best	2.682	2.222	0.550	−1.931	0.318	0.077
(2) Screening	2.145	2.150	−3.524	−0.021	0.090	0.011
(3) Screening + Incentive	2.463	2.101	−3.548	−0.072	0.325	0.013
(4) Status Quo + Incentive	0.679	0.243	1.568	2.522	0.017	0.007

Notes: This table shows key outcomes from counterfactual contract schedules. All outcomes are based on ex-ante expectations over students. Counterfactuals vary FFS the incremental net cost of flagship universities for each student, enforcing capacity and budget constraints. All outcomes are incremental, relative to the status quo. Unless otherwise indicated, outcomes are in units of $10,000. Revenue is a conditional average among students who counterfactually enroll at a flagship university.

With perfect information about students' graduation chances and demand for flagship enrollment, the regulator could improve welfare by $26,820 per admitted student. Fixing the net cost of the outside option, the regulator offers each student a net cost for flagship enrollment that conditions on all observed characteristics and the student's latent graduation chance. 82 percent of the welfare increase represents student surplus. The remainder largely represents the regulator's valuation of graduation externalities. By varying the difference in net cost between flagship universities and their alternatives, the regulator increases graduation by 0.55 percentage points. Graduation increases at all universities – not just flagships, because non-flagship students have relatively low graduation effects of switching to a flagship. Despite greater efficiency, enrollment at flagships decreases by nearly 2 percentage points. Estimated welfare does not reflect that educational quality might improve with lower student-faculty ratios.

Even without perfect information, the regulator can achieve 80 percent of the first-best welfare increase by targeting financial aid based on students' ex-ante observed characteristics ("Screening"). In the status quo, financial aid only conditions on expected family contribution, which reflects a relatively small share of the variation in student demand. Moreover, as discussed in Section 3.6, universities' privately optimal financial aid schedules target students with relatively low graduation value-added. With improved screening, consumer surplus is nearly identical to first-best. Even though the regulator values graduation, the schedule lowers graduation by 3.5 percentage points relative to the status quo. Two patterns drive this large and counterintuitive decrease. First, there is substantial unobserved variation in graduation chances. Second, conditioning on observed characteristics, students with high graduation value-added have relatively low consumer surplus or low net cost elasticity.

Combining screening with graduation incentives can achieve 92 percent of the first-best welfare increase. With incentives, students receive an additional discount on net cost when graduating. For comparability to the status quo, I fix this discount at a size similar to status quo BOT loans. Efficiency may further improve if the incentive varied optimally with student observed characteristics. With any incentive versus screening alone, the net cost is higher for students who do not graduate so that the budget balances. Average outcomes are similar to screening alone because of the correlation between graduation value-added and net cost elasticity. Graduation incentives lower flagships' expected net cost and improve welfare for targeted students, but these discounts may not be cost-effective. Greater welfare among targeted students with high value-added may not outweigh lower welfare among non-targeted students who have a strong preference to attend a flagship despite low value-added.[20] These non-targeted students need to pay a higher net cost or attend a less preferred university.

Without screening, graduation incentives alone can achieve 25 percent of the first-best welfare increase. Relative to the status quo, prior counterfactuals sometimes lead to large changes in expected net cost among students who have the same EFC. Such changes may be infeasible under institutional constraints. For example, Texas has a legal requirement that tuition set-asides benefit students below a threshold EFC.

[20]These non-targeted students have a large surplus when the net cost does not include graduation incentives.

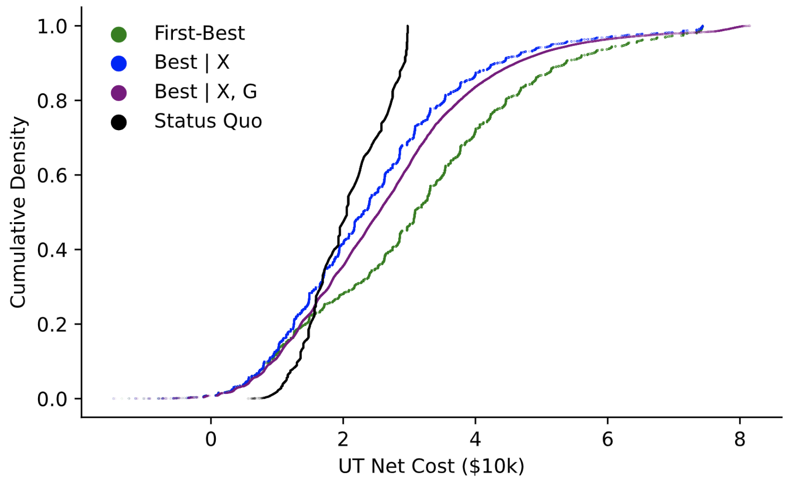
Now, I fix average net cost by EFC at the status quo level while adding a graduation incentive. This approach prevents further redistribution across income but allows redistribution across match quality. Students' consumer surplus increase is an order of magnitude smaller than prior counterfactuals, but statewide graduation rates increase three times as much as in first best: 1.6 percentage points.[21] With limited variation in unobserved value-added, screening on observed characteristics (third-degree price discrimination) and graduation incentives (second-degree price discrimination) can be thought of as substitutes. Welfare increases from only screening or only incentives are much larger than the combination of screening and incentives.

Figure 3.5 shows how counterfactual financial aid schedules largely improve welfare through cross-subsidization. Given the social objective, especially the budget and capacity constraints, students with relatively inelastic demand should face higher net costs. Greater revenue from these students permits much lower net costs for the relatively elastic students who most increase social welfare by enrolling at a flagship. Among counterfactual schedules, greater average welfare corresponds to a flatter cumulative distribution function and more redistribution. For example, the first-best schedule is the most progressive because perfect information enables more precise price discrimination.

Counterfactual financial aid schedules also lead to large compositional changes, which generally benefit students from historically underrepresented backgrounds. Students' incremental consumer surplus and graduation rates correlate with demographic characteristics, so counterfactual financial aid changes the composition of flagship enrollment. Table 3.8 compares the difference in average characteristics between flagship universities and alternatives. As before, these numbers rely on the sample of Top 10% students admitted to multiple universities, and this sample is not necessarily

[21]Comparisons of this counterfactual to others should be interpreted with caution. Greater enrollment drives part of the increased graduation and welfare. After fixing expected net cost and satisfying the budget constraint, there are no degrees of freedom to bound enrollment below capacity.

Figure 3.5: Cumulative Distributions of Expected Net Cost Across Counterfactual Schedules



Notes: This figure shows the empirical cumulative distributions of net cost of attendance across students in the estimation sample for each counterfactual financial aid schedule. In First-Best, each student receives a personalized financial aid offer to each flagship university that maximizes the social objective given ex-ante student observed characteristics X and graduation likelihood g.

representative of Texas high school students or overall university enrollment. In the status quo, flagship students are relatively white, low income, female, and low family education, with fewer AP courses. Flagship students come from high schools with quite similar college graduation rates. Besides income, differences are relatively small. First best financial aid most increases flagship enrollment among students from historically underrepresented backgrounds: students of color, first-generation students, students with few AP courses, and students from low-graduation high schools. With cross-subsidization, the gap by income approaches zero. Composition is qualitatively similar across counterfactuals except for graduation incentives alone, which especially incentivize flagship enrollment among relatively low-income and male students.

Table 3.8: Counterfactual Demographic Composition: Flagship - Alternatives

	(SQ)	(1)	(2)	(3)	(4)
White	0.004	−0.041	−0.030	−0.025	−0.044
Wealth (EFC Percentile)	−0.056	0.003	−0.004	0.002	−0.087
Female	0.007	0.010	0.022	0.029	−0.031
College-Educated Parents	−0.007	−0.083	−0.073	−0.055	−0.108
Advanced Courses	−0.406	−1.207	−1.349	−1.281	−0.686
College Graduation by HS	−0.001	−0.051	−0.046	−0.041	−0.027

Notes: Each cell is a difference in demographic composition between flagship universities and other 4-year public universities. For example, 0.4pp more students are white at flagship universities in the status quo. Column numbers index counterfactuals: (1) First-Best, (2) Screening, (3) Screening + Incentive, and (4) Status Quo + Incentive, relative to (SQ) Status Quo. College graduation by HS is the average 4-year graduation (at public 4-year universities) among students that graduated from the same high school.

3.8 Conclusion

This paper demonstrates how revenue-neutral financial aid design can improve both efficiency and equity by allocating limited spots at selective universities to the students who most benefit. Graduation incentives are expedient at a moment when federal and state policies constrain universities from increasing diversity through admissions. First, graduation incentives exploit students' private information about match quality to lower the price for relatively well-matched students and increase their enrollment. Second, although graduation incentives are race-neutral (and neutral with respect to all demographics), they happen to also increase enrollment at selective universities among historically underrepresented students. These features may be increasingly useful if future policy further restricts universities from considering student race or its proxies, e.g., via essays or scholarships.

The benefits of financial aid design require changing students' choice of university. Combining evidence from a discontinuity in financial need and a structural model of college choice, I show how historical graduation incentives failed to improve welfare. B-On-Time loans could have improved match quality by changing the relative price of selective universities. Instead, it introduced uniform incentives that aimed to increase effort. In this setting, the marginal student was too constrained to change behavior.

B-On-Time's limitations are instructive for contemporary national and state policies. To target selection and improve match quality, programs must change relative incentives to attend low versus high-quality universities. For example, as of 2024, the College Cost Reduction Act aims to reduce higher education spending by expanding income-based repayment plans, limiting per-person federal loans, and reducing dispersion in list prices. Around the same time, several university systems announced zero-tuition policies for students below an income threshold. In such cases, unintended consequences for match quality might diminish the benefits of lower average costs. Tighter bounds on net cost can limit opportunities for screening on price and for redistribution.

Appendix A: Supplementary Materials to Chapter 1

A.1 Data Sources

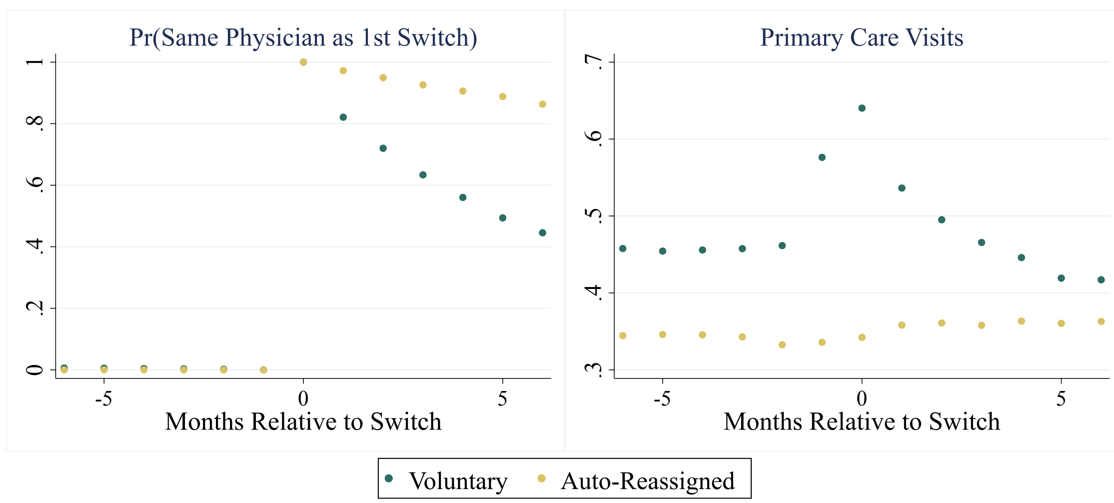
I use several data sources to construct the estimation sample. The Norwegian Control and Payment of Health Reimbursements Database (KUHR) tracks reimbursement for outpatient claims organized at the level of bill line, i.e., reimbursement code, and identifies most patients and physicians. The Norwegian Patient Registry (NPR) is a database of reimbursement for inpatient claims organized at the level of encounter. I use ICD-10 and ICPC-2 codes from both sources to classify chronic illness. I identify avoidable hospitalizations following Table A1 from Page et al. (2007). Base payments come from a basic subsidy rate dataset. Various datasets from the Norwegian GP Registry identify spells when patients are registered to patient lists and when physicians are contracted to provide care to those patient lists. The physician-list dataset also identifies contract details: the maximum number of registered patients and indicators for shared lists and fixed-salary reimbursement. I use anonymous identifiers for physicians, lists, and patients to link datasets and convert spells into monthly panels. Patients' birth date, gender, disability payment receipt, and income come from tax records.

A.2 Additional Figures and Tables

Figure A.1, Panel A shows that in every month after auto-reassignment, some auto-reassigned patients switch again to a third physician, but on average, 83 percent remain for at least six months, so auto-reassignment represents a persistent change in enrollment. For comparison, voluntary switchers are more likely to remain with their new physician, consistent with stronger preferences.[1] Panel B shows that auto-reassigned patients' average treatment intensity is higher after switching physicians, consistent with Kwok (2018), and potentially crowding out incumbents. Compared to auto-reassigned switchers, voluntary switchers have persistently higher treatment intensity, including a sudden large increase in the month before switching. Voluntary switchers' treatment intensity declines after switching but remains much higher than before, and the change in treatment intensity is larger than the change among auto-reassigned switchers. The different trajectories of voluntary and auto-reassigned switchers suggest that endogeneity from patient sorting is important in this context. Even if a health shock prompts voluntary switchers to find new physicians rather than wait to be auto-reassigned, they choose physicians that they will remain with for longer and visit more often for at least six months. In the United States, patients of exiting physicians must all choose a new physician and the choice of physician likely incorporates unobserved determinants of treatment intensity like match quality even if the timing is exogenous. As examples of match quality, Kristiansen and Sheng (2022) show that low-SES patients assigned to a physician with a low-SES background have lower mortality, and Dahlstrand (2022) shows that matching high-risk patients to high-quality physicians lowers avoidable hospitalizations.

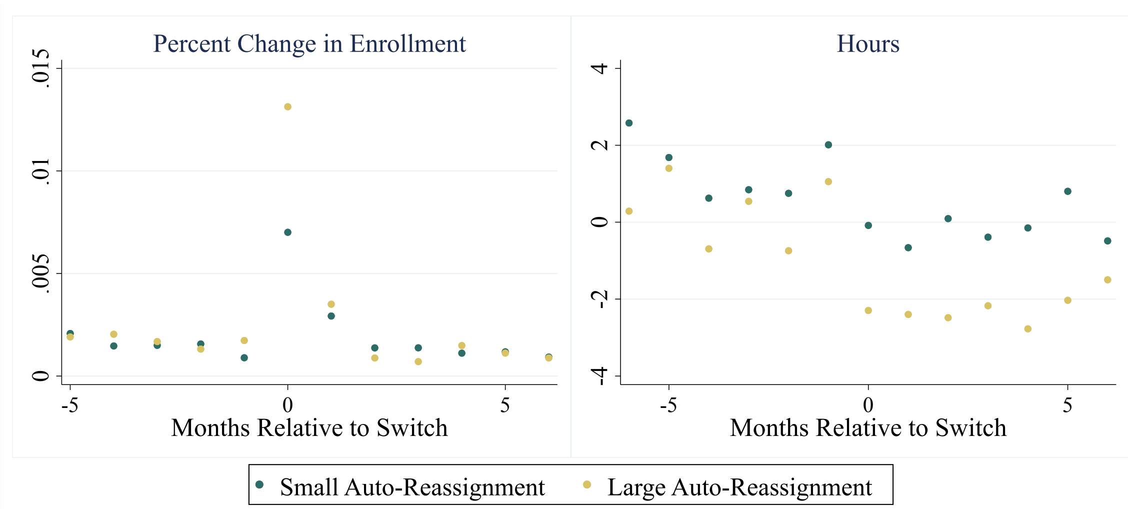
[1] Voluntary switchers do not directly contribute to the identifying variation of this paper. However, they might choose the same physicians to which others are auto-reassigned.

Figure A.1: Patients of Exiting Physicians: Voluntary Switches vs. Auto-Reassigned



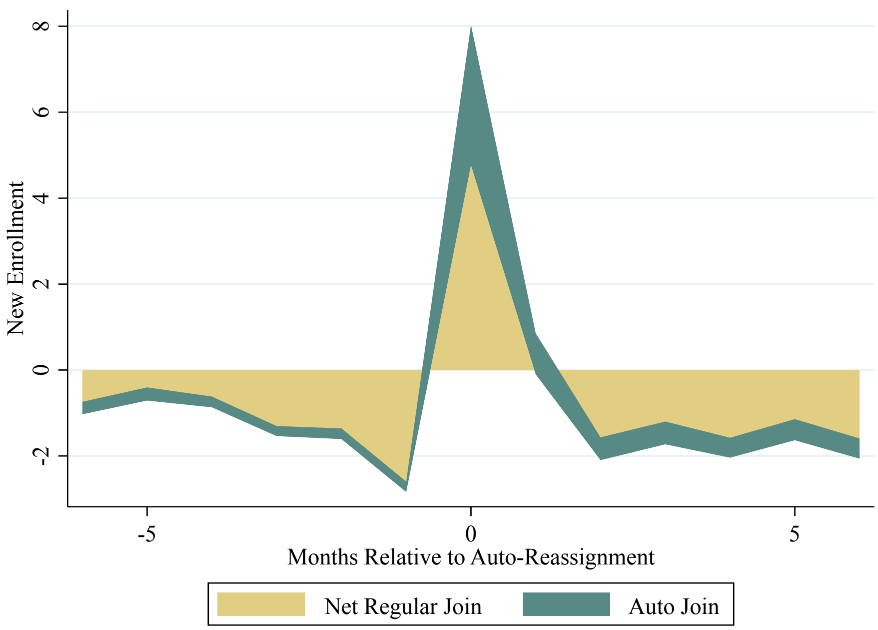
Notes: Both panels reflect raw means among patient-months within 13-month spells, based on switching patients of exiting physicians whose spells are not necessarily balanced. For auto-reassigned patients, Month 0 is the first month after they are administratively reassigned to an incumbent's physician. For voluntary switching patients, Month 0 is the first month with a new physician, up to 5 months before their previous physician exits. The left panel shows the share of patients registered to their first physician after the previous one exited. The right panel shows average days per month with a physician encounter.

Figure A.2: Trends in Outcomes among Incumbent Patients



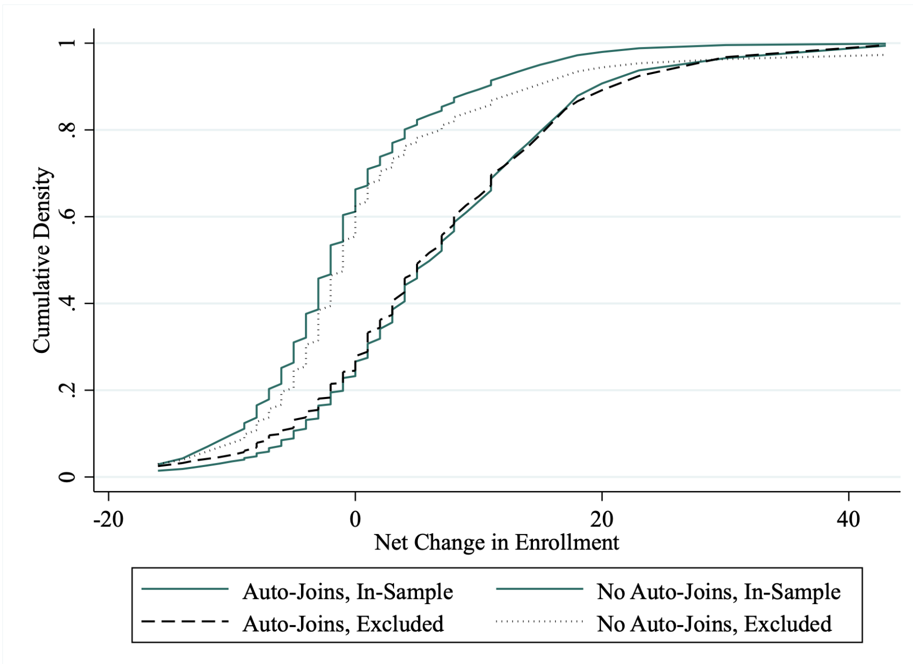
Notes: Both panels reflect means among physician-months within 13-month spells. Month 0 is the first month with auto-reassigned patients. Panel (a) shows the average percent change in the total number of registered patients. Panel (b) shows seasonally adjusted physician hours among incumbent patients, the residual from a regression on a time trend and calendar month fixed effects. Large Auto-Reassignment refers to spells with more than one patient auto-reassigned in month 0 (the mean is 5.5).

Figure A.3: Average Net Change to Enrollment



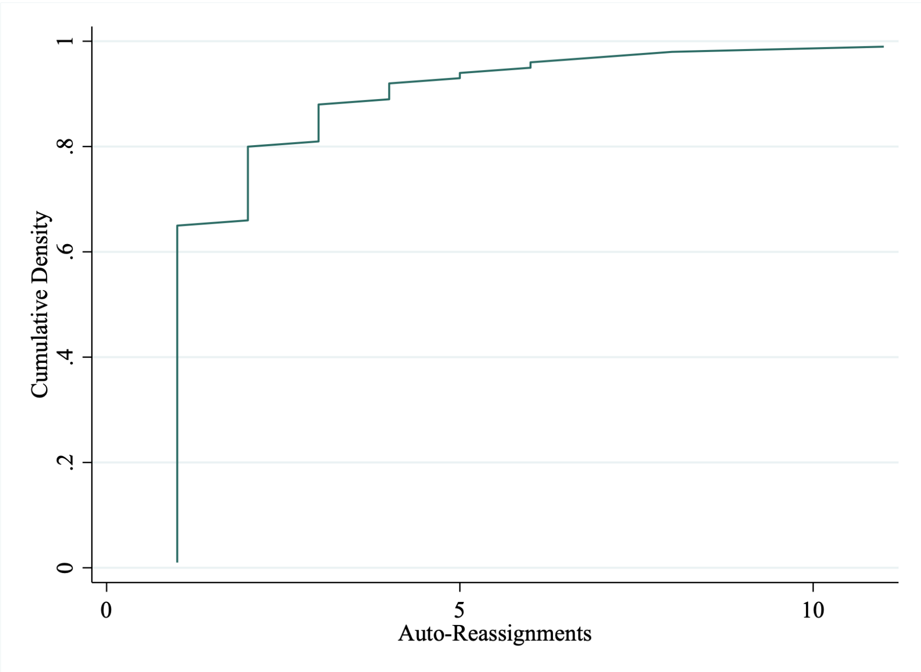
Notes: This figure decomposes the average one-month net change in enrollment between net regular joins and auto-joins. "Net Regular Join" is the average of the difference between patients that voluntarily leave and patients that voluntarily join. "Auto Join" is the average new registrations of patients that are administratively reassigned from other physicians. I average over physicians in the analysis sample in each month relative to the first auto-reassignment. Auto-reassignment in event-months -6 through -1 are from non-exiting physicians.

Figure A.4: Cumulative Distribution of Enrollment Changes



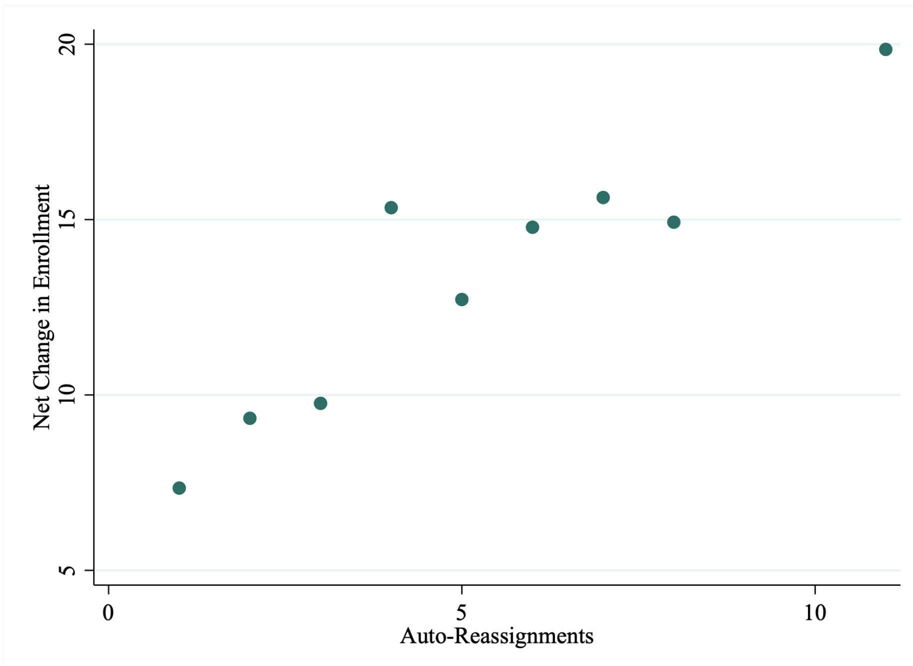
Notes: This figure shows the cumulative distribution of one-month changes in the number of patients for four subsamples of physicians. "Auto-Joins, In-Sample" includes months with any auto-reassignment to a physician in the analysis sample. "No Auto-Joins, In Sample" reflects other months in the analysis sample. Likewise, out-of-sample physician-months are split between the remaining two categories based on whether an auto-reassigned patient registered with the physician. Out-of-sample indicates with an out-of-sample physician or an out-of-sample month for an in-sample physician. Each curve is truncated at the 1st and 99th percentiles.

Figure A.5: Cumulative Distribution of Auto-Reassignments



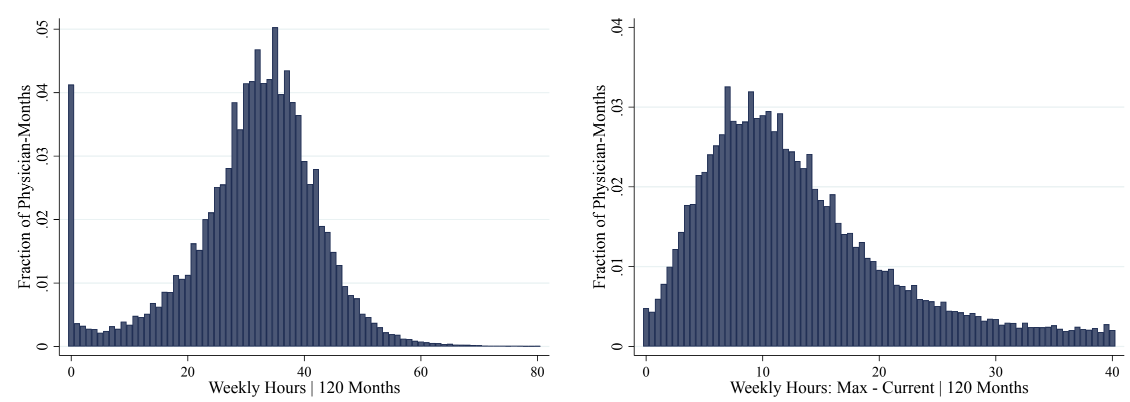
Notes: This figure shows the cumulative distribution of the number of patients auto-reassigned among physicians in the analysis sample in the first month of auto-reassignment (event month 0). The curve is truncated at the 99th percentile.

Figure A.6: Auto-Reassignments and Enrollment Changes



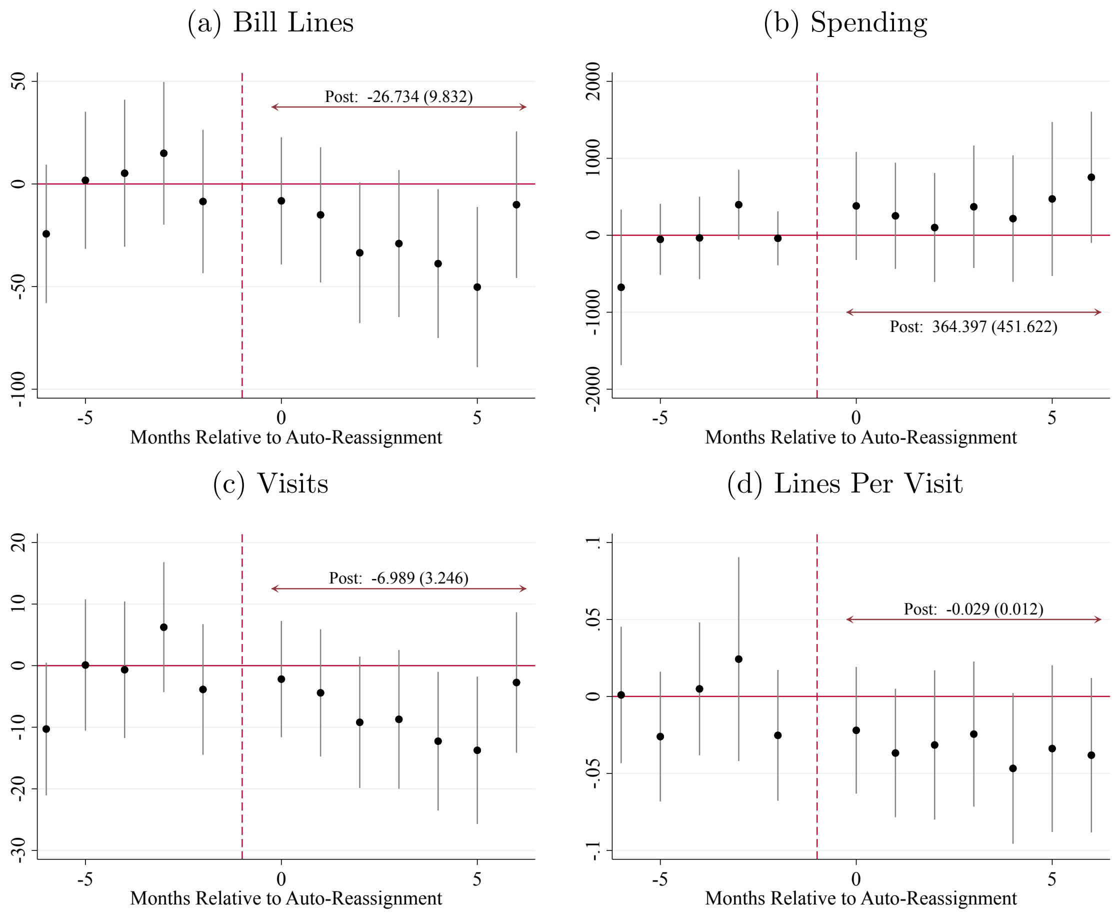
Notes: This figure shows the correlation between the number of patients auto-reassigned and the net change in enrollment. Each point is average among physicians for a percentile of the x-axis. When multiple percentiles have the same value, I use an average of averages. This figures uses the analysis sample in the first month of auto-reassignment (event month 0). The plot is truncated at the 99th percentile.

Figure A.7: Capacity Constraints and Bunching of Workload



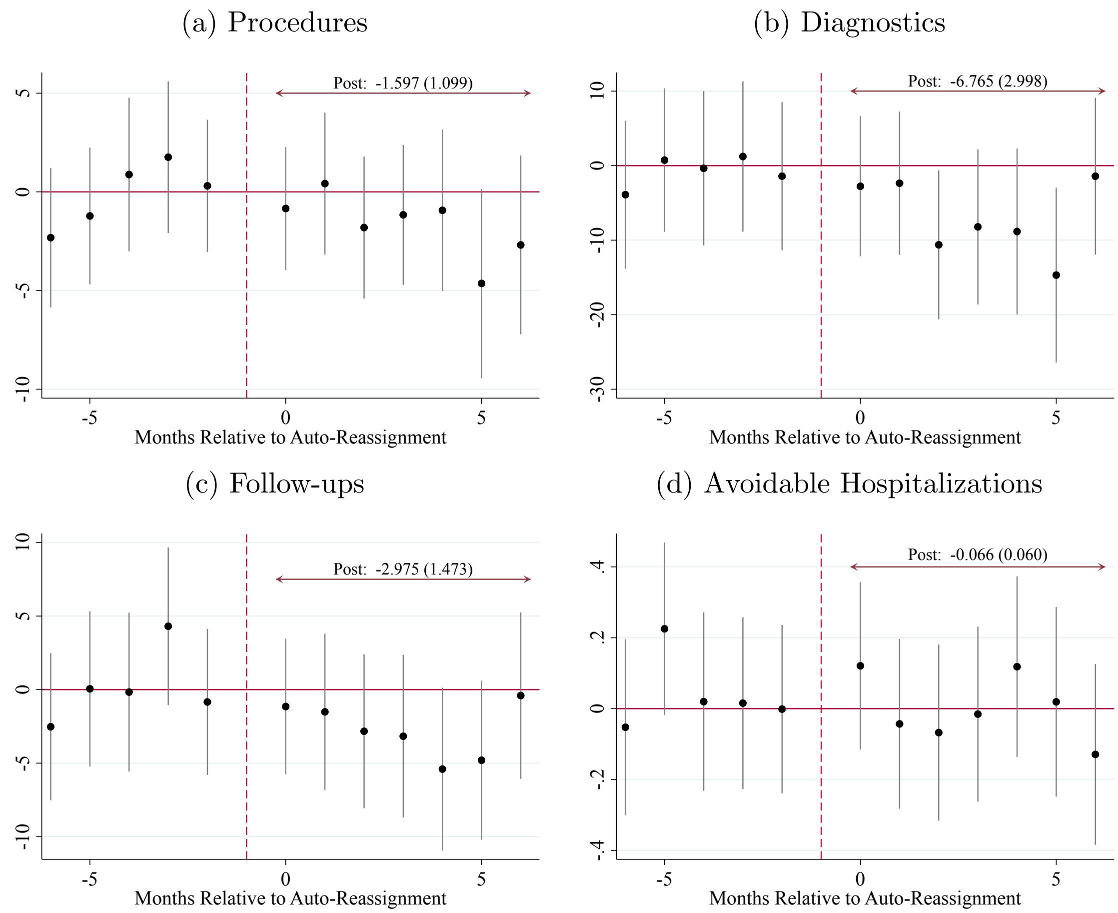
Notes: Both figures reflect weekly hours for physicians observed for the full sample period, 2008–2017. Weekly hours equal hours per month divided by weeks per month. The left figure shows the untransformed distribution of weekly hours, truncated at 80 hours. The right figure shows the distribution of transformed weekly hours ($\tilde{M}_{jt}$). The transformation is $\max_t M_{jt} - M_{jt}$. The x-axis is truncated at 40 and I exclude the first month when a physician works the maximum number of hours.

Figure A.8: Treatment Measures Among Incumbent Patients: Large versus Small Auto-Reassignments



Notes: This figure shows estimates and 95% confidence intervals of the difference between physicians with large and small auto-reassignments in each month relative to auto-reassignment. The omitted month is -1. I define large auto-reassignments as those with two or more reassigned patients (969 of 2,722 physician-spells). Each panel represents a separate regression with the dependent variable as indicated. The physician-month sample reflects aggregate treatment intensity among a fixed set of incumbent patients. Standard errors are clustered at the physician level. Estimates correspond to β_{1t} in the following regression: $Y_{jt} = \beta_{0t} + \beta_{1t}\text{AutoHigh}_j + \beta_j + \epsilon_{jt}$. "Post" indicates the average difference in outcomes after auto-reassignment between physicians with large and small auto-reassignments. Standard errors are in parentheses, and the coefficient is β_1 in the following regression: $Y_{jt} = \beta_{0t} + \beta_1\text{AutoHigh}_j \times 1[t \geq 0] + \beta_j + \epsilon_{jt}$.

Figure A.9: Treatment Types Among Incumbent Patients: Large versus Small Auto-Reassignments



Notes: This figure shows estimates and 95% confidence intervals of the difference between physicians with large and small auto-reassignments in each month relative to auto-reassignment. The omitted month is -1. I define large auto-reassignments as those with two or more reassigned patients (969 of 2,722 physician-spells). Each panel represents a separate regression with the dependent variable as indicated. The physician-month sample reflects aggregate treatment intensity among a fixed set of incumbent patients. Standard errors are clustered at the physician level. Estimates correspond to β_{1t} in the following regression: $Y_{jt} = \beta_{0t} + \beta_{1t}\text{AutoHigh}_j + \beta_j + \epsilon_{jt}$. "Post" indicates the average difference in outcomes after auto-reassignment between physicians with large and small auto-reassignments. Standard errors are in parentheses, and the coefficient is β_1 in the following regression: $Y_{jt} = \beta_{0t} + \beta_1\text{AutoHigh}_j \times 1[t \geq 0] + \beta_j + \epsilon_{jt}$.

Table A.1: Sample Selection

Sample Restriction	Physicians	Physician-Months
Full Sample	8,198	538,362
Received auto-reassignments and nearby months	6,190	126,933
Balanced 13-month spells with identifier	3,723	89,732
No change to primary physician	3,693	89,115
$\geq$ 500 incumbent patients	3,452	83,343
No shared list or temporary physician	3,189	77,207
No downsizing	2,362	45,968
No change to reimbursement rate	2,156	40,495
First balanced spell if overlapping	2,065	35,386
Final Sample	2,065	35,386

Notes: This table shows how the number of observation decreases with each of the sample conditions described in Section 1.2.2. Conditions are applied in the order shown.

Table A.2: Test of Conditionally Random Assignment

	(1)	(2)
Availablity	0.028	0.027
	(0.007)	(0.007)
	[0.000]	[0.000]
Same Municipality	10.542	10.562
	(1.283)	(1.288)
	[0.000]	[0.000]
Share Female		1.387
		(1.138)
		[0.223]
Share Chronic		0.398
		(1.097)
		[0.717]
Share Age 65+		−0.324
		(0.761)
		[0.671]
# Observations	10,777	10,777

Notes: This table shows regressions of the number of auto-reassignments on the availability (maximum enrollment - lagged enrollment), an indicator for the physician residing in the same municipality as the exiting physician, fixed effects for the combination of the exiting physician and month of exit (not shown), and in column (2), average characteristics of incumbent patients. The unit of observation is a physician-month for physicians receiving auto-reassignments.

Table A.3: Effect of Enrollment Among All Registered Patients

	Enrollment	Hours	Bill Lines	Spending	Visits
	(1)	(2)	(3)	(4)	(5)
Cuml. Auto-Joins	1.083				
	(0.069)				
	[<0.001]				
Enrollment		0.003	0.136	11.538	0.391
		(0.052)	(0.285)	(8.667)	(0.331)
		[0.954]	[0.634]	[0.183]	[0.239]
Dep. Var. Mean	1274.917	141.206	1547.545	13941.826	502.655
New Pat. Mean		0.124	1.313	131.205	0.400
F-Statistic	112.037	12.268	24.786	8.944	23.365
Observations	35,386	33,411	33,411	33,411	33,411
	Lines Per Visit	Procedures	Diagnostics	Follow-ups	Avoidable Hosp.
	(6)	(7)	(8)	(9)	(10)
Enrollment	−0.001	−0.086	0.107	0.150	0.007
	(0.001)	(0.050)	(0.094)	(0.080)	(0.001)
	[0.327]	[0.084]	[0.257]	[0.059]	[<0.001]
Dep. Var. Mean	2.922	97.827	367.454	179.930	4.292
New Pat. Mean	3.244	0.096	0.293	0.165	0.004
F-Statistic	84.653	48.531	37.662	23.279	6.405
Observations	33,411	33,411	33,411	33,411	35,386

Notes: This table displays estimates of coefficients from regressions of Equations 1.1 and 1.2. Each column represents a separate regression with the dependent variable as indicated in the table. The physician-month sample reflects aggregate treatment intensity among all currently registered patients. Standard errors clustered at the physician level are reported in parentheses, and p-values are reported in brackets. "Cuml. Auto-Joins" indicates the number of patients auto-reassigned to a physician since the start of the spell. New Pat. Mean is the average of the dependent variable among auto-reassigned patients, averaged across physicians in the six months after auto-reassignment. Regressions also include fixed effects for spell and event-month.

Table A.4: Patients in Sample Experience Gap in Treatment Intensity

	Hours (1)	Avoidable Hosp. (2)
In-Sample	-0.006	0.017
	(0.001)	(0.006)
	[<0.001]	[0.002]
Dep. Var. Mean	0.122	0.345
F-Statistic	22.280	9.157
Observations	57,760,517	62,013,800

Notes: This table displays estimates of coefficients from regressions of the indicated dependent variable on an indicator for observations in-sample and fixed effects for each combination of month, 5-year age bin, gender, an indicator for recent registration, and indicators for primary and secondary chronic illnesses. The sample includes all patients in Norway for each month in 2015 and outcomes are patient-specific. Standard errors clustered at the physician level are reported in parentheses, and p-values are reported in brackets. Patient Hours are approximated by multiplying physician-level hours by the share of reimbursement. Avoidable Hospitalization is scaled up by 100.

Table A.5: Heterogeneity in the Effect of Auto-Reassignment on Enrollment

	Capacity		Fee Level		Schedule	
	Slack (1)	Binds (2)	Low (3)	High (4)	Part-Time (5)	Full-Time (6)
Cuml. Auto-Joins	5.674	1.037	1.041	2.699	1.006	1.206
	(0.789)	(0.039)	(0.048)	(0.498)	(0.036)	(0.142)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Dep. Var. Mean	1184.388	1356.589	1178.162	1343.490	1150.030	1341.655
F-Statistic	33.577	134.240	83.519	46.769	100.088	66.967
Observations	16,783	18,603	14,677	20,709	12,324	23,062
	Age		Diagnoses		Gender	
	Under 65 (7)	Over 65 (8)	Healthy (9)	Chronic (10)	Male (11)	Female (12)
Cuml. Auto-Joins	1.083	1.083	1.083	1.083	1.083	1.083
	(0.069)	(0.069)	(0.069)	(0.069)	(0.069)	(0.069)
	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Dep. Var. Mean	1274.917	1274.917	1274.917	1274.917	1274.917	1274.917
F-Statistic	112.037	112.037	112.037	112.037	112.037	112.037
Observations	35,386	35,386	35,386	35,386	35,386	35,386

Notes: This table displays estimates of coefficients from regressions of Equation 1.1. Each column represents a separate regression among a subsample with the dependent variable as indicated in the table. Columns (1-6) reflect aggregate treatment intensity among all incumbent patients for subsets of physicians. Columns (7-12) reflect aggregate treatment intensity among subsets of incumbent patients. For example, Column (1) includes physicians with enrollment consistently less than 99 percent of initial stated capacity, and Column (2) includes all other physicians. Standard errors clustered at the physician level are reported in parentheses, and p-values are reported in brackets. High-fee physicians receive supplementary reimbursement for each visit due to a certificate from additional training. Full-time physicians treat patients for at least six hours per weekday, on average, during the six months prior to auto-reassignment. Regressions also include fixed effects for spell and event-month.

Table A.6: Heterogeneity in the Effect of Enrollment on Bill Lines

	Capacity		Fee Level		Schedule	
	Slack (1)	Binds (2)	Low (3)	High (4)	Part-Time (5)	Full-Time (6)
Enrollment	−0.028	−0.376	−0.326	−0.110	−0.445	−0.233
	(0.272)	(0.053)	(0.053)	(0.454)	(0.030)	(0.070)
	[0.919]	[<0.001]	[<0.001]	[0.809]	[<0.001]	[<0.001]
Dep. Var. Mean	1032.164	1211.539	808.396	1351.888	792.144	1305.120
1st Stage F-Stat.	33.577	134.240	83.519	46.769	100.088	66.967
Observations	16,783	18,603	14,677	20,709	12,324	23,062
	Age		Diagnoses		Gender	
	Under 65 (7)	Over 65 (8)	Healthy (9)	Chronic (10)	Male (11)	Female (12)
Enrollment	−0.316	−0.027	−0.265	−0.078	−0.093	−0.251
	(0.031)	(0.037)	(0.042)	(0.020)	(0.017)	(0.042)
	[<0.001]	[0.455]	[<0.001]	[<0.001]	[<0.001]	[<0.001]
Dep. Var. Mean	783.404	343.060	604.972	521.492	481.626	644.839
1st Stage F-Stat.	112.037	112.037	112.037	112.037	112.037	112.037
Observations	35,386	35,386	35,386	35,386	35,386	35,386

Notes: This table displays estimates of coefficients from regressions of Equation 1.2. Each column represents a separate regression among a subsample with the dependent variable as indicated in the table. Columns (1-6) reflect aggregate treatment intensity among all incumbent patients for subsets of physicians. Columns (7-12) reflect aggregate treatment intensity among subsets of incumbent patients. For example, Column (1) includes physicians with enrollment consistently less than 99 percent of initial stated capacity, and Column (2) includes all other physicians. Standard errors clustered at the physician level are reported in parentheses, and p-values are reported in brackets. High-fee physicians receive supplementary reimbursement for each visit due to a certificate from additional training. Full-time physicians treat patients for at least six hours per weekday, on average, during the six months prior to auto-reassignment. Regressions also include fixed effects for spell and event-month.

Table A.7: Heterogeneity in the Effect of Enrollment on Visits

	Capacity		Fee Level		Schedule	
	Slack (1)	Binds (2)	Low (3)	High (4)	Part-Time (5)	Full-Time (6)
Enrollment	0.059	−0.003	0.015	0.074	−0.191	0.259
	(0.087)	(0.156)	(0.157)	(0.137)	(0.012)	(0.054)
	[0.496]	[0.983]	[0.925]	[0.591]	[<0.001]	[<0.001]
Dep. Var. Mean	333.016	403.671	296.589	422.302	253.387	432.563
1st Stage F-Stat.	33.577	134.240	83.519	46.769	100.088	66.967
Observations	16,783	18,603	14,677	20,709	12,324	23,062
	Age		Diagnoses		Gender	
	Under 65 (7)	Over 65 (8)	Healthy (9)	Chronic (10)	Male (11)	Female (12)
Enrollment	0.008	−0.003	−0.002	0.007	0.012	−0.007
	(0.127)	(0.022)	(0.108)	(0.040)	(0.048)	(0.100)
	[0.947]	[0.876]	[0.986]	[0.865]	[0.802]	[0.945]
Dep. Var. Mean	261.099	109.062	212.772	157.388	155.915	214.246
1st Stage F-Stat.	112.037	112.037	112.037	112.037	112.037	112.037
Observations	35,386	35,386	35,386	35,386	35,386	35,386

Notes: This table displays estimates of coefficients from regressions of Equation 1.2. Each column represents a separate regression among a subsample with the dependent variable as indicated in the table. Columns (1-6) reflect aggregate treatment intensity among all incumbent patients for subsets of physicians. Columns (7-12) reflect aggregate treatment intensity among subsets of incumbent patients. For example, Column (1) includes physicians with enrollment consistently less than 99 percent of initial stated capacity, and Column (2) includes all other physicians. Standard errors clustered at the physician level are reported in parentheses, and p-values are reported in brackets. High-fee physicians receive supplementary reimbursement for each visit due to a certificate from additional training. Full-time physicians treat patients for at least six hours per weekday, on average, during the six months prior to auto-reassignment. Regressions also include fixed effects for spell and event-month.

Table A.8: Heterogeneity in the Effect of Enrollment on Reimbursement

	Capacity		Fee Level		Schedule	
	Slack (1)	Binds (2)	Low (3)	High (4)	Part-Time (5)	Full-Time (6)
Enrollment	−0.839	1.649	0.968	7.135	−2.916	6.976
	(2.419)	(4.526)	(4.257)	(11.029)	(1.978)	(1.852)
	[0.729]	[0.716]	[0.820]	[0.518]	[0.140]	[<0.001]
Dep. Var. Mean	9242.519	11252.667	7232.838	12472.557	7517.379	11785.897
1st Stage F-Stat.	33.577	134.240	83.519	46.769	100.088	66.967
Observations	16,783	18,603	14,677	20,709	12,324	23,062
	Age		Diagnoses		Gender	
	Under 65 (7)	Over 65 (8)	Healthy (9)	Chronic (10)	Male (11)	Female (12)
Enrollment	1.094	0.174	0.617	0.652	0.707	0.561
	(3.590)	(0.640)	(3.026)	(1.141)	(1.339)	(2.829)
	[0.761]	[0.785]	[0.839]	[0.568]	[0.597]	[0.843]
Dep. Var. Mean	7349.652	2949.635	5782.335	4516.951	4466.973	5832.313
1st Stage F-Stat.	112.037	112.037	112.037	112.037	112.037	112.037
Observations	35,386	35,386	35,386	35,386	35,386	35,386

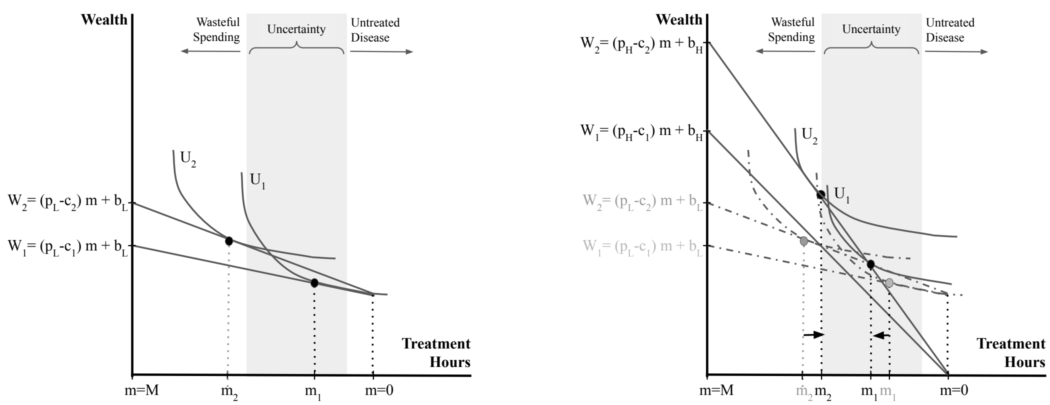
Notes: This table displays estimates of coefficients from regressions of Equation 1.2. Each column represents a separate regression among a subsample with the dependent variable as indicated in the table. Columns (1-6) reflect aggregate treatment intensity among all incumbent patients for subsets of physicians. Columns (7-12) reflect aggregate treatment intensity among subsets of incumbent patients. For example, Column (1) includes physicians with enrollment consistently less than 99 percent of initial stated capacity, and Column (2) includes all other physicians. Standard errors clustered at the physician level are reported in parentheses, and p-values are reported in brackets. High-fee physicians receive supplementary reimbursement for each visit due to a certificate from additional training. Full-time physicians treat patients for at least six hours per weekday, on average, during the six months prior to auto-reassignment. Regressions also include fixed effects for spell and event-month.

Appendix B: Supplementary Materials to Chapter 2

B.1 Additional Analysis

B.1.1 Additional Figures

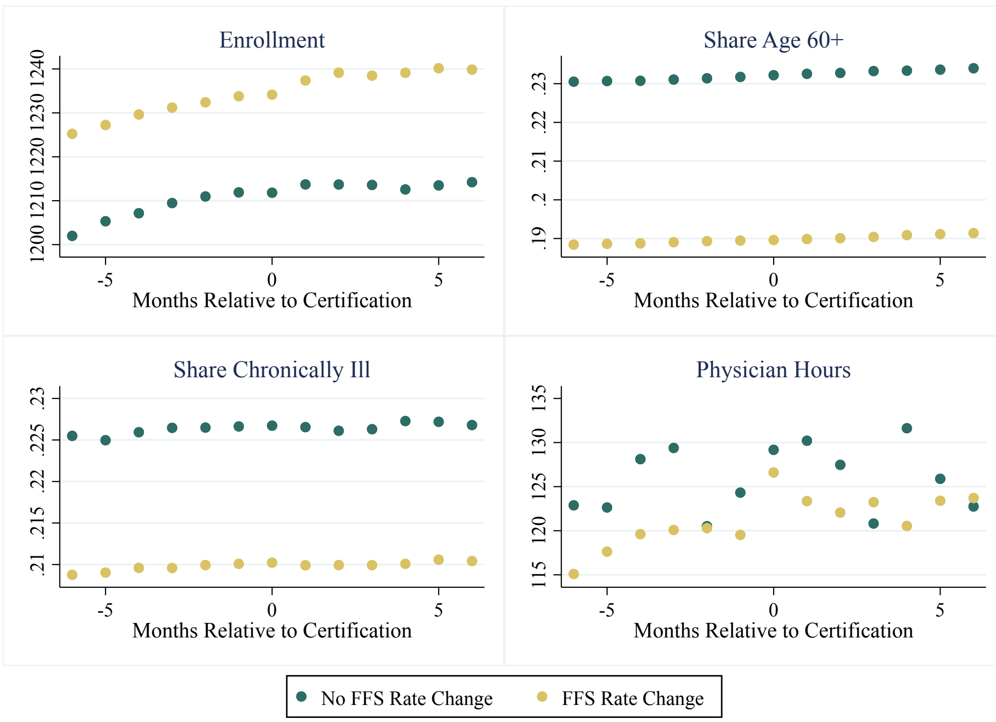
Figure B.1: A Uniform Contract May Be Efficient



(a) Initial Uniform Contract (b) Better Uniform Contract

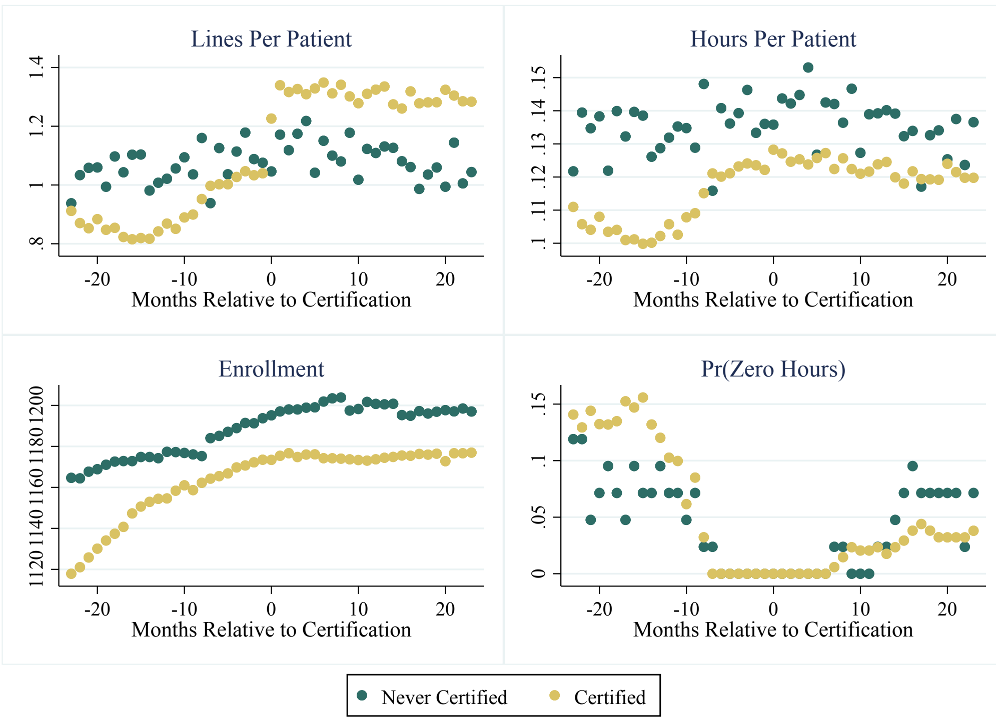
Notes: This figure shows a stylized example with two physicians, in which a a uniform contract is efficient. The x-axis plots treatment intensity $m \in [0, M]$ from right to left. Each panel shows the indifference curves of these physicians and the budget constraint(s) implied by simple reimbursement contract(s) with a base payment and an hourly wage. The shaded region includes the efficient level of labor supply which is unobserved to the regulator. In the left panel, the single status quo contract is efficient only for Physician 1. In the right panel, the new uniform contract has high marginal reimbursement p and is efficient for both physicians.

Figure B.2: Raw Means of Characteristics Relative to Certification



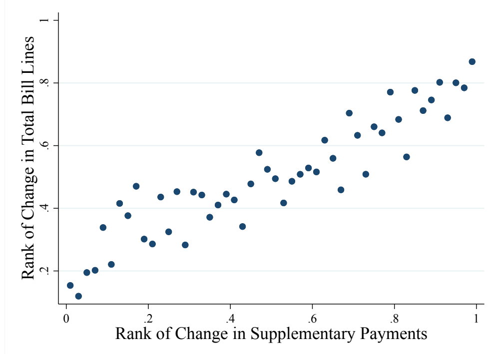
Notes: These plots show averages of treatment intensity outcomes across patient-months in the estimation and control samples in each month relative to certification. Each sample is a balanced panel of patients, and in the estimation sample, Month 0 is the first month in which the registered physician received a certification supplement.

Figure B.3: Long-Run Means Relative to Certification

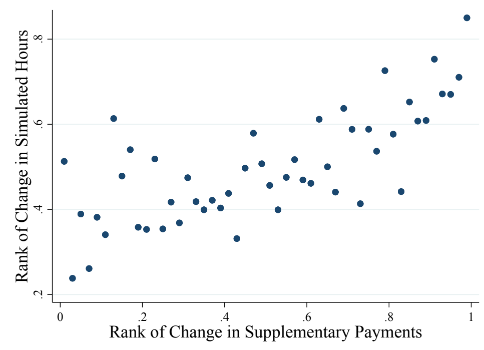


Notes: These plots show average across physicians in the estimation and control samples in each month relative to certification. Each sample is a panel of physicians, and in the estimation sample, Month 0 is the first month in which the registered physician received a certification supplement. Unlike in other analyses like Figures 2.2 and B.2, each observation used to generate plots reflects aggregate labor supply of physicians, rather than the subset of treatment in a balanced sample of registered patients. Physicians must be registered to at least one list in each of the 49 months, but that list may change and labor supply may be zero in a given month. Bill lines include the certification supplement. Per Patient indicates that the aggregate is divided by the total number of registered patients.

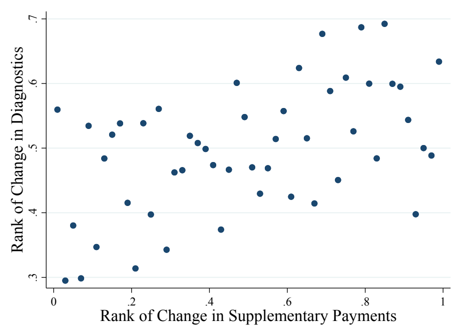
Figure B.4: Similarity of Physician-Specific Responses to Certification Across Treatment Types



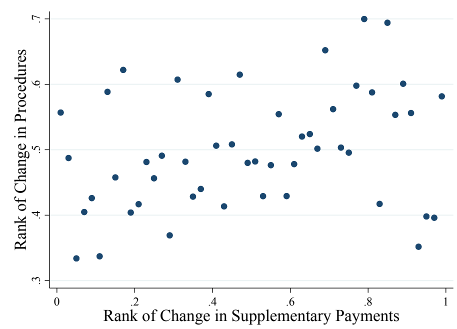
(a) Total Bill Lines



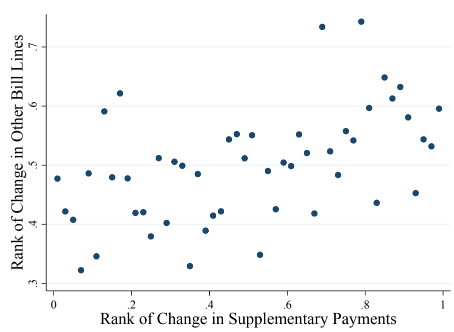
(b) Simulated Hours



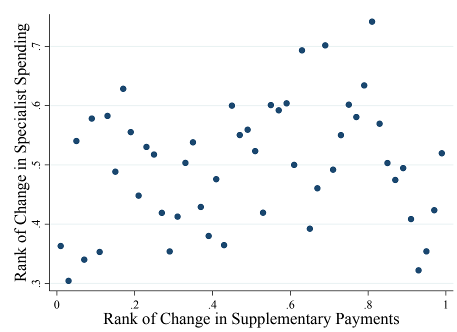
(c) Diagnostics



(d) Procedures



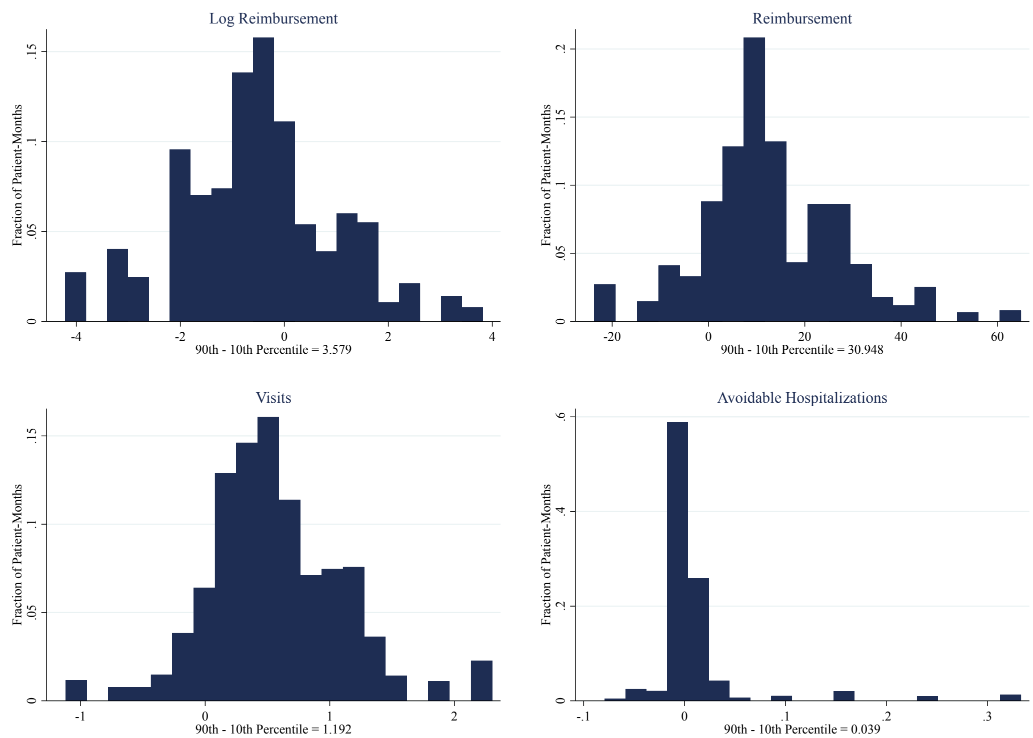
(e) Other Bill Lines



(f) Specialist Spending

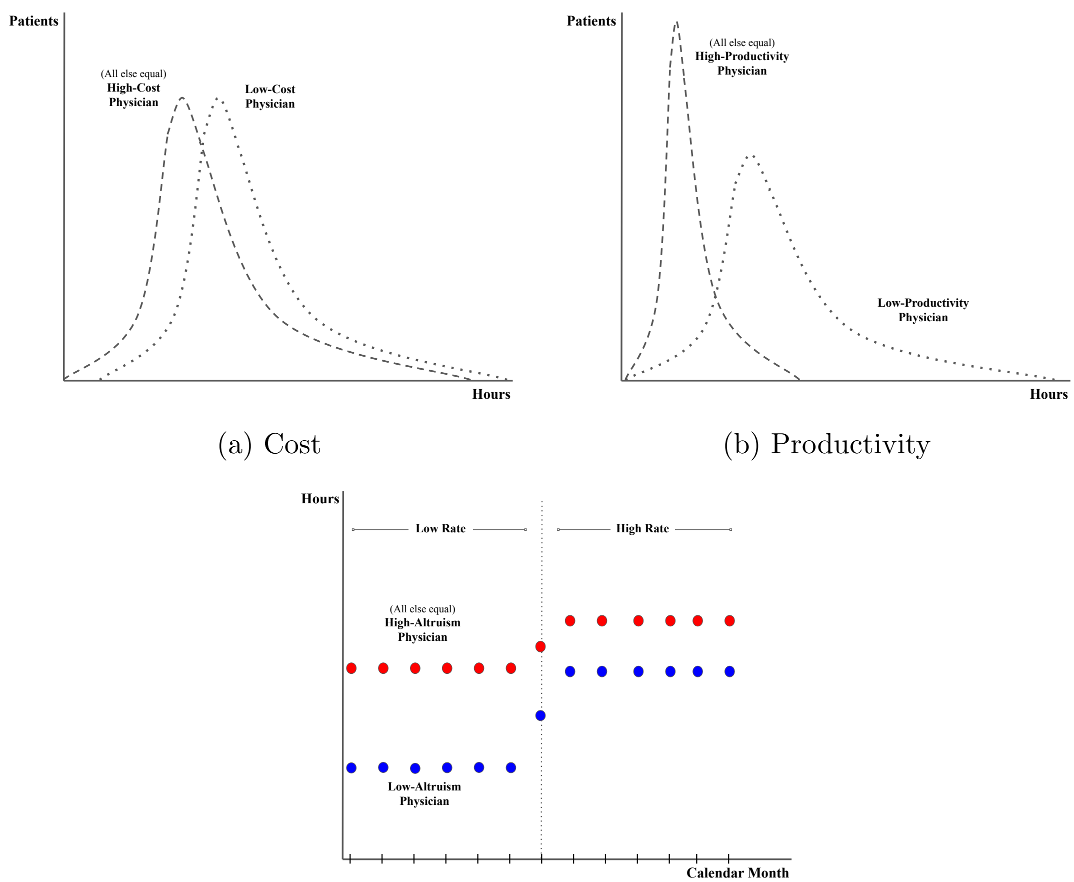
Notes: These plots show that physicians with relatively large increases in supplementary payments post-certification also have relatively large post-certification increases in other measures of treatment intensity. I first take means across patient-months in the six months pre-certification and the six months post-certification, for each physician. Next, I calculate the percentile rank across physicians of $Post - Pre$. Each panel use a different treatment measure to construct the y-axis. Each point is a mean for one of 50 quantiles along the x-axis. The sample includes certified physicians.

Figure B.5: Shrunk Assignment Effects for Certified Physicians



Notes: These histograms show the distribution assignment effects among physicians in the main estimation sample. Following Ginja et al. (2022), I estimate assignment effects by comparing patients from the same exiting physician who are conditionally randomly assigned to various focal physicians. Assignment effects are focal physician fixed effects from a regression including fixed effects for the exiting physician and calendar year. To reflect conditional randomness, I add controls for focal physician availability and an indicator for the same municipality. All estimates are shrunk to the mean using Empirical Bayes, where within- and across-physician variance are estimated using the full list of patients. All dependent variables are per-patient monthly averages during the (up to) six months after assignment.

Figure B.6: Stylized Example of Identification Intuition



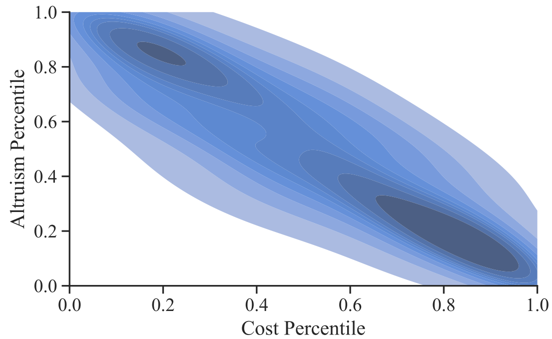
(a) Cost

(b) Productivity

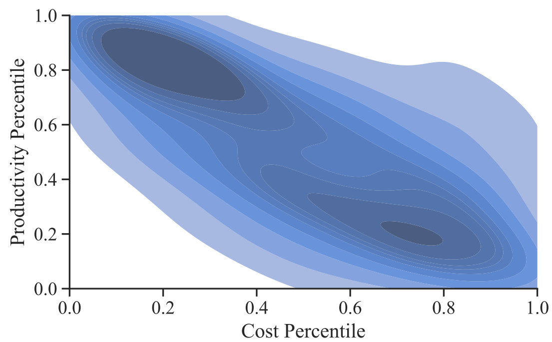
(c) Altruism

Notes: These plots illustrate the identification intuition of physician heterogeneity for the main specification ($\sigma = 0$). All else equal, cost represents a level shift in the distribution of treatment intensity, productivity increases the dispersion of that distribution, and altruism lowers responsiveness to fee-for-service rates.

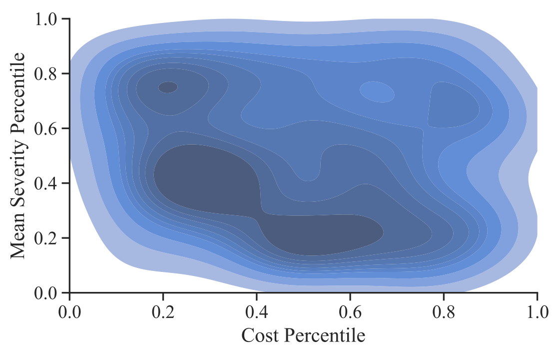
Figure B.7: Distribution of Physician Heterogeneity



(a) Cost and Altruism



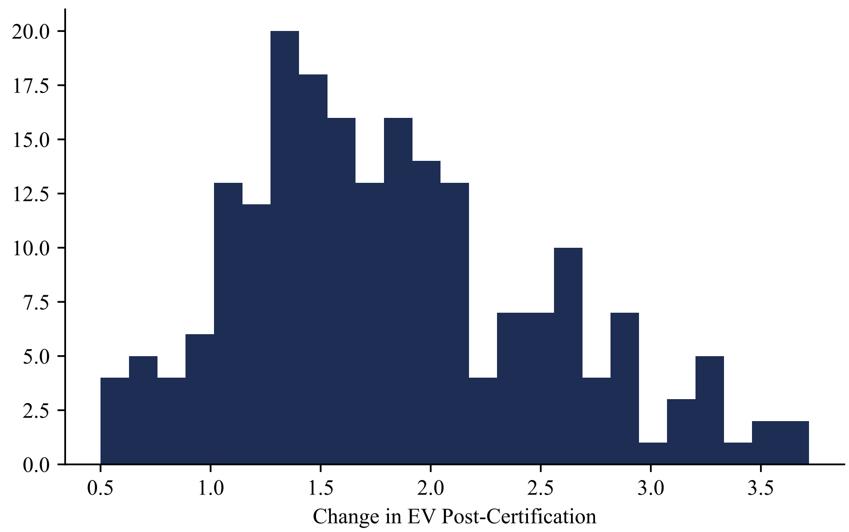
(b) Cost and Productivity



(c) Cost and Severity

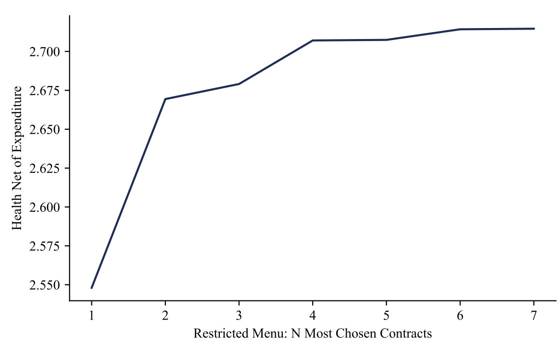
Notes: These plots summarize the joint distribution of estimated cost, altruism, and productivity across the estimation sample. Darker regions indicate higher density.

Figure B.8: Change in Expected Indirect Utility from Certification



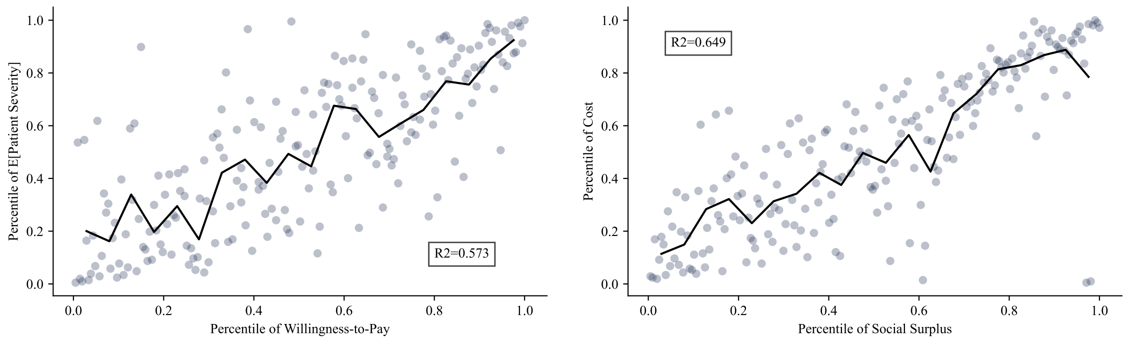
Notes: The y-axis is the count of physicians in each bin. The x-axis is the difference in average expected indirect utility (per patient-month) after certification minus before certification. Integration uses 6 quadrature nodes.

Figure B.9: Restricted Menus Achieve Less Welfare



Notes: The y-axis is expected scaled health production net of expenditure. The x-axis is the number of contracts per menu. For each menu, I re-solve for optimal base payments. I focus the search on the optimal menu's N most chosen contracts. I restrict this function to be non-decreasing when setting the base payment for the marginal contract.

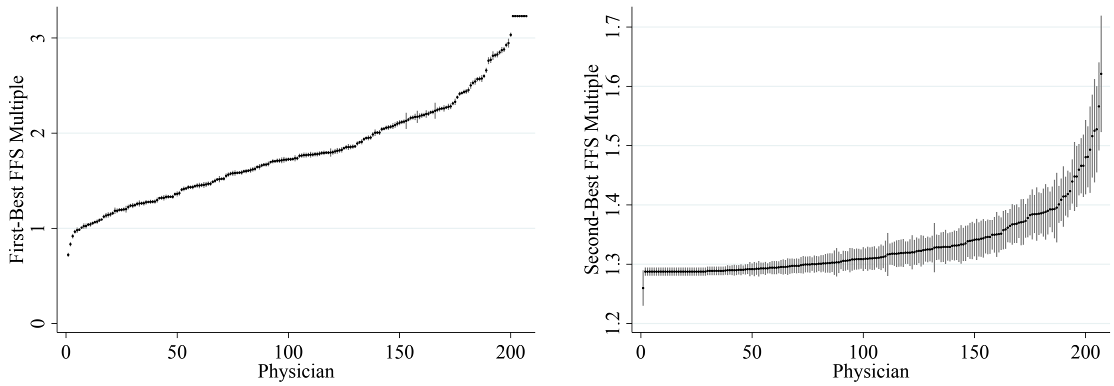
Figure B.10: Two-Contract Menus: Correlations with Physician Type



(a) Private Gain and Mean Patient Severity (b) Social Surplus and Cost

Notes: This figure plots, across physicians, the correlation between each incremental outcome from the two-contract menu in Figure 2.8a and its strongest predictor by bivariate R^2. I separately regress the outcomes (private gain and social surplus) on percentiles of each dimension (cost, altruism, production). The R^2 statistics for private gain are 0.038 for c, 0.010 for α, and 0.041 for γ. The R^2 statistics for social surplus are 0.588 for α, 0.097 for λ, and 0.096 for mean patient severity. Private gain is the difference in expected indirect utility between the high- and low-fee-for-service contracts. Social surplus is the difference between contracts in expected (scaled) health production minus expenditure.

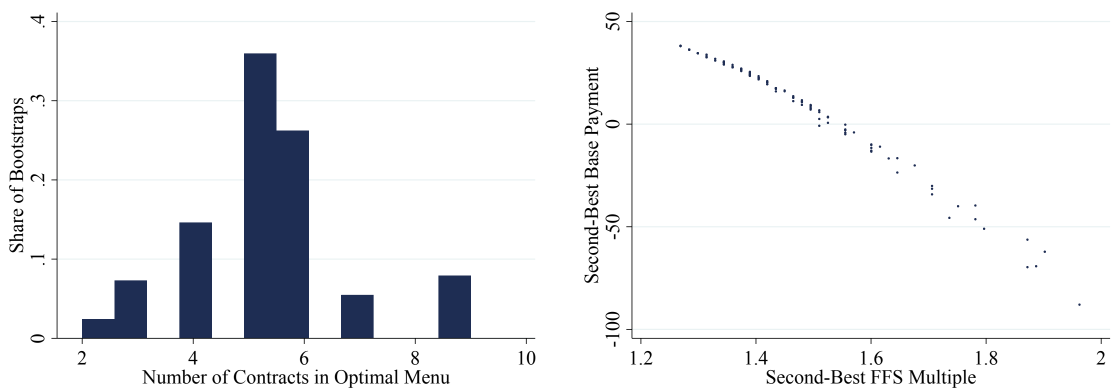
Figure B.11: Physician-Specific Contracts Across Bootstrap Samples



(a) First-Best Fee-For-Service Multiple By Physician (b) Second-Best Fee-For-Service Multiple By Physician

Notes: This figure plots the distribution of fee-for-service multiples across bootstrap samples for each physician. The x-axis is sorted separately for each panel, by mean Fee-For-Service Multiple. Error bars represent the bootstrapped 95 percent confidence interval. In each bootstrap sample, patient-months are randomly selected to maintain the original sample size per physician. In First-Best, the regulator has perfect information about physician types and offers each physician her efficient rate. In Second-Best, given imperfect information, the regulator designs the optimal menu of contracts and each physician self-selects a contract.

Figure B.12: Optimal Menu Across Bootstrap Samples

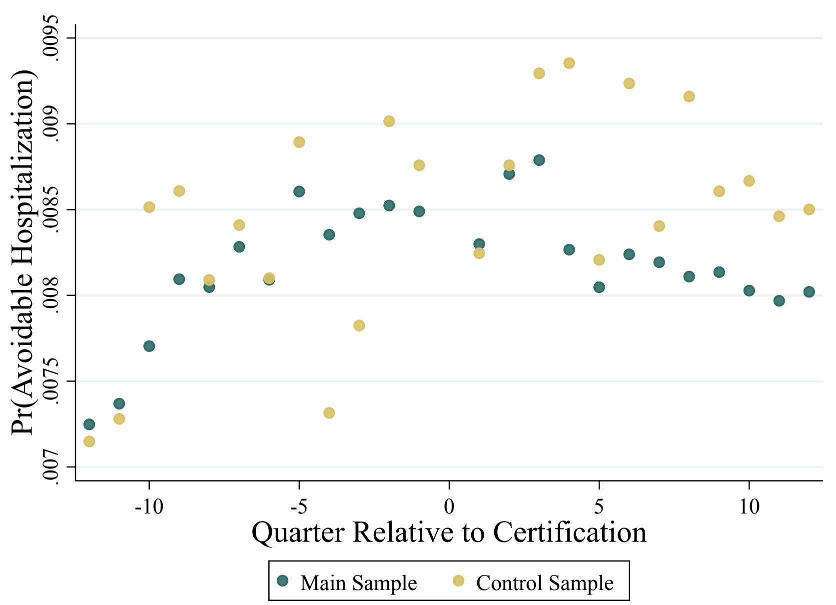


(a) Number of Contracts

(b) Base Payment Schedule

Notes: The left panel plots the distribution of the number of contracts in the optimal menu across bootstrap samples. Only contracts selected by two or more physicians are included. The right panel plots every optimal menu contract selected by two or more physicians, pooled across bootstrap samples. A contract is a pair of the reimbursement rate multiple and base payment. In each bootstrap sample, patient-months are randomly selected to maintain the original sample size per physician. In Second-Best, given imperfect information, the regulator designs the optimal menu of contracts and each physician self-selects a contract.

Figure B.13: Certification and Avoidable Hospitalization



Notes: The plot shows the share of patients with an avoidable hospitalization in each quarter over six years. Each point is a mean across physicians. The underlying data includes patients who were consistently registered in the six months surrounding certification. The denominator of the physician-quarter patient share is the number of patients who are alive. Outside of Quarters -2 to 2, patients are not necessarily registered to the focal physician. The Main Sample includes certified physicians and the Control Sample includes non-certified physicians with randomly selected focal months. Both samples are restricted to physicians with certification dates in 2011-2014 so that utilization data exists in across all 72 months.

B.1.2 Additional Tables

Table B.1: Types of Reimbursement Codes

	Volume	Count	Examples
Time/Talking	48%	10	Consultation with GP; Supplement for 20+ min visit; Remote patient contact
Testing	22%	8	Taking lab samples; Immunological CRP test; Glucose dry chemical analysis; Thrombotest/INR test
Materials	4%	4	Local anesthetic; Equipment for Category 2 (e.g., ECG)
Procedures	1%	1	Major surgical procedures; Minor surgical procedures
Other	18%	3	Continuing educ. supplement
Infrequently Used	8%	163	Surcharge for biopsy; Finger; Wrist region; Travel Supplement

Notes: This table classifies the top 26 reimbursement codes by volume into categories. All other codes representing 8 percent of volume are included in the final row. Volume is the share of reimbursement lines and Count is the number of unique codes in each category. Examples include a selection of translated descriptions for reimbursement codes.

Table B.2: Sample Selection

	Physicians	Patients
Total Personnel	12,677	
Registered to Patient List	8,928	
Linkable to Utilization	7,956	
Overlapping Certification	1,288	
Fixed and Present Physician	1,269	
Balanced 13-Month Spell	714	799,083
Balanced Patient Panel	619	643,363

Notes: This table shows the number of remaining physicians after each sample selection criterion which are applied cumulatively. The utilization for a particular physician-patient pair is available if both the physician and patient are citizens or permanent residents with tax identifiers. Not all certified physicians receive their certification during the sample period. "Fixed and Present" indicates that each physician is linked to exactly one patient-list with every month at a single location throughout the spell, so the list has no change in the associated list; moreover, neither the physician nor the list exits within or immediately after the spell. Spells are balanced when all prior conditions are met for the six months before and after certification, rather than in at least one month.

Table B.3: Means by Patient Type

	Patients	Share	Age	Chronic	Spend	FFS Rate	Hours
1	147,775	0.115	10.484	0.000	2.646	32.973	0.081
2	96,503	0.075	32.094	0.099	5.080	47.685	0.107
3	83,275	0.065	40.384	0.122	5.765	45.822	0.126
4	54,410	0.042	37.941	0.055	8.752	45.807	0.192
5	65,015	0.051	41.193	0.001	9.331	46.662	0.201
6	51,919	0.040	43.938	0.041	10.248	46.466	0.222
7	50,825	0.039	59.143	0.501	11.671	47.772	0.246
8	35,968	0.028	66.521	0.760	15.302	45.837	0.336
9	33,473	0.026	59.451	1.000	18.823	48.723	0.388
10	24,200	0.019	72.333	1.000	25.271	50.351	0.504

Notes: Summary statistics reflect patients' monthly totals six months before certification in the estimation sample. Monetary measures are in USD. Hours are total reimbursement divided by a wage index.

Table B.4: Registered Patient Summary Statistics

	Control Sample	Estimation Sample					
	Mean	Mean	Std. Dev.	% > 0	10th	50th	90th
Patient Characteristics							
Reimbursement	8.59	8.33	25.49	20.74	0.00	0.00	30.92
Simulated Hourly Rate	43.82	43.76	6.86	100.00	32.38	45.49	50.95
Simulated Hours	0.19	0.18	0.56	20.74	0.00	0.00	0.68
Base Payment	4.03	4.01	0.11	100.00	3.84	4.02	4.13
Visits	0.37	0.34	0.84	20.76	0.00	0.00	1.00
Hours	0.11	0.10	0.29	20.78	0.00	0.00	0.33
Reimbursement Lines	0.90	0.87	2.59	20.79	0.00	0.00	3.00
Procedures	0.06	0.07	0.57	3.55	0.00	0.00	0.00
Diagnostics	0.24	0.22	0.99	8.04	0.00	0.00	0.00
Extra Time	0.10	0.08	0.45	5.03	0.00	0.00	0.00
Clinic Reimbursement	2.49	2.84	101.22	7.43	0.00	0.00	0.00
Specialist Reimbursement	19.84	19.24	86.66	22.88	0.00	0.00	59.67
Acute Hospitalizations	0.02	0.02	0.22	1.38	0.00	0.00	0.00
Age	40.54	37.57	22.78	100.00	6.67	36.58	69.00
Female	0.48	0.50	0.50	50.42	0.00	1.00	1.00
Chronic Illness	0.23	0.21	0.41	21.03	0.00	0.00	1.00
New Patient	0.20	0.10	0.29	9.59	0.00	0.00	0.00
Disability	0.07	0.06	0.25	6.42	0.00	0.00	0.00
Physician Characteristics							
Enrollment	1201.99	1225.23	299.93	100.00	867.00	1197.00	1589.00
Max Enrollment	1268.60	1273.48	293.21	100.00	900.00	1220.00	1600.00
Physician Hours/Week	28.36	26.56	9.44	100.00	13.13	27.33	37.27
Female Physician	0.45	0.43	0.49	42.94	0.00	0.00	1.00
Physician Age	42.87	40.23	5.92	100.00	34.08	38.83	48.67
Migrant Physician	0.27	0.28	0.45	27.82	0.00	0.00	1.00
Pr(Diagnostic)	0.81	0.76	0.10	100.00	0.63	0.77	0.87
Ever Fixed-Salary	0.01	0.03	0.17	2.82	0.00	0.00	0.00
Patients Age 60+	0.23	0.19	0.10	100.00	0.07	0.18	0.32
Patients with Chronic Illness	0.23	0.21	0.06	100.00	0.14	0.20	0.29
Patients	131800	643363					
Physicians	136	619					

Notes: Summary statistics reflect patients' monthly totals six months before certification (or the control month 0 for the control sample). % > 0 indicates the share of patients with a strictly positive measure (row). Other columns reflect the mean, standard deviation, and 10^{th} , 50^{th} , and 90^{th} percentiles. Monetary measures are in USD. Physician Characteristics are also averaged across patients. The last two Physician Characteristics reflect shares of registered patients.

Table B.5: Registered Patient Summary Statistics versus Population

	Population	Estimation Sample					
	Mean	Mean	Std. Dev.	% > 0	10th	50th	90th
Patient Characteristics							
Age	38.436	37.225	22.684	1.000	6.417	36.250	68.417
Female	0.495	0.505	0.500	0.505	0.000	1.000	1.000
Chronic Illness	0.200	0.210	0.407	0.210	0.000	0.000	1.000
Disability	0.060	0.064	0.244	0.064	0.000	0.000	0.000
Physician Characteristics							
Enrollment	1297.232	1235.749	314.715	1.000	880.000	1197.000	1592.000
Female Physician	0.356	0.438	0.496	0.438	0.000	0.000	1.000
Physician Age	49.000	39.777	6.123	1.000	33.500	38.083	49.500
Migrant Physician	0.215	0.226	0.418	0.226	0.000	0.000	1.000
Patients	5525876	215529					
Physicians	4769	207					

Notes: Summary statistics reflect patients' monthly totals. The Population column reflects all Norwegian patients in 2012. All other columns reflect patients in the estimation sample six months before certification (or the control month 0 for the control sample). % > 0 indicates the share of patients with a strictly positive measure (row). Other columns reflect the mean, standard deviation, and 10^{th} , 50^{th} , and 90^{th} percentiles. Physician Characteristics are also averaged across patients.

Table B.6: Distribution of Patient Severity

	Estimate	Std. Err.
Patient Type 2	0.039	(0.001)
Patient Type 3	0.053	(0.001)
Patient Type 4	0.083	(0.001)
Patient Type 5	0.091	(0.001)
Patient Type 6	0.092	(0.001)
Patient Type 7	0.091	(0.001)
Patient Type 8	0.109	(0.001)
Patient Type 9	0.111	(0.001)
Patient Type 10	0.129	(0.002)
February	0.030	(0.001)
March	0.011	(0.001)
April	0.020	(0.001)
May	0.010	(0.001)
June	0.018	(0.001)
July	0.014	(0.001)
August	−0.059	(0.001)
September	0.013	(0.001)
October	0.017	(0.001)
November	0.017	(0.001)
December	0.018	(0.001)
$\log(1 + m_{t-1})$	0.024	(0.000)
$m_{t-1} = 0$	0.050	(0.001)
Cancer	0.010	(0.002)
Diabetes	0.028	(0.002)
COPD	0.031	(0.002)
Asthma	0.018	(0.002)
CVD	0.035	(0.002)
1+ Chronic Illness	0.014	(0.002)
2+ Chronic Illnesses	−0.005	(0.002)
Female	0.001	(0.000)
Disability Receipt	0.055	(0.001)
Income Percentile	−0.013	(0.001)
Recent Acute ER Visit	0.022	(0.001)
Recent Acute ER Visit 2+	0.032	(0.001)
Time Trend	0.009	(0.002)
New Patient	0.006	(0.001)
$\log \sigma_\lambda$	−0.389	(0.003)
$P(\lambda > 0) : d_0$	−3.389	(0.019)
$P(\lambda > 0) : d_1$	11.462	(0.132)

Notes: This table shows model estimates with asymptotic standard errors calculated using the approximate Hessian. Unobserved patient severity is distributed $\ln \lambda \sim N(\beta_\lambda X_\lambda, \sigma_\lambda) \,|\, \lambda > 0$ and $Pr(\lambda > 0) = f(d_0 + d_1 \beta_\lambda X_\lambda)$, where $f(z) = \frac{\exp z}{1+\exp z}$. The first set of estimates corresponds to β_λ.

Table B.7: Counterfactual Outcomes by Physician Location

Physicians		Efficient Contracts		Menu of Contracts		
Type	Share	$\Delta E[h(m)]$	$\Delta E[p\,m+b]$	$\Delta E[h(m)]$	$\Delta E[p\,m+b]$	$\Delta E[V(p)]$
Most Urban: 1	0.11	6.09	1.72	2.10	2.18	1.68
2	0.31	8.90	2.30	3.08	2.42	1.80
3	0.34	7.32	1.99	2.65	2.23	1.68
4	0.16	9.22	2.46	2.57	2.15	1.65
5	0.04	11.11	2.58	3.43	2.51	1.88
Most Rural: 6	0.04	13.50	2.69	4.36	2.84	2.07

Notes: This table shows average outcomes for efficient (personalized) contracts and the optimal menu of contracts, disaggregated across groups of physicians (rows). Physicians are grouped by the centrality index of their municipality. $\Delta E[h(m)]$ represents the change in health production relative to the status quo, for efficient contracts and the optimal menu of contracts. Likewise, $\Delta E[p\,m+b]$ represents incremental expected expenditure and ΔEV represents incremental expected indirect utility. Outcomes are averages across patients within each group, measured in USD. I assume that the less than 1 percent of physicians who do not have a linked municipality are in the most urban category.

Table B.8: Counterfactual Outcomes: Menu for each Patient Type

	$\Delta SS_{Efficient}$	$\Delta SS_{Uniform}$		ΔSS_{Menu}		Menu $\succ$ Uniform
	Level	Level	Share of Eff.	Level	Share of Eff.	
Baseline	8.396	2.548	0.303	2.714	0.323	✓
Patient Type 1	3.190	0.877	0.275	0.977	0.306	✓
Patient Type 2	4.560	1.264	0.277	1.332	0.292	✓
Patient Type 3	6.343	1.928	0.304	1.990	0.314	✓
Patient Type 4	7.810	2.447	0.313	2.520	0.323	✓
Patient Type 5	9.802	2.701	0.276	2.892	0.295	✓
Patient Type 6	11.868	3.389	0.286	3.554	0.299	✓
Patient Type 7	11.844	3.321	0.280	3.505	0.296	✓
Patient Type 8	15.291	4.328	0.283	4.511	0.295	✓
Patient Type 9	19.851	5.593	0.282	5.975	0.301	✓
Patient Type 10	25.702	6.842	0.266	7.185	0.280	✓
All Patient Types	8.586	2.433	0.283	2.569	0.299	✓

Notes: This table compares key outcomes between counterfactual contract menus. All outcomes are based on ex-ante expectations over patient-months using estimated distributions of G and F, weighted across physicians by enrollment. Outcomes are summarized by the change in social surplus, defined as the change in health production versus pre-certification minus the change in expenditure versus post-certification. Share of Eff. divides the change in levels of social surplus for the optimal menu by the change in levels for efficient contracts. Relative to Table 2.4 (included as "Baseline"), each row after the first summarizes a separate analysis for each observed patient type. Analyses are separate in the sense of unique benchmarks, menus, and weighting across physicians. All Patient Types weights the type-specific counterfactual outcomes by share of the main estimation sample.

Table B.9: Counterfactual Outcomes with Perturbations

	$\Delta SS_{Efficient}$	$\Delta SS_{Uniform}$		ΔSS_{Menu}		Menu $\succ$ Uniform
	Level	Level	Share of Eff.	Level	Share of Eff.	
Baseline	8.396	2.548	0.303	2.714	0.323	✓
$0 \times Var(c)$	7.885	2.122	0.269	2.464	0.313	✓
$\frac{1}{2} \times c$	3.423	2.183	0.638	2.184	0.638	✓
$2 \times c$	5.560	1.194	0.215	1.332	0.240	✓
$2 \times Var(c)$	15.123	2.361	0.156	2.361	0.156	
$0 \times Var(\alpha)$	8.664	2.606	0.301	2.921	0.337	✓
$\frac{1}{2} \times \alpha$	5.838	2.005	0.343	2.040	0.349	✓
$2 \times \alpha$	11.188	2.791	0.249	3.178	0.284	✓
$2 \times Var(\alpha)$	9.978	2.327	0.233	2.327	0.233	
$0 \times Var(\gamma)$	8.645	2.564	0.297	2.652	0.307	✓
$\frac{1}{2} \times \gamma$	2.892	0.881	0.305	0.933	0.322	✓
$2 \times \gamma$	22.371	5.519	0.247	6.030	0.270	✓
$2 \times Var(\gamma)$	8.733	2.542	0.291	2.733	0.313	✓
Uncorrelated c, α, γ	10.215	2.117	0.207	2.176	0.213	✓
Drop Outliers of c, α, γ	8.993	2.576	0.286	2.802	0.312	✓
$\frac{1}{2} \times Var(\theta_k), \theta_k \in c, \alpha, \gamma$	8.416	2.721	0.323	2.998	0.356	✓
$0 \times Var(\gamma), 0 \times Var(\alpha)$	8.680	2.763	0.318	2.991	0.345	✓
$0 \times Var(c), 0 \times Var(\alpha)$	7.622	2.466	0.324	2.819	0.370	✓
$0 \times Var(c), 0 \times Var(\gamma)$	8.318	2.124	0.255	2.421	0.291	✓
$\frac{1}{2} \times \sigma_\lambda$	6.446	1.732	0.269	1.803	0.280	✓
$2 \times \sigma_\lambda$	23.791	5.530	0.232	6.456	0.271	✓
$\frac{1}{2} \times \alpha_G$	4.449	1.324	0.298	1.310	0.294	
$2 \times \alpha_G$	16.599	4.991	0.301	5.667	0.341	✓
Add Control Sample	9.681	4.010	0.414	4.161	0.430	✓
Constrain Capacity	17.524	2.063	0.118	4.376	0.250	✓
Exclude Part-Time Physicians	8.781	2.559	0.291	2.730	0.311	✓
Only Urban Physicians	8.374	2.561	0.306	2.737	0.327	✓
Only Rural Physicians	9.360	2.644	0.282	2.788	0.298	✓
Alt. Health Parameterization	8.426	2.561	0.304	2.737	0.325	✓

Notes: This table compares key outcomes between counterfactual contract menus. All outcomes are based on ex-ante expectations over patient-months using estimated distributions of G and F, weighted across physicians by enrollment. Outcomes are summarized by the change in social surplus, defined as the change in health production versus pre-certification minus the change in expenditure versus post-certification. Share of Eff. divides the change in levels of social surplus for the optimal menu by the change in levels for efficient contracts. Relative to Table 2.4 (included as "Baseline"), each row perturbs one or more parameters before repeating counterfactual analyses. The parameters are marginal cost c, altruism α, productivity γ^{-1}, standard deviation of the log patient severity σ_λ, and altruism of the regulator α_R. $0 \times Var(c)$ fixes c at the sample mean. $\frac{1}{2} \times c$ multiplies c by 0.5 for all physicians. $2 \times Var(c)$ uses the following function: $f(c) = \bar{c} + \sqrt{2} \times (c - \bar{c})$. Outliers are below the 1st percentile or above the 99th of c, α, or γ. In one perturbation, I impose a capacity constraint on simulated hours per physician-month and approximate the shadow cost of capacity (see Appendix B.1.3 for details). Rural physicians are in municipalities with low centrality indexes. Part-Time physicians spend less than 25 hours per week with patients in the six months before certification.

Table B.10: Test for Patient Sorting

	Predicted Health (SDs) (1)	Switch (2)	Hospitalization (3)	Mortality (4)
Post-Certification	0.133^{***}			
	(0.044)			
Predicted Health (SDs)		-0.001	-0.003^{***}	-0.001^{*}
		(0.004)	(0.001)	(0.001)
Age	-0.014^{***}	-0.001^{***}	-0.000^{***}	-0.045^{***}
	(0.002)	(0.000)	(0.000)	(0.002)
Asthma	-0.200^{***}	-0.005	0.052^{***}	0.016
	(0.075)	(0.009)	(0.014)	(0.015)
Age^2	0.000^{***}	0.000^{***}	0.000^{***}	0.000^{***}
	(0.000)	(0.000)	(0.000)	(0.000)
Cancer	-0.183^{***}	0.006	-0.025^{*}	0.072^{***}
	(0.070)	(0.008)	(0.014)	(0.015)
COPD	-0.291^{***}	0.033^{***}	0.181^{***}	0.108^{***}
	(0.087)	(0.009)	(0.015)	(0.016)
CVD	-0.341^{***}	0.007	-0.034^{**}	0.025^{*}
	(0.089)	(0.008)	(0.014)	(0.015)
Diabetes	-0.400^{***}	-0.004	-0.036^{***}	0.035^{**}
	(0.113)	(0.008)	(0.014)	(0.015)
Female	-0.164^{***}	0.009^{**}	-0.001	-0.001
	(0.033)	(0.004)	(0.001)	(0.002)
Income Percentile	0.099^{***}	0.022^{***}	-0.010^{***}	0.085^{***}
	(0.034)	(0.005)	(0.001)	(0.006)
1+ Chronic Illness	-0.111	-0.009	0.038^{***}	-0.037^{**}
	(0.086)	(0.008)	(0.014)	(0.015)
2+ Chronic Illnesses	-0.046	-0.000	0.040^{***}	-0.032^{*}
	(0.069)	(0.009)	(0.014)	(0.017)
Observations	2583264	215272	215272	54192
R^2	0.102	0.022	0.043	0.133
Outcome mean	-0.013	0.060	0.021	0.042

Notes: This table shows estimates of the correlation between patients' model-predicted health production and outcomes of interest measured after the estimation sample. Health production is normalized to standard-deviation units within the estimation sample. All specifications include year and calendar month fixed effects and cluster at the physician level. Column (1) includes the entire spell and regresses expected health production given parameter estimates on an indicator for months after specialization and patient covariates. Columns (2)-(4) are cross-sectional regressions using expected health production as the treatment variable of interest. The dependent variables are an indicator for switching to a new physician within 6 months, an indicator for an avoidable hospitalizations in the next 12 months, and an indicator for mortality within the next 24 months. Column (4) includes patients over 45 years old. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level.

B.1.3 Income Effects and Capacity Constraints

This section extends the main model to cases with decreasing returns to treatment intensity from higher reimbursement rates. The first case is lower marginal utility of marginal reimbursement for high-workload physicians: income effects. High workload is driven by differences between physicians in the number of patients ("enrollment") and those patients' expected severity ("composition"). Moreover, income effects introduce complementary between the treatment intensity decisions of various patients. For example, increasing the treatment intensity for patient 1 may increase the marginal utility of leisure, lowering treatment intensity for patient 2. To tractably model this dynamic, I assume that patients arrive sequentially and only short-term future treatment intensity affects the marginal utility of leisure.[1] Equivalently, a physician will treat a patient slightly less intensively if that physician expects to work many hours over the next month treating other patients. As before, for each patient $i \in 1, ..., N$, the health shock is realized only when that patient arrives. The private objective becomes:

$$EV(x; \lambda_i, F, \theta) = \max_{m_i \geq 0} \; x(m_i) - c(m_i) + \sigma E\left[l\left(\sum_{i'=1}^{N} m_{i'}^*\right) \mid F(\lambda_{i'})\right] + \alpha h(m_i, \lambda_i)\,, \tag{B.1}$$

The additional term ($\sigma E\left[l\left(\sum_{i'=1}^{N} m_{i'}^*\right)\right]$) represents the money-metric distaste for expected workload. The expectation enters because, before arrival, each future patient i' has uncertain severity.

The key assumption is that the expected (but not realized) treatment of one patient may affect the privately optimal choice for another patient of the same physician: $\frac{dm'}{dm} = 0$. Physicians anticipate the effect of making similar choices on the marginal utility of leisure. With this assumption, each patient's likelihood depends

[1]Alternatively or additionally, I could relax the assumption that the marginal utility of net income equals 1 by introducing curvature, but that approach unnecessarily complicates the expression for physicians' private gain from switching contracts.

on an independent draw of their own severity, along with the contract and the number and composition of other patients. In estimation, I assume quadratic preferences, $l(x) = -\frac{(x)^2}{2}$, so the marginal utility of leisure is strictly positive and increases exponentially in the expected number of hours worked, and I substitute observed average treatment intensity for expected treatment intensity since the two should coincide at true parameters. The privately optimal level of treatment intensity becomes:

$$m^*(p, \lambda, (N-1)\bar{m}) = \max\{0, \frac{p - c - \sigma(N-1)\bar{m} + \alpha\gamma\lambda}{\alpha + \sigma}\} \tag{B.2}$$

and the likelihood is constructed as before by inverting for ϵ_λ.

For identification intuition, it is helpful to first discuss two reduced-form parameters. Given *any* distribution of patient severity and additive quadratic health production, the first-order condition can be simplified to $m = \max\{0, \beta_0 + \beta_1\lambda\}$ where the level β_0 and slope β_1 are specific to a combination of physician and time period. It could also be specific to patient observables. Generally, to identify β_0 and β_1, these quantities need to be independent of (the random component of) λ. To separably identify β_1 from parameters governing $F(\lambda)$, a physician needs to be observed for at least two periods with the same distribution of patients and no model-predicted change to β_1. In that case, repeated draws of λ drive variation in m, so conditional moments of m match the corresponding moments $F(\lambda)$. Linear separability between utility from net income and health production implies that β_0 and β_1 are constant for a physician if the reimbursement rate and the set of patients are constant. Given β_1 and the distribution of λ, β_0 is identified by the responsiveness of a physician's average treatment intensity (over patients), relative to other physicians or time periods.

The marginal rate of substitution between leisure and net income σ is identified by the responsiveness of β_0 to the number (N) and composition ($\bar{m}$) of patients within physician over time. Given σ and practice characteristics, the responsiveness of β_0 and β_1 to fee-for-service over time within-physician identifies altruism. Critically, this requires observing treatment intensity choices for the same physicians at different fee-for-service rates, which only occurs in the certification sample. Persistent residual variation in β_1 identifies productivity and persistent residual variation in β_0 identifies cost. Only altruism must be time-invariant; all other parameters can be both physician-specific and time-varying, including curvature of preferences over leisure. However, for estimation, I assume time-invariance and symmetric σ because implied β_0 and β_1 may be noisy even with large data leading to overestimation of across-time variance in physician heterogeneity.

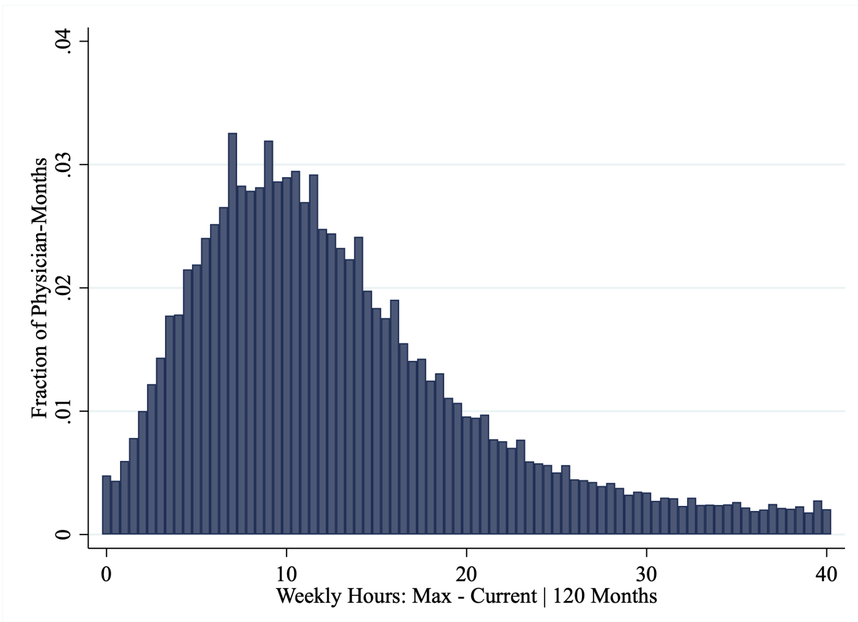
Consistent with prior studies that find treatment intensity increases in marginal reimbursement, likelihood ratio tests fail to find evidence of income effects.[2] Although simulated hours of treatment do not increase with fee-for-service rates for some physicians, high altruism and large variance in patient health shocks better explain this pattern than income effects – marginal utility of leisure increasing in the expected workload.

In addition to income effects, capacity constraints may limit counterfactual treatment intensity from greater fee-for-service rates. For example, physicians may only be able to treat patients up until a threshold number of hours each month ($\sum_{i=1}^{N} m_i \leq \bar{M}$). If capacity constraints sometimes bind, then over a long period (120 months) with idiosyncratic variation in enrollment, composition, and realized severity, some physicians' monthly total treatment intensity should bunch near the maximum. I instead find that the distribution of treatment intensity relative to a physician-specific maximum is relatively smooth near the maximum.

The main findings are also robust to imposing capacity constraints (See Table B.9). Intuitively, adding a capacity constraint means reducing each treated patient's treatment intensity by a fixed amount per physician-month – excess total hours per treated patient – where excess total hours is the difference between unconstrained total hours and capacity. The more general first-order condition is $m^*(p) = \max\{0, \frac{p-c-\mu_c}{\alpha} + \gamma\lambda\}$. Substituting this condition into the capacity constraint

[2] In estimation, I search over positive scaled values of σ.

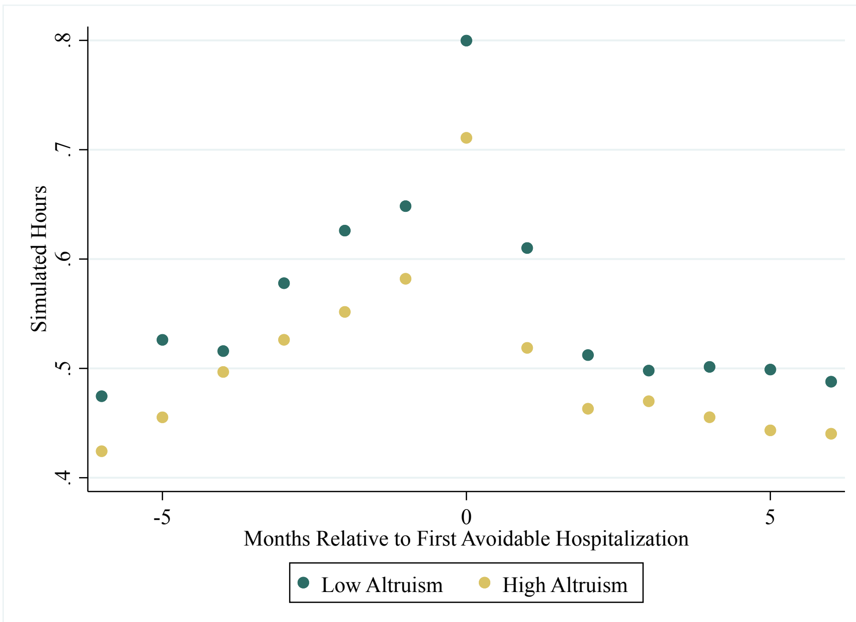
Figure B.14: Capacity Constraints: Hours Do Not Bunch Near Each Physician's Maximum



Notes: This figure shows the distribution of transformed hours per week ($\tilde{M}_{jt}$) across physician-months ($j-t$). The transformation is $\max_t M_{jt} - M_{jt}$. The x-axis is truncated at 40 and I exclude the first month when a physician works the maximum number of hours. According to the theoretical framework, $M_{jt} = \sum_i^{N_{jt}} \text{argmax}\, u(x(m_{ijt}) - c(m_{ijt})) + \alpha h(m_{ijt}, \gamma\lambda_{ijt})$, s.t. $\sum_i^{N_{jt}} m_{ijt} \leq \bar{M}_j$, where λ_{ijt} is stochastic. If capacity binds and $F(\lambda)$ is continuous, then $Pr(M_{jt} = \bar{M}_j \equiv \max_t M_{jt}) >> Pr(M_{jt} = \bar{M}_j - \epsilon)$ for small $\epsilon > 0$.

pins down the shadow cost of capacity, $\mu_c = \alpha(\frac{\sum_i \max\{0, \frac{p-c-\mu_c}{\alpha} + \gamma\lambda\} - \bar{M}_j}{N_{jt} Pr(m^*(p)>0)})$. An exact μ_c is a fixed point of this equation which varies for each physician-month pair. This fixed point may not converge with quadrature, so for the robustness check, I approximate it as $\hat{\mu}_c = \alpha(E[m^0_{ijt} \mid m^0_{ijt} > 0] - \frac{\bar{M}}{N_{jt} Pr(m^0_{ijt}>0)})$ where $m^0_{ijt} = \max\{0, \frac{p-c}{\alpha} + \gamma\lambda\})$ is the unconstrained treatment intensity.[3]

Figure B.15: Treatment Intensity Responds to Health Shocks



Notes: This figure shows average simulated hours across patient months in the six months before and after each patient's first avoidable hospitalization. The sample includes pre-certification patient-months for a balanced panel of consistently registered patients and is subset by whether the registered physician's estimated altruism is above or below the sample median.

Finally, I conclude that altruism estimates are not biased because high-altruism physicians are not contained from increasing treatment intensity when a patient has an avoidable hospitalization. Estimates of high altruism reflect that some physicians are less responsive to increased reimbursement rates. These estimates may

[3] Two further adjustments help limit approximation error. First, I bound the denominator below by 1. Second, I use 0.5 as a threshold for m^0_{ijt} in the denominator to avoid over-correcting based on mass near zero treatment intensity. This threshold represents approximately the 90^{th} percentile of status quo treatment intensity.

be biased if the low response reflects some unobserved constraint rather than altruism. B.15 shows that the mean treatment intensity of high-altruism and low-altruism physicians is similarly responsive to the shock of a first avoidable hospitalization.

B.1.4 Selection into Certification

To empirically estimate the model outlined above, I rely on plausibly exogenous within-physician variation in reimbursement rates generated by receiving certification as a general practitioner. 80 percent of physicians receive this certification at some point in their career, and the estimation sample includes a fraction of these. If certified physicians in the estimation sample are selected on unobserved heterogeneity, then counterfactuals lack external validity for the full population of physicians. This section extends the model to account for potential selection and test its implications. Although this extended model could be fully estimated, I find that estimates using the subset of physicians are similarly predictive of treatment intensity in a control sample of never-certified physicians, and conclude that selection is not a first-order concern for the main research question.

Physicians choose to become certified if the increase in indirect expected utility outweighs the cost of certification and difference in iid taste shocks:

$$\max_{S,NS} \left\{ E_\lambda V(p + p_S; \theta, F(\lambda)) - C_s + \epsilon_S, E_\lambda V(p; \theta, F(\lambda)) + \epsilon_{NS} \right\} .$$

I include taste shocks for certification choice but not counterfactual contract choice because certification requires additional training with idiosyncratic benefits and costs, rather than a purely financial change with impacts fully characterized by physician type. The key assumptions here are the constant cost of certification and independence between taste shocks, physician type, and patient severity. These might be violated if, e.g., only some physicians have binding time constraints outside of work with registered patients. Another assumption is that certification (with required training) does not impact health production, but this can be relaxed. Consistent with empirical findings, this model of certification assumes that certification does not change the distribution of registered patients F or the number of patients. If the cost of certification is large relative to taste shocks, then the distribution of types who become certified will differ from the unconditional distribution.

This model helps guide intuition about how physicians in the estimation sample might be selected on unobserved heterogeneity. Larger draws of taste shocks might drive certification, which would not impact external validity. However, if the costs of certification are relatively large, then certified physicians have greater private gain from switching to the certified fee-for-service rate. Section B.3.1 shows that such physicians have relatively low cost, high altruism, and high productivity. As a result, estimates should be less predictive of treatment intensity out-of-sample. To test this, I follow a similar estimation procedure to recover all parameters besides the set of α in the control sample. I use the correlation between $\ln \alpha$ and observed physician characteristics to predict α in the control sample and then hold those values fixed. Table B.11 shows regression of actual treatment intensity m on predicted $E[m]$. Although the differences between the samples are precise, they are small. The coefficient on $E[m]$ is just as far from 1 in both samples but in opposite directions, and disappears with fixed effects, suggesting that selection on unobserved heterogeneity is minimal.

Estimates are consistent with physicians rationally choosing to become certified. All physicians experience an increase in expected indirect utility (EV). B.8 shows the distribution of this change in EV across physicians. The large average increase in EV and a symmetric (rather than left-skewed) distribution suggest minimal selection on unobserved heterogeneity.[4]

B.1.5 Optimal Nonlinear Uniform Contract

This paper primarily investigates contracts in which revenue is a linear function of treatment intensity. This structure nests the ways healthcare providers are typically

[4]Since most physicians in the sample waited several years to become certified despite large potential increases in EV, taste shocks of certification must be large relative to costs.

Table B.11: Test for Selection on Unobserved Physician Heterogeneity

	Certified	Non-Certified	Certified and Non-Certified		
	(1)	(2)	(3)	(4)	(5)
$E[m]$	1.041***	1.025***	1.032***	1.041***	1.087***
	(0.002)	(0.005)	(0.002)	(0.002)	(0.003)
$E[m] \times$ Control				−0.016***	−0.018***
				(0.005)	(0.005)
Control				−0.001	
				(0.001)	
Female					−0.013***
					(0.001)
Age					−0.000***
					(0.000)
Chronic Illnesses					−0.021***
					(0.001)
Intercept	−0.007***	−0.008***		−0.007***	
	(0.001)	(0.001)		(0.001)	
Physician FEs			✓		✓
Observations	2013672	385416	2399088	2399088	2399088
R^2	0.113	0.108	0.114	0.112	0.114

Notes: All regressions use observed treatment intensity as the dependent variable. The control (Non-Certified) sample is constructed identically to the main estimation (Certified) sample, except that the starting pool of physicians is a random subset of those that never become certified. The last three columns pool both samples. $E[m]$ is calculated based on parameter estimates given observable characteristics. Control is an indicator for the control sample.

reimbursed in most settings. Linear contracts may be common because they are relatively easy to implement and understand.[5] However, larger welfare gains may be possible when revenue is a flexible function of treatment intensity. For example, after a large amount of treatment, the marginal return to health may be small, so low marginal reimbursement can limit relatively inefficient spending.

The optimal nonlinear uniform contract performs about half as well as efficient linear contracts in terms of improving patient health, but the participation constraint requires large increases in expenditure, so the gains to social surplus are

[5]For example, Norway uses a survey of physicians' costs to inform service-level uniform reimbursement rates.

small. Figure B.16 shows that relative to the best linear uniform contract, the nonlinear contract lowers the marginal rates of low levels of treatment intensity and increases the marginal rates of high levels of treatment intensity. Figure B.17 shows that this means redistributing away from patients with relatively low severity to most other patients with relatively high severity. The distribution of treatment intensity more closely resembles efficient linear contracts with the nonlinear uniform contract that with menu of linear contracts. Figure B.18 shows corresponding distribution of health production and expenditure. However, without a base payment, the nonlinear uniform contract is not directly comparable to other counterfactuals because the nonlinear contract redistributes from physicians towards patients. Figure B.19 illustrates this tradeoff: although physicians are on average equally well off under the nonlinear contract relative to the status quo, 56 percent are individually worse off, and these losses represent up to 5 percent of status quo revenue. Physicians with losses tend to have high productivity and low patient severity, i.e., less initial under-treatment. If these physicians eventually exit the system, the risk of under-treatment for unmatched patients grows. For all physicians to be weakly better off, as required in other counterfactuals, the nonlinear contract needs \$3.32 in base payments, which is slightly higher than average base payments under the menu. After this adjustment, the gain in social surplus is only 23 percent of first best while the menu of linear contracts achieves 32 percent.[6]

Institutional differences also help explain the different impacts of a non-linear uniform contract in this setting relative to Gaynor et al. (2023). With primary care and the large estimated dispersion in unobserved patient severity, there does not seem to be a narrow range of medically appropriate treatment intensity for a non-linear

[6] Alternatively, one could moderate the participation constraint by incorporating physician exit: if indirect utility is lower than the threshold, then the corresponding patients have zero treatment intensity and zero public expenditure. However, this form of the constraint would only negligibly change counterfactual outcomes. Across counterfactuals, health production net of expenditure exceeds health production with zero treatment intensity by more than the base payment for either all physicians or all but one.

contract to target. Moreover, my estimates imply that marginal health production is nearly universally positive, so decreasing treatment intensity is not generally efficient. In Gaynor et al. (2023), more than half of observed treatment intensity was high enough to damage health based on a known cutoff.[7]

I use a demand profile approach to derive the nonlinear contract, similar to Gaynor et al. (2023), while also drawing on intuition from Chade et al. (2022). The demand profiling approach approximates the global design problem – avoiding the need to simultaneously optimize the social objective for a continuum of marginal reimbursement. I approximate the continuum with finitely many reimbursement rates and independently optimize one rate at a time. Each rate applies to a fixed segment of treatment intensity. An increase in one rate corresponds to incremental treatment intensity, and in turn, incremental health production, private costs, and public expenditure. These changes only occur among the patients with marginal treatment intensity.

I discretize the support of treatment intensity into T intervals.[8] For each interval $[m_t, m_{t+1})$, I find the corresponding reimbursement rate p_t that maximizes scaled incremental health production net of incremental private costs:

$$E_{\theta,\lambda}\left[(\alpha_G + \alpha)(h(m^*, \gamma\lambda) - h(m_t, \gamma\lambda)) - c\,(m^* - m_t) \mid m^* \geq m_t\right]$$

The interior of the expectation is a transformation of the social objective. Recall that the regulator maximizes expected health production subject to budget and participation constraints as well as privately optimal treatment intensity m^* which depends on the contract x, type θ, and patient severity. Based on equivalence of the first-order conditions after fixing shadow costs, I maximize a weighted sum of health production,

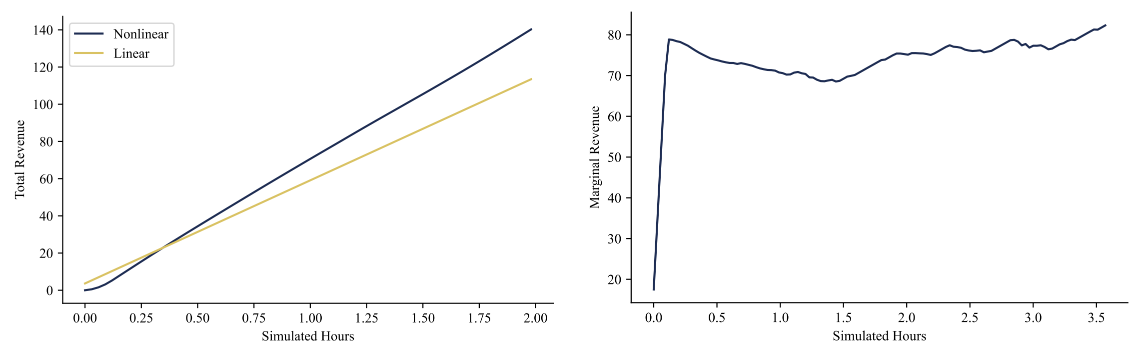
[7] These characterizations mostly refer to Figure 3 in that paper, which is based on a patient with median observed severity.

[8] Each interval spans approximately 2 simulated minutes of treatment intensity between 0 and 2 simulated hours.

private indirect utility, and public expenditure.[9]

In the objective's conditional expectation, I focus on patients and physicians who contribute to incremental social surplus, i.e., those whose treatment intensity varies with p_t. I also restrict the space of contracts so that marginal reimbursement $p(m)$ only crosses effective marginal cost $c - \alpha h_m(m, \gamma\lambda)$ for a unique value of treatment intensity $m^*(x = \{p_t\}_t)$.[10] If marginal reimbursement p_t exceeds effective marginal cost at m_t (or equivalently, $m^*(p_t) \geq m_t$), then the same is true at all lower levels of treatment.

Figure B.16: Optimal Nonlinear Uniform Contract



(a) Total Revenue

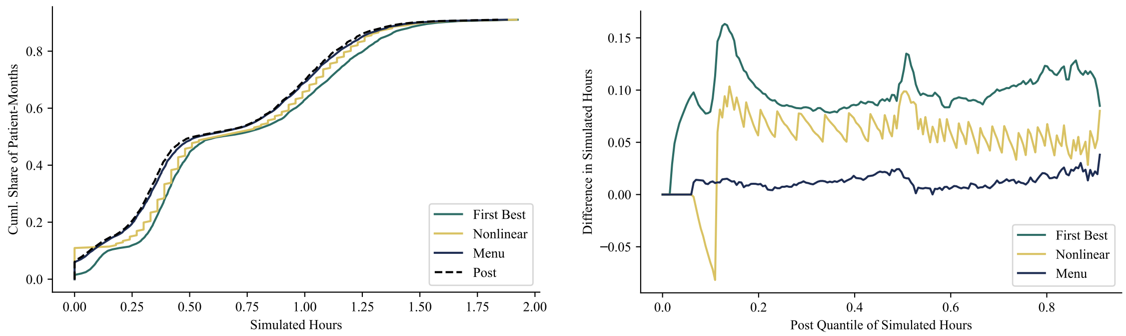
(b) Marginal Revenue

Notes: These plots illustrate the optimal nonlinear uniform contract which provides revenue as a flexible function of treatment intensity (along the x-axis). Linear indicates the optimal linear uniform contract including the base payment. In this figure, the nonlinear contract does not include the base payment.

[9]I fix the shadow costs of expenditure (μ_B) and participation ($\mu_{P,\theta}$ at $\frac{1}{\alpha_G}$ before rescaling the objective by α_G. Equivalently, α_G is the regulator's altruism and the regulator is willing to sacrifice \$1 of expenditure to either increase scaled health production or private indirect utility by \$1, so public expenditure and private revenue add to zero.

[10]Equivalently, the physician objective is concave and single-peaked for all physicians and patients. Marginal revenue cannot increase at a greater rate than marginal health production: $p_t - p_{t-1} < \alpha \, \forall t, \alpha$. I also focus on contracts where revenue is weakly increasing in treatment intensity. Neither of these constraints binds at the solution.

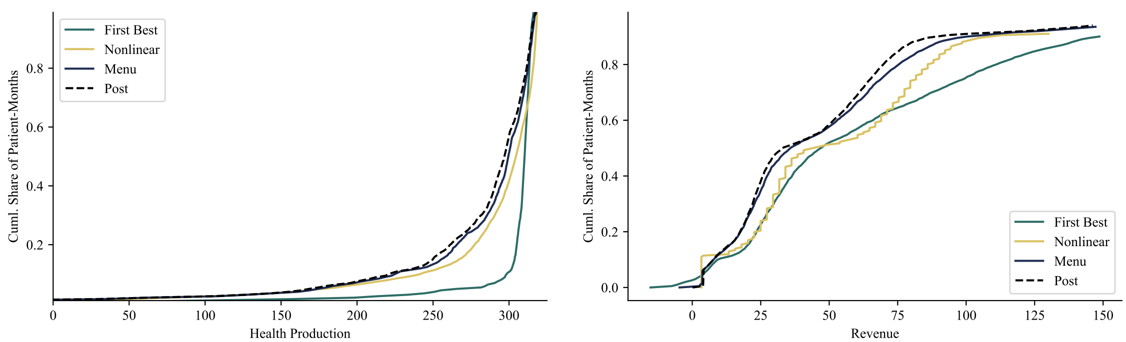
Figure B.17: Distribution of Treatment Intensity Across Contract Types



(a) Cumulative Distribution (b) Difference in Cumulative Distributions

Notes: These plots compare the optimal nonlinear uniform contract to a menu of linear contracts and first-best (full information) linear contracts. The right panel illustrates how treatment intensity compares to the status quo, under the higher reimbursement rate of certified physicians, by subtracting corresponding quantiles, e.g., the 5th percentile of treatment intensity under first-best contracts minus the 5th percentile of treatment intensity under the status quo. Post-Cert. indicates the status quo with a high reimbursement rates from certification. First Best indicates efficient linear contracts under full information. These plots condition on a positive health shock.

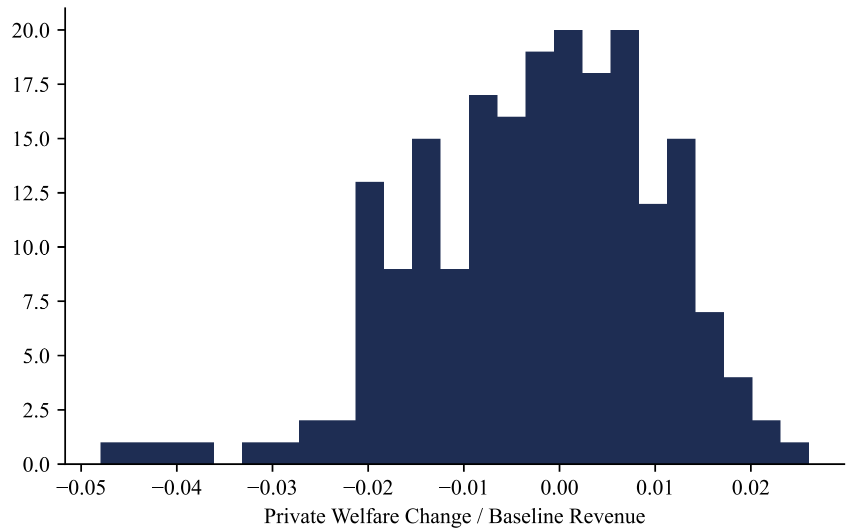
Figure B.18: Distribution of Health Production and Expenditure Across Contract Types



(a) Cumulative Distribution of Health Production (b) Cumulative Distribution of Expenditure

Notes: These plots compare the optimal nonlinear uniform contract to a menu of linear contracts and first-best (full information) linear contracts. Post-Cert. indicates the status quo with a high reimbursement rates from certification. First Best indicates efficient linear contracts under full information. These plots condition on a positive health shock.

Figure B.19: Some Physicians Worse Off with the Optimal Nonlinear Uniform Contract



Notes: This plot shows that many physicians experience moderate losses in private indirect utility under the optimal nonlinear uniform contract, as a share of status quo revenue. The reference point is each physician's pre-certification status quo. The y-axis is a count of physicians. The underlying data is an expectation over patients the distribution of severity.

B.2 Data and Estimation Details

B.2.1 Data Sources

I use several data sources to construct the estimation sample. The Norwegian Control and Payment of Health Reimbursements Database (KUHR) tracks reimbursement for outpatient claims organized at the level of bill line, i.e., reimbursement code, and identifies most patients and physicians. The Norwegian Patient Registry (NPR) is a database of reimbursement for inpatient claims organized at the level of encounter. I use ICD-10 and ICPC-2 codes from both sources to classify chronic illness. I identify avoidable hospitalizations following Table A1 from Page et al. (2007). base payments come from a basic subsidy rate dataset. Various datasets from the Norwegian GP Registry identify periods when patients are registered to patient lists and when physician are contracted to provide care to those patient lists. The physician-list dataset also identifies contract details: the maximum number of registered patients and indicators for shared lists and fixed-salary reimbursement. I use anonymous identifiers for physicians, lists, and patients to link datasets and convert periods into monthly panels. Physicians' birth date, gender, and birth country come from a personnel file. Patients' birth date, gender, disability payment receipt, and income come from tax records.

B.2.2 Construction of Treatment Intensity

I classify each patient into an observed type based on the combination of gender, 5-year age bins, and indicators for first and second prior chronic diagnosis, including cancer, diabetes, COPD, CVD, or asthma. I sort these 108 initial groups based on average reimbursement and further aggregate them into 10 types. Each aggregated type represents approximately 10 percent of aggregate spending in the estimation sample because treatment intensity is distributed approximately log-normally. The lowest type includes 23 percent of patient-months and the highest type represents 4 percent of patient-months.

For each patient type, I use all Norwegian patients to calculate the average bundle of services received and the average hours required to provide that bundle. I attribute time to encounters and reimbursement codes based on the share of reimbursement within an hour in the utilization data, e.g., 1-2 pm on January 1, 2010. I multiply each non-certification reimbursement code by the current administrative reimbursement rate. I average across codes, weighting where the number of lines per patient type per month. After certification, this numerator also includes current certification supplementary payments for an average number of visits per patient type. Finally, I divide by average hours per patient-type to calculate the simulated wage p_{kt}, i.e., the reimbursement per hour a physician would receive for providing the average bundle of services to a patient of type k in month t. Treatment intensity m_{ijt} equals patient-month fee-for-service revenue divided by marginal reimbursement and roughly corresponds to hours of treatment per patient-month ("simulated hours").

B.2.3 Counterfactual Analysis

This section reviews the technical assumptions underlying counterfactual analysis. I first describe the process for quantifying counterfactual outcomes given contracts. Then, I detail the algorithms that identify each set of contracts: efficient contracts, the optimal uniform contract, the optimal two-contract menu, and the optimal menu of contracts.

I measure all counterfactual outcomes as ex-ante expectations over registered patients of certified PCPs. I simulate patient severity for 60 patient simulants for each physician in the sample: 10 patient observed types multiplied by 6 quadrature nodes. For each of the 10 patient types per physician, I use averages of β_λ and $Pr(\lambda > 0)$, which aggregate over in-sample patients' observed characteristics like chronic illnesses and age. From the physician's first-order condition, treatment intensity is a function of simulated severity, estimated physician type, and contract. Likewise, indirect utility is a function of predicted treatment intensity, simulated severity, and the contract. Within a given menu, each physician's privately optimal contract maximizes average indirect utility. Ex-ante expectations reflect three levels of aggregation.[11] First, I average across quadrature nodes using quadrature weights to approximate the integral of normally distributed log patient severity. Second, I average across patient types, weighting by the observed number of patients in the estimation sample per physician. Third, I average across physicians, weighting by total registered patients six months before certification.

Scaled health production per simulated patient equals $H - \frac{1}{2}\alpha_R(m^* - \gamma\lambda))^2$. α_R can be thought of as the regulator's altruism or the inverse of the shadow cost of expenditure. I calibrate it with a revealed preference assumption. When setting supplementary reimbursement for certification, the regulator values incremental health production exactly as much as incremental expenditure. Expenditure equals $pm^*(p; \lambda, \theta) + b$, i.e., privately optimal treatment intensity multiplied by fee-for-service rates plus the base payment. I generally report incremental expected health production which subtracts the pre-certification expected value.

To focus on the role of reimbursement in treatment intensity, I fix total registered patients, the share of patient types for each physician, pre-certification fee-for-service rates, and status quo base payments at values six months before certification. For example, this removes variation in patient severity from seasonality and the time trend, so counterfactual treatment intensity at post-certification fee-for-service rates will typically be higher than observed in the data. To be consistent, I simulate all post-certification outcomes following the same process as counterfactuals, using the immediate change in the fee-for-service rate.

I enforce budget and participant constraints in counterfactuals when possible. I assume post-certification expected expenditure is the budget. Likewise, for participation constraints, I use expected indirect utility during the sample period to construct physician-specific participation thresholds. Physicians continue to work throughout

[11] When calculating expected indirect utility per physician per contract, I only aggregate over quadrature nodes and patient types.

the sample period at those levels of indirect utility, so they might reasonably be expected to continue in counterfactuals. All physicians prefer their post-certification contract, so I aggregate participation constraints by requiring that the same share of physicians weakly prefer counterfactual contracts over the lesser of their pre- or post-certification contract.

I solve the regulator's objective numerically for the set of physicians in my sample. All counterfactuals use a grid of 200 equally spaced points between 0.5 and 3.5. Each point reflects a multiple of pre-certification fee-for-service rates, which vary across physicians and patient types. The optimal uniform contract maximizes overall expected health production while satisfying global constraints. The other counterfactuals involve a large number of control variables and constraints. The global budget constraint also creates complementarity across physicians. Constrained maximization algorithms do not work well in this context. Instead, I enforce the participation constraints directly and search for contracts that maximize social surplus, i.e., incremental expected scaled health production minus incremental expected expenditure.

Efficient contracts are personalized to each physician with counterfactual perfect information about physician types. I identify efficient contracts by solving physician-specific problems. I select the fee-for-service rate that maximizes a physician's social surplus conditional on also satisfying her participation constraint. I minimize base payments so that participation constraints bind given the efficient contract and privately optimal treatment intensity. This solution is approximate because physicians have different numbers of patients and the weighted average of differences does not equal the difference of weighted averages. In some robustness checks, I take an additional step to enforce the global budget constraint. I lower the fee-for-service rate multiple by one grid point for one physician at a time to produce the smallest reduction in social surplus while lowering expenditure until the budget is slack.

For the optimal menu of contracts, I use a line-search algorithm. The algorithm finds the optimal base payment for each fee-for-service multiple on the grid, one at a time, while holding base payments for other fee-for-service multiples fixed. For stability, I search over discrete values of base payments rather than use an optimization routine. I also run the line-search algorithm twice. The first iteration uses a broad grid of base payments specific to each contract that covers a wide range of potential participation in that contract: $dEV > 0$ for each of $1, 2, ..., I$ physicians in a uniform contract. The second iteration searches locally for improvements using a grid of quadrature nodes. I enforce the participation constraint by always including the uniform contract in the menu, but the global budget constraint is difficult to strictly enforce with this method, so I maximize health production net of expenditure and penalize increased expenditure over the budget. In particular, the objective is $\Delta E[h(m^*)|b(p)] - \min\{0, \Delta R\} + \max\{0, \Delta R\}^2$ where $R \equiv E[pm^* + b(p)|b(p)]$ and Δ subtracts the reference values from counterfactual outcomes.

B.3 Derivations

B.3.1 Comparative Statics

This section characterizes how multi-dimensional heterogeneity contributes to the feasibility and efficiency of a menu of contracts relative to a uniform contract. Building on the exposition in Section 1.3.2, it is convenient to substitute the regulator's constraints into the objective. I assume that the shadow cost of the budget constraint $\mu_B \equiv \frac{1}{\alpha_R}$ is constant and that base payment $b(p)$ is large enough to satisfy all participation constraints.[12] Then, a realization of money-metric social surplus has the following expression:

$$SS(p, b, \lambda) = \alpha_R h\left(m^*, \gamma\lambda\right) - \left(pm^* + b(p)\right) .$$

I also assume that health production is twice continuously differentiable: returns to treatment are sometimes positive, strictly decreasing in treatment, and weakly decreasing in weighted patient severity $\gamma\lambda$.

With perfect information, base payment b_{FB} is set so that the participation constraint binds: $V(p, b, \lambda) = \underline{\mathrm{V}}$. This results in a special case of social surplus:

$$\begin{aligned} SS^{FB}(p, b, \lambda) &= \alpha_R h\left(m^*, \gamma\lambda\right) - pm^* + V(p, \lambda) - \underline{\mathrm{V}} \\ &= (\alpha_R + \alpha) h\left(m^*, \gamma\lambda\right) - cm^* - \underline{\mathrm{V}} . \end{aligned}$$

In this case, the first-best reimbursement rate p^{FB} satisfies the first-order condition:

$$\frac{d}{dp} SS(p, b, \lambda) = \left((\alpha_R + \alpha) h_m\left(m^*, \gamma\lambda\right) - c\right) m_p^* = 0 .$$

Equivalently, private cost equals marginal health production, scaled by both social and private altruism, at the privately optimal level of treatment intensity. Substituting the parameterization for health production, the efficient rate is proportional to private cost, and decreasing in private altruism: $p^{FB} = \frac{\alpha_R}{\alpha + \alpha_R} c$. As the regulator

[12] α_R can be interpreted as the regulator's altruism.

relaxes the budget constraint by increasing the weight on health production relative to expenditure ($\alpha_R \to \infty$), $p_{FB} \to c$.[13]

Next, consider the second-best framing from Section 2.1.3. Starting from a uniform contract, when is it efficient to add a second contract with greater fee-for-service to the menu? This requires a comparison of incremental indirect utility ("private benefits" or "WTP") and incremental surplus, so let $\Delta f(p) \equiv f(p_H) - f(p_L)$ and focus on realizations of patient severity λ large enough for positive treatment intensity. How does WTP vary with physician type, all else equal? Since $\frac{d\Delta \int V(p)dF(\lambda)}{d\theta_k} = \Delta \int \frac{dV(p)}{d\theta_k} dF(\lambda)$, I first derive $\frac{dV(p)}{d\theta_k}$ using the envelope theorem:

$$\begin{aligned}
\frac{dV(p)}{dc} &= \frac{d}{dc}\left((p-c)m(p) + \alpha h(m(p), \gamma\lambda)\right) = -m(p) \\
\frac{dV(p)}{d\alpha} &= h(m(p), \gamma\lambda) \\
\frac{dV(p)}{d\gamma} &= \alpha h_{(\gamma\lambda)}(m(p), \gamma\lambda)\lambda \\
\frac{dV(p)}{d\lambda} &= \alpha h_{(\gamma\lambda)}(m(p), \gamma\lambda)\gamma
\end{aligned}$$

From $h_{mm} < 0$, the physician's first-order condition implies that $m(p)$ is strictly increasing, so $\Delta \frac{d}{dc} V(p) < 0$. Next, $\Delta \frac{d}{d\alpha} V(p) > 0$ when health production increases in treatment intensity. Finally, from $h_{(\lambda\gamma)m} \geq 0$, $\Delta \frac{d}{d\gamma} V(p) \geq 0$ and $\Delta \frac{d}{d\lambda} V(p) \geq 0$.

Before proceeding, it is useful to derive statics of treatment intensity with

[13] Conversely, with altruistic physicians and an extreme budget constraint ($\alpha_R = 0$), the efficient rate approaches 0.

respect to physician type by differentiating the physician's first-order condition:

$$
\begin{aligned}
\frac{dV}{dm} &= \frac{d}{dm}\left((p-c)m + \alpha h(m,\gamma\lambda)\right) \\
&= p - c + \alpha h_m(m,\gamma\lambda) && = 0 \\
\frac{d^2V}{dpdm} &= 1 + \alpha h_{mm}(m,\gamma\lambda)\frac{dm}{dp} && = 0 \\
\frac{d^2V}{dcdm} &= -1 + \alpha h_{mm}(m,\gamma\lambda)\frac{dm}{dc} && = 0 \\
\frac{d^2V}{d\alpha dm} &= \alpha h_{mm}(m,\gamma\lambda)\frac{dm}{d\alpha} + h_m(m,\gamma\lambda) && = 0 \\
\frac{d^2V}{d\gamma dm} &= \alpha h_{mm}(m,\gamma\lambda)\frac{dm}{d\gamma} + \alpha h_{m(\gamma\lambda)}(m,\gamma\lambda)\lambda && = 0 \\
\frac{d^2V}{d\lambda dm} &= \alpha h_{mm}(m,\gamma\lambda)\frac{dm}{d\lambda} + \alpha h_{m(\gamma\lambda)}(m,\gamma\lambda)\gamma && = 0
\end{aligned}
$$

Then,

$$
\begin{aligned}
\frac{dm}{dp} &= \frac{-1}{\alpha h_{mm}(m(p),\gamma\lambda)} \\
\frac{dm}{dc} &= \frac{1}{\alpha h_{mm}(m(p),\gamma\lambda)} \\
\frac{dm}{d\alpha} &= \frac{-h_m(m(p),\gamma\lambda)}{\alpha h_{mm}(m(p),\gamma\lambda)} \\
\frac{dm}{d\gamma} &= \frac{-\lambda h_{m(\gamma\lambda)}(m(p),\gamma\lambda)}{h_{mm}(m(p),\gamma\lambda)} \\
\frac{dm}{d\lambda} &= \frac{-\gamma h_{m(\gamma\lambda)}(m(p),\gamma\lambda)}{h_{mm}(m(p),\gamma\lambda)}
\end{aligned}
$$

For $\frac{d}{d\theta_k}SS(p)$:

$$
\begin{aligned}
\frac{dSS(p)}{dc} &= \frac{d}{dc}\left(\alpha_R h\left(m^*,\gamma\lambda\right) - (pm^* + b(p))\right) \\
&= (\alpha_R h_m(m(p),\gamma\lambda) - p)\frac{dm(p)}{dc} \\
\frac{dSS(p)}{d\alpha} &= (\alpha_R h_m(m(p),\gamma\lambda) - p)\frac{dm(p)}{d\alpha} \\
\frac{dSS(p)}{d\gamma} &= (\alpha_R h_m(m(p),\gamma\lambda) - p)\frac{dm(p)}{d\gamma} + \alpha_R h_{(\gamma\lambda)}(m(p),\gamma\lambda)\lambda
\end{aligned}
$$

Since $\frac{dm(p)}{dc} < 0$ and $h_{mm} < 0$, $\Delta \frac{dSS(p)}{dc} > 0$. If h is increasing over the relevant support, then $\frac{dm(p)}{d\alpha} > 0$ and $(\alpha_R h_m(m(p), \gamma\lambda) - p)$ is decreasing in p, so $\Delta \frac{dSS(p)}{d\alpha} < 0$. From $h_{m(\lambda\gamma)} \leq 0$, $\frac{dm(p)}{d\gamma} < 0$, so $\Delta \frac{dSS(p)}{d\gamma} > 0$ and $\Delta \frac{dSS(p)}{d\lambda} > 0$.

In summary, given assumptions and all else equal, low-cost, high-altruism, high-productivity (low γ), and low-severity (low $E[\lambda]$) physicians are relatively likely to choose a high-fee-for-service contract, but this choice produces relatively small increases in social surplus. The feasibility and efficiency of a separating equilibrium sometimes require correlation in cost, altruism, and productivity.

B.3.2 Likelihood

The likelihood is based on the random component of patient severity. Treatment intensity m may equal zero either because the underlying severity is zero or because it is too low for a privately optimal choice of $m > 0$. Since $\frac{dm}{d\lambda} > 0$, I can split cases based on $\tilde{\lambda}$, the minimum λ such that $m \geq 0$.

$$\begin{aligned} l(m \mid \theta, x, X_\lambda) &= l(m \mid \lambda \leq \tilde{\lambda}) Pr(\lambda \leq \tilde{\lambda}) + l(m \mid \lambda > \tilde{\lambda}) Pr(\lambda > \tilde{\lambda}) \\ &= 1[m = 0] Pr(\lambda \leq \tilde{\lambda}) + 1[m > 0] Pr(\lambda = \lambda^{-1}(m) \mid \lambda > \tilde{\lambda}) Pr(\lambda > \tilde{\lambda}) \left| \frac{d\epsilon}{d\lambda} \frac{d\lambda}{dm} \right| . \end{aligned}$$

For $\tilde{\lambda} > 0$,[14] denoting the CDF of $\lambda \mid \lambda > 0$ as F_λ, the two-stage process for λ can be decomposed:

$$\begin{aligned} Pr(\lambda \leq \tilde{\lambda}) &= Pr(\lambda = 0) + Pr(\lambda > 0) F_\lambda(\tilde{\lambda}) \\ Pr(\lambda > \tilde{\lambda}) &= (1 - F_\lambda(\tilde{\lambda})) Pr(\lambda > 0) . \end{aligned}$$

[14] If $\tilde{\lambda} = 0$, then $Pr(\lambda > \tilde{\lambda}) = Pr(\lambda > 0)$.

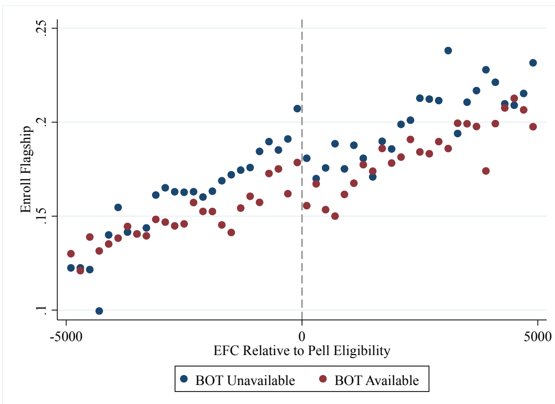
Under parametric assumptions,

$$
\begin{aligned}
\lambda^{-1}(m) &= \frac{m-\beta_0}{\beta_1} && \text{if } m > 0 \\
0 \leq \lambda^{-1}(m) \leq \tilde{\lambda} &\equiv \max\left\{0, \frac{-\beta_0}{\beta_1}\right\} && \text{if } m = 0 \\
\beta_0 &= \frac{p-c-\sigma(N-1)E[m']}{\alpha+\sigma} \\
\beta_1 &= \frac{\alpha\gamma}{\alpha+\sigma} = \frac{dm}{d\lambda} \\
Pr(\lambda > 0) &= \frac{\exp d_0 + d_1\beta_\lambda X_\lambda}{1+\exp d_0 + d_1\beta_\lambda X_\lambda} \\
Pr(\lambda = \lambda^{-1}(m) \mid \lambda > \tilde{\lambda}) &= (1-F_\lambda(\tilde{\lambda}))^{-1}\phi\left(\frac{\log \lambda^{-1}(m) - \beta_\lambda X_\lambda}{\sigma_\lambda}\right) \\
F_\lambda(\tilde{\lambda}) &= 1[\tilde{\lambda} > 0]\Phi\left(\frac{\log \tilde{\lambda} - \beta_\lambda X_\lambda}{\sigma_\lambda}\right) \\
\frac{d\epsilon}{d\lambda} &= \frac{1}{\sigma_\lambda \lambda}
\end{aligned}
$$

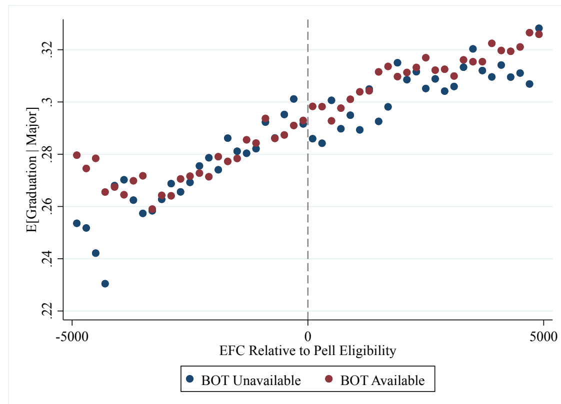
where Φ and ϕ are the CDF and PDF of a standard normal.

Appendix C: Supplementary Materials to Chapter 3

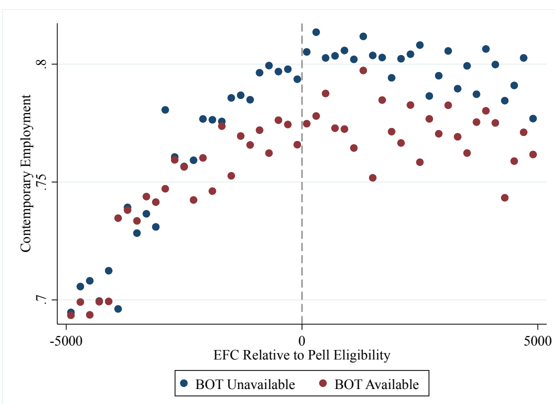
Figure C.0.1: Pell Eligibility, Program Choice, and Effort



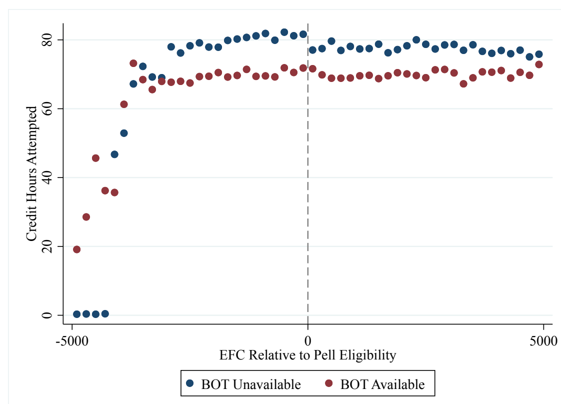
(a) Enroll Flagship



(b) E[Graduation | Major]



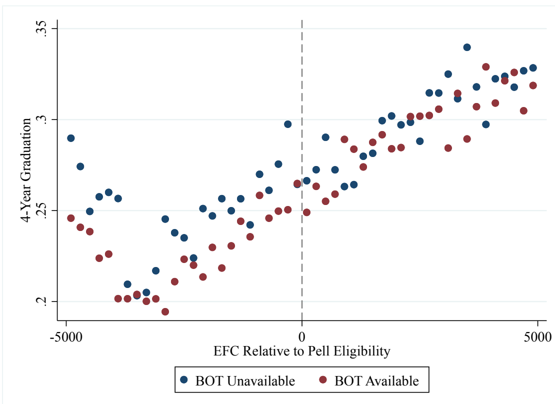
(c) Contemporary Employment



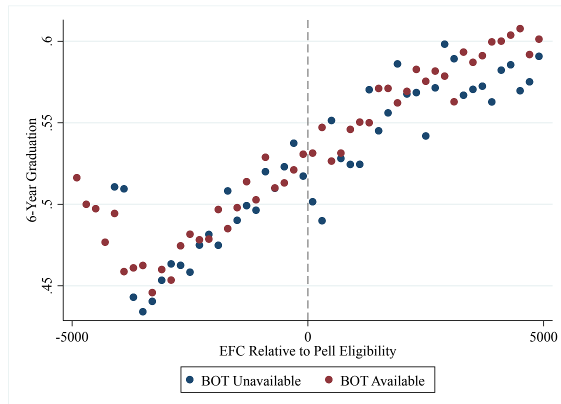
(d) Credit Hours Attempted

Notes: The x-axis shows expected family contribution (EFC) minus the maximum EFC for Pell Grant eligibility, which varies by year. Points represent averages of students within $500 bins of EFC. BOT Available includes students with initial financial aid offers in years when BOT loans were widely available.

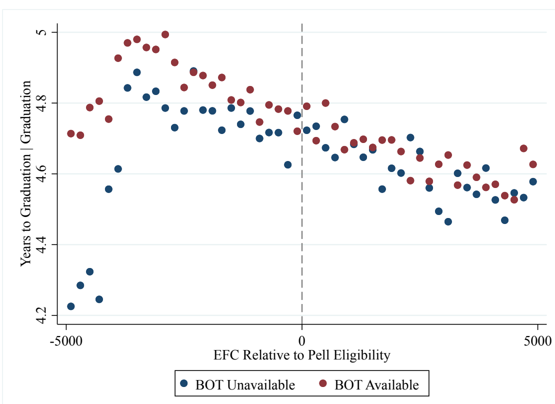
Figure C.0.2: Pell Eligibility and Outcomes



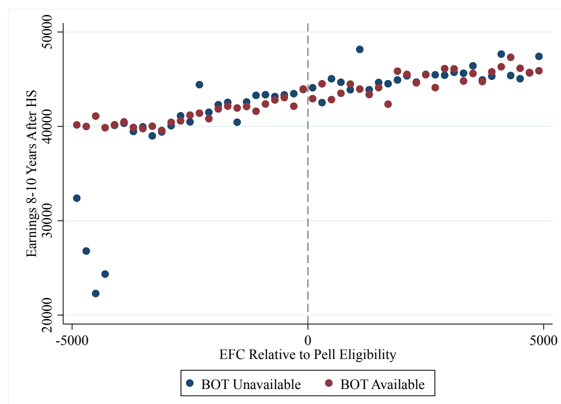
(a) Graduation Within 4 Years



(b) Graduation Within 6 Years



(c) Years To Graduation | Any Graduation



(d) Annual Earnings 8-10 Years After High School

Notes: The x-axis shows expected family contribution (EFC) minus the maximum EFC for Pell Grant eligibility, which varies by year. Points represent averages of students within $500 bins of EFC. BOT Available includes students with initial financial aid offers in years when BOT loans were widely available.

Table C.0.1: Difference-in-Discontinuity: Pell Ineligible

	Estimate	Std. Err.	P-Value	Outcome Mean	R^2	Obs.
Financial Aid						
Federal Aid	−0.334	(0.015)	[<0.001]	1.661	0.837	256,984
Any B-On-Time Loans	−0.002	(0.000)	[<0.001]	0.036	0.102	256,984
Traditional Loans	0.614	(0.107)	[<0.001]	3.987	0.231	256,984
Net Cost of Attendance	1.284	(0.149)	[<0.001]	12.115	0.609	256,984
Program Choice						
Enrolled at a Flagship	0.001	(0.002)	[0.560]	0.169	0.933	256,984
E[Graduation | Major]	−0.008	(0.004)	[0.054]	0.288	0.104	256,984
Effect of Major on Earnings	−0.196	(0.110)	[0.076]	−0.248	0.043	256,981
Effort Proxies						
Credit Hours Attempted	−1.105	(0.800)	[0.168]	68.077	0.587	256,984
Contemporary Employment	−0.005	(0.008)	[0.548]	0.759	0.039	256,984
Max Class Rank	−0.053	(0.020)	[0.009]	2.301	0.671	256,796
Outcomes						
Graduation Within 4 Years	−0.009	(0.009)	[0.334]	0.243	0.089	256,979
Graduation Within 6 Years	−0.033	(0.012)	[0.008]	0.513	0.119	227,552
Earnings After 8-10 Years	−1.224	(0.719)	[0.090]	73.061	0.365	256,984
Balance						
Top 10% Within HS	0.001	(0.007)	[0.886]	0.213	0.210	269,037
Top 25% Within HS	0.014	(0.008)	[0.105]	0.402	0.140	269,037
Advanced HS Courses	0.049	(0.056)	[0.382]	3.906	0.252	284,046
White	−0.001	(0.008)	[0.908]	0.421	0.221	284,046
Free-Lunch	0.001	(0.007)	[0.867]	0.244	0.127	284,046
Discipline Days	0.031	(0.106)	[0.768]	1.834	0.027	284,046

Notes: This table shows estimates of the uninteracted discontinuity at the Pell eligibility threshold (from years without B-On-Time loans), i.e., β_1 from Equation (3.2). The specification corresponds to Column (4) in Table 3.1: a local quadratic difference-in-discontinuity with controls for students' initial covariates, final net cost of attendance, and fixed effects for the financial aid year, among students with an EFC within \$5000 of the Pell eligibility threshold. All monetary measures are scaled down by 1000.

Works Cited

Jason Abaluck, Leila Agha, Chris Kabrhel, and Ali Raja. The determinants of productivity in medical testing: Intensity and allocation of care - proquest, 12 2016. URL https://www.proquest.com/openview/a7ee8aca39ceab054e5986a9b86db102/1?pq-origsite=gscholar&cbl=42182.

Marie Allard, Izabela Jelovac, and Pierre-Thomas Léger. Payment mechanism and gp self-selection: capitation versus fee for service. *International Journal of Health Care Finance and Economics*, 14:143–160, 03 2014. doi: 10.1007/s10754-014-9143-z.

R. J. Andrews, J. Li, and M. F. Lovenheim. Quantile treatment effects of college quality on earnings. *Journal of Human Resources*, 51:200–238, 01 2016. doi: 10.3368/jhr.51.1.200.

Eduardo M. Azevedo and Daniel Gottlieb. Perfect competition in markets with adverse selection. *Econometrica*, 85:67–105, 2017. doi: 10.3982/ecta13434.

Ivan Badinski, Amy Finkelstein, Matthew Gentzkow, and Peter Hull. Geographic variation in healthcare utilization: The role of physicians. 10 2023. doi: 10.3386/w31749.

Jori Barash. Do new patients displace existing patients' treatment?, 04 2024. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4343073.

Victoria Barham and Olga Milliken. Payment mechanisms and the composition of physician practices: Balancing cost-containment, access, and quality of care. *Health Economics*, 24:895–906, 07 2014. doi: 10.1002/hec.3069.

Sanjay Basu, Seth A Berkowitz, Robert L Phillips, Asaf Bitton, Bruce E Landon, and Russell S Phillips. Association of primary care physician supply with population mortality in the united states, 2005-2015. *JAMA internal medicine*, 179:506–514, 2019. doi: 10.1001/jamainternmed.2018.7624. URL `https://www.ncbi.nlm.nih.gov/pubmed/30776056/`.

Charles Bellemare and Bruce Shearer. Multidimensional heterogeneity and the economic importance of risk and matching: evidence from contractual data and field experiments. *The RAND Journal of Economics*, 44:361–389, 06 2013. doi: 10.1111/1756-2171.12023.

Simon Bensnes and Ingrid Huitfeldt. Rumor has it: How do patients respond to patient-generated physician ratings? *Journal of Health Economics*, 76:102415, 03 2021. doi: 10.1016/j.jhealeco.2020.102415.

Barbara Biasi. The labor market for teachers under different pay schemes. *American Economic Journal: Economic Policy*, 13:63–102, 08 2021. doi: 10.1257/pol.20200295.

Tamara Bischof and Boris Kaiser. Who cares when you close down? the effects of primary care practice closures on patients. *Health Economics*, 05 2021. doi: 10.1002/hec.4287.

Øyvind Bjertnæs, Hilde Hestad Iversen, Rebecka Norman, and Jose M. Valderas. Web-based public ratings of general practitioners in norway: Validation study. *JMIR Formative Research*, 7:e38932, 03 2023. doi: 10.2196/38932. URL `https://formative.jmir.org/2023/1/e38932`.

Sandra E. Black, Jeffrey T. Denning, and Jesse Rothstein. Winners and losers? the effect of gaining and losing access to selective colleges on education and labor market outcomes, 03 2020. URL `https://www.nber.org/papers/w26821`.

Zachary Bleemer. Affirmative action, mismatch, and economic mobility after california's proposition 209. *The Quarterly Journal of Economics*, 137, 09 2021. doi: 10.1093/qje/qjab027.

College Board. Trends in student aid highlights, 06 2019. URL `https://research.collegeboard.org/trends/student-aid/highlights`.

Kurt R. Brekke, Tor Helge Holmås, Karin Monstad, and Odd Rune Straume. Do treatment decisions depend on physicians' financial incentives? *Journal of Public Economics*, 155:74–92, 11 2017. doi: 10.1016/j.jpubeco.2017.09.012.

Zach Y. Brown, Christopher Hansman, Jordan Keener, and Andre F. Veiga. Information and disparities in health care quality: Evidence from gp choice in england, 03 2023. URL `https://www.nber.org/papers/w31033?utm_campaign=ntwh&utm_medium=email&utm_source=ntwg14`.

Marika Cabral, Colleen Carey, and Sarah Miller. The impact of provider payments on health care utilization: Evidence from medicare and medicaid, 11 2021. URL `https://policycommons.net/artifacts/1894340/the-impact-of-provider-payments-on-health-care-utilization/2644327/`.

James H. Cardon and Igal Hendel. Asymmetric information in health insurance: Evidence from the national medical expenditure survey. *The RAND Journal of Economics*, 32:408, 2001. doi: 10.2307/2696362.

Colleen M. Carey, Sarah Miller, and Laura R. Wherry. The impact of insurance expansions on the already insured: The affordable care act and medicare. *American Economic Journal: Applied Economics*, 12:288–318, 10 2020. doi: 10.1257/app.20190176. URL `https://www.nber.org/system/files/working_papers/w25153/w25153.pdf`.

Hector Chade, Victoria R. Marone, Amanda Starc, and Jeroen Swinkels. Multidimensional screening and menu design in health insurance markets, 10 2022. URL `https://www.nber.org/papers/w30542`.

David C. Chan. The efficiency of slacking off: Evidence from the emergency department. *Econometrica*, 86:997–1030, 2018. doi: 10.3982/ecta13565.

David C Chan, Matthew Gentzkow, and Chuan Yu. Selection with variation in diagnostic skill: Evidence from radiologists. *The Quarterly Journal of Economics*, 137:729–783, 01 2022. doi: 10.1093/qje/qjab048.

David C. Chan, Jr and Yiqun Chen. The productivity of professions: Evidence from the emergency department, 10 2022. URL `https://www.nber.org/papers/w30608`.

Chiang-Hua Chang, A. James O'Malley, and David C. Goodman. Association between temporal changes in primary care workforce and patient outcomes. *Health Services Research*, 52:634–655, 06 2016. doi: 10.1111/1475-6773.12513.

Benjamin Chartock. Quality disclosure, demand, and congestion: Evidence from physician ratings. 02 2023.

Raj Chetty, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole, and Nils Wernerfelt. Social capital ii: determinants of economic connectedness. *Nature*, page 1–13, 08 2022. doi: 10.1038/s41586-022-04997-3. URL `https://www.nature.com/articles/s41586-022-04997-3`.

Philippe Choné and Ching-to Albert Ma. Optimal health care contract under physician agency. *Annals of Economics and Statistics*, page 229, 2011. doi: 10.2307/41615481.

Jeffrey Clemens and Joshua D. Gottlieb. Do physicians' financial incentives affect medical treatment and patient health? *American Economic Review*, 104: 1320–1349, 04 2014. doi: 10.1257/aer.104.4.1320. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4144420/`.

Mark Colas, Sebastian Findeisen, and Dominik Sachs. Optimal need-based financial aid. *Journal of Political Economy*, 129:492–533, 02 2021. doi: 10.1086/711952.

Jesse M. Cunha, Trey Miller, and Emily Weisburst. Information and college decisions: Evidence from the texas go center project. *Educational Evaluation and Policy Analysis*, 40:151–170, 11 2017. doi: 10.3102/0162373717739349.

Amanda Dahlstrand. Defying distance? the provision of services in the digital age, 01 2022. URL `https://www.dropbox.com/s/skusfyobfaoku2c/Dahlstrand_JMP.pdf?dl=0`.

S. B. Dale and A. B. Krueger. Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics*, 117:1491–1527, 11 2002. doi: 10.1162/003355302320935089.

Stacy B. Dale and Alan B. Krueger. Estimating the effects of college characteristics over the career using administrative earnings data. *Journal of Human Resources*, 49:323–358, 2014. doi: 10.3368/jhr.49.2.323. URL `https://www.nber.org/papers/w17159.pdf`.

Jeffrey T. Denning, Benjamin M. Marx, and Lesley J. Turner. Propelled: The effects of grants on graduation, earnings, and welfare. *American Economic Journal: Applied Economics*, 11:193–224, 07 2019. doi: 10.1257/app.20180100.

Xavier D'Haultfœuille and Philippe Février. The provision of wage incentives: A structural estimation using contracts variation. *Quantitative Economics*, 11: 349–397, 2020. doi: 10.3982/qe597.

Cauê Dobbin, Nano Barahona, and Sebastián Otero. The equilibrium effects of subsidized student loans, 2022. URL `https://sebotero.github.io/papers/DBO_draft.pdf`.

Rudy Douven, Minke Remmerswaal, and Robin Zoutenbier. Do altruistic mental health care providers have better treatment outcomes? *Journal of Human Resources*, 54:310–341, 09 2017. doi: 10.3368/jhr.54.2.0716.8070r1.

Joseph J. Doyle, Steven M. Ewer, and Todd H. Wagner. Returns to physician human capital: Evidence from patients randomized to physician teams. *Journal of Health Economics*, 29:866–882, 12 2010. doi: 10.1016/j.jhealeco.2010.08.004.

Liran Einav, Amy Finkelstein, and Mark R. Cullen. Estimating welfare in insurance markets using variation in prices*. *Quarterly Journal of Economics*, 125:877–921, 08 2010. doi: 10.1162/qjec.2010.125.3.877.

Liran Einav, Amy Finkelstein, Stephen P Ryan, Paul Schrimpf, and Mark R Cullen. Selection on moral hazard in health insurance. *American Economic Review*, 103:178–219, 02 2013. doi: 10.1257/aer.103.1.178.

Liran Einav, Amy Finkelstein, and Neale Mahoney. Provider incentives and healthcare costs: Evidence from long-term care hospitals, 06 2018. URL `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3110133`.

Liran Einav, Amy Finkelstein, Yunan Ji, and Neale Mahoney. Voluntary regulation: Evidence from medicare payment reform. *The Quarterly Journal of Economics*, 09 2021. doi: 10.1093/qje/qjab035.

Paul J. Eliason, Paul L. E. Grieco, Ryan C. McDevitt, and James W. Roberts. Strategic patient discharge: The case of long-term care hospitals. *American Economic Review*, 108:3232–3265, 11 2018. doi: 10.1257/aer.20170092.

Randall P. Ellis and Thomas G. McGuire. Provider behavior under prospective reimbursement. *Journal of Health Economics*, 5:129–151, 06 1986. doi: 10.1016/0167-6296(86)90002-0.

Dennis Epple, Richard Romano, Sinan Sarpçad, and Holger Sieg. A general equilibrium analysis of state and private colleges and access to higher education in the u.s. *Journal of Public Economics*, 155:164–178, 11 2017. doi: 10.1016/j.jpubeco.2016.09.003. URL `https://www.sciencedirect.com/science/article/abs/pii/S004727271630130X`.

Dennis Epple, Richard Romano, Sinan Sarpça, Holger Sieg, and Melanie Zaber. Market power and price discrimination in the us market for higher education. *RAND Journal of Economics (Wiley-Blackwell)*, 50:201–225, 2019. doi: 10.1111/1756-2171.12267. URL `http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=1&sid=547c4139-9617-4b3c-a8aa-b1396400b1b0%40sessionmgr4008`.

Dennis Epple, Francisco Martinez-Mora, and Richard Romano. Charter school practices and student selection: An equilibrium analysis, 12 2021. URL `https://www.nber.org/papers/w29529`.

Andrew J. Epstein and Sean Nicholson. The formation and evolution of physician treatment styles: An application to cesarean sections. *Journal of Health Economics*, 28:1126–1140, 12 2009. doi: 10.1016/j.jhealeco.2009.08.003.

Brent J. Evans and Tuan D. Nguyen. Monetary substitution of loans, earnings, and need-based aid in postsecondary education: The impact of pell grant eligibility. *Economics of Education Review*, 70:1–19, 06 2019. doi: 10.1016/j.econedurev.2019.02.004.

Itzik Fadlon and Jessica Van Parys. Primary care physician practice styles and patient care: Evidence from physician exits in medicare. *Journal of Health Economics*, 71:102304, 05 2020. doi: 10.1016/j.jhealeco.2020.102304.

Elena Falcettoni. The determinants of physicians' location choice: Understanding the rural shortage. *SSRN Electronic Journal*, 2018. doi: 10.2139/ssrn.3493178.

Hanming Fang and Zenan Wu. Multidimensional private information, market structure, and insurance markets. *The RAND Journal of Economics*, 49: 751–787, 2018. URL `https://www.jstor.org/stable/45147791`.

Ian Fillmore. Price discrimination and public policy in the us college market. *The Review of Economic Studies*, 07 2022. doi: 10.1093/restud/rdac051.

Amy Finkelstein, Matthew Gentzkow, and Heidi Williams. Place-based drivers of mortality: Evidence from migration. *American Economic Review*, 111:2697–2735, 08 2021. doi: 10.1257/aer.20190825.

Ian Forde, Carol Nader, Karolina Socha-Dietrich, Jillian Oderkirk, and Francesca Colombo. Primary care review of denmark, 2016. URL `https://www.oecd.org/health/health-systems/Primary-Care-Review-of-Denmark-OECD-report-December-2016.pdf`.

Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer. Labour supply, service intensity, and contracts: Theory and evidence on physicians. *Journal of Applied Economics*, 36:686–702, 2021. doi: 10.1002/jae.2840. URL `https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.2840`.

Seth Freedman, Ezra Golberstein, Tsan-Yao Huang, David J. Satin, and Laura Barrie Smith. Docs with their eyes on the clock? the effect of time pressures on primary care productivity. *Journal of Health Economics*, 77: 102442, 05 2021. doi: 10.1016/j.jhealeco.2021.102442.

Matteo M Galizzi, Time Tammi, Geir Godager, Ismo Linnosmaa, and Daniel Wiesen. Provider altruism in health economics. 2015. URL https://www.julkari.fi/bitstream/handle/10024/125787/urn_isbn_978-952-302-429-8.pdf?sequence=1.

Craig L Garthwaite. The doctor might see you now: The supply side effects of public health insurance expansions. *American Economic Journal: Economic Policy*, 4:190–215, 08 2012. doi: 10.1257/pol.4.3.190.

Martin Gaynor, Nirav Mehta, and Seth Richards-Shubik. Optimal contracting with altruistic agents: Medicare payments for dialysis drugs. *The American Economic Review*, 113:1530–1571, 06 2023. doi: 10.1257/aer.20210208.

Rita Ginja, Julie Riise, Barton Willage, and Alexander Willén. Does your doctor matter? doctor quality and patient outcomes. *SSRN Electronic Journal*, 2022. doi: 10.2139/ssrn.4130163.

Geir Godager and Daniel Wiesen. Profit or patients' health benefit? exploring the heterogeneity in physician altruism. *Journal of Health Economics*, 32: 1105–1116, 12 2013. doi: 10.1016/j.jhealeco.2013.08.008. URL https://www.sciencedirect.com/science/article/abs/pii/S0167629613001070.

Sara Goldrick-Rab, Robert Kelchen, Douglas N. Harris, and James Benson. Reducing income inequality in educational attainment: Experimental evidence on the impact of financial aid on college completion. *American Journal of Sociology*, 121:1762–1817, 05 2016. doi: 10.1086/685442.

Joshua Gottlieb, Maria Polyakova, Kevin Rinz, U Census Bureau, Hugh Shiplett, and Victoria Udalova. Who values human capitalists' human capital? healthcare spending and physician earnings, 07 2020. URL https://www2.census.gov/ces/wp/2020/CES-WP-20-23.pdf.

Gautam Gowrisankaran, Keith A. Joiner, and Pierre-Thomas Léger. Physician practice style and healthcare costs: Evidence from emergency departments, 12 2017. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3095116.

Trude Gunnes, Lars J. Kirkebøen, and Marte Rønning. Financial incentives and study duration in higher education. *Labour Economics*, 25:1–11, 12 2013. doi: 10.1016/j.labeco.2013.04.010.

Eric A. Hanushek, Jin Luo, Andrew J. Morgan, Minh Nguyen, Ben Ost, Steven G. Rivkin, and Ayman Shakeel. The effects of comprehensive educator evaluation and pay reform on achievement, 03 2023. URL https://www.nber.org/papers/w31073.

Matthew C. Harris, Yinan Liu, and Ian McCarthy. Capacity constraints and time allocation in public health clinics. *Health Economics*, 29:324–336, 01 2020. doi: 10.1002/hec.3984.

Heike Hennig-Schmidt, Reinhard Selten, and Daniel Wiesen. How payment systems affect physicians' provision behaviour—an experimental investigation. *Journal of Health Economics*, 30:637–646, 07 2011. doi: 10.1016/j.jhealeco.2011.05.001.

Kate Ho and Robin S Lee. Health insurance menu design for large employers. *The RAND Journal of Economics*, 54:598–637, 10 2023. doi: 10.1111/1756-2171.12452.

Thomas P. Hoe. Does hospital crowding matter? evidence from trauma and orthopedics in england. *American Economic Journal: Economic Policy*, 14: 231–262, 05 2022. doi: 10.1257/pol.20180672.

Renee Y. Hsia and Yu-Chu Shen. Emergency department closures and openings: Spillover effects on patient outcomes in bystander hospitals. *Health Affairs*, 38:1496–1504, 09 2019. doi: 10.1377/hlthaff.2019.00125.

Ulla Hämäläinen, Kristian Koerselman, and Roope Uusitalo. Financial incentives for timely graduation. 03 2016. URL `https://2016.economicsofeducation.com/user/pdfsesiones/027.pdf`.

Tor Iversen and Hilde Lurås. Patient switching in general practice. *Journal of health economics*, 30:894–903, 09 2011. doi: 10.1016/j.jhealeco.2011.07.008.

William Jack. Purchasing health care services from providers with unknown altruism. *Journal of Health Economics*, 24:73–93, 01 2005. doi: 10.1016/j.jhealeco.2004.06.001.

Anupam B. Jena, N. Clay Mann, Leia N. Wedlund, and Andrew Olenski. Delays in emergency care and mortality during major u.s. marathons. *New England Journal of Medicine*, 376:1441–1450, 04 2017. doi: 10.1056/nejmsa1614073.

Yunan Ji. Can competitive bidding work in health care? evidence from medicare durable medical equipment, 2021. URL `https://scholar.harvard.edu/files/yunan/files/dme.pdf`.

Astrid Kiil and Kurt Houlberg. How does copayment for health care services affect demand, health and redistribution? a systematic review of the empirical evidence from 1990 to 2011. *The European Journal of Health Economics*, 15: 813–828, 08 2013. doi: 10.1007/s10198-013-0526-8.

Marte Kjøllesdal, Thor Indseth, Hilde Hestad Iversen, and Oyvind Bjertnaes. Patient experiences with general practice in norway: a comparison of immigrant groups and the majority population following a national survey. *BMC Health Services Research*, 20, 12 2020. doi: 10.1186/s12913-020-05963-3.

Roxanne Kovacs and Mylene Lagarde. Does high workload reduce the quality of healthcare? evidence from rural senegal. *Journal of Health Economics*, 82: 102600, 03 2022. doi: 10.1016/j.jhealeco.2022.102600.

Ida Lykke Kristiansen and Sophie Yanying Sheng. Doctor who? the effect of physician patient match on the ses-health gradient. *Working Paper*, 05 2022. URL https://www.econ.ku.dk/cebi/publikationer/working-papers/CEBI_WP_05-22.pdf.

Jennifer Kwok. How do primary care physicians influence healthcare? evidence on practice styles and switching costs from medicare, 2018. URL https://static1.squarespace.com/static/5bd6632951f4d49caf7eecd5/t/5dcc79af9653dd567508688f/1573681584132/JenniferKwok_PrimaryCarePhysicians_JMP.pdf.

Soonwoo Kwon. Optimal shrinkage estimation of fixed effects in linear panel data models, 03 2023. URL https://soonwookwon.github.io/files/Kwon_OptShrFE_Mar23.pdf.

A. Lans Bovenberg and Bas Jacobs. Redistribution and education subsidies are siamese twins. *Journal of Public Economics*, 89:2005–2035, 12 2005. doi: 10.1016/j.jpubeco.2004.12.004.

Nicholas Lawson. Liquidity constraints, fiscal externalities and optimal tuition subsidies. *American Economic Journal: Economic Policy*, 9:313–343, 11 2017. doi: 10.1257/pol.20150079.

Andrew Lee. Flat fee compensation, lawyer incentives, and case outcomes in indigent criminal defense, 2021. URL https://andrewlee543.github.io/files/AndrewLee_JMP.pdf.

Den Norske Legeforening. Normaltariffen, 2022. URL https://normaltariffen.legeforeningen.no/book/Fastlegetariffen-2022.

Edwin Leuven, Hessel Oosterbeek, and Bas van der Klaauw. The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, 8:1243–1265, 2010. URL https://www.jstor.org/stable/40961558.

Adam Looney and Constantine Yannelis. What went wrong with federal student loans? *The Journal of Economic Perspectives*, 38:209–236, 08 2024. doi: 10.1257/jep.38.3.209.

Lovdata. Forskrift om fastlegeordning i kommunene, 09 2017. URL https://lovdata.no/dokument/SF/forskrift/2012-08-29-842#KAPITTEL_3.

Michael Lovenheim and Jonathan Smith. Returns to different postsecondary investments: Institution type, academic programs, and credentials. *Handbook of the economics of education*, 6:187–318, 01 2023. doi: 10.1016/bs.hesedu.2022.11.006.

James Macinko, Barbara Starfield, and Leiyu Shi. Quantifying the health benefits of primary care physician supply in the united states. *International journal of health services : planning, administration, evaluation*, 37:111–26, 2007. doi: 10.2190/3431-G6T7-37M8-P224. URL https://www.ncbi.nlm.nih.gov/pubmed/17436988.

Davide Malacrino, Samuel Nocito, and Raffaele Saggio. Do reforms aimed at reducing time to graduation work? evidence from the italian higher education system, 07 2024. URL https://www.nber.org/papers/w32659?utm_campaign=ntwh&utm_medium=email&utm_source=ntwg14.

Victoria R. Marone and Adrienne Sabety. When should there be vertical choice in health insurance markets? *American Economic Review*, 112:304–342, 01 2022. doi: 10.1257/aer.20201073.

Kifmann Mathias, Heineck Martin, and Lorenz Normann. A duration analysis of the effects of tuition fees for long-term students in germany. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 226:82–109, 2006. URL https://ideas.repec.org/a/jns/jbstat/v226y2006i1p82-109.html.

Thomas G McGuire and Mark V Pauly. Physician response to fee changes with multiple payers. *Journal of Health Economics*, 10:385–410, 01 1991. doi: 10.1016/0167-6296(91)90022-f.

Jack Mountjoy and Brent R. Hickman. The returns to college(s): Relative value-added and match effects in higher education, 09 2021. URL https://www.nber.org/papers/w29276.

Richard Murphy and Gill Wyness. Testing means-tested aid. *Journal of Labor Economics*, 03 2022. doi: 10.1086/719995.

Florence Naegelen and Michel Mougeot. Power of incentives with motivated agents in public organizations. *Journal of Public Economic Theory*, 13:391–416, 05 2011. doi: 10.1111/j.1467-9779.2011.01505.x.

Tuan D. Nguyen, Jenna W. Kramer, and Brent J. Evans. The effects of grant aid on student persistence and degree attainment: A systematic review and meta-analysis of the causal evidence. *Review of Educational Research*, 89: 831–874, 09 2019. doi: 10.3102/0034654319877156.

Statistics Norway. Gps and emergency primary health care, 05 2023. URL https://www.ssb.no/en/helse/helsetjenester/statistikk/allmennlegetjenesten.

Norwegian Ministry of Health and Welfare. Time for action — the personnel in a sustainable health and care service, 02 2023. URL https://www.regjeringen.no/no/dokumenter/nou-2023-4/id2961552/?ch=6.

Anthea Page, Sarah Ambrose, John Glover, and Diana Hetzel. Atlas of avoidable hospitalisations in australia: ambulatory care-sensitive conditions, summary, 04 2007. URL https://www.aihw.gov.au/reports/hospitals/atlas-avoidable-hospitalisations-australia/summary.

Veronica Rattini. The effects of financial aid on graduation and labor market outcomes: New evidence from matched education-labor data. *Economics of Education Review*, 96:102444, 10 2023. doi: 10.1016/j.econedurev.2023.102444. URL https://www.sciencedirect.com/science/article/pii/S0272775723000912.

Judith Rosta, Olaf G Aasland, and Magne Nylenna. Changes in job satisfaction among doctors in norway from 2010 to 2017: a study based on repeated surveys. *BMJ Open*, 9:e027891, 09 2019. doi: 10.1136/bmjopen-2018-027891.

Rachel B. Rubin. The pell and the poor: A regression-discontinuity analysis of on-time college enrollment. *Research in Higher Education*, 52:675–692, 01 2011. doi: 10.1007/s11162-011-9215-6.

Adrienne Sabety. The value of relationships in healthcare. *Journal of Public Economics*, 225:104927, 09 2023. doi: 10.1016/j.jpubeco.2023.104927. URL https://www.sciencedirect.com/science/article/abs/pii/S0047272723001093.

Adrienne H. Sabety, Anupam B. Jena, and Michael L. Barnett. Changes in health care use and outcomes after turnover in primary care. *JAMA Internal Medicine*, 181:186, 02 2021. doi: 10.1001/jamainternmed.2020.6288.

Ity Shurtz, Alon Eizenberg, Adi Alkalay, and Amnon Lahad. Physician workload and treatment choice: The case of primary care, 09 2018. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3247229.

Marianne Simonsen, Lars Skipper, Niels Skipper, and Peter Rønø Thingholm. Discontinuity in care: Practice closures among primary care providers and patient health care utilization. *Journal of Health Economics*, 80:102551, 12 2021. doi: 10.1016/j.jhealeco.2021.102551.

Ellen Rabben Svedahl, Kristine Pape, Marlen Toch-Marquardt, Lena Janita Skarshaug, Silje-Lill Kaspersen, Johan Håkon Bjørngaard, and Bjarne Austad. Increasing workload in norwegian general practice – a qualitative study. *BMC Family Practice*, 20, 05 2019. doi: 10.1186/s12875-019-0952-5. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6530128/.

Arthur Taburet, Alberto Polo, and Quynh-Anh Vo. Screening using a menu of contracts: A structural model of lending markets. *SSRN Electronic Journal*, 01 2024. doi: 10.2139/ssrn.4720778.

Roosa Tikkanen and Melinda Abrams. U.s. health care from a global perspective, 2019: Higher spending, worse outcomes?, 01 2020. URL https://www.commonwealthfund.org/sites/default/files/2020-01/Tikkanen_US_hlt_care_global_perspective_2019_OECD_db_v2.pdf.

Laura Vallejo-Torres and Stephen Morris. Primary care supply and quality of care in england. *The European Journal of Health Economics*, 19:499–519, 2018. doi: 10.1007/s10198-017-0898-2. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5913392/.

Wilbert van der Klaauw. Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach*. *International Economic Review*, 43:1249–1287, 11 2002. doi: 10.1111/1468-2354.t01-1-00055. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-2354.t01-1-00055.

Benjamin Vatter. Quality disclosure and regulation: Scoring design in medicare advantage, 09 2022. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4250361.

Joel Waldfogel. First degree price discrimination goes to school. *The Journal of Industrial Economics*, 63:569–597, 12 2015. doi: 10.1111/joie.12085.

Yaping Wu, Yijuan Chen, and Sanxi Li. Optimal compensation rule under provider adverse selection and moral hazard. *Health Economics*, 27:509–524, 09 2017. doi: 10.1002/hec.3590.

Yunchou Wu. Essays in health economics: The design of primary care incentives. 2020. URL https://livrepository.liverpool.ac.uk/3090824/1/200907640_June2020.pdf.

Jia Xiang. Physicians as persuaders: Evidence from hospitals in china. *Available at SSRN 3906909*, 2021.

Xuan Zhang. The effects of physician retirement on patient outcomes: Anticipation and disruption. *Journal of Public Economics*, 207:104603, 03 2022. doi: 10.1016/j.jpubeco.2022.104603.

Vita

Jori Barash

Education

B.S. in Economics and Mathematics, University of Southern California	*2017*
M.S. in Economics, University of Texas at Austin	*2021*
Ph.D. in Economics, University of Texas at Austin	*2025*

Subfields of Study

Empirical Industrial Organization, Market Design, Labor Economics

Address: joribarash@gmail.com