Should Physicians Choose Their Reimbursement Rate? Menu Design for Physician Payment Contracts*

Jori Barash†

July 2024

Click here for the most recent version.

Abstract

By screening physicians with differentiated contracts, healthcare payers might better address under- and over-treatment. I characterize how patient health and public expenditure depend on the dispersion and correlation of physicians' marginal cost, altruism, and productivity. In the setting of Norwegian primary care, I find novel reduced-form evidence that physicians vary along these dimensions. To simulate outcomes under counterfactual menus of linear contracts, I estimate the joint distribution of physician heterogeneity, exploiting a sudden increase in marginal reimbursement and subsequent changes to treatment intensity. Relative to the status quo uniform contract, a budget-neutral menu would increase patient health by approximately $33 million per year, driven by greater treatment intensity among physicians with low altruism and high cost, particularly among high-severity patients in rural areas.

Keywords: physician agency, self-selection
JEL Codes: D04, D47, H51, I11, J33

*This research has received support from the Research Council of Norway (grant #256678). Data made available by Statistics Norway and the Norwegian Patient Registry have been essential for this project. I am grateful to my advisors Bob Town, Victoria Marone, and Michael Geruso for guidance and to Jason Abaluck, Hande Celebi, Jeffrey Clemens, Gue Sung Choi, Martin Gaynor, Geir Godager, Zijing He, Yumin Hong, Jordan Jones, Hayden Parsley, and Margarita Petrusevich for helpful suggestions. Comments from participants at UT Austin seminars, the Norwegian Health Economics Seminar, IIOC, KHEC, and ASHEcon are gratefully acknowledged. Thank you to Alexander Vu for excellent research assistance.

†Department of Economics, University of Texas at Austin. Email: joribarash@utexas.edu.

1 Introduction

Health insurance programs need to incentivize an appropriate level of treatment when contracting with physicians. Since physicians' treatment decisions often respond to reimbursement rates, contracting can reduce under-treatment by increasing rates and over-treatment by decreasing rates.[1] Across the developed world, a large share of healthcare spending is paid according to physician fee schedules, so improved schedule design may have substantial aggregate impacts on health equity and cost.[2] Insurance programs tend to fixate on the level of reimbursement for a single fee schedule. However, allowing physicians to instead choose one of several fee schedules might induce more appropriate levels of treatment.

This paper presents the first empirical evidence that replacing a single fee schedule with a shared menu of contracts would result in large cost-effective gains to health. I show how the distribution of unobserved physician heterogeneity determines whether the shared menu should include more than one contract. I estimate this distribution in the context of Norwegian primary care, evaluate the social cost of informational asymmetry, and derive the menu of linear contracts that maximizes patient health.

I present a model of physician decision-making to quantify the expenditure and health impacts of counterfactual reimbursement schemes. In the model, physicians choose a contract and then each patient's level of treatment. Each contract consists of a base payment per patient-month and a reimbursement rate per unit of treatment intensity ("fee-for-service rate," abbreviated "FFS"), e.g., an hour of patient interactions.[3] Physicians choose treatment intensity to maximize a weighted sum of their private net income and patient health production (e.g., as in Ellis and McGuire, 1986). I supplement this model with heterogeneity in physicians' marginal cost of effort, altruism, and productivity, motivated by novel evidence from plausibly causal reduced-form research designs. In the model, altruism is the relative weight on patient health relative to private profit, cost lowers profit, and productivity augments treatment intensity in producing health. Relative to a regulator, physicians have private information about both their heterogeneity and patients' initial illness severity.

This model sheds light on when it is efficient for a regulator to offer more than one contract. The socially efficient menu of contracts maximizes expected health production subject to constraints on public

[1] Brekke et al. (2017) and Brekke et al. (2020) show physicians responding to financial incentives in this paper's empirical setting. Also, higher rates might have impacts beyond treatment choices, e.g., physician entry.

[2] 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).

[3] 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."

expenditure and physicians' participation. One important aspect is the correlation between a physician's cost, altruism, and productivity. For example, physicians with high costs of effort tend to provide less treatment, so the socially efficient reimbursement rate increases in cost. A uniform reimbursement rate may be too low for some high-cost physicians and too high for others. Instead, a menu can separate high-cost physicians into contracts with high reimbursement rates when higher base payments compensate low-cost physicians for accepting low rates. If physicians only varied in cost, separation is unlikely because all else equal, increasing the reimbursement rate leads to a relatively small increases in the private objective of high-cost physicians (low "willingness-to-pay"). If instead, high-cost physicians have high willingness-to-pay due to another dimension of heterogeneity like altruism, then these physicians might choose a high-reimbursement contract and efficiently increase treatment intensity.[4]

Drawing on comparative statics of the model, I find novel reduced-form evidence that Norwegian primary care physicians vary along multiple dimensions. Consistent with heterogeneity in cost of effort, treatment intensity varies 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 intensity responds heterogeneously across physicians to increased reimbursement rates.[7] Such multi-dimensional heterogeneity reinforces the potential for a menu of contracts to increase efficiency. However, 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 use maximum likelihood estimation and a balanced sample of registered patients to rationalize observed treatment intensity. Identifying variation comes from a large and sudden increase in the reimbursement rate. Local regulations rule out several sources of 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 universe of procedure-level public healthcare utilization in Norway. With patient records outside of the estimation sample, I can relax and test assumptions that may be necessary in other settings.[8]

[4] High-altruism physicians most value the gains in health from increased treatment. They may also have high cost because, e.g., prosocial physicians take on additional work in clinics.

[5] Figure A.6 illustrates the identification intuition. A physician with high cost of effort treats all types of patients less than an otherwise similar low-cost physician.

[6] Productivity augments treatment intensity in health production, so treatment intensity is 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 have less scope to vary treatment intensity. At any reimbursement rate, these physicians sacrifice profit to provide greater health production.

[8] For example, I test whether physicians' hours bunch at capacity constraints, whether patients systematically sort toward

I estimate considerable heterogeneity in physicians' marginal cost, altruism, and productivity, implying large social costs of imperfect information. Parameter estimates accurately predict treatment intensity both in- and out-of-sample, across physicians and across time for each physician. With perfect information, the regulator would offer a different contract to each physician. Physician heterogeneity corresponds to widely dispersed efficient rates which increase welfare by \$8.39 per patient-month, or approximately 70 percent of baseline spending.[9]

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 FFS rates for lower base payments. Gains are largest among high-severity patients of physicians with high cost and low altruism – those with relatively low status-quo treatment intensity who are most responsive to higher FFS rates. The menu also helps narrow rural-urban health disparities. Relative to the status quo, the gains from a menu of linear contracts are striking because menus are rarely featured in physician contract design. Relative to full-information contracts, the menu's impact is somewhat modest, highlighting the significance of informational asymmetry and the potential for further flexibility in contracting.

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 beyond Norwegian primary care.

This paper synthesizes 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, Jelovac and Léger, 2014; Barham and Milliken, 2014; Wu, Chen and Li, 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 that menu for Norwegian primary care physicians to provide the first empirical evidence that any uniform contract is less efficient. This paper also extends the empirical literature on socially optimal menu design with multi-dimensional

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.

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

consumer heterogeneity in insurance to a new selection market – physician labor supply – with unique dimensions of heterogeneity (Fang and Wu, 2018; Marone and Sabety, 2022; Ho and Lee, 2023). Estimating a joint distribution of agent types and characterizing the relative efficiency of a uniform contract is similar to the study of health insurance menus in Marone and Sabety (2022). In a parallel exercise, I use the graphical framework from Einav, Finkelstein and Cullen (2010) to provide intuition for how a two-contract menu can increase efficiency for some distributions of physicians.

I contribute to the literature documenting heterogeneity among physicians' altruism (Hennig-Schmidt, Selten and Wiesen, 2009; Godager and Wiesen, 2013; Douven, Remmerswaal and Zoutenbier, 2017; Galizzi et al., 2015) and practice style (Epstein and Nicholson, 2009; Chan and Chen, 2022; Doyle, Ewer and Wagner, 2010; Gowrisankaran, Joiner and Léger, 2017) by simultaneously estimating three key correlated dimensions of heterogeneity. Policies that assume physicians vary along only one dimension may result in unintended consequences.[10] This paper reinforces prior findings that treatment intensity increases in marginal reimbursement (Brekke et al., 2017; Einav, Finkelstein and Mahoney, 2018; Eliason et al., 2018; Clemens and Gottlieb, 2014; Cabral, Carey and Miller, 2021; Xiang, 2021). I show heterogeneity in this response, and decompose that heterogeneity into structural physician types and variation in patient treatment need. Einav et al. (2021) document hospitals' selection into bundled contracts on levels (increased revenue absent behavior change) and slopes (increased revenue from behavior change). Documenting similar selection on levels and slopes, I show how the further decomposition of physician types enables welfare analysis in contexts where selection affects both expenditure and health outcomes.

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, Mehta and Richards-Shubik (2023) estimate distributions of cost and altruism of dialysis clinics and derive the optimal non-linear uniform contract for an anti-anemia drug. I extend that paper's framework with unobserved patient severity and heterogeneity in productivity; these extensions substantially alter the optimal menu of contracts.

Going forward, Section 2 presents the theoretical model and characterizes when offering two contracts

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

is more efficient than one. Section 3 describes the empirical setting and presents novel reduced-form evidence consistent with multi-dimensional physician heterogeneity. Section 4 discusses the parameterization and identification to recover the estimates which are summarized in Section 5. Section 6 demonstrates the efficiency of a counterfactual menu of contracts, provides intuition, and evaluates robustness. Section 7 concludes.

2 Theoretical Framework

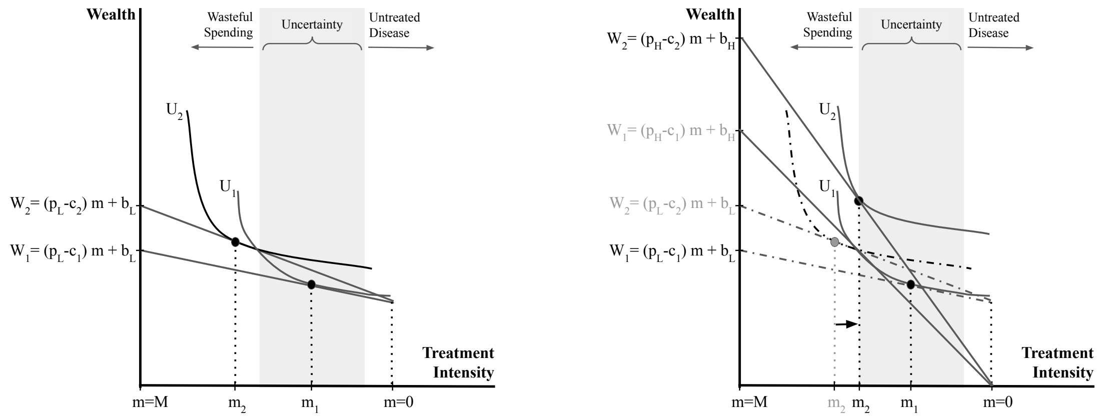
2.1 Graphical Intuition for a Menu of Contracts

Before discussing the details of the theoretical framework, I use a stylized graphical example to illustrate how a uniform contract may be inefficient when physicians are heterogeneous. 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 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 1 plots wealth W against treatment intensity 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 underlying treatment need is not observed by the regulator, 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 to untreated disease. The shaded region reflects the regulator's uncertainty about patient severity, and in turn, which level of labor supply is efficient. 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 A.1 shows a counterexample where a uniform contract is efficient. If the physicians are nearly identical, then the differences between their

Figure 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 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 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.

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.

2.2 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 draws severity 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 patient severity jointly determine patient health outcomes.

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.[11] Contracts can be thought of as ordered by p, in which case $-b(p)$ is the price of each contract. A menu of contracts is characterized by the function $b(p)$ that maps each potential reimbursement rate $p' \in [\underline{p}, \overline{p}]$ to a base payment. A menu may consist of a uniform contract, in which case all other reimbursement rates are excluded by setting sufficiently low corresponding base payments.

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 $\theta = \{c, \alpha, \gamma\}$, which is distributed in the population according to $G(\theta)$. Private cost c includes both financial and opportunity costs. Altruism α is the marginal rate of substitution between utility derived from patient health production and utility derived from net income. Both intrinsic and extrinsic forces may motivate physicians to value patient health, e.g., prosociality and reputation. Productivity γ is a measure of physician skill that determines how much treatment intensity is needed to produce a given health benefit. 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, with which a low-skill physician may under-diagnose a patient and treat less intensively (e.g., as in Abaluck et al., 2016).

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 preferences over patient health production $h(m, \gamma\lambda)$:

$$V(p; \lambda, \theta) \equiv \max_{m \geq 0} (p-c)\,m + b(p) + \alpha h(m, \gamma\lambda)\,. \tag{1}$$

After selecting a contract, the physician observed 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.[12]

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

[11] Sections 3.2 and B.2 describe how I calculate reimbursement per hour using data on higher-resolution services, e.g., visits, procedures, and diagnostics.

[12] Appendix A.3 relaxes and tests the assumption of linear preferences over net income with, e.g., taste for leisure or a constraint on aggregate treatment intensity.

health production subject to a global budget constraint and each physician's participation constraint.[13] Total payments to physicians ("expenditure") cannot exceed the budget threshold, which incorporates the 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:

$$\begin{aligned} \max_{b(p)} \quad & \int_\theta E[h(m^*(p^*_\theta;\theta), \gamma\lambda; \theta) \mid \lambda \sim F]\, dG(\theta) && (2) \\ \text{s.t.} \quad & \int_\theta E[p^*_\theta m^*(p^*_\theta;\theta) + b(p^*_\theta) \mid \lambda \sim F]\, dG(\theta) \le \bar{B} && [\mu_B, \text{ Budget}] \\ & E[V(p^*_\theta;\theta) \mid \lambda \sim F] \ge \bar{v}(\theta),\ \forall\theta && [\mu_{P,\theta}, \text{ Participation}] \end{aligned}$$

where μ_B and $\mu_{P,\theta}$ are the shadow costs of expenditure and participation.[14] The social objective partially coincides with the physician objective because of altruism and a binding participation constraint, 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 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 informational asymmetry about physician types θ. In this case, the regulator sets a personalized contract for each physician $p^{FB}_\theta(m;\theta)$, which corresponds to the efficient level of treatment intensity, $m^*(p^{FB}_\theta;\theta)$ ("first-best").

[13] Equivalently, the regulator maximizes a weighted sum of expectations over health production, expenditure, and physician indirect utility.

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

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 in physicians' marginal cost and decreases in altruism (See Appendix C.1). As the budget constraint relaxes, this level converges to private marginal cost.

2.3 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 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 is it efficient to introduce a second contract? 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 with relatively low cost, high altruism, and high productivity have the highest willingness to pay for a greater reimbursement rate. Rate increases 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-FFS 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

$$E[\Delta h(m(p),\gamma\lambda) \mid \Delta E\,V(p) \geq 0, \Delta E[pm(p) + b(p)] \leq 0] \geq 0\,. \tag{3}$$

All physicians who choose p_H will increase treatment intensity relative to p_L. If h is locally monotonic and concave, 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 FFS reimbursement? Necessarily, physicians choosing the high-FFS contract must value incremental health production more than incremental costs on average. Importantly, physician contract choice is a selection market – the average cost of the high-FFS 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 to FFS expenditure among physicians who choose the high-FFS contract:

$$E\left[\Delta\left(pm(p,\lambda)+b\right) \mid \Delta E[V(p,b,\lambda)] \geq 0\right] \leq 0 \tag{4}$$

Comparing the partial derivatives of indirect utility and expenditure highlights the roles of correlation and dispersion.[15] Physicians are more likely to choose the high-FFS contract if they have low cost, high altruism, high productivity, or high patient severity $E\lambda$.[16] If physicians only vary along one of these four dimensions, self-selection leads to more positive incremental expenditure, potentially violating the budget constraint. In direct contrast, physicians are most likely to decrease expected FFS 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-FFS contract does not necessarily generalize to the broader question of menu design with any number of contracts. For example, if the FFS rate of the reference contract is lower than the optimal uniform contract, it may be efficient to add a higher-FFS 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-FFS 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 subsidize efficiently higher FFS rates and health production for other physicians. Nevertheless, the sufficiency condition's intuition may inform reframing the problem as a sequence of two-contract menus that span a large set of reimbursement rates.[17]

3 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. Going forward, I extend the framework to estimate

[15]See Appendix C.1 for derivations and a similar discussion with weaker assumptions.

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

[17]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.

such a distribution, derive the optimal menu, and measure its impacts. I first explore several necessary assumptions in the setting of Norwegian primary care. Section 3.1 presents institutional details, which support that the focal variation in treatment intensity is driven by physician heterogeneity and contracts rather than patient composition. Section 3.2 details the construction of a balanced estimation sample of patients that further removes potentially confounding variation. Section 3.3 introduces reduced-form evidence consistent with physician heterogeneity in cost, altruism, and productivity, which suggests that the status-quo uniform contract may be inefficient.

3.1 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 greater FFS revenue.[18] 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.[19] Once the training is completed, physicians can apply for the certificate, which they typically receive within three months of application. Supplementary payments begin around that time and continue for five years. Before 2017, 80 percent of physicians received this certificate during their careers.[20]

Apart from certification, physicians face nationally uniform reimbursement incentives. On average, physicians receive 70 percent of revenue from FFS payments governed by an administratively set schedule of rates.[21] 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).[22] 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,

[18] 24 percent reflects an average within the estimation sample, including reimbursement for other services provided during consultations.

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

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

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

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

highlighting the importance of physicians' discretion in choosing treatment intensity (See Table A.1).[23] The other 30 percent of revenue comes from base payments of approximately $4 per registered patient per month. Both FFS 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. Beneath the contracted maximum number of patients, physicians must accept any patients who choose to register. National guidance states that physicians must be accessible to registered patients within contracted opening hours, e.g., 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 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 among physicians with fewer patients than the contracted maximum. The choice set infrequently changes due to the national licensing system, which fixes the total number of local physicians in the short term. Long-term relationships between physicians and patients help construct a representative balanced panel for the estimation sample.

3.2 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").[24] 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 the Norwegian Directorate of

[23] In the United States, most claims for primary care consultations also include one of a small number of procedure codes.
[24] 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.

Health and Statistics Norway.[25] 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 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 A.2 provides more detail on sample selection. For robustness analyses, I construct a similarly defined control sample using patients whose physicians do not experience sudden changes in reimbursement.[26]

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 FFS 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.[27] I inflate all money-metric variables by Norway's monthly all-goods-and-services CPI to January 2023 USD.

Summary statistics suggest that the final estimation sample is approximately representative. The estimation sample includes 643,363 patient-spells (13 months each) at 619 unique physicians.[28] Table 1 describes the distribution of selected characteristics and outcomes six months before certification, and three facts stand out.[29] First, most patients do not visit their physician during a typical month. Second, the average physician spends 28 hours per week with registered patients (90th percentile = 40) suggesting that with sufficient reimbursement, physicians can increase treatment intensity. Third, there

[25] See Appendix B.1 for additional details on data sources.

[26] 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.

[27] See Appendix B.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 A.3 shows average characteristics and sample share separately for each patient type, including the simulated wage.

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

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

Table 1: Registered Patient Summary Statistics

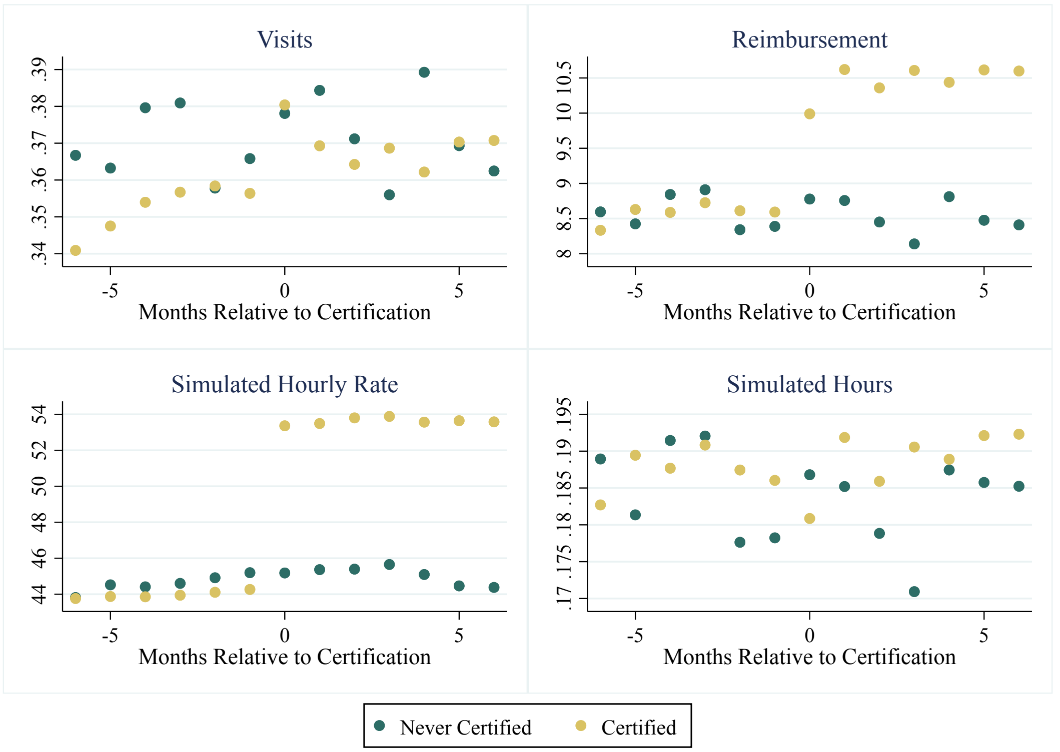
	Control Sample	Estimation Sample					
	Mean	Mean	Std. Dev.	% > 0	10th	50th	90th
Patient Characteristics							
Reimbursement	8.59	8.33	25.49	20.74	0.00	0.00	30.92
Simulated Hourly Rate	43.82	43.76	6.86	100.00	32.38	45.49	50.95
Simulated Hours	0.19	0.18	0.56	20.74	0.00	0.00	0.68
Capitation Payment	4.03	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	30.72	28.77	10.23	100.00	14.23	29.61	40.38
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 10th, 50th, and 90th percentiles. Monetary measures are in USD. Physician Characteristics are also averaged across patients. The last two Physician Characteristics reflect shares of registered patients.

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 plots the trend in raw means, showing that visits, total reimbursement, and simulated hours all increase suddenly 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 A.2). These plots and most subsequent analyses reflect short-run variation (13 months) around the sudden change in incentives, months after physicians tend to 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 A.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.

Figure 2: Raw Means of Treatment Intensity Relative to Certification



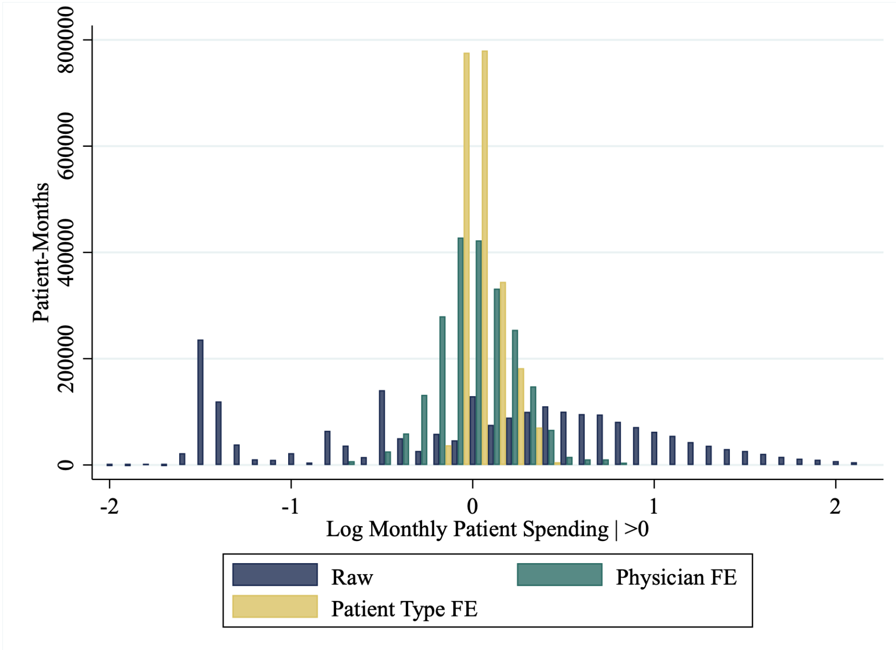
Notes: These plots show averages of treatment intensity outcomes across patient-months in the estimation and control samples in each month relative to certification. Each sample is a balanced panel of patients, and in the estimation sample, Month 0 is the first month in which the registered physician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

3.3 Stylized Facts

A necessary condition for physician self-selection is that physicians vary across at least one dimension. I show novel reduced-form evidence consistent with heterogeneity in physicians' cost, altruism, and productivity. First, I show descriptively that some physicians treat observably similar patients more than others, driving a large share of variation in treatment intensity. Second, I exploit quasi-random patient assignment among a subset of physicians to suggest that this heterogeneity is not driven by patient selection and distinguish the roles of cost and productivity. Third, with a stacked differences-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 3 shows the persistent variation across physicians in how inten-

Figure 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.

sively 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.[30] 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. By contrast, physician fixed effects are widely dispersed, highlighting the large role of physicians in treatment intensity, similar to recent work like 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

[30]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.

et al. (2022).[31] As shown in Figure A.5, these physician effects are similarly dispersed even after shrinking effects to account for estimation error, reinforcing the importance of persistent physician heterogeneity. If patients selected physicians based on unobserved type, then the assignment effects would be less dispersed. For example, high-severity patients might have large gains to health from choosing low-cost physicians, producing a negative covariance. Limited patient selection is consistent with evidence from Norway that both treatment intensity and patient star ratings are uncorrelated with causal reductions in 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 A.5 shows that both sets of assignment effects are highly dispersed. For example, moving from the 10th to 90th 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 health production can vary among patients with similar severity receiving similar treatment intensity. Other natural experiments show dispersion across physicians in notions of productivity like resource use and skill, e.g., avoiding hospital readmissions (Doyle, Ewer and Wagner, 2010; Gowrisankaran, Joiner and Léger, 2017; Chan, Gentzkow and Yu, 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.[32] To evaluate the effect of higher reimbursement from

[31] 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.

[32] For any health production function, 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. With the parameterization used below, $\frac{dm}{dp} = \frac{1}{\alpha}$.

certification on treatment intensity, I estimate the following stacked differences-in-differences regression:

$$Y_{ijt} = \beta_1 Post_{jt} \times Certified_j + \boldsymbol{\beta}_x \boldsymbol{X}_{jt} + \gamma_i + \gamma_{y(t)} + \gamma_{m(t)} + \epsilon_{ijt} \quad (5)$$

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 greater treatment in the six months after certification than 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 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 showing that average reimbursement does not trend differently for certified versus non-certified physicians in the months before certification.[33]

Table 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 visits and in-visit services.[34] Simulated hours, which combines all categories of treatment, increases 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 at 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.[35] I do not find evidence that certification immediately affects specialist treatment or acute hospitalizations.

[33] See Section 3.2 for discussion of the long-run variation shown in Figures A.3 and A.13.

[34] 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.

[35] If registered physicians' valuation of health production does not fully internalize substitution with other providers, then changes to health production might also be underrepresented.

Table 2: Main Effects of Certification on Treatment Intensity

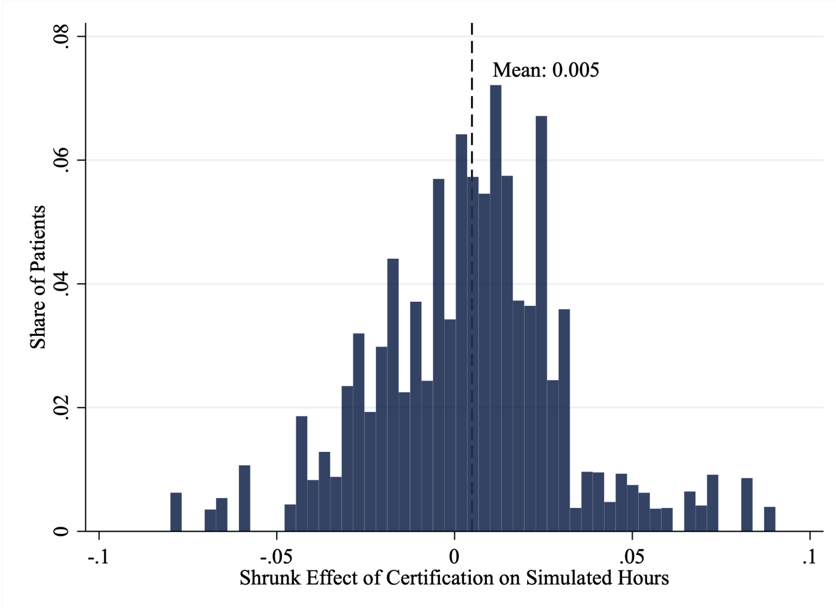
	Post × Certified		Mean (Pre)	R^2	Obs.
Visits	0.015***	(0.001)	0.355	0.401	9,301,956
Reimbursement	2.093***	(0.106)	8.581	0.213	9,301,956
Simulated Hours	0.006**	(0.002)	0.187	0.186	9,301,956
Procedures	−0.001	(0.001)	0.071	0.237	9,301,956
Diagnostics	0.009***	(0.002)	0.229	0.266	9,301,956
Extra Time 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 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 become certified for registered patients, among complete spells. Unless otherwise indicated, all outcomes are specific to a pair of physician and patient with registration numbers, and zeroes are included. Visits includes any in-person encounter. Reimbursement indicates FFS revenue. Simulated Hours is reimbursement divided by a price index as described in Section 3.2. 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.

Consistent with dispersion in physicians' altruism, I find heterogeneity in the effect of certification on treatment intensity. I extend the difference-in-difference analysis to include a post-certification indicator for each physician. Figure 4 plots 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 6.2. 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, Selten and Wiesen, 2009). To interpret estimated elasticities exclusively as altruism, physicians must not vary in their ability to increase treatment intensity. Section 6.2 discusses several tests of this assumption. For example, Figure A.14 provides descriptive evidence that capacity constraints do not bind in this setting and Figure A.15 shows that high-altruism and low-altruism physicians similarly respond to observed shocks to patient health..

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. Section 2.3 illustrates how physicians with high efficient reimbursement rates must have relatively high willingness-to-pay for higher rates. Before estimating the

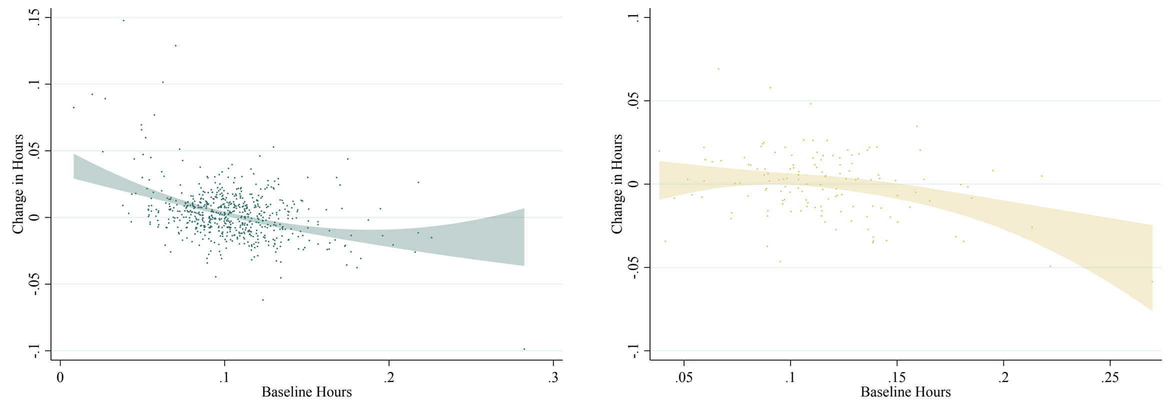
Figure 4: Distribution of Physician-Level Effect of Certification on Simulated Hours



Notes: This histogram shows estimates of β_{1j} from equation 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 structure for physician types in the next section, I check for a consistent pattern in the raw data. Figure 5 shows that physicians who initially provide low treatment intensity (e.g., from high cost) tend to most increase treatment intensity after certification (e.g., from low altruism). That pattern does not clearly hold among physicians without reimbursement rate variation.

Figure 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 of 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.

4 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. This section reviews additional assumptions to support estimation as well as the intuition for which patterns in the data help recover each parameter.

4.1 Parameterization

I estimate the joint distribution of physician heterogeneity and the distribution of patient severity by maximizing the likelihood of observed treatment intensity. Privately optimal treatment intensity equates marginal net income with 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.[36] To generate a likelihood, I make two parametric assumptions that I later relax in Section 6.2. First, since economies of scale are unlikely in this setting, I assume that costs increase linearly in treatment intensity: $c(m) = c\,m$.[37] Second, health production is quadratic in the distance between treatment intensity and productivity-scaled severity: $h(m, \lambda; \gamma) =$

[36] Sections 3.3 and 6.2 discuss evidence supporting this assumption.

[37] 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.

$H - \frac{1}{2}(m - \gamma\lambda)^2$. 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:

$$m^*(p, \lambda, F) = \max\{0, \frac{p-c}{\alpha} + \gamma\lambda\}\,. \tag{6}$$

Gaynor, Mehta and Richards-Shubik (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: $(\ln \lambda \mid \lambda > 0) \sim N(\beta_\lambda X_\lambda, \sigma_\lambda)$.[38] 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 C.2 presents the full expression of the conditional likelihood.

4.2 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. 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. Figure A.6 shows stylized visual examples of these patterns. Residual of this variation in physician heterogeneity, the 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.

[38] 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

4.3 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: $l(m \mid \theta_i, p, F)$.[39] 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. Estimated parameters are sometimes simple transformations of model parameters.[40] 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.[41] To accommodate computer resource constraints, I separately estimate parameters for three subsamples: spells starting in 2008-2010, 2011-2013, and 2014-2016. I use the 2011-2013 subsample for counterfactuals because estimates best predict treatment intensity.

5 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 6 shows a correlation of nearly 1 for both the estimation sample and a control sample of never-certified physicians.[42] Estimates predict treatment intensity well both across physicians and over time for particular physicians. Table A.12 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 intensity.[43] 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 A.8 shows the distribution of this change in EV across physicians.

The correlation between estimated cost, altruism, and productivity reinforces the potential for efficient self-selection. Figure A.7 summarizes the joint distribution of physician heterogeneity graphically. Overall, high-cost physicians tend to have low altruism and low productivity, while high-altruism physi-

[39] I use L-BFGS-B with the Python module JAX to calculate the analytic gradients of the log-likelihood objective.

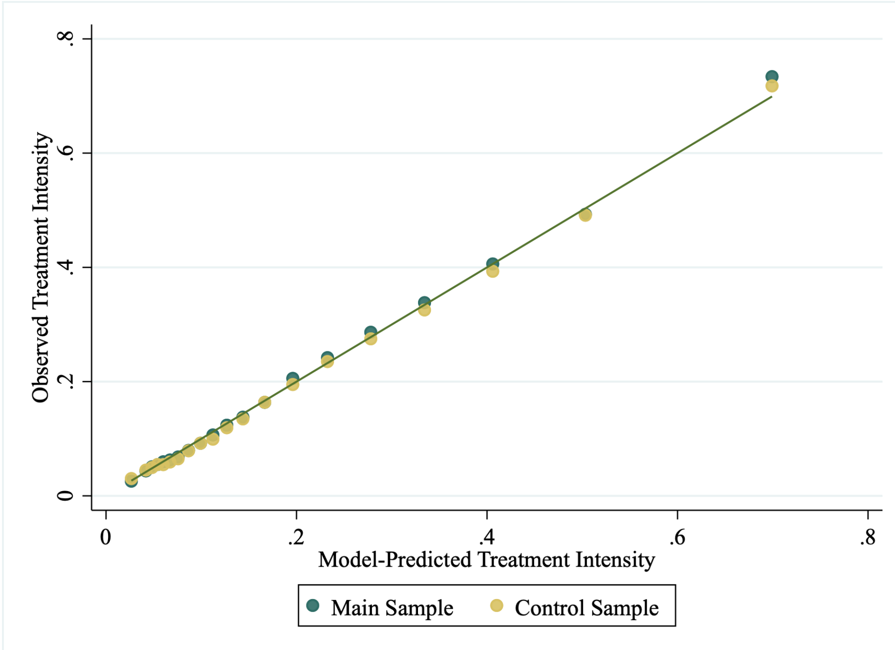
[40] c is a multiple of the FFS rate six months before certification, α is scaled by 1000, and σ_λ is exponentiated. The transformation of c implies that marginal cost varies across patients of the same physician.

[41] 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.

[42] 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 3.2).

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

Figure 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.

cians tend to have high productivity. The upper panel of Table 3 shows that observed characteristics explain some of this variation.[44] For example, productive physicians tend to be younger and born in Norway. They hold larger lists of patients, make greater use of diagnostics relative to procedures,and historically worked under fee-for-service contracts. Physicians that work in rural municipalities tend to have high cost and low altruism. The bottom panel shows residual variation in physician types is also widely dispersed and correlated..[45]

Estimates of the distribution of patient severity imply that patient observable characteristics explain a moderate share of the variation in treatment intensity (See Table A.6). Seasonality and particular chronic illnesses are major determinants of patients' treatment need. 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. A 1-standard deviation increase in lagged treatment only increases the health shock about as much as the average difference

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

[45] 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.

Table 3: Correlates of Physician Heterogeneity

	$\ln c$	$\ln \alpha$	$\ln \gamma$
Constant	0.852***	7.046***	−0.315***
	(0.207)	(0.424)	(0.044)
Age	0.009	0.040***	0.021***
	(0.007)	(0.014)	(0.002)
Max Enrollment	0.019**	−0.003	−0.004**
	(0.008)	(0.015)	(0.002)
Pr(Diagnostic)	−0.041***	−0.014	−0.080***
	(0.009)	(0.016)	(0.002)
Ever Fixed-Salary	0.051	0.093	0.094***
	(0.042)	(0.087)	(0.009)
Female	0.020	−0.066**	0.002
	(0.015)	(0.030)	(0.003)
Migrant	0.013	0.030	0.043***
	(0.017)	(0.035)	(0.003)
Rural Municipality	0.044**	−0.070*	0.004
	(0.018)	(0.037)	(0.004)
Trend	−0.207	−0.520	0.204***
	(0.241)	(0.494)	(0.050)
S.D. Residual	0.170***	0.365***	0.140***
	(0.010)	(0.009)	(0.002)
$\rho(\ln c, \ln \alpha)$	0.429***		
	(0.035)		
$\rho(\ln c, \ln \gamma)$	0.542***		
	(0.038)		
$\rho(\ln \alpha, \ln \gamma)$	−0.141***		
	(0.046)		

Notes: This table regresses log physician-level estimates of cost c, altruism α, and 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.

between January and April. This small coefficient reinforces the assumption that the distribution of health shocks is conditionally independent across months within each patient. Residual of the rich set of patient covariates, variation in patient severity is large. For example, Table A.7 shows that replacing σ_λ with zero would lower the average treatment intensity by 30 percent and the variance in treatment intensity by 77 percent. For comparison, the variance would be 9 percent lower with patients uniformly distributed across physicians.

6 Counterfactual Menus of Contracts

6.1 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 informational 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, 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, 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.[46] Table 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 FFS rates. Appendix B.3 provides additional detail on how I measure counterfactual outcomes and search for counterfactual menus.

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 FFS 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 7 shows substantial heterogeneity in the efficient reimbursement rates.[47] On average, efficient rates are 84 percent above the initial status quo rate with substantial variation (SD =

[46] For comparison, Gaynor, Mehta and Richards-Shubik (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.

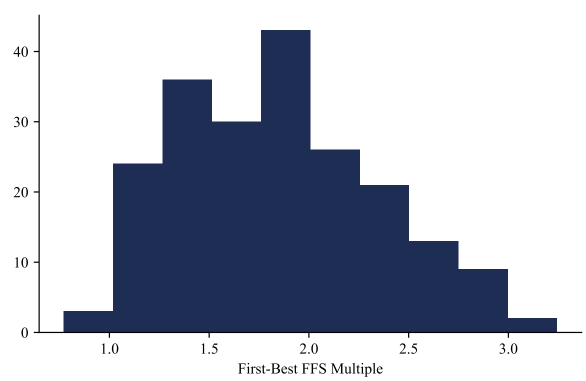
[47] Throughout this section, I discuss counterfactual reimbursement rates multiples. For example, 1.2 indicates 120 percent of the initial FFS rate. This approach preserves variation in FFS rates across patients while allowing simple graphical comparisons across counterfactuals. In a robustness check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

Table 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	138.9	0.264	138.9	113.6
	(0.4)	(0.001)	(0.4)	(0.4)
Efficient Contracts	525.5	1.000	137.0	0.0
	(3.0)	(0.000)	(0.6)	(0.0)
Optimal Uniform Contract	153.6	0.292	132.5	103.6
	(2.1)	(0.003)	(0.5)	(0.5)
Optimal Menu of Contracts	176.0	0.335	144.6	108.9
	(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 FFS rates, and base payments are fixed at values six months before certification. Post-certification FFS rates are fixed at values in the month after certification. Counterfactuals vary FFS 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 FFS 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 A.11 and A.12 further illustrate the distribution of counterfactual contracts across bootstrap samples.

Figure 7: Dispersion in Efficient Reimbursement Rates



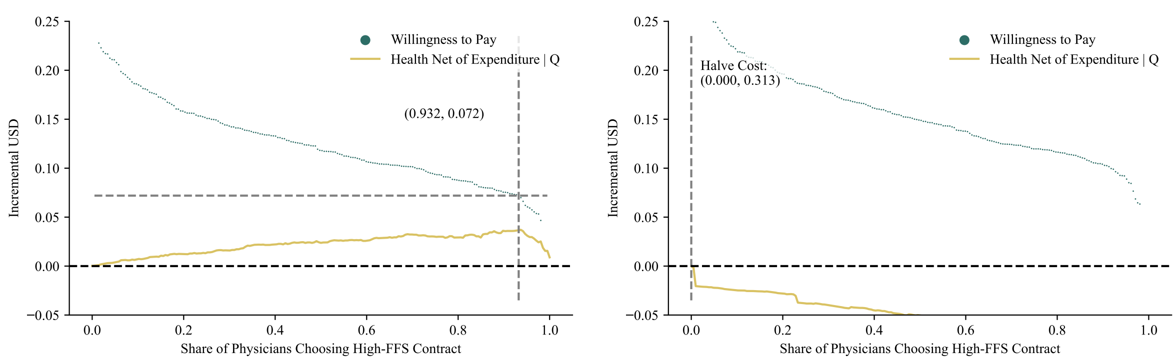
Notes: The y-axis is the count of physicians in each bin. The x-axis is a multiple of pre-certification FFS that maximizes scaled health production subject to strict physician-level participation constraints and a global budget constraint (average expenditure must be less than status quo post-certification). The grid of FFS 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 percent). Efficient rates are large because a large share of physicians have high cost or low altruism.

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 large 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 2.3, this intermediate 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, Finkelstein and Cullen (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 greater reimbursement rate. If p_H requires accepting a relatively low base payment, then only a fraction of physicians with high willingness to pay will choose it, where $WTP \equiv EV(p_H, 0) - EV(p_L, 0)$. High-WTP physicians have relatively low cost, high altruism, and high productivity (See Appendix C.1). However, these characteristics also predict relatively large increases in expenditure which might outweigh the corresponding increase in treatment intensity, especially if cost is low relative to altruism.

Figure 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 FFS rate and a FFS rate that is incrementally higher while varying the difference in the base payment between these contracts. The x-axis reflects the continuum of physicians, ordered by decreasing willingness to pay (WTP) for the high-FFS contract, where WTP is the difference in expected indirect utility between the high- and low-FFS contracts. The green line is incremental social surplus for each percentile of WTP: expected (scaled) health production minus expenditure among all patients (and all physicians). Grey dashed lines indicate the optimal share of physicians choosing the high-FFS 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 8a shows the tradeoff between increased health production and increased expenditure across physicians, ordering physicians by decreasing WTP. The WTP curve is like a demand curve, indicating participation in the high-FFS contract for various prices Δb. I also summarize welfare as incremental social surplus: expected health production minus expected expenditure, relative to the low-FFS contract, where expenditure reflects both FFS and base payment changes in equilibrium.[48] For each share of physicians choosing high-FFS, 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-FFS contract with a $0.07 lower base payment. With smaller differences in the base payment, more physicians would choose the high-FFS contract and expenditure would outweigh incremental health production. Figure A.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 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 so that all physicians choose the low-FFS contract.

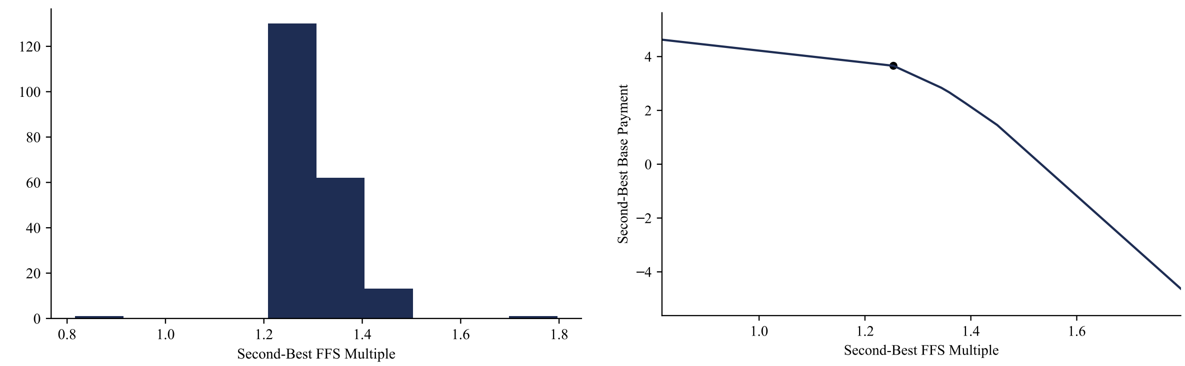
The optimal 7-contract menu achieves large efficiency gains by separating some physicians into high-FFS contracts: $33 million per year more than the status quo or 34 percent of first-best. To search for this menu, I adapt the line-search algorithm from Marone and Sabety (2022) and Azevedo and Gottlieb (2017). Most physicians choose just one of three contracts (Figure 9a) and the optimal base payment decreases concavely in the FFS rate (Figure 9b).[49] Perhaps a smaller menu would involve lower implementation costs; Figure A.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.

In part, redistribution across patients drives the gain in average efficiency from efficient contracts and the optimal menu. To explore redistribution, I first examine dispersion in treatment intensity. Table A.7 shows that relative to the status quo, the overall variance of treatment intensity increases for both efficient contracts and the optimal menu. Without considering multidimensional heterogeneity, it may seem counterintuitive for both welfare and dispersion in treatment intensity to increase. For example, the

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

[49] Moreover, Figures A.12 shows that across bootstrap samples, the optimal menu consistently lies approximately on the same curve of Base Payment versus FFS Multiple.

Figure 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 FFS 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 FFS rates for the optimal menu. The point indicates the optimal uniform contract.

place-based effects literature (e.g., Badinski et al., 2023) often considers counterfactuals aimed at lowering dispersion in utilization, perhaps based on the intuition that dispersion reflects excessive treatment for some and inadequate treatment for others. To better understand the equity implications of dispersion, I disaggregate counterfactual outcomes across physician types. Table 5 categorizes physicians into 16 groups based on whether cost, altruism, productivity, and expected patient severity are each 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. These physicians tend to have high WTP. Incremental expenditure is typically less than incremental health production among these physicians.

6.2 Robustness

Relaxing restrictions on model assumptions and sample construction suggests that the efficiency of self-selection is not driven by 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. Motivating this analysis, Table 1 shows that non-certified physicians have slightly higher treatment intensity which might not be fully explained by observed differences, e.g., more patients that are slightly older and more chronically ill. Non-certified physicians are also older, less likely to be migrants, and more likely to use diagnostics. To explore unobserved differences for non-certified physicians, I estimate the distribution of unobserved heterogeneity based on the relatively weak

Table 5: Counterfactual Outcomes by Physician Type

Physicians		Efficient Contracts		Menu of Contracts		
Type	Share	$\Delta E[h(m)]$	$\Delta E[pm+b]$	$\Delta E[h(m)]$	$\Delta E[pm+b]$	$\Delta E[V(p)]$
$c_L, \alpha_H, \gamma_L, F_L$	0.171	1.316	0.558	0.854	1.506	1.178
$c_H, \alpha_L, \gamma_H, F_L$	0.160	14.155	3.414	2.980	1.544	1.161
$c_H, \alpha_L, \gamma_H, F_H$	0.155	19.983	4.475	5.768	3.129	2.296
$c_L, \alpha_H, \gamma_L, F_H$	0.154	1.439	0.621	1.151	2.544	2.011
$c_L, \alpha_L, \gamma_H, F_L$	0.049	3.861	1.368	1.741	1.960	1.527
$c_H, \alpha_L, \gamma_L, F_H$	0.047	21.536	4.810	4.904	2.010	1.412
$c_L, \alpha_H, \gamma_H, F_H$	0.045	3.200	1.251	2.272	3.964	3.140
$c_H, \alpha_H, \gamma_L, F_L$	0.037	5.662	1.972	1.722	1.567	1.206
$c_L, \alpha_H, \gamma_H, F_L$	0.033	2.181	0.852	1.218	2.031	1.630
$c_H, \alpha_H, \gamma_L, F_H$	0.031	6.611	2.367	2.451	2.522	1.951
$c_H, \alpha_H, \gamma_H, F_L$	0.025	5.997	2.052	1.942	2.065	1.654
$c_H, \alpha_L, \gamma_L, F_L$	0.022	7.346	2.343	2.156	1.366	0.977
$c_H, \alpha_H, \gamma_H, F_H$	0.019	6.271	2.311	3.404	3.949	3.041
$c_L, \alpha_L, \gamma_L, F_L$	0.017	5.736	1.273	6.488	2.202	1.129
$c_L, \alpha_L, \gamma_L, F_H$	0.017	6.693	1.690	10.592	4.987	2.855
$c_L, \alpha_L, \gamma_H, F_H$	0.017	4.335	1.624	3.106	3.863	2.855

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, 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, $\Delta E[pm + b]$ represents incremental expected expenditure and ΔEV represents incremental expected indirect utility. Outcomes are averages across patients within each group, measured in USD.

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 within-physician variation in FFS 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 3). Likewise, Table A.12 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.[50]

Second, I find evidence for external validity outside of Norway: even large perturbations of estimates

[50]See Table A.10. 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.

rarely change the main finding. In Table A.10, I first perturb cost c, altruism α, and productivity γ. 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.[51] When doubling the variance of cost or altruism, I cannot find efficiency gains from a menu. Although this perturbation increases the variation in efficient rates, it weakens the correlation between incremental efficiency and willingness-to-pay from increased rates. 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 4), physician-specific contracts (Figure A.11), and the relationship between reimbursement rate and base payment in the optimal menu (Figure A.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 if enrollment is lower than its 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 A.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.[52] Enrollment and the share of patients with higher treatment need should increase if patients are sorting towards physicians with greater health production after certification due to increased treatment intensity. Likewise, as shown in Section 3.3, physicians' fixed effects in treatment intensity are similarly 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,

[51] The exception is halving cost of effort.

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

and fixed effects for year and calendar month.[53] Column (2) of Table A.11 shows that the correlation between health and switching to a new physician is imprecise with point estimates small in magnitude. By contrast, expected health production is predictive of (lower) future avoidable hospitalizations and mortality. In Figure A.13, raw means also suggest a decline in avoidable hospitalization after three years. Likewise, cumulative mortality is 36 percent lower than among patients 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 lower treatment intensity by a small amount in response to newly registered patients (Barash, 2023) or an increase in reimbursement rates (Figure 4). To estimate physicians' marginal disutility of expected workload, I extend the theoretical framework and estimation strategy with additional assumptions, detailed in Appendix A.3. If income effects do exist, they seem too small relative to unobserved variation in patient severity to be economically meaningful. Figure A.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. By contrast, if some physicians occasionally reached capacity due to idiosyncratic variation in the number of patients or realized severity, then monthly treatment intensity would bunch at high values. Next, Figure A.15 shows that the treatment intensity of high-altruism and low-altruism physicians is similarly 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.[54] I do not find evidence that patients of high-altruism physicians are more likely to seek treatment elsewhere.[55] Finally, as shown in Table A.10, the optimal menu of contracts leads to similar welfare gains over a uniform contract when I impose a capacity constraint and repeat counterfactuals.[56] 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

[53] 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.

[54] 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).

[55] 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.

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

valuable but unmeasured.[57]

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.[58] 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 1, and health production should initially increase in treatment. I estimate $h_1 = 0.073$ and the likelihood ratio test rejects $h_1 = 0$.

6.3 Extensions

Even when considering more flexible contracts, a menu of contracts tends to dominate a uniform contract and the gap remains large between a menu and personalized contracts under full information. 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 A.9 shows that welfare is similar when contracts depend on observed patient severity.[59] Higher-spending patient types tend to have larger gains from efficient rates and the optimal menu while the difference between the optimal menu and the optimal uniform contract is similar.

Second, Table A.10 shows that counterfactual outcomes are similar to Baseline when separating the analysis between urban and rural patients.[60] The limited benefits of regional contracts is surprising because in the status quo, regional variation is one of few exceptions to nationally uniform reimbursement.[61] Baseline (national) contracts might perform relatively well because within-region physician heterogeneity is larger than across-region heterogeneity. Consistent with prior literature, Table A.8 shows that health disparities among rural patients remain a pressing concern. Relative to the status quo, eliminating informational asymmetry about physicians improves patient health by $13 for the most rural patients and

[57] Table A.10 shows similar gains to a menu when excluding physicians that initially work part-time, i.e., those who spend less than 25 hours per week with registered patients.

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

[59] I use the same procedure as before, but separately for each of ten observed patient types. Contracts include a FFS rate (as a level rather than multiple of status quo) and a base payment. The final row aggregates across counterfactuals weighting by overall sample share of patient types. Each aggregated number in the final row is a weighted mean of weighted means, so it is not directly comparable to Baseline.

[60] 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.

[61] 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.

$6 for most urban patients. A 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. I search for an optimal uniform contract that is quadratic rather than linear in treatment intensity: $x(m) = b + p_1 m + p_2 m^2$. Institutional differences may explain why a non-linear uniform contract achieves large welfare gains in Gaynor, Mehta and Richards-Shubik (2023) but not in this setting. 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, Mehta and Richards-Shubik (2023), more than half of observed treatment intensity was high enough to damage health based on a known cutoff.[62]

The regulator might further improve patient health through policies that complement contracts by shifting the allocation of patients across physicians. For example, in Table 5, the decomposition across physician and patient types suggests that perhaps, high severity patients be not registered with high-cost low-altruism physicians. 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 locations. For example, adjusting base payments for the observed health of nearby patients may induce better match quality. 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, I find suggestive evidence that it may be a fruitful path for future research. Table A.7 shows that the variance of treatment intensity would be 9 percent lower if patients were uniformly distributed across physicians. Such differences in variance highlight the influence of patient-physician sorting on treatment. This finding roughly mirrors Badinski et al. (2023) which is based on US Medicare beneficiaries' annual utilization.[63] Location explains some of this variation: Table A.8 shows that welfare gains are larger for rural patients. I also find that combining efficient reimbursement rates with optimal patient switches can increase incremental social surplus by 17

[62] These characterizations mostly refer to Figure 3 in that paper, which is based on a patient with median observed severity. The non-parametric optimal contract is approximately quadratic over the distribution of realized treatment, suggesting that this parametric robustness check may be informative. With my model, it is straightforward to simulate counterfactual outcomes with quadratic contracts because the first-order condition is still linear in treatment intensity and severity. I search over a base payment, linear FFS multiple, and a uniform quadratic term, using the trust gradient algorithm to enforce constraints, initializing parameters at the optimal uniform linear contract.

[63] Badinski et al. (2023) estimate that removing variation across regions in persistent physician heterogeneity would reduce the gap in utilization across above-median and below-median regions by 20 percent. This number is not directly comparable in part because place-based effects may partially reflect persistent reimbursement differences across regions in the United States.

percent relative to efficient reimbursement rates alone.[64]

So far, while fixing the distribution of physicians, the cost of informational asymmetry remains large despite added 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; performance benchmarks might increase altruism, and promoting long-term patient-physician relationships might increase productivity via soft knowledge. For example, benchmarks can increase information adding information and facilitating learning, particularly about past patients that have since left the list. Physicians current do not observe all the long-term impacts of treatment like utilization and avoidable hospitalizations.

7 Conclusion

This paper presents 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. To show this, 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

[64] This exercise involves a stylized example of two vertically differentiated physicians at the 10th and 90th percentile of (initial) efficient FFS rates. I begin by counterfactually assigning the average patient distribution, corresponding FFS rates, and average enrollment to both physicians, simulate each patient, then alternate 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.

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 suggestive evidence of 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.[65]

Beyond Norway, this paper's framework for evaluating the efficiency of self-selection is broadly applicable to settings featuring heterogenous 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. This paper's framework makes use of reimbursement variation, which often exists in these settings, but even cross-sectional data can be sufficient with additional assumptions.[66]

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

This paper also explores 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. For example, making physicians identical would not reduce variance, while eliminating variation in reimbursement

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

[66] 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.

across patients would reduce variance by 13 percent. Second, consistent with existing evidence that patients imperfectly perceive physician quality, I find suggestive evidence that patients are not 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. Third, I do not find evidence of income effects or capacity constraints in Norway, but these features may add nuance to contracting in related settings.

References

Abaluck, Jason, Leila Agha, Chris Kabrhel, and Ali Raja. 2016. "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care - ProQuest."

Allard, Marie, Izabela Jelovac, and Pierre-Thomas Léger. 2014. "Payment mechanism and GP self-selection: capitation versus fee for service." *International Journal of Health Care Finance and Economics*, 14: 143–160.

Azevedo, Eduardo M., and Daniel Gottlieb. 2017. "Perfect Competition in Markets With Adverse Selection." *Econometrica*, 85: 67–105.

Badinski, Ivan, Amy Finkelstein, Matthew Gentzkow, and Peter Hull. 2023. "Geographic Variation in Healthcare Utilization: The Role of Physicians."

Barash, Joridan. 2023. "Practice Size and Short-Run Utilization: Evidence from Norway."

Barham, Victoria, and Olga Milliken. 2014. "Payment Mechanisms and the Composition of Physician Practices: Balancing Cost-Containment, Access, and Quality of Care." *Health Economics*, 24: 895–906.

Bensnes, Simon, and Ingrid Huitfeldt. 2021. "Rumor has it: How do patients respond to patient-generated physician ratings?" *Journal of Health Economics*, 76: 102415.

Biasi, Barbara. 2021. "The Labor Market for Teachers under Different Pay Schemes." *American Economic Journal: Economic Policy*, 13: 63–102.

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

Brekke, Kurt R., Tor Helge Holmås, Karin Monstad, and Odd Rune Straume. 2020. "How Does The Type of Remuneration Affect Physician Behavior?" *American Journal of Health Economics*, 6: 104–138.

Brown, Zach Y., Christopher Hansman, Jordan Keener, and Andre F. Veiga. 2023. "Information and Disparities in Health Care Quality: Evidence from GP Choice in England."

Cabral, Marika, Colleen Carey, and Sarah Miller. 2021. "The Impact of Provider Payments on Health Care Utilization: Evidence from Medicare and Medicaid."

Cardon, James H., and Igal Hendel. 2001. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *The RAND Journal of Economics*, 32: 408.

Chade, Hector, Victoria R. Marone, Amanda Starc, and Jeroen Swinkels. 2022. "Multidimensional Screening and Menu Design in Health Insurance Markets."

Chan, David C, Matthew Gentzkow, and Chuan Yu. 2022. "Selection with Variation in Diagnostic Skill: Evidence from Radiologists." *The Quarterly Journal of Economics*, 137: 729–783.

Chan, Jr, David C., and Yiqun Chen. 2022. "The Productivity of Professions: Evidence from the Emergency Department."

Chartock, Benjamin. 2023. "Quality Disclosure, Demand, and Congestion: Evidence from Physician Ratings."

Choné, Philippe, and Ching-to Albert Ma. 2011. "Optimal Health Care Contract under Physician Agency." *Annals of Economics and Statistics*, 229.

Clemens, Jeffrey, and Joshua D. Gottlieb. 2014. "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?" *American Economic Review*, 104: 1320–1349.

Douven, Rudy, Minke Remmerswaal, and Robin Zoutenbier. 2017. "Do Altruistic Mental Health Care Providers Have Better Treatment Outcomes?" *Journal of Human Resources*, 54: 310–341.

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

Einav, Liran, Amy Finkelstein, and Mark R. Cullen. 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices*." *Quarterly Journal of Economics*, 125: 877–921.

Einav, Liran, Amy Finkelstein, and Neale Mahoney. 2018. "Provider Incentives and Healthcare Costs: Evidence from Long-Term Care Hospitals."

Einav, Liran, Amy Finkelstein, Stephen P Ryan, Paul Schrimpf, and Mark R Cullen. 2013. "Selection on Moral Hazard in Health Insurance." *American Economic Review*, 103: 178–219.

Einav, Liran, Amy Finkelstein, Yunan Ji, and Neale Mahoney. 2021. "Voluntary Regulation: Evidence from Medicare Payment Reform." *The Quarterly Journal of Economics*.

Eliason, Paul J., Paul L. E. Grieco, Ryan C. McDevitt, and James W. Roberts. 2018. "Strategic Patient Discharge: The Case of Long-Term Care Hospitals." *American Economic Review*, 108: 3232–3265.

Ellis, Randall P., and Thomas G. McGuire. 1986. "Provider behavior under prospective reimbursement." *Journal of Health Economics*, 5: 129–151.

Epstein, Andrew J., and Sean Nicholson. 2009. "The formation and evolution of physician treatment styles: An application to cesarean sections." *Journal of Health Economics*, 28: 1126–1140.

Fang, Hanming, and Zenan Wu. 2018. "Multidimensional private information, market structure, and insurance markets." *The RAND Journal of Economics*, 49: 751–787.

Galizzi, Matteo M, Time Tammi, Geir Godager, Ismo Linnosmaa, and Daniel Wiesen. 2015. "Provider altruism in health economics."

Gaynor, Martin, Nirav Mehta, and Seth Richards-Shubik. 2023. "Optimal Contracting with Altruistic Agents: Medicare Payments for Dialysis Drugs." *The American Economic Review*, 113: 1530–1571.

Ginja, Rita, Julie Riise, Barton Willage, and Alexander Willén. 2022. "Does Your Doctor Matter? Doctor Quality and Patient Outcomes." *SSRN Electronic Journal*.

Godager, Geir, and Daniel Wiesen. 2013. "Profit or patients' health benefit? Exploring the heterogeneity in physician altruism." *Journal of Health Economics*, 32: 1105–1116.

Gottlieb, Joshua, Maria Polyakova, Kevin Rinz, U Census Bureau, Hugh Shiplett, and Victoria Udalova. 2020. "Who Values Human Capitalists' Human Capital? Healthcare Spending and Physician Earnings."

Gowrisankaran, Gautam, Keith A. Joiner, and Pierre-Thomas Léger. 2017. "Physician Practice Style and Healthcare Costs: Evidence from Emergency Departments."

Hanushek, Eric A., Jin Luo, Andrew J. Morgan, Minh Nguyen, Ben Ost, Steven G. Rivkin, and Ayman Shakeel. 2023. "The Effects of Comprehensive Educator Evaluation and Pay Reform on Achievement."

Hennig-Schmidt, Heike, Reinhard Selten, and Daniel Wiesen. 2009. "How Payment Systems Affect Physicians' Provision Behavior - An Experimental Investigation." *SSRN Electronic Journal*.

Ho, Kate, and Robin S Lee. 2023. "Health insurance menu design for large employers." *The RAND Journal of Economics*, 54: 598–637.

Iversen, Tor, and Hilde Lurås. 2011. "Patient switching in general practice." *Journal of health economics*, 30: 894–903.

Jack, William. 2005. "Purchasing health care services from providers with unknown altruism." *Journal of Health Economics*, 24: 73–93.

Ji, Yunan. 2021. "Can Competitive Bidding Work in Health Care? Evidence from Medicare Durable Medical Equipment."

Kwon, Soonwoo. 2023. "Optimal Shrinkage Estimation of Fixed Effects in Linear Panel Data Models."

Lee, Andrew. 2021. "Flat Fee Compensation, Lawyer Incentives, and Case Outcomes in Indigent Criminal Defense."

Legeforening, Den Norske. 2022. "Normaltariffen."

Lovdata. 2017. "Forskrift om fastlegeordning i kommunene."

Marone, Victoria R., and Adrienne Sabety. 2022. "When Should There Be Vertical Choice in Health Insurance Markets?" *American Economic Review*, 112: 304–342.

McGuire, Thomas G, and Mark V Pauly. 1991. "Physician response to fee changes with multiple payers." *Journal of Health Economics*, 10: 385–410.

Naegelen, Florence, and Michel Mougeot. 2011. "Power of Incentives with Motivated Agents in Public Organizations." *Journal of Public Economic Theory*, 13: 391–416.

Page, Anthea, Sarah Ambrose, John Glover, and Diana Hetzel. 2007. "Atlas of avoidable hospitalisations in Australia: ambulatory care-sensitive conditions, Summary."

Vatter, Benjamin. 2022. "Quality Disclosure and Regulation: Scoring Design in Medicare Advantage."

Wu, Yaping, Yijuan Chen, and Sanxi Li. 2017. "Optimal compensation rule under provider adverse selection and moral hazard." *Health Economics*, 27: 509–524.

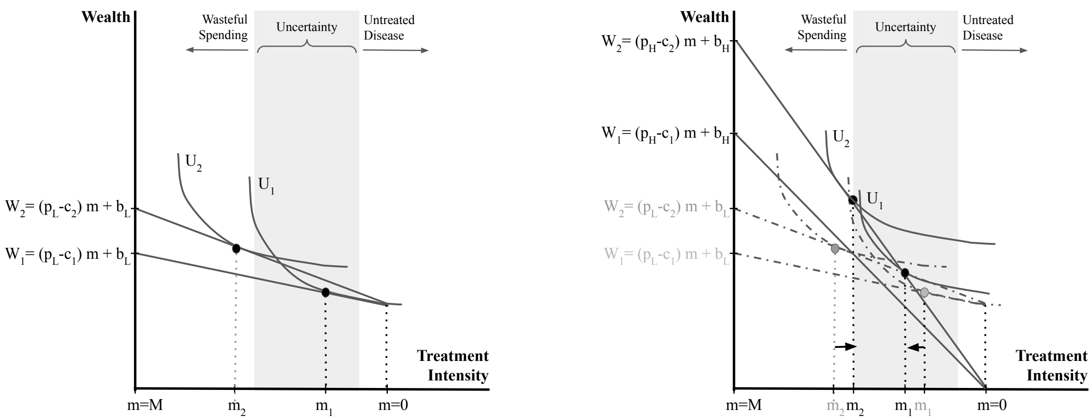
Wu, Yunchou. 2020. "Essays in Health Economics: The Design of Primary Care Incentives."

Xiang, Jia. 2021. "Physicians as Persuaders: Evidence from Hospitals in China." *Available at SSRN 3906909*.

A Additional Analysis

A.1 Additional Figures

Figure A.1: A Uniform Contract May Be Efficient

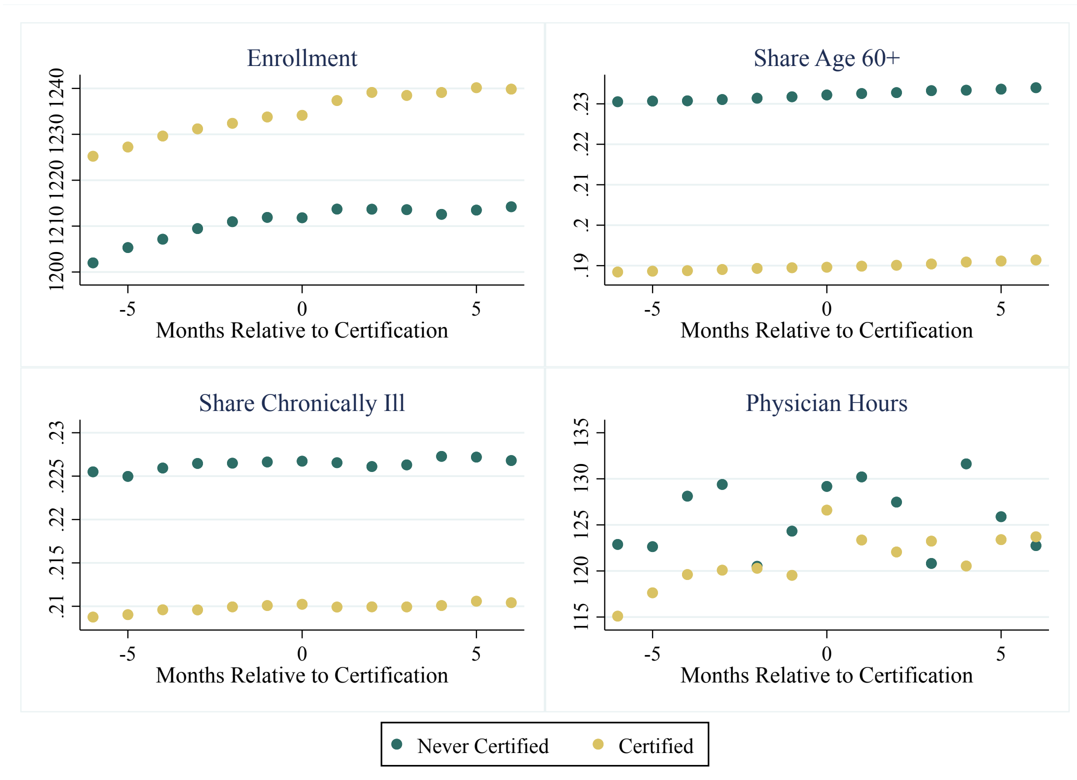


(a) Initial Uniform Contract

(b) Better Uniform Contract

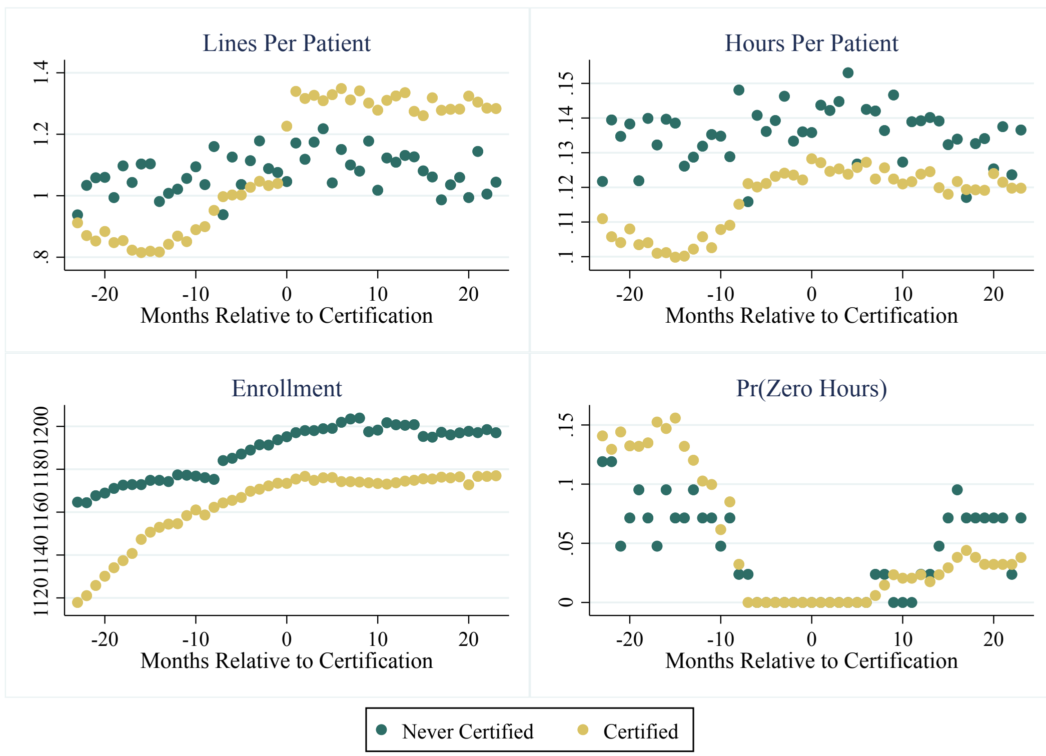
Notes: This figure shows a stylized example with two physicians, in which a a uniform contract is efficient. The x-axis plots treatment intensity $m \in [0, M]$ from right to left. Each panel shows the indifference curves of these physicians and the budget constraint(s) implied by simple reimbursement 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 A.2: Raw Means of Characteristics Relative to Certification



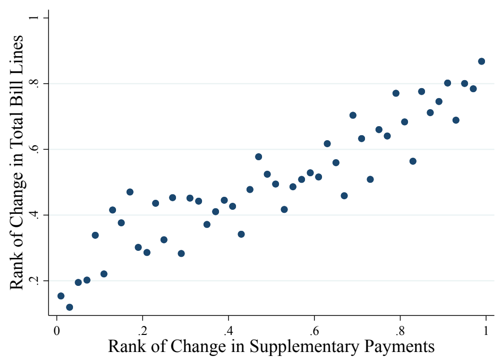
Notes: These plots show averages of treatment intensity outcomes across patient-months in the estimation and control samples in each month relative to certification. Each sample is a balanced panel of patients, and in the estimation sample, Month 0 is the first month in which the registered physician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

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

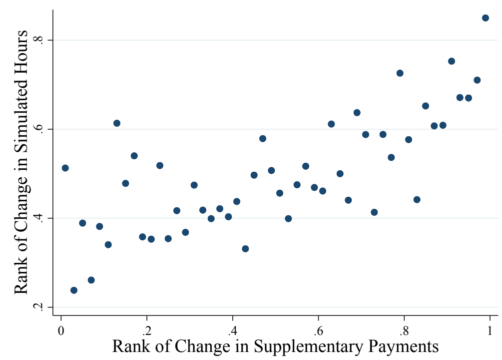


Notes: These plots show average across physicians in the estimation and control samples in each month relative to certification. Each sample is a panel of physicians, and in the estimation sample, Month 0 is the first month in which the registered physician received a certification supplement. Unlike in other analyses like Figures 2 and A.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.

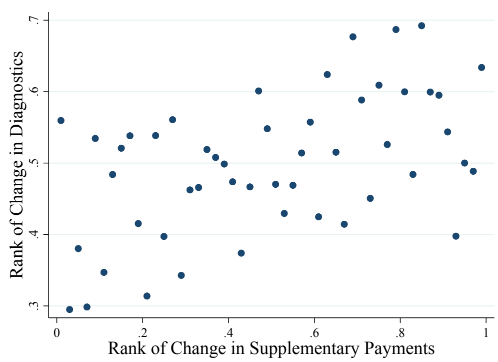
Figure A.4: Similarity of Physician-Specific Responses to Certification Across Treatment Types



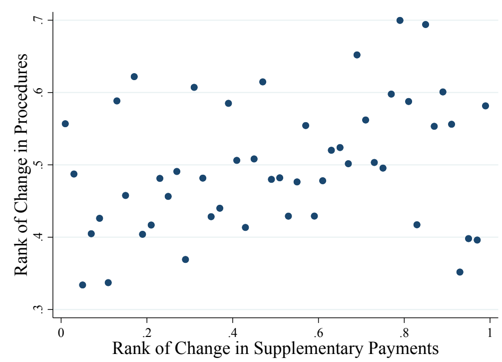
(a) Total Bill Lines



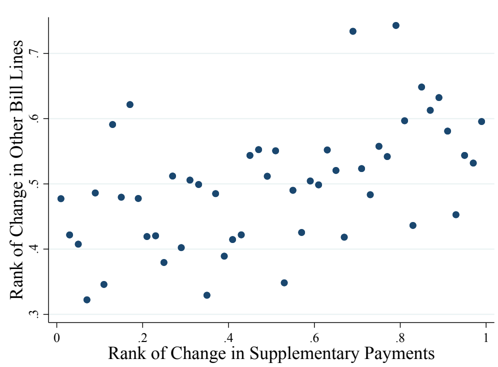
(b) Simulated Hours



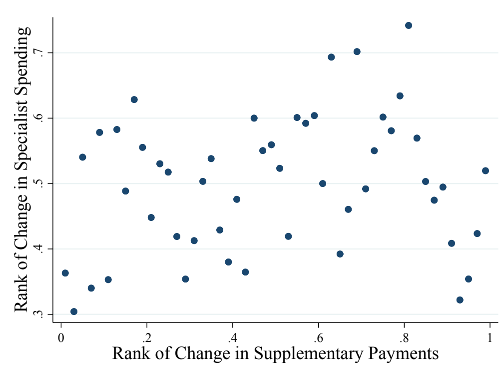
(c) Diagnostics



(d) Procedures



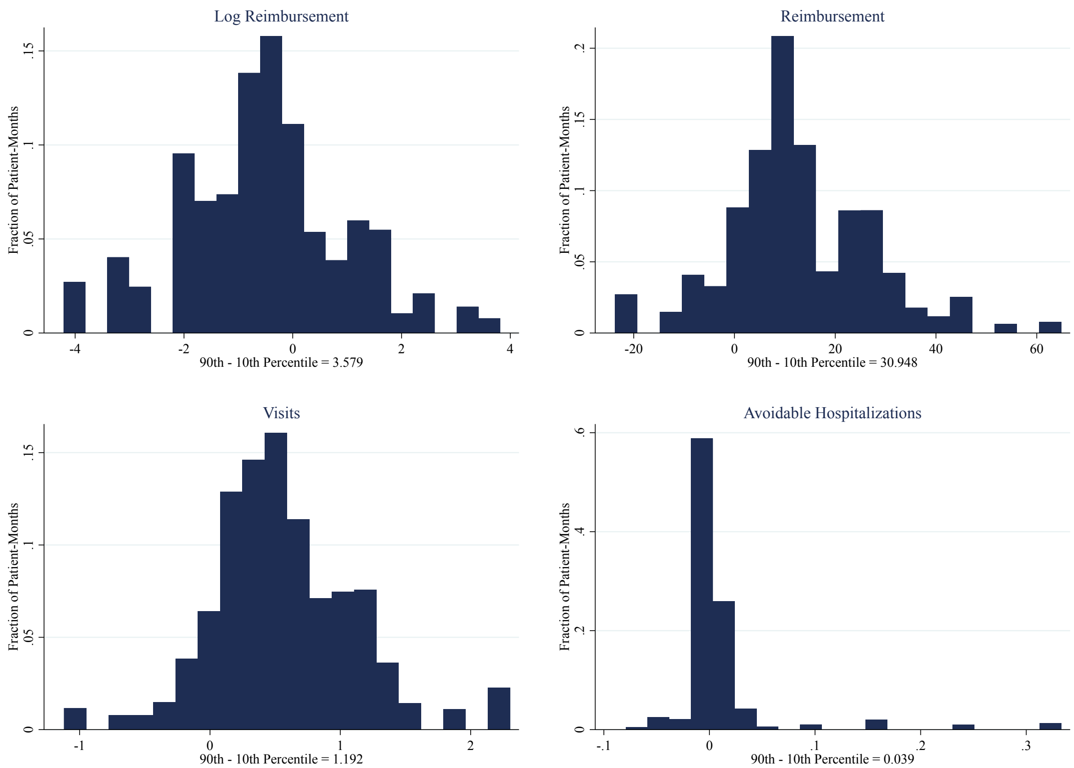
(e) Other Bill Lines



(f) Specialist Spending

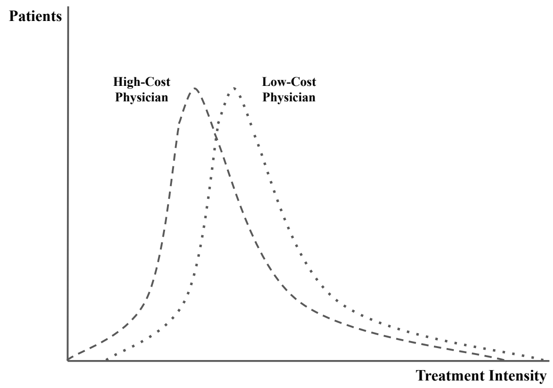
Notes: These plots show that physicians with relatively large increases in supplementary payments post-certification also have relatively large post-certification increases in other measures of treatment intensity. I first take means across patient-months in the six months pre-certification and the six 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 A.5: Shrunk Assignment Effects for Certified Physicians

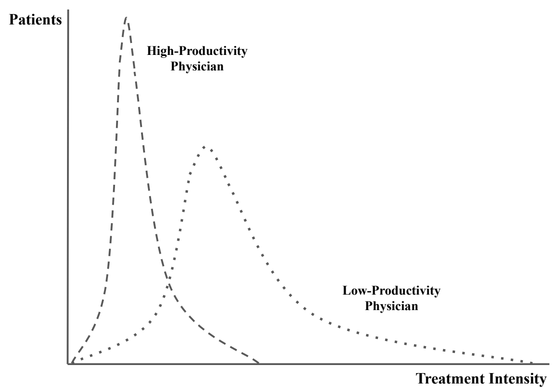


Notes: These histograms show the distribution assignment effects among physicians in the main estimation sample. Following Ginja et al. (2022), I estimate assignment effects by comparing patients from the same exiting physician who are conditionally randomly assigned to various focal physicians. Assignment 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.

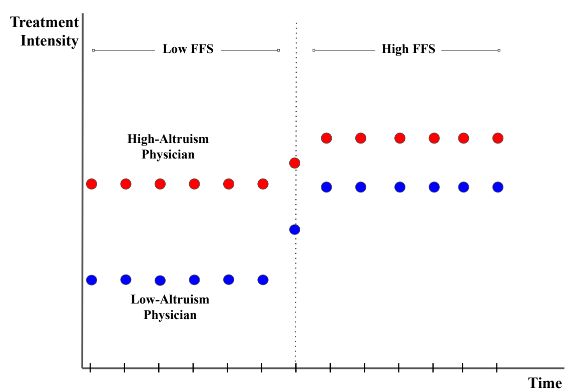
Figure A.6: Stylized Example of Identification Intuition



(a) Cost



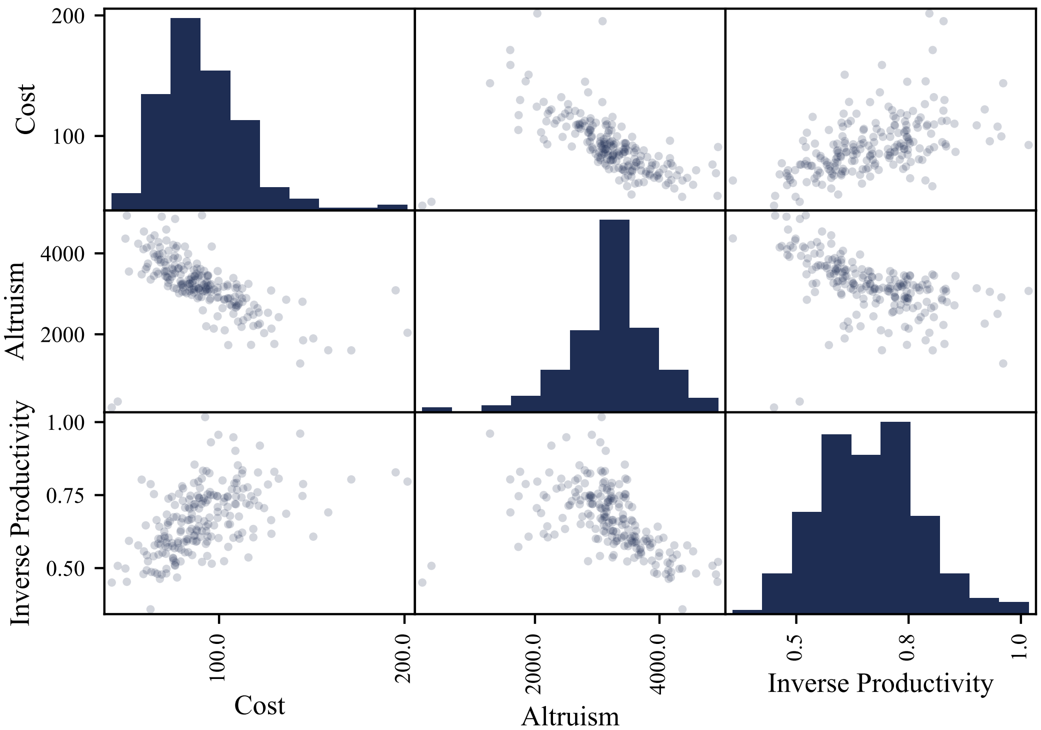
(b) Productivity



(c) Altruism

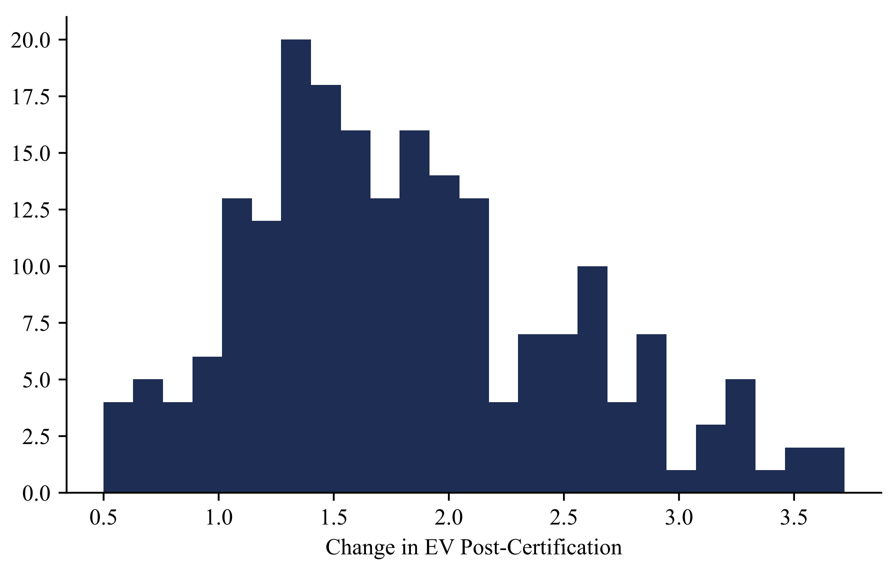
Notes: These plots illustrate the identification intuition of physician heterogeneity for the main specification ($\sigma = 0$). All else equal, cost represents a level shift in the distribution of treatment intensity, productivity increases the dispersion of that distribution, and altruism lowers responsiveness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

Figure A.7: Distribution of Physician Heterogeneity



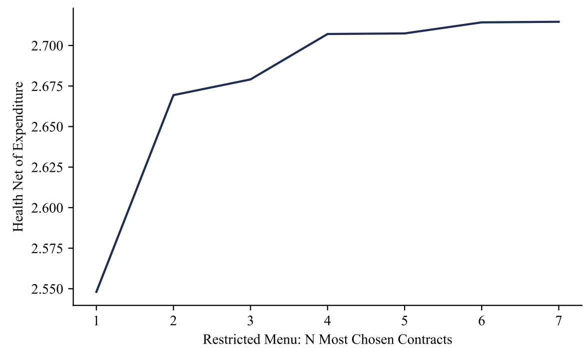
Notes: These plots summarize the distribution of estimated cost, altruism, and productivity across the full estimation sample. Plots on the diagonals are histograms and plots off the diagonals are two-way correlations.

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



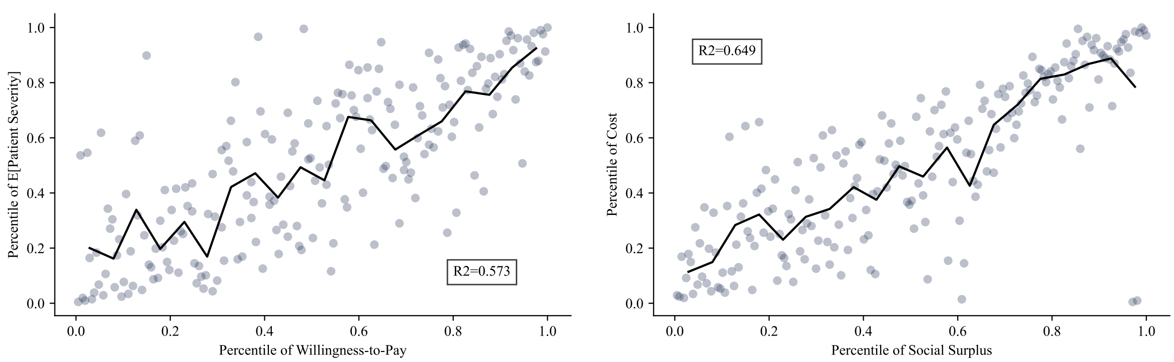
Notes: The y-axis is the count of physicians in each bin. The x-axis is the difference in average expected indirect utility (per patient-month) after certification minus before certification. Integration uses 6 quadrature nodes.

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

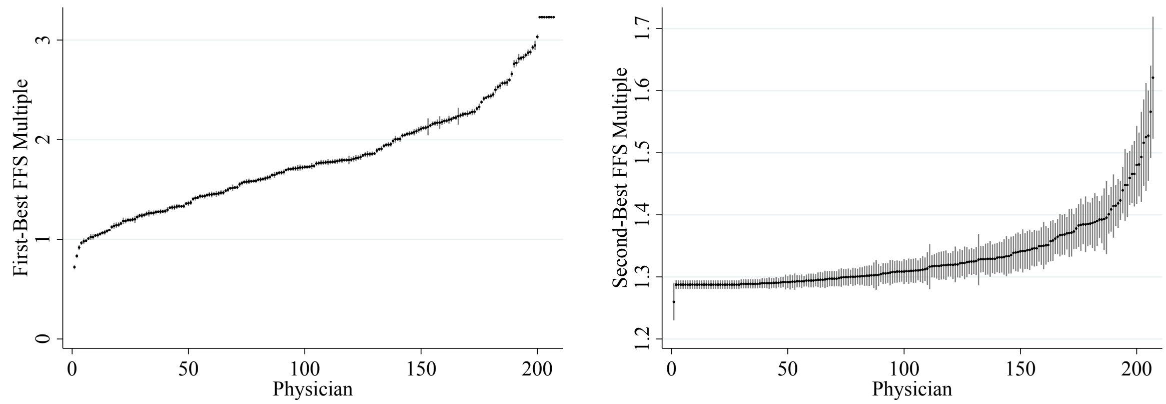


(a) Willingness to Pay and Mean Patient Severity

(b) Social Surplus and Cost

Notes: This figure plots, across physicians, the correlation between each incremental outcome from the two-contract menu in Figure 8a and its strongest predictor by bivariate R^2. I separately regress the outcomes (WTP and social surplus) on percentiles of each dimension (cost, altruism, production). The R^2 statistics for WTP 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. WTP is the difference in expected indirect utility between the high- and low-FFS contracts. Social surplus is the difference between contracts in expected (scaled) health production minus expenditure.

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

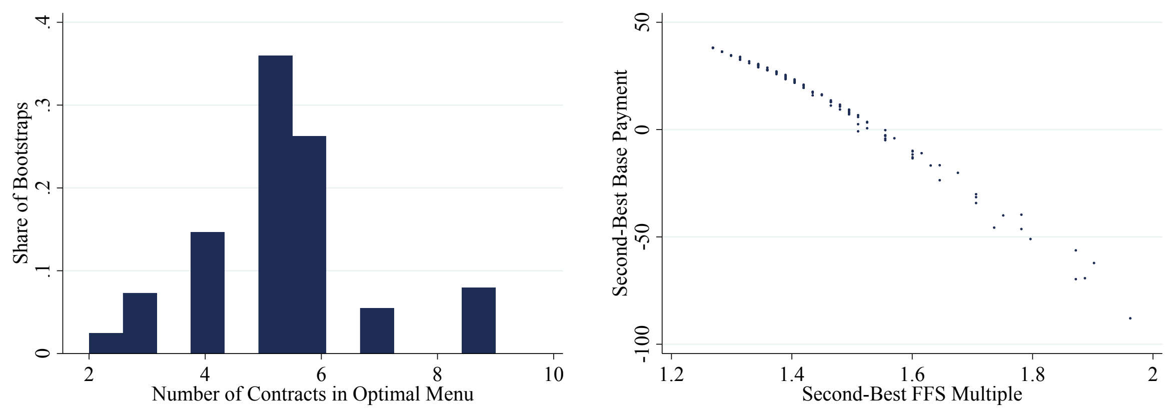


(a) First-Best FFS Multiple By Physician

(b) Second-Best FFS Multiple By Physician

Notes: This figure plots the distribution of FFS multiples across bootstrap samples for each physician. The x-axis is sorted separately for each panel, by mean FFS 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 A.12: Optimal Menu Across Bootstrap Samples

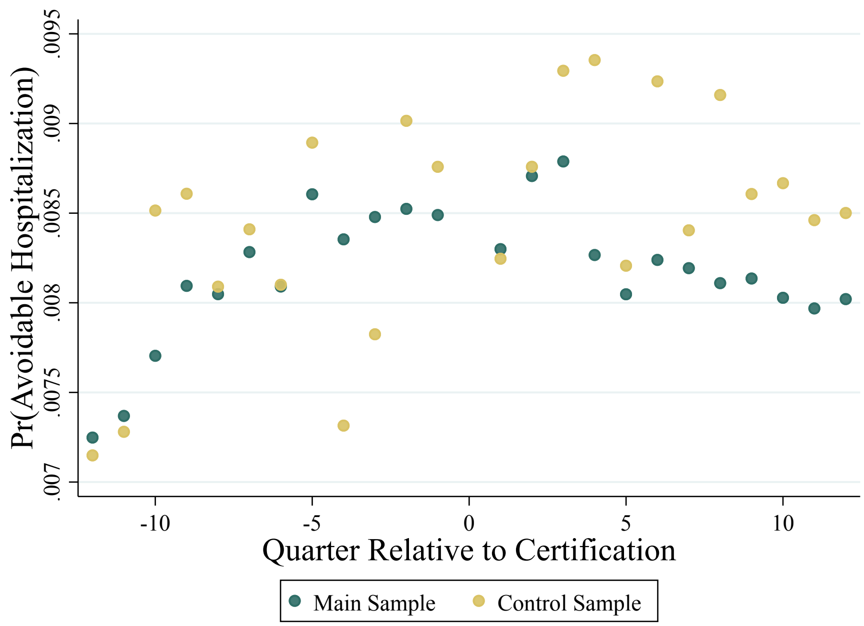


(a) Number of Contracts

(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 A.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 among their patients who were consistently registered in the six months surrounding certification. Outside of Quarters -2 to 2, patients are not necessarily registered to the focal physician, but the denominator is the number of those patients who are alive in a given quarter. The plot reflects physicians with certification dates in 2011-2014 in order for utilization data to exist in across all 72 months.

A.2 Additional Tables

Table A.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 A.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 A.3: Means by Patient Type

	Patients	Share	Age	Chronic Illness	Reimbursement	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 A.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
Capitation 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	30.72	28.77	10.23	100.00	14.23	29.61	40.38
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 10th, 50th, and 90th percentiles. Monetary measures are in USD. Physician Characteristics are also averaged across patients. The last two Physician Characteristics reflect shares of registered patients.

Table A.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 10th, 50th, and 90th percentiles. Physician Characteristics are also averaged across patients.

Table A.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 A.7: Treatment Intensity Variance Decomposition

	$E[m]$		$Var[m]$	
	Level	Share of Baseline	Level	Share of Baseline
Baseline	0.169	1.000	0.251	1.000
Fix $\sigma_\lambda = 0$	0.119	0.702	0.057	0.226
Fix $F(\lambda)$ at Mean	0.164	0.968	0.228	0.908
Fix $G(\theta)$ at Mean	0.175	1.034	0.252	1.006
Fix FFS at Mean	0.145	0.859	0.219	0.874
First-Best	0.199	1.175	0.293	1.168
Second-Best	0.180	1.062	0.265	1.057

Notes: This table calculates moments of treatment intensity (simulated hours) under counterfactual parameters, six months prior to certification. Variance is adjusted for weighting. Fix $F(\lambda)$ at Mean assumes all physicians have the same distribution of observed characteristics among patients, maintaining average observed heterogeneity across patients. Fix $G(\theta)$ at Mean assumes all physicians have the same cost, altruism, and productivity, using the sample mean. Fix FFS at Mean assumes all patients at all physicians correspond to the same weighted average simulated wage regardless of observed characteristics.

Table A.8: 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	2.09	2.18	1.68
2	0.31	8.89	2.30	3.06	2.41	1.80
3	0.34	7.31	1.99	2.64	2.23	1.68
4	0.16	9.22	2.46	2.56	2.14	1.65
5	0.04	11.10	2.57	3.43	2.50	1.87
Most Rural: 6	0.04	13.49	2.68	4.41	2.86	2.06

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 A.9: 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	8.393	2.546	0.303	2.710	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 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 A.10: 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 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, 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 A.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 A.11: 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.

A.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.[67] 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:

$$EV(x; \lambda_i, F, \theta) = \max_{m_i \geq 0} \; x(m_i) - c(m_i) + \sigma E\left[l\left(\sum_{i'=1}^{N} m_{i'}^* \right) \mid F(\lambda_{i'}) \right] + \alpha h(m_i, \lambda_i) \,, \tag{7}$$

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 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:

$$m^*(p, \lambda, (N-1)\bar{m}) = \max\{0, \frac{p - c - \sigma(N-1)\bar{m} + \alpha\gamma\lambda}{\alpha + \sigma}\} \tag{8}$$

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*

[67] 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' willingness to pay for contracts.

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 FFS over time within-physician identifies altruism. Critically, this requires observing treatment intensity choices for the same physicians at different FFS 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.

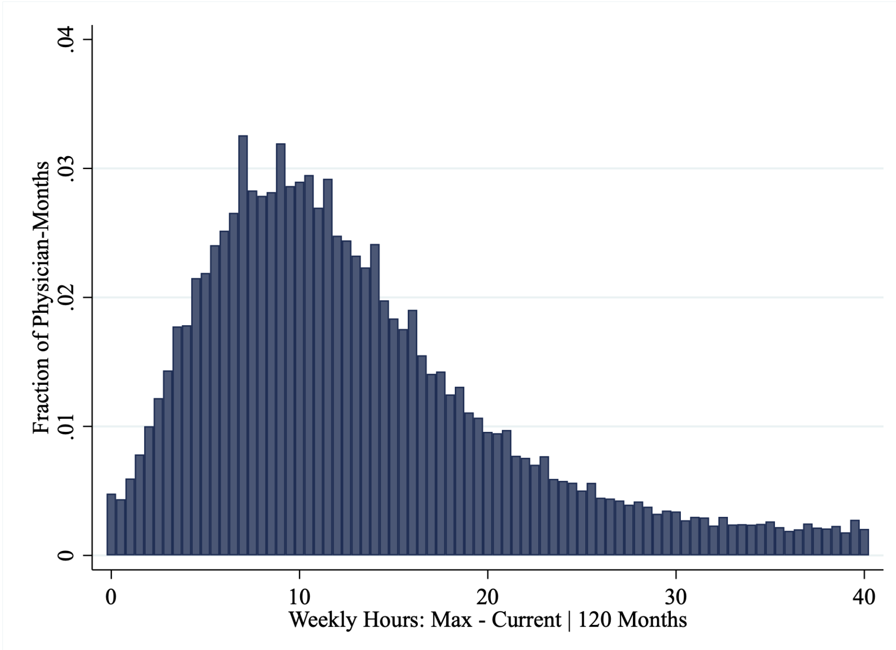
Consistent with prior studies that find treatment intensity increases in marginal reimbursement, likelihood ratio tests fail to find evidence of income effects.[68] Although simulated hours of treatment do not increase with FFS rates for some physicians, high altruism and large variance in patient health 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 FFS 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

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

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.

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

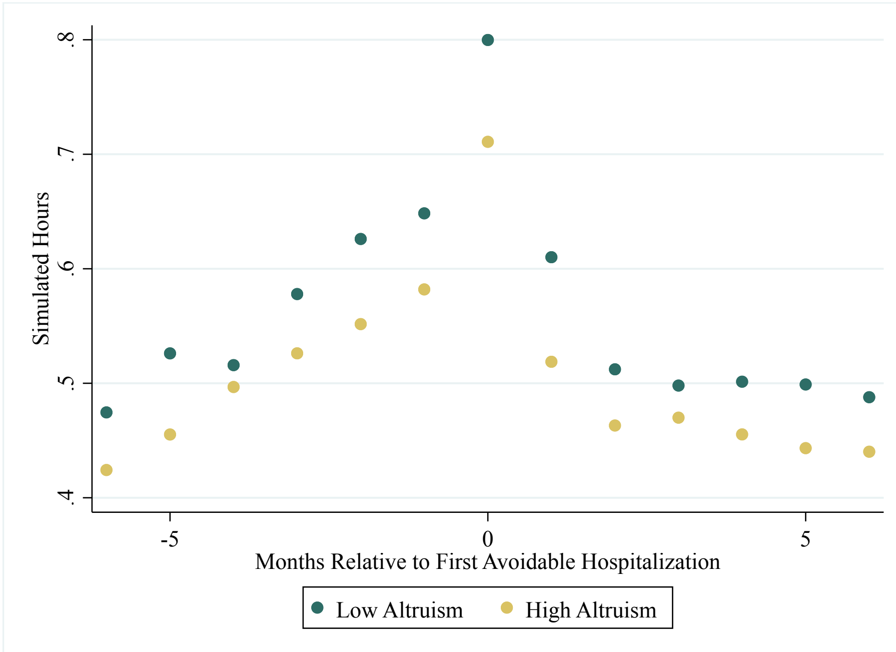


Notes: This figure shows the distribution of transformed hours per week ($\tilde{M}_{jt}$) across physician-months ($j-t$). The transformation is $\max_t M_{jt} - M_{jt}$. The x-axis is truncated at 40 and I exclude the first month when a physician works the maximum number of hours. According to the theoretical framework, $M_{jt} = \sum_i^{N_{jt}} \operatorname{argmax} u(x(m_{ijt}) - c(m_{ijt})) + \alpha h(m_{ijt}, \gamma\lambda_{ijt})$, 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 $\epsilon > 0$.

The main findings are also robust to imposing capacity constraints (See Table A.10). 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 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 pair of physician and month. 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.[69]

[69]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 90th percentile of status quo treatment intensity.

Figure A.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 be biased if the low response reflects some unobserved constraint rather than altruism. A.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.

A.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:

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

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 willingness to pay for the certified FFS rate. Section C.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 A.12 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). A.8 shows the distribution of this change in

Table A.12: 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.

EV across physicians. The large average increase in EV and a symmetric (rather than left-skewed) distribution suggest minimal selection on unobserved heterogeneity.[70]

[70]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.

B Data and Estimation Details

B.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 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 FFS revenue divided by marginal reimbursement and roughly corresponds to hours of treatment per patient-month ("simulated hours").

B.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.[71] 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 FFS rates plus the base payment. I calibrate the level normalization H as 10% of the value of a statistical life-month, which only impacts the variance decomposition. I generally report incremental expected health production which subtracts the pre-certification expected value.

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

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 FFS 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 FFS 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 FFS 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 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 FFS 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 FFS 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 FFS 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 FFS multiple on the grid, one at a time, while holding base payments for other FFS 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.

C Derivations

C.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 2.2, 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.[72] 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: $V(p, b, \lambda) = \underline{V}$. This results in a special case of social surplus:

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

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 relaxes the budget constraint by increasing the weight on health production relative to expenditure ($\alpha_R \to \infty$), $p_{FB} \to c$.[73]

Next, consider the second-best framing from Section 2.3. Starting from a uniform contract, when is it efficient to add a second contract with greater FFS to the menu? This requires a comparison of

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

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

incremental indirect utility ("willingness-to-pay" 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 $h_{m(\lambda\gamma)} \leq 0$, $\Delta \frac{d}{d\gamma} V(p) \leq 0$ and $\Delta \frac{d}{d\lambda} V(p) < 0$.

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

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

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-FFS 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.

C.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 $\tilde{\lambda} > 0$,[74] denoting the CDF of $\lambda \mid \lambda > 0$ as F_λ, the two-stage process for λ can be decomposed:

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

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.

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