# Access to Care in Equilibrium*

Nathaniel Mark†

December 5, 2020

Link to the latest version

**Abstract**

This paper studies consumer access to medical care as an equilibrium outcome of a market without prices. I use data from the Northern Ontario primary care market to estimate an empirical matching model where patients match with physicians. The market is cleared by a non-price mechanism: the effort it takes to find a physician. I use the model to study the distribution and determinants of access to care and to evaluate the effectiveness of policy remedies. I find that access to care is low and unevenly distributed. 26% of patients who would see a physician in a full access environment do not receive care in 2014. The issue is particularly acute in rural areas. Further, physicians discriminate in favor of patients with higher expected utilization, thereby increasing access for older and sicker patients while decreasing access for younger and healthier patients. The estimated model is used to evaluate two policies: grants to incentivize physicians to practice in low-access areas and a payment reform that provided incentives for physicians to increase the numbers of patients on their books. While both policies are partially successful, the model suggests potential improvements.

*I am grateful for the support and guidance from Bernard Salanié, Kate Ho, Ashley Swanson, and John Asker. I thank Simon Lee, Michael Riordan, Tobias Salz, Andrea Prat, Adam Sacarny, José Luis Montiel Olea, Doug Almond, and seminar participants at the IO colloquium, the IO workshop, the econometrics colloquium, and the Student Research Breakfast at Columbia University, as well as participants at the Health Economics and Policy Seminar at the Columbia Mailman School of Public Health for their useful comments. Without the infinite patience and dedication of the staff at the Institute of Clinical and Evaluative Studies, this paper would not have been possible. To this end, I especially thank Erin Graves, Eliane Kim, Stefana Jovanovska, and Refik Saskin. For sticking with me through the data acquisition process, I thank Sophia Johnson. For financial support, I thank the Program for Economic Research. All errors are my own.
This study made use of de-identified data from the ICES Data Repository, which is managed by the Institute for Clinical Evaluative Sciences with support from its funders and partners: Canada's Strategy for Patient-Oriented Research (SPOR), the Ontario SPOR Support Unit, the Canadian Institutes of Health Research and the Government of Ontario. The opinions, results and conclusions reported are those of the authors. No endorsement by ICES or any of its funders or partners is intended or should be inferred. Parts of this material are based on data and information compiled and provided by CIHI. However, the analyses, conclusions, opinions and statements expressed herein are those of the author, and not necessarily those of CIHI.

†Columbia University (Ph.D Candidate). Email: ndm2125@columbia.edu

# 1 Introduction

Access to care is easy to spot when it is missing. In 2011, a Globe and Mail journalist, Gloria Galloway, moved to the Ottawa area. At first, Ms. Galloway went without a regular primary care physician because she could not find one that was accepting patients. After a change in health status, she decided to put more effort into the search. She "phoned all 84 doctors who were listed as practising within 10 kilometres of [her] home." Every single physician rejected her. Finally, after months of searching, a new clinic opened. The clinic was far from her home – 17 km – but was accepting new patients. Ms Galloway went in person to register on opening day. She found hundreds of people in line, but waited regardless (Galloway, 2011). It was the only option.

In this paper, I use medical care utilization data to study access to care. Utilization is determined by both access and patient preferences. To measure access, I must account for preferences. Ms. Galloway's story illustrates this challenge. The data reflect only that Ms. Galloway did not go to the physician for a span of time, then went to a physicians who was 17 km away. Without imposing further structure, the data analyst cannot determine whether these outcomes were driven by low access to care. Ms. Galloway may have simply preferred to avoid the doctor.

I exploit observed matching patterns between patients and physicians to parse out the effects of access to care and patient preferences for care. For example, I can attribute a difference between regions in the share of patients who see a doctor to lower in access to care if, ceteris paribus, one region has lower physician supply. In practice, the exercise is complicated by interactions between regions, heterogeneous patients and physicians, and competition among market participants.

I take each of these factors into account. To do so, this paper measures access to care as an equilibrium outcome of a matching market between patients and physicians. Access to care is defined as the share of patients who would attain care in a full access environment who already attain care in the current equilibrium. Patients have heterogeneous preferences over physicians and physicians have heterogeneous preferences over patients. I assume that observed matchings are generated by a Rationing-by-Waiting equilibrium, which was introduced in the theoretical matching literature by Galichon and Hsieh (2019). In a Rationing-by-Waiting equilibrium, the market is cleared without prices by adjusting the effort it takes for one side of the market to match with the other side. Patients, for example, must expend effort by waiting on wait lists or frequently calling physicians in order to attain care.

The empirical setting is the primary care market in Northern Ontario, Canada. This setting has several attractive features for the study of access to care. First, patients face zero user fees. Thus, prices are exogenous to the equilibrium matching. Second, distinct markets for primary care can be easily defined in Northern Ontario. Almost all primary care is provided by family physicians, private practice is rare, and markets are geographically isolated. Third, the province makes detailed deidentified billings data available to researchers. The main dataset consists of a panel of patient and physician-level observations, importantly including measures of health and healthcare utilization, physician characteristics, and patient-physician matches from 2004-2014.

Additionally, Northern Ontario contains rich variation in policies intended to increase access to care. These allow me to trace out physician preferences and provide interesting settings for policy counterfactuals. Specifically, the Ontario government implemented reforms to how they pay primary care physicians from 2002 to 2006. A primary goal of these reforms was to increase access by decreasing unnecessary care and increasing the ease with which patients could attain care (Hutchison et al., 2001; Coopers, 2001). The reforms introduced alternative payment models for family physicians, including capitation payment models and enhanced fee-for-service payment models. Alternative payment models are voluntary: family physicians are at liberty to participate in any payment model they qualify for. The alternative payment models increased the average revenue a physician received per visit, incentivizing physicians to accept more patients. Additionally, both alternative payment models have capitation components (payments per patients regardless of utilization). The capitation payments are coarsely risk-adjusted. Therefore, the alternative payment models increased the incentive for physicians to select patients based on their characteristics. In the empirical specification, I account for selection into payment models by modeling physician choice behavior.

Access to care is defined as the share of patients who would attain care in a full access environment who already attain care in the current equilibrium. Defining a full access environment is subjective. In the main specification, I define a full access environment as the hypothetical choice conditions that patients would face if they lived in the largest city in Northern Ontario, Sudbury, and all physicians in Sudbury were accepting patients. Sensitivity analyses are conducted by comparing results under alternative definitions of full access. For ease of exposition, I define *access loss* as $1 -$ access, or the share of patients who would attain care in the full access environment who *do not* attain care in the current equilibrium.

This paper uses the proposed measure of access to care to study three topics. First, I show the distribution of access to care across regions and patient types in Northern Ontario in 2014. Second, I decompose estimated access loss into its determinants. Last, I discuss the policy implications of the model. I use the estimated model to analyze whether grants for physicians to locate in low access areas are justified and I study the impact of the physician payment reforms on access to care.

Access loss is large and unequally distributed. I find that 26% of patients who would see a physician in a full access environment do not receive care in 2014. Healthier, younger, and more rural patients have higher access loss than their counterparts. Rural patients with no comorbidities and aged 0-34 have access loss of 50%, while urban patients with comorbidities and aged 65+ have access loss of 5%. Competitive effects drive the access loss of younger and healthier patients. I find that physicians discriminate in favor of patients with higher expected utilization. This increases the effort it takes for healthier and younger patients to attain care – driving them out of the market.

The determinants of access loss vary by type of patient. In urban and semi-urban areas, access loss is primarily driven by capacity constraints of physicians. I estimate that if physicians had enough capacity to accept all patients, urban access loss would fall from 12.25% to 0.68%. In rural areas, physician capacity constraints explain less of the access loss. Removing capacity constraints would change access loss from 43.04% to 31.47% in rural areas. The remaining access loss is primarily caused by distances needed to travel to a physician.

These findings have policy implications. Ontario, for example, provides a grant of up to $117,600 for primary care physicians who begin to practice in low access areas. A similar program in the United States provides medical school loan repayments of up to $50,000. In high access, high population areas, physicians can attain high revenues after entry by stealing patients from incumbent physicians. Thus, an entry into these locations does not significantly increase access to care. In low access, low population areas, an entrant physician may attract few patients but increase access significantly. Therefore, providing financial incentives for physicians to locate in low access areas is justified.

Lastly, I assess the impact of the physician payment reforms on access to care. I find that the alternative payment models increase access to care by 5 percentage points (pp). Physicians accept more patients when they are in an alternative payment model. This explains most of the impact of the alternative payment models on access. In line with the existing literature, physicians are relatively unresponsive to revenue when they are selecting patients (Rudoler et al., 2015a; Kantarevic and Kralj, 2014). Thus, the effect of the alternative payment models on selection of patients only accounts for 0.26pp of the access gains.

**Related Literature**

This paper contributes to three main literatures. First, I advance the literature on measuring access to care and decomposing the determinants of access to care by accounting for equilibrium impacts. Second, I contribute to the literature on estimating decentralized non-transferable utility (NTU) empirical matching models by applying a recent theoretical advancement to an empirical setting. Third, I advance the study of how physician payment models impact access to care.

The first set of results in this paper describe the distribution of access to care across patient characteristics and geography. The literature on measuring the distribution of access to care is wide and interdisciplinary.

Measuring the distribution of access across patient characteristics is complex. Without estimating patient preferences, it is difficult to compare access across patient characteristics, as patient characteristics are highly correlated with whether a patient wants to attain care. Thus, research on this topic is sparse. An important exception is the body of literature that studies the distribution of access to care across socioeconomic status. This work focuses on determining the extent to which patients with the same healthcare needs receive different levels of healthcare (Wagstaff et al., 1991; Kakwani et al., 1997; Pulok et al., 2020). Rather than estimate patient preferences for care, the literature uses regression techniques to decompose healthcare utilization into utilization explained by health needs and utilization explained by non-need factors. The object of interest is a measure similar to the Gini index for income inequality: a horizontal inequity index (HI).

Measuring access across geography is a developed field. The matching model used in this paper, although methodologically distinct, is related to gravity models used by public health scholars to measure the spatial distribution of access. Gravity models, such as the two-stage floating catchment model, define access as the ratio of supply to demand. Supply (demand) in each location is determined by the observables of physicians (patients). Supply and demand are redistributed according to a distance decay function such that patients demand nearby physicians at higher rates than distant physicians (Luo and Wang, 2003; Luo and Qi, 2009; McGrail and Humphreys, 2015;

Kim et al., 2018).

The empirical matching model used in this paper similarly allows demand to decay according to distance and uses patient and physician characteristics to estimate supply and demand. However, the matching model allows me to estimate the decay function rather than pre-specifying it. Further, the matching model allows for greater complexity in access patterns. My measure of access to care can be heterogeneous across patients in the same location, and provides insight into the determinants of access to care. However, the empirical matching model is more demanding than gravity models in terms of computational complexity and data needs.

This paper adds to the empirical literature on decentralized matching markets with non-transferable utility. Most empirical applications of decentralized non-transferable utility matching markets use centralized non-transferable utility techniques to clear the market, then estimate by simulated moment matching or analogous Bayesian techniques (Hitsch et al., 2010; Boyd et al., 2013; Vissing, 2017; Agarwal, 2015; Matveyev, 2013). They assume that equilibrium matchings are pairwise stable. However, pairwise stable equilibria are generally not unique, so practitioners further assume that matchings are generated by a process that mimics a centralized matchmaker's algorithm, such as the Simulated Gale-Shapley Deferred Acceptance algorithm. Simpler algorithms can be used when one side of the market is assumed to have vertical preferences (Agarwal, 2015; Gazmuri, 2019).

By applying recent developments in the theoretical matching literature to an empirical application, this paper estimates a non-transferable utility matching model without imposing the deferred acceptance algorithm. Further, the Rationing-by-Waiting equilibrium allows for the estimation of equilibrium effort costs – an object of interest.[1]

Lastly, this paper contributes to the literature on the effect of physician payment models on access to care. Much of this literature focuses on the impact of payment models on physician productivity (Kantarevic et al., 2011) and patient selection (Alexander; Rudoler et al.; Kantarevic and Kralj). Due to the wide range of policies studied world-wide, results have been mixed. Within studies of the Ontario payment reforms, Kantarevic et al., 2011 find that physicians who switch to an enhanced fee-for-service model increase the number of patients they accept. Rudoler et al. (2016) finds no selection of patients based on risk (expected utilization). Unlike these reduced form studies, I am able to account for the equilibrium effects of the payment models. Encouragingly, my results largely confirm the reduced form conclusions.

A related subject is the effect of insurance reimbursement rates in the United States on access to physician services (Shen and Zuckerman, 2005; Chen, 2014; Alexander and Schnell, 2018; Benson, 2018). The literature has found that physicians are responsive to reimbursement rates when choosing to accept patients. In a recent paper, Alexander and Schnell (2018) exploit a quasi-exogenous shift

[1]A related literature focuses on estimating demand with unobserved capacity constraints. Goeree (2008), for example, models the probability that a personal computer is in a consumer's choice set at the time of a choice. This methodology could be used to model the probability that a physician is willing to accept a patient. de Palma et al. (2007) showed that this demand system could equivalently be modeled as consumers paying equilibrium effort costs to attain a capacity constrained. Other studies of demand with capacity constraints use data to determine bounds on capacity (Conlon and Mortimer, 2013) or choice sets (Gaynor et al., 2016).

in medicaid reimbursement rates caused by the Affordable Care Act. They find that physicians increase their willingness to accept medicaid patients when medicaid payments increase. In my empirical setting, I also find that physicians increase the probability of accepting a patient if payment increases, but by much lower magnitudes. Alexander and Schnell (2018) find an elasticity of physician willingness to accept a patient of .83. I find an average elasticity of physician willingness to match with a patient of .16.

# 2 Access To Care in Equilibrium

Many people in Northern Ontario do not attain primary care. This is the outcome of a market for primary care in equilibrium. To provide insight on how much this is determined by lack of access and what mechanisms impact access to care, I present a model of the market for primary care. To mirror the conditions of Northern Ontario, the market is cleared by a non-price mechanism: the effort it takes for a patient (physician) to match with a physician (patient).

## 2.1 Equilibrium Outcomes

Figure 1 shows the percent of Northern Ontarians in 2014 who did not see a comprehensive care primary care physician. 40% of the population do not attain care. As expected, younger and healthier populations are less likely to attain care. Rural Ontarians are less likely to attain care than their urban counterparts. By themselves, these results do not imply that access to care is low. To determine whether people do not attain care due to access to care, we must account for patient preferences.

Survey evidence suggests that access to care does contribute to the high percentage of people who do not attain care. In 2014, 56.1% of Canadians who stated that they did not have a regular doctor attributed it to inaccessibility. They stated that either no doctor in their area was taking new patients, their doctor had retired or left the area, or there were no doctors in their area (Statistics Canada, 2007). In 2005, a survey found that only 11.4% of physicians were accepting new patients in Ontario (News Staff, 2006). In 2017, 14.5% of patients in Ontario stated that they had to wait 8 days or longer to see their doctor. Waiting times are worse in Northern Ontario. 40.7% of patients in the North East and 36.4% of patients in the North East reported wait times over 8 days (Health Quality Ontario, 2018).[2]

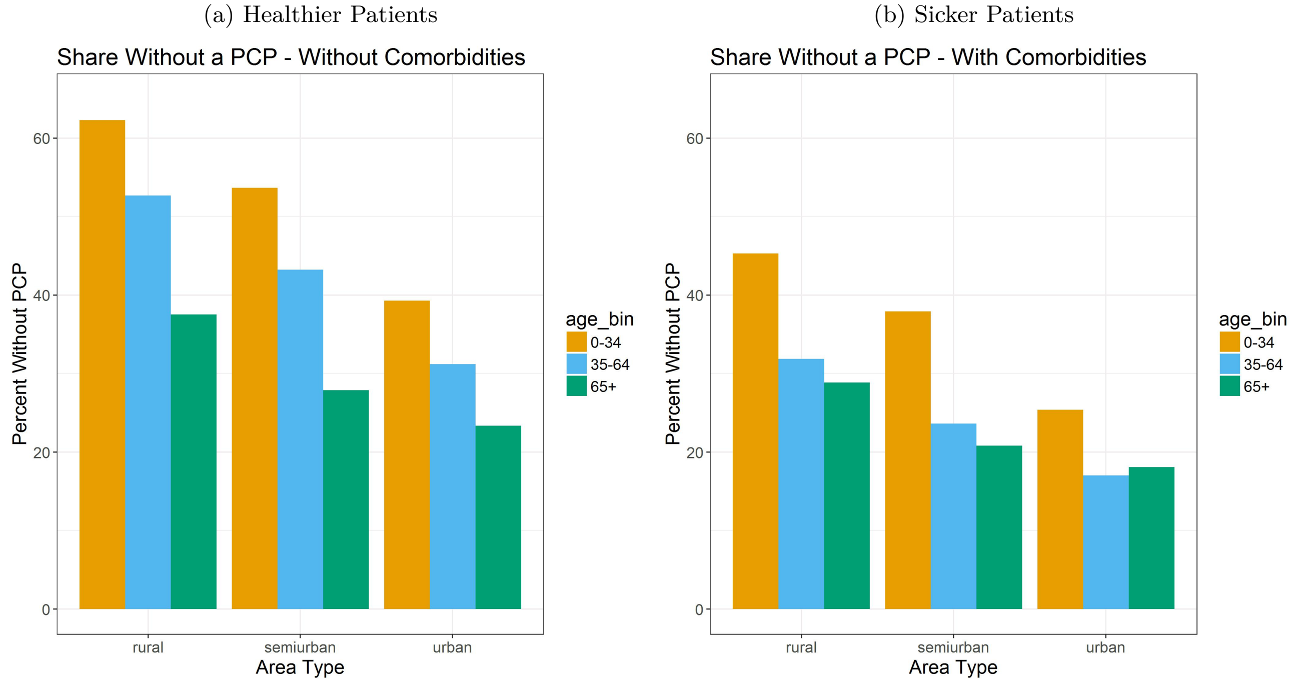
## 2.2 A Simplified Model: Effort as a Non-Price Market Clearing Mechanism

This section presents a simplified model of supply and demand in a market with non-price rationing. This model provides a stylized representation of how insufficient supply lowers access to care. It also presents the main methodological concepts used in the rest of this paper.

I define patient demand as the number of patients who demand care. Patient demand depends on the characteristics that define the market $\boldsymbol{X}, \boldsymbol{Y}$ and the amount of effort a patient must expend

[2] North East and North West refer to Local Integration Health Networks.

Figure 1: Percent of Patients Who Do Not See a Primary Care Physician in 2014




in order to attain care $\tau^u$. Effort $\tau^u$ only includes effort spent in the search process to attain care. Market characteristics include the locations and characteristics of patients, $\boldsymbol{X}$, and the location and characteristics of physicians, $\boldsymbol{Y}$. The effort a patient expends to attain care $\tau^u$ can be interpreted as waiting time, though other forms of effort may be present. It is assumed that demand is downward sloping in $\tau^u$.

Supply is similarly defined as a function of market characteristics and the amount of effort it takes a physician to attract one patient $\tau^v$. The effort a physician expends can be interpreted as advertising costs, though this object can represent many different types of effort. Supply is downward sloping in $\tau^v$.

$$\text{Demand: } Q^D = D(\tau^u; \boldsymbol{X}, \boldsymbol{Y})$$
$$\text{Supply: } Q^S = S(\tau^v; \boldsymbol{X}, \boldsymbol{Y})$$

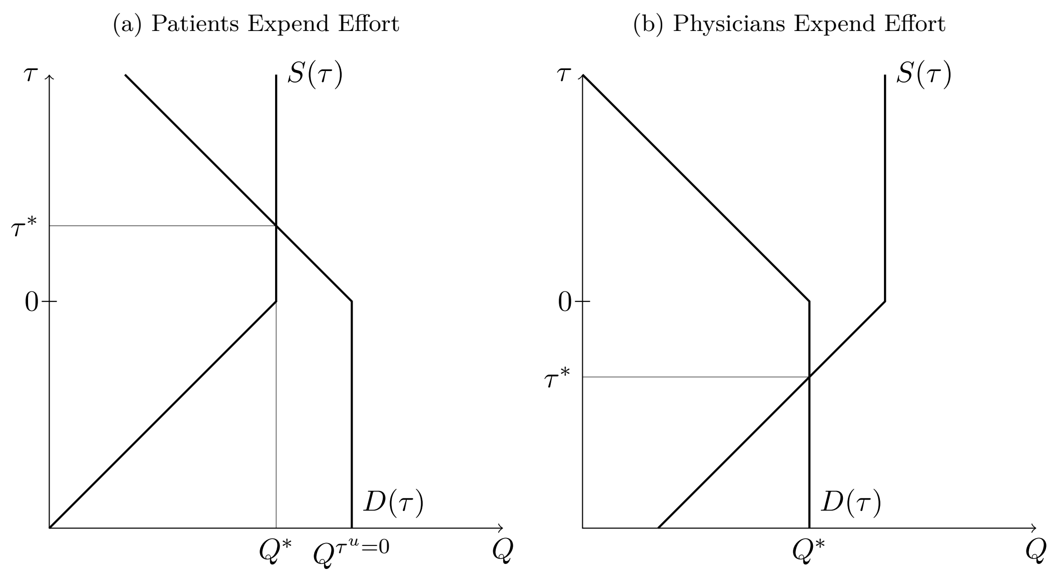
In equilibrium, two conditions hold. First, effort costs are such that the market clears. Second, only one side of the market expends effort to match with the other. That is, if patients are expending effort to match with physicians, then physicians are not expending effort to attract patients. In the context of a waiting line, the intuition is clear. A line of patients forms. Physicians arrive at the line, taking patients one at a time. If the rate of physician arrival at exceeds the rate of patient arrival, then the line of patients is soon replaced with a line of physicians waiting for patients to arrive.

$$\text{Market Clearing: } Q^D = Q^S$$
$$\text{One-Sided Effort Condition: } 0 = min\{\tau^u, \tau^v\}$$

Figure 2 presents the model graphically. To do so, I define net patient effort $\tau$ as the difference between the effort expended by patients and the effort expended by physicians, $\tau = \tau^u - \tau^v$.

Importantly, transfers do not exist in this market. Physicians do not attain utility when patients expend effort. Therefore, when physician capacity constraints are binding and $\tau > 0$, supply is perfectly inelastic with respect to net patient effort. Similarly, demand is perfectly inelastic with respect to net patient effort when all patients who want care are receiving it. Thus, in all equilibria, either physicians are operating at full capacity and patients are expending effort, or all patients who want care are receiving care and physicians are expending effort.

Figure 2: Supply and Demand with Non-Price Market Clearing

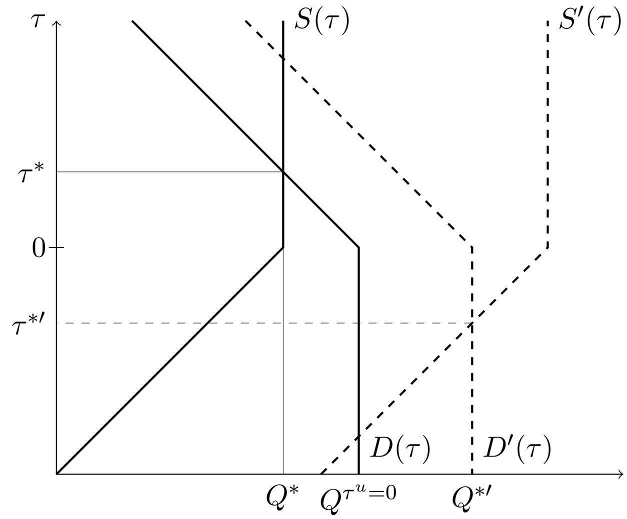



**Shifts in Supply are Correlated with Shifts in Demand** When physician locations and characteristics change, both supply and demand are affected. Take, for example, the entry of a new physician into a town that previously did not have a doctor. Adding the new physician shifts supply out trivially. Additionally, in the town where the entry occurred, patients have expanded willingness to expend effort. The value of the care offered is higher now for these patients, thus shifting demand.

Other forms of changes in physician characteristics only affect supply. For example, if an individual physician increases their capacity, demand is unaffected. Demand for physician services are orthogonal to the number of services each physician is willing to provide.

Figure 3 presents a shift in supply and demand graphically. Allow a change in physician characteristics from $\boldsymbol{Y}$ to $\boldsymbol{Y'}$. Demand then shifts from $D(\tau) = D(\tau; \boldsymbol{X}, \boldsymbol{Y})$ to $D'(\tau) = D(\tau; \boldsymbol{X}, \boldsymbol{Y'})$ and supply shifts from $S(\tau) = S(\tau; \boldsymbol{X}, \boldsymbol{Y})$ to $S'(\tau) = S(\tau; \boldsymbol{X}, \boldsymbol{Y'})$. Both shifts expand the number of patients who attain care.

Figure 3: An Increase in Supply




**Access to Care** I define access to care as the share of patients who would attain care in a full access environment who already attain care in the current equilibrium. Access loss is defined as the share of patients who would attain care in a full access environment who *do not* attain care in the current equilibrium. In terms of model outputs, these objects are defined:

$$\text{Access to Care} = \frac{Q^*}{Q^{FE}}$$
$$\text{Access Loss} = \frac{Q^{FE} - Q^*}{Q^{FE}}$$

Where $Q^{FE}$ is the equilibrium number of patients who attain care under market characteristics $\boldsymbol{X}, \boldsymbol{Y}^{FE}$ and $Q^*$ is the equilibrium number of patient who attain care under the current market characteristics.

In a market with non-price rationing, insufficient supply lowers access to care through two mechanisms. First, in environments with insufficient supply, doctors are few and far between.

Patients must travel long distances to the nearest physician's office and patients have difficulties finding a physician who fits their needs. This decreases patient willingness to go to a doctor. Second, insufficient supply makes it harder to attain care. Physicians are at capacity and thus cannot take on any more patients. Care must be rationed. Patients face long waiting times for appointments, long waiting lists to get on doctor's rosters, and frustrating searches for physicians who are accepting patients.

Using the simplified model, access loss can be graphically decomposed into its two mechanisms. The effects of rationing-by-effort costs can be interpreted as the share of patients who do not attain care because the effort it would take to attain care is greater than their willingness to expend effort, $\frac{Q^{\tau^u=0}-Q^*}{Q^{FE}}$. This effect could also be interpreted as the impact of physician capacity constraints on access loss. As alluded to above, increasing capacity of existing physicians shifts the supply curve but not the demand curve. Thus, removing physician capacity constraints would lead to a shift in supply sufficient to allow all patients who want care to attain care. The remainder of access loss, $\frac{Q^{FE}-Q^{\tau^u=0}}{Q^{FE}}$, can be attributed to the sparsity of the geographic distribution of physicians.

### 2.3 Complicating the Model: Heterogeneous Effort Costs

This simple model is complicated by the fact that not all patients face the same effort to attain care. Physicians discriminate between patients, making it easier for some patients to attain care and harder for others. Additionally, local variation in market characteristics affect effort.

Physicians discriminate between patients. Altruistic tendencies of physicians cause some discrimination (Hennig-Schmidt et al., 2011). Physicians prefer to treat patients who need care, and will provide special accommodations for those in particular need (McGuire, 2000). This behavior, called positive prioritization by Gravelle and Siciliani (2008), can be explained by a model of physician agency, where physicians treat patients with the highest expected benefit of treatment until their capacity is reached. Under a general set of assumptions, positive prioritization is welfare improving relative to non-discriminatory rationing (Iversen and Siciliani, 2011).

Physicians may also discriminate to attain more profitable patients. The literature is large and mixed on this topic.[3] The implications for discrimination based on profitability depends on the specifics of how physicians are paid. If physicians are paid by a fee-for-services model, physicians will discriminate in favor of patients who are expected to high margin services. If physicians are paid by capitation models, physicians will discriminate in favor of low cost patients, conditional on risk-adjustments.

In addition to physician discrimination, heterogeneity in patient effort arises from local variation in market characteristics. Variation in the density of primary care physicians across geographic space is large (Green et al., 2017). Further, physician practices are very heterogeneous in size (see figure 17). Thus, two identical patients in neighboring towns may face very different market conditions. Importantly, however, the towns cannot be treated as separate markets, as patients are

[3] Primary Care: In the affirmative (US): Benson (2018), in the negative (Ontario): Rudoler et al. (2016); Kantarevic and Kralj (2014); Hospitals in the affirmative (US): Alexander (2020)

willing to travel for care.

These complications suggest a natural extension of the simple model. Assume that patients can be split into discrete types $\theta \in \Theta$. Types can be determined by health needs, location, and other characteristics. A type $\theta$ patient must expend effort $\tau^u_{\theta j}$ to match with physician $j$. Patients choose among the physicians for whom their willingness to exert effort is greater than the effort cost. Symmetrically, physician $j$ chooses patients according to their preferences and the effort costs $\tau^v_{\theta j}$.

Define $\mu_{\theta j}$ to be the number of patients $\theta$ who attain care from physician $j$. Supply and demand can be written as a function of effort costs and market characteristics.

$$\text{Demand: } \mu^D_{\theta j} = D_{\theta j}(\tau^u_{\theta j}; \boldsymbol{\tau}^u_{\theta -j}, \boldsymbol{X}, \boldsymbol{Y})$$
$$\text{Supply: } \mu^S_{\theta j} = S_{\theta j}(\tau^v_{\theta j}; \boldsymbol{\tau}^v_{-\theta j}, \boldsymbol{X}, \boldsymbol{Y})$$

As before, an equilibrium exists when the market is cleared and the one-sided effort conditions hold.

$$\text{Market Clearing: } \mu^D_{\theta j} = \mu^S_{\theta j} \ \forall \theta, j$$
$$\text{One-Sided Effort Condition: } 0 = min\{\tau^u_{\theta j}, \tau^v_{\theta j}\} \ \forall \theta, j$$

### 2.4 A Note on Markets with Prices and Non-Price Rationing

Before turning to the matching model, it is useful to comment on the relationship between prices and access to care in the context of this model. In markets where prices are flexible, they may adjust to stimulate supply.[4] In the United States, for example, family medicine physician wages are 6.00% higher in non-metropolitan areas than metropolitan areas[5], suggesting some role for price adjustments in increasing supply (Lee, 2010; Newhouse et al., 1982). However, even in the multi-payer system of the US, over 39% of physician and clinical service expenditures are by large public payers (CMS, 2019).[6] Medicare, which accounts for 23.5% of expenditures, sets prices according to input costs and the value of physician effort. Prices vary geographically according to differences in the cost of supplying care, not the level of access (Chan and Dickstein, 2019), with

[4] Of course, prices add an additional mechanism through which access is restricted. Indeed, much of the literature on access in the United States focuses on the negative impact of prices on access to care. As the focus of this paper is on markets without flexible prices, I do not wade into the argument on the relative merits of regulated prices for access to care.

[5] Author's calculation using BLS Occupational Employment Statistics (U.S. Bureau of Labor Statistics, 2019). Average wages in metropolitan and non-metropolitan areas are calculated as weighted averages of annual wages, weighted by employment. 48 areas have low numbers of employment (below 30). Employment for these areas is assumed to be 15. If these low employment areas are removed, non-metropolitan wages are found to equal 6.31%.

[6] In 2018, 8.4% of physician and clinical services expenditures were out-of pocket payments. 43% were private insurance payments, 23.5% Medicare, 10.7% Medicaid, 4.8% Other federal Health Insurance Programs, and 9.7% other third party payers, some of which are also funded with federal funds.

the exception of periodic targeted add-ons (Mroz et al., 2020).[7] Further, Medicare's price setting practices influence prices paid by private payers market (Clemens and Gottlieb, 2017).

In this paper's empirical setting – Ontario, Canada – prices are inflexible and set by the government at the provincial level. Globally, inflexible prices are the norm. Of the 29 countries in a 2009 OECD report, only 5 were found to have some aspect of primary care prices negotiated at the local level (Paris et al., 2010). Two were found to set resource-based relative values scales at a centralized level, with some negotiation at the local level. Without flexible prices, insufficient supply necessitates rationing by other means. In medical care markets without a centralized rationing system, services are distributed according to effort: those willing and able to wait in long lines or spend time calling physicians are those who attain services (Iversen and Siciliani, 2011).

# 3 Matching Model: Patient and Physician Matching

This section formalizes the model described in section 2 using theory from the non-transferable utility matching market literature. It follows a modified version of the Rationing-by-Waiting framework developed by Galichon and Hsieh (2019).

## 3.1 Patient and Physician Preferences

Patients are indexed by $i$ and are members of a patient type, $\theta \in \Theta_t$. Patients of type $\theta$ share observable characteristics and location. Patient type $\theta$ has $n_{\theta t}$ members in market $t$. Physicians are indexed by $j \in \mathcal{J}_t$. Each physician can match with a maximum of $m_{jt}$ patients in market $t$, which varies by physician. $m_{jt}$ is the maximum number of patients that could match with physician $j$. I call this the physician's *panel capacity*.

Patients derive utility from matching with physician $j$ in market $t$ as a function of distance $d_{\theta jt}$ and match observables $\boldsymbol{x}_{\theta jt}$. Match observables are physician characteristics, patient characteristics, or interactions. Lastly, each patient $i$ of type $\theta$ has an additive taste shock, $\epsilon_{i\theta jt}$, for physician $j$ in market $t$. Without loss of generality, patients derive zero mean latent utility if they do not match with a physician. I use the empty set index to denote a match with no physicians.

$$u_{i\theta jt} = u(d_{\theta jt}, \boldsymbol{x}_{\theta jt}) + \epsilon_{i\theta jt}$$
$$u_{i\theta\emptyset t} = \epsilon_{i\theta\emptyset t}$$

Physicians derive utility independently from each potential space in their patient panel, $q \in \{1, ..., m_{jt}\}$. The utility a physician derives from matching in panel space $q$ with a patient of type $\theta$ in market $t$ depends on observables $\boldsymbol{y}_{\theta jt}$. These observables are a function of patient and physician characteristics. Physicians have an additive taste shock, $\eta_{\theta jqt}$, for each panel space $q$, patient type $\theta$, and market $t$.

[7]The adjustments based on geography, Geographic Adjustment Factors, are negatively correlated with rurality, but have not been found to be correlated with survey measures of access to care for medicare beneficiaries (Committee on GAFs, 2011).

$$v_{\theta jqt} = v(\boldsymbol{y}_{\theta jt}) + \eta_{\theta jqt}$$
$$v_{\emptyset jqt} = \eta_{\emptyset jqt}$$

Note that the model does not allow physicians to have idiosyncratic taste shocks over individual patients. Physician latent utility is equal over all patients in the same patient type. This assumption is common in the transferable utility matching literature (Choo and Siow, 2006), is fairly reasonable in the empirical application, and provides the basis of a tractable functional form of match probabilities between patient and physician types.

### 3.2 Markets and Matchings

A market is defined by the characteristics and preferences of patients and physicians, the number of patients of each type, $n_{\theta t}$, and the each physician's panel capacity, $m_{jt}$. If preferences can be parameterized by a vector $\boldsymbol{\beta}$, then a market is formally a tuple $(\boldsymbol{\beta}, \boldsymbol{n}_t, \boldsymbol{m}_t, \boldsymbol{D}_t, \boldsymbol{X}_t, \boldsymbol{Y}_t)$ where $\boldsymbol{n}_t = \{n_{\theta t}\}_{\forall \theta}$, $\boldsymbol{m}_t = \{m_{jt}\}_{\forall j}$, $\boldsymbol{D}_t = \{d_{\theta jt}\}_{\forall \theta j}$, $\boldsymbol{X}_t = \{\boldsymbol{x}_{\theta jt}\}_{\forall \theta j}$, and$\boldsymbol{Y}_t = \{\boldsymbol{y}_{\theta jt}\}_{\forall \theta j}$.

A matching, $\mu_t$, defines which type of patients are matched with which physicians. Formally, a matching is a measure over the set of patient types and the set of physicians.

$$\begin{aligned} \mu_t &: \{\Theta_t, \emptyset\} \times \{\mathcal{J}_t, \emptyset\} \to \mathbb{N} \\ s.t. & \sum_{j \in \{\mathcal{J}_t, \emptyset\}} \mu_t(\theta, j) = n_{\theta_t} \\ & \sum_{\theta \in \{\Theta_t, \emptyset\}} \mu_t(\theta, j) = m_{jt} \end{aligned}$$

The realization of a matching at a point $(\theta, j)$ is the number of patients of type $\theta$ that are matched with physician $j$.

### 3.3 Equilibrium

**Stability** Most of the decentralized non-transferable utility matching market literature focuses on stable matchings. A matching is stable if:

1. No patient prefers to be unmatched than be matched with their current physician.

2. No physician prefers to leave a panel space empty that is currently occupied by a patient.

3. No patient and physician in one panel space would both prefer to match with each other than keep their current matches.

The assumption that matchings are stable in a decentralized market is justified by Adachi (2003). Adachi shows that a realistic dynamic search model produces the same set of equilibria as the stability assumption. Specifically, the set of equilibria of a two-sided search model in a large market converges to the set of stable matchings as search costs converge to zero.[8]

[8]This is shown, however, only in the case where a threshold strategy is used.

**Existing Methods to Predict a Unique Matching** The stability assumption alone does not predict a unique matching, which is needed for point-identified estimation techniques. The empirical literature tends to use the Gale-Shapley deferred-acceptance (DA) algorithm to choose a unique matching from the set of stable matchings (Gale and Shapley, 1962). Hitsch et al. (2010) show that the DA algorithm performs well predicting outcomes in the online dating market.

Though conceptually simple, using the DA algorithm as an equilibrium concept has drawbacks. First, the DA algorithm chooses a stable matching which is optimal for one side of the market. This assumption can affect welfare predictions, as the DA algorithm chooses a one-sided optimal matching. Second, the DA algorithm is computationally demanding. If used within an estimation routine, the DA algorithm must be run many times at each iteration of the optimization algorithm. In a large market such as the Ontario primary care market, this can be intractable (Chiappori and Salanié, 2016; See Hsieh (2012) for a discussion). Third, the DA algorithm has the unattractive feature that it allocates differing equilibrium utilities to observably identical agents.

Alternatives to the DA algorithm to model decentralized markets have relied on assumptions of very large markets and/or strictly vertical preferences on one side of the market (Azevedo and Leshno, 2016; Agarwal, 2015; Gazmuri, 2019; Menzel, 2015).[9] Such assumptions are unreasonable in the Ontario primary care market. Markets in less populated parts of Northern Ontario are not large. Physician preferences are horizontal, as physicians are paid under varying payment models and homophily in age and sex are possible.

### 3.3.1 The Rationing-By-Waiting Equilibrium

I employ the Rationing-by-Waiting equilibrium concept proposed by Galichon and Hsieh (2019) to predict unique matchings. To my knowledge, I am the first to apply this equilibrium concept to an empirical setting. I first present the equilibrium concept, then I discuss its attributes.

#### Definition: Rationing-By-Waiting Equilibrium

A tuple $(\mu_t, \boldsymbol{\tau}_t^v, \boldsymbol{\tau}_t^u)$ is a Rationing-by-Waiting Equilibrium if

$$\mu_t : \{\Theta, \emptyset\} \times \{\mathcal{J}, \emptyset\} \rightarrow \mathbb{N} \ \forall t \tag{1}$$

$$s.t. \ \sum_j \mu_{\theta jt} + \mu_{\theta\emptyset t} = n_{\theta t} \ \forall\theta\forall t \tag{2}$$

$$\sum_\theta \mu_{\theta jt} + \mu_{\emptyset jt} = m_{jt} \ \forall j\forall t \tag{3}$$

$$\begin{aligned}\mu_{\theta jt} &= n_{\theta t} P(u_{\theta jt} - \tau^u_{\theta jt} > u_{\theta j't} - \tau^u_{\theta j't} \ \forall j' \neq j \wedge u_{\theta jt} - \tau^u_{\theta jt} > u_{\theta\emptyset t}) \\ &= m_{jt} P(v_{\theta jt} - \tau^v_{\theta jt} > v_{\theta' jt} - \tau^\gamma_{\theta' jt} \ \forall\theta' \neq \theta \wedge v_{\theta jt} - \tau^v_{\theta jt} > v_{\emptyset jt}) \ \forall\theta\forall j\forall t\end{aligned} \tag{4}$$

$$min(\tau^u_{\theta jt}, \tau^v_{\theta jt}) = 0 \ \forall\theta\forall j\forall t \tag{5}$$

[9] Agarwal (2015); Galichon and Hsieh (2019): Vertical preferences; Menzel (2015): Infinitely Large Markets; Azevedo and Leshno (2016): Both

$\mu_t$ is a matching in market $t$. $\mu_{\theta jt}$ is shorthand for number of matches $\mu_t(\theta, j)$. $\tau^u_{\theta jt} \in \boldsymbol{\tau}^u_t$ is the additive effort cost (in utils) that a patient of type $\theta$ must incur to match with physician $j$. $\tau^v_{\theta jt} \in \boldsymbol{\tau}^v_t$ is the effort cost that physician $j$ must incur to match with a patient of type $\theta$. The interpretation of effort costs is discussed extensively in section §2.

Equations 1,2, and 3 define a matching. Equation (4) is the market clearing condition. In equilibrium, the number of patients of type $\theta$ who choose physician $j$ must be equal to the number of patients of type $\theta$ that physician $j$ chooses. Equation (5) is the one sided waiting condition: either patients of type $\theta$ wait for physician $j$'s spaces *or* physician $j$ waits for patients of type $\theta$.

Rationing-by-Waiting equilibria are stable and unique (see section §F).[10] Therefore, the assumption that matchings are produced by Rationing-by-Waiting equilibria allows for point-identified estimation of preferences. Further, computing a Rationing-by-Waiting equilibrium is computationally easier than implementing than the DA algorithm, it fits the empirical setting well, and it has an intuitive interpretation.

# 4 Empirical Setting: Primary Care in Northern Ontario

Northern Ontario was chosen as the empirical setting of this paper for three reasons. First, the prices that are paid to physicians are known and patients face zero user fees. Prices are therefore exogenous to the matching process. Second, primary care markets are easy to define, both according to physician specialties and geography. Lastly, detailed patient-level data on healthcare utilization is available for use by researchers.

Reforms to physician payment models are also advantageous for this research. These policy reforms provide variation that helps separately identify physician preferences and patient preferences. Additionally, the reform is an interesting policy to study using the estimated matching model.

The remainder of this section details the institutional environment of primary care in Northern Ontario then discusses the data used in the analysis.

## 4.1 Institutional Background

**Zero User Fees** Primary Health Care in Ontario is a single payer, private practice system. The Ontario Health Insurance Plan (OHIP) pays for all diagnostic and physician services that are deemed necessary care. There are zero user fees for these services. The services are financed through provincial and federal income taxation. Pharmaceutical and non-essential services are not covered by the single payer, and are often funded by private insurance (Marchildon, 2013; Glazier et al., 2019).

**Definable Markets** An insignificant private primary health care market exists alongside the publicly funded market. Private payment for necessary care is illegal, except in special circumstances (Ontario Legislature, 2004). Users of the private primary health care system must pay out-of-pocket

[10]See Galichon and Hsieh (2019) for a fuller discussion of the Rationing-by-Waiting equilibrium's properties.

(Starfield, 2010). Primary care clinics that sell access are rare in Ontario. There were 7 so-called "boutique" clinics in 2007 and 6 in 2017 (Mehra, 2008, 2017). None of these clinics are in Northern Ontario.[11]

Primary care in Ontario is provided almost exclusively by family physicians, with the exception of some pediatricians and nurse practitioners. Unlike in the United States, internists "rarely provide primary care, as they're trained to be hospitalists and subspecialists" (Bernstein, 2013). Geriatricians, likewise, do not provide primary care services (Frank and Wilson, 2015). Pediatricians are "encouraged to pursue subspecialties and act as 'consultants' instead of providing primary care" (Bernstein, 2013). In 2004, 30-40% of children's primary care visits were to a pediatrician. These visits were primarily in large urban areas rather than Northern Ontario (Canadian Paediatric Society Public Education Subcommittee, 2004). Nurse Practitioners work alongside family physicians in many primary care clinics. In these clinics, physicians are responsible for billing and the Nurse Practitioners are salaried (NPAO, 2020). A small number of nurse practitioner-led clinics do exist in Ontario in 2017, with the first such clinic opening in 2006 (College of Nurses of Ontario, 2008; Heale and Butcher, 2010). Pediatricians who provide primary care and nurse practitioners from nurse practitioner-led primary care clinics are excluded from the sample, as I am unable to distinguish them from their non-primary care colleagues.

**Alternative Payment Models** Before 2002, family physicians in Ontario were paid on a fee-for-service basis. For each service attained (e.g. an office visit, a flu shot), a physician would attain a pre-specified fee from the Ontario Health Insurance Plan (OHIP). From 2002 to 2006, the province introduced alternative payment models for family physicians. Participation in alternative payment models is voluntary and family physicians are able to participate in any payment model they qualify for. There were two main categories of alternative payment models: capitation and enhanced fee-for-service. Section §B provides a detailed description of all alternative payment models.

Capitation payment models provide yearly and monthly payments for each patient on a physician's roster, regardless of the services provided to the patient.[12] These capitation payments are risk-adjusted according to a patient's sex and age in five-year bins. Physicians in capitation payment models are paid a fraction of the fee-for-service fee for a specified basket of services. Also, physicians in capitation payment models receive bonuses for providing specific services to rostered patients. The first capitation payment model was introduced in April 2002, and the most popular capitation payment model was introduced in November 2006.

To incentivize physicians from rostering patients who do not actually go to their practice (and go elsewhere), the capitation payment models have an "Access Bonus" structure: A physician group is provided an access bonus of 18.59 percent of all capitation payments. If rostered patients receive services from physicians outside of the group, the cost of those services are deducted from the access

[11] In 2017, all were in the highly urbanized Golden Horseshoe region.

[12] Note that I refer to the terms "panel" and "roster" to describe slightly different objects. A physician's panel is all patients who primarily go to that physician. A physician's roster is all patients that are formally signed up to be the physician's patient. In the empirical work, these are assumed to be equivalent. They are, however, different with regards to the institutional details.

bonus (Glazier et al., 2019).

Enhanced fee-for-service (EFFS) payment models provide small monthly capitation payments for rostered patients and the same bonuses as physicians in the capitation models. In the more popular enhanced fee-for-service model, physicians are paid fees that are 10% or 15% higher than the fee-for-service fee for certain services provided to rostered patients. The most popular enhanced fee-for-service payment model was introduced in October 2005 (Buckley et al., 2014b).

The primary goal of these payment models was to increase access. Secondary goals were to control costs and increase quality of care (Hutchison et al., 2001). Physicians who joined the alternative payment models are usually mandated to join as a group (at least 3 physicians), though "physicians practicing in groups do not need to be co-located or share the same electronic medical record" (Kiran et al., 2018). This led to a decrease in solo practitioners from 37.4% in 2001 to 24.9% in 2010 (Hutchison and Glazier, 2013). As a group, physicians were expected to offer more after-hours visit opportunities and provide more continuity of care.

**Incentives Generated by the Alternative Payment Models** Alternative payment models increase the incentive for physicians to add patients and increase the incentive for patient selection. Figure 4b illustrates both incentives. In this figure, the horizontal axis is the expected utilization of a patient attaining care from a fee-for-service physician. The vertical axis is the revenue a physician would expect to attain per patient visit. The plotted curves describe the relationship between expected number of visits and expected revenue per visit in each payment model.[13]

The alternative payment models provide an incentive for physicians to expand their practices, as long as physician supply is upward sloping in revenue. A physician attains more revenue per visit from a patient if she is in an enhanced fee-for-service model or a capitation model, regardless of the utilization of the patient.[14] Physicians have a further incentive to decrease per-patient utilization in order to expand the number of patients in the panel. These incentives are a feature of the alternative payment models – the policymakers hoped they would increase access to care.

The alternative payment models also changed the incentive to differentially select patients. Physicians in the alternative payment models are paid partially on a per-patient basis. These capitation payments are risk adjusted on sex and five-year age bins. In the capitation payment models, capitation payments make up the majority of physician revenue. Therefore, there is a large incentive for physicians to select patients with low expected utilization, conditional on sex and age. This can be seen graphically in figure 4b. The slope of expected revenue per visit with respect to expected utilization is particularly steep for the capitation payment model.

The details of the payment models imply that incentives for selection along other dimensions exists. The access bonus structure of the capitation model provides an incentive for physicians to select patients who are less likely to visit other physicians. The enhanced fee structure of the

[13]In this analysis, I hold patient utilization fixed when comparing the payment models. Section §A.5 describes how expected number of visits and expected revenue is estimated.

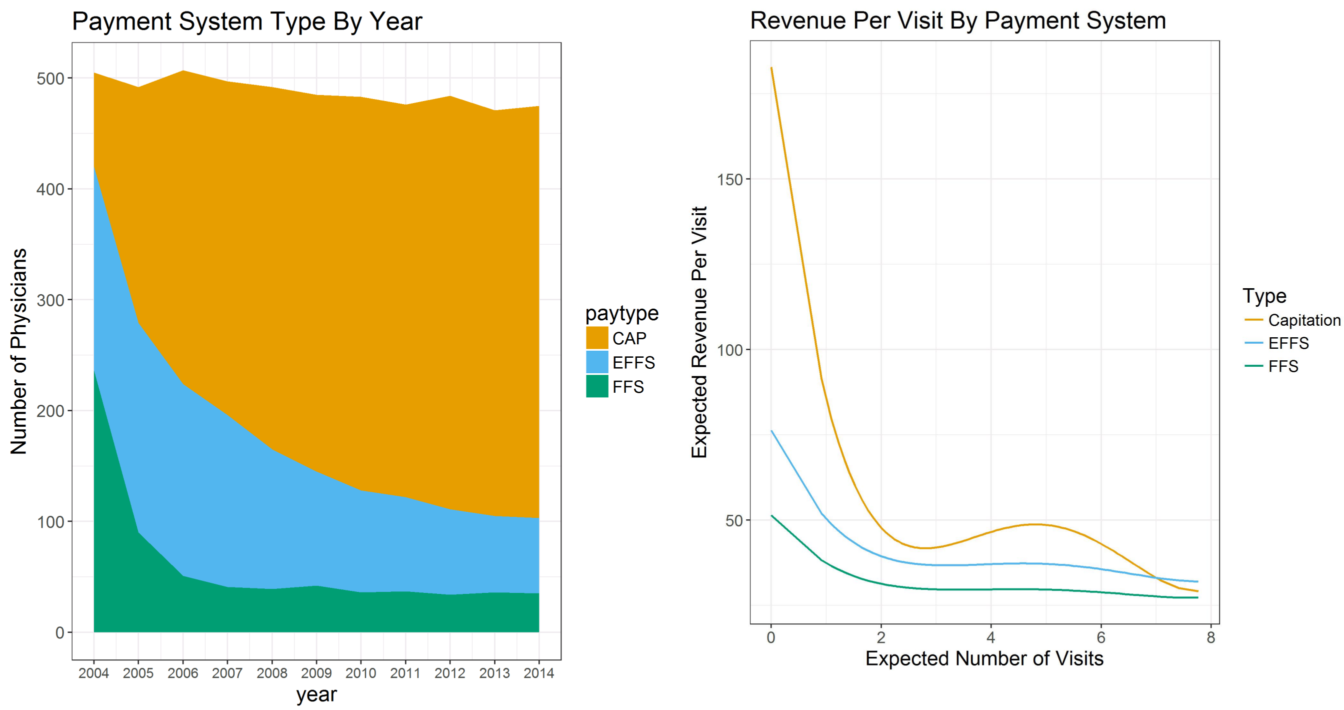
[14]An upward sloping supply curve cannot be assumed. Physicians may have backwards-bending supply curves or may be unmotivated by financial incentive.

enhanced fee-for-service model provides an incentive for physicians to select patients who are likely to attain services that have higher margins. I account for these incentives in the model.

Figure 4: Alternative Payment Models

(a) Alternative Payment Models Over Time

(b) Incentives For Selection and Expansion




**Selection into Alternative Payment Models** Physicians select into alternative payment models based on their characteristics and potential revenue gains. Table 1 shows summary statistics of physicians in each payment model. Enhanced fee-for-services physicians look roughly similar to fee-for-service physicians, with the exception that they have more patients. Capitation physicians are younger, more likely to be female, have older patients, and have more patients. Importantly, capitation physicians have patients who are particularly profitable in a capitation payment model. This suggests either that patients select into capitation models based on the makeup of their patient panel, or that they select patients in response to incentives that are built into the capitation model.

Table 1: Payment Model Selection Dataset Summary Statistics

| Payment Model | Capitation | EFFS | FFS |
|---|---|---|---|
| Number of Physician-Years (N) | 3,750 | 1,389 | 712 |
| Age (mean (sd)) | 49.30 (10.59) | 51.00 (11.05) | 50.65 (13.15) |
| Male (%) | 2496 (66.6) | 1053 (75.8) | 506 (71.1) |
| Total Patients (mean (sd)) | 816.86 (422.46) | 1390.64 (749.27) | 853.75 (575.53) |
| Percent Patients over 50 (mean (sd)) | 0.51 (0.13) | 0.43 (0.13) | 0.42 (0.17) |
| Percent Female Patients (mean (sd)) | 0.56 (0.09) | 0.55 (0.08) | 0.55 (0.10) |
| Percent Pats W. Comorbids (mean (sd)) | 0.25 (0.07) | 0.24 (0.07) | 0.23 (0.10) |
| FFS Revenue per visit (mean (sd)) | 29.54 (1.09) | 28.99 (0.84) | 29.34 (1.01) |
| EFFS Revenue per visit (mean (sd)) | 36.44 (1.65) | 35.70 (1.22) | 36.15 (1.49) |
| CAP Revenue per visit (mean (sd)) | 44.39 (4.82) | 41.57 (3.33) | 42.43 (4.26) |
| Area Type (%) | | | |
| Rural | 1773 (47.3) | 277 (19.9) | 271 (38.1) |
| Semiurban | 1219 (32.5) | 240 (17.3) | 159 (22.3) |
| Urban | 758 (20.2) | 872 (62.8) | 282 (39.6) |
| Pay Model Last Year (%) | | | |
| CAP | 3001 (80.0) | * | * |
| EFFS | 153 (4.1) | 1074 (77.3) | * |
| FFS | 596 (15.9) | * | *(>90) |

Statistics are calculated at the physician-year observation level. The sample used to create this table is the sample used to estimate unobserved taste for revenue. It includes physicians in Northern Ontario from 2004-2015. *Excluded due to small bin sizes.

### 4.2 Data

The primary dataset used in this analysis is a panel of patient-physician matches at a yearly frequency from 2004-2014. Data with information on patient characteristics, physician characteristics, and location characteristics are also used in the analysis. These data were compiled from several administrative datasets at the Institute of Clinical and Evaluative Sciences. In this subsection, I describe the datasets used in the main analysis and present summary statistics. Section §A provides more detail on how the datasets were constructed.

**Patient Characteristics** Patient characteristics and patient-physician matches were collected from Ontario Health Insurance Plan (OHIP) billings data. This data is a record of every billable service provided to patients in Ontario. From these data, I collect patient-level information on characteristics, comorbidities, and expected utilization. I also infer the number of patients who do not attain care from census data.

The main analysis requires patient characteristics at the time that they make their choice of physician. For characteristics that do not change over a year, these are taken to be the most common value within the year. Location, however, may change in the year and is endogenous to the timing of choice. I define the location of a patient to be the location they had when attaining services from their matched physician in that year.[15]

Patient comorbidities are defined using the ICD-9 Royal College of Surgeons' Charlson Comorbidity Mapping developed by Brusselaers and Lagergren (2017), with adjustments for the small differences between the Ontario Health Insurance Plan diagnoses codes and ICD-9 codes. Section A.2 provides details on how comorbidities are constructed. Twenty percent of patients have comorbidities (see table 14).

Expected revenue and expected visits are estimated using risk-adjustment methodology (Kautter et al., 2014). Section §A.5 details this estimation procedure.

Expected visits is the expected utilization of a patient conditional on comorbidities and characteristics. Expected visits are estimated assuming a patient has a fee-for-service physician. For each patient, three estimated revenues are computed: one for each payment model. Expected revenue is estimated as the expected revenue a physician would attain from the patient, assuming that the patient's utilization remains at fee-for-service levels.

Patients who do not attain any services are not observed in the data. I infer the number of these patients from patient data in adjacent years and census data. First, patients who are absent from the data in one year but attain services in adjacent years are added as patients who did not attain care. Second, census data is used to infer the remaining unobserved patients. Census data provide counts (up to the nearest 5 persons) of the number of persons for each location-age-sex category in the years 2001, 2006, 2011, and 2016. For non-census years, I assume that the count is the weighted average of the two closest census years. I subtract the number of patients in each location-age-sex category that we observe in the data from the corresponding census counts to infer the number of patients who did not attain care.

Patient types are discrete bins of patients based on characteristics. Characteristics that define patient types are: age in 15-year bins, sex, location, whether the patient has a comorbidity, and whether the patient is above or below a cutoff of expected revenue in a fee-for-service payment model. I use the median expected revenue in the age, sex, location, comorbidity bin as the cutoff.

Table 5a presents the patient characteristics summary statistics. 20% of patients have a comorbidity, and 33% do not attain care in a given year. Patients provide more revenue for physicians in the alternative payment models than in the fee-for-service system. There are 37,873 discrete patient types, with an average of 230.2 patients in each type.

**Physician Characteristics** Physician characteristics were collected from two datasets. The Corporate Provider Database provides physician characteristics, including physician sex, age, specialty, group affiliations, and locations. Physician payment models are from the Client Agency Program Enrolment dataset.

Physician choices are made at the "potential panel space" level $q$. For each potential panel space, physicians choose a patient type or choose to leave the panel space open. The total number of potential panel spaces, $m_j^a$, is unobserved in the data. I assume that the total number of potential

[15] Precisely, the location is the most common location associated with billings with their chosen physician.

panel spaces for each physician is the maximum number of patients they have over all years. Lastly, I calculate the number of potential panel spaces in market $t$, $m_{jt}$, by subtracting the number of patients outside of market $t$ who match with the physician from $m_j^a$.

Physicians who match with less than 300 patients and physicians who are not comprehensive care primary care physician are excluded from the sample. Table 5b provides physician characteristics summary statistics of physicians in the sample. These summary statistics are broken out by geographical market in table 15. Physicians have 1,170 patients and 4,894 visits per year on average. Physicians who are in the sample for the entire 11-year panel make up most physician-year observations.

Figure 5: Patient and Physician Summary Statistics

(a) Patient Summary Statistics

| | Summary Statistics |
|---|---|
| N | 8,718,090 |
| N Types | 37,873 |
| Male (%) | 48.9 |
| Age (%) | |
| 0-14 | 16.1 |
| 15-34 | 23.5 |
| 35-49 | 20.5 |
| 50-64 | 22.0 |
| 65+ | 17.9 |
| Has Comorbidity (mean (sd)) | 0.20 (0.40) |
| Revenue (mean (sd)) | |
| Capitation | 175.41 (81.99) |
| EFFS | 147.50 (64.49) |
| FFS | 119.63 (54.67) |
| Area Type (%) | |
| Rural | 3,020,672 (34.6) |
| Semiurban | 2,482,297 (28.5) |
| Urban | 3,215,121 (36.9) |
| Unmatched (mean (sd)) | 0.33 (0.18) |

Unit: Patient-Year; Panel: 2004-2014

(b) Physician Summary Statistics

| | Summary Statistics |
|---|---|
| N | 5,374 |
| Male (%) | 70.1 |
| Age (mean (sd)) | 49.82 (11.01) |
| N Years in sample | 10.31 (1.86) |
| N Patients | 1,169.51 (700.86) |
| Unfilled Capacity | 0.16 (0.16) |
| Group Status (%) | |
| Independent | 15.8 |
| Multiple groups | 10.3 |
| One Group | 73.9 |
| Payment Model (%) | |
| CAP | 62.8 |
| EFFS | 24.6 |
| FFS | 12.6 |
| Area Type (%) | |
| Rural | 40.2 |
| Semiurban | 27.7 |
| Urban | 32.2 |
| Visits (mean (sd)) | 4,893.79 (3,175.55) |
| Offers Walkins (%) | 29.6 |

Unit: Physician-Year; Panel: 2004-2014

**Matches and Drivetime** A patient matches with the physician they visit most in a year. A visit is defined in section 6.1. In the case of a tie, the patient matches with the physician with whom their visits had the highest value.[16] This selects physicians who provided more comprehensive

[16] Physicians who are more than 150 km away from a patient are excluded from the set of possible physicians to match with. Visits to those physicians are likely during temporary residence.

visits.[17] This simple definition of a match was adopted to allow for a longer panel and for sake of simplicity. When a more complex algorithm was used to determine matches, results were similar.

The distance between a patient and physician is estimated as the population-weighted drive time between their locations. Locations are known up to the census subdivision, which are municipalities. Population-weighted average drive times between census subdivisions are derived using census population data and the Bing Maps Distance Matrix API. Each census subdivision was split into Canadian Census Dissemination Areas (DA). A dissemination area is the smallest standard geographic area for Canada. Dissemination areas are loosely uniform in terms of population (400-700 persons). Statistics Canada has calculated the representative point (latitude/longitude) of population for each DA. I calculate the drive time between all DA representative points in a market. The average drivetime between two census subdivisions is calculated as the population-weighted average drivetime between all combination of DAs.

Match pattern summary statistics presented in table 2 reflect preferences of patients and physicians. The average drive time between patients and physicians is 17.23 minutes. The percent of patients who attain care scale with urbanity, age, and number of comorbidities. Evidence of homophily also exists: female patients are more likely to match with female doctors than male patients.

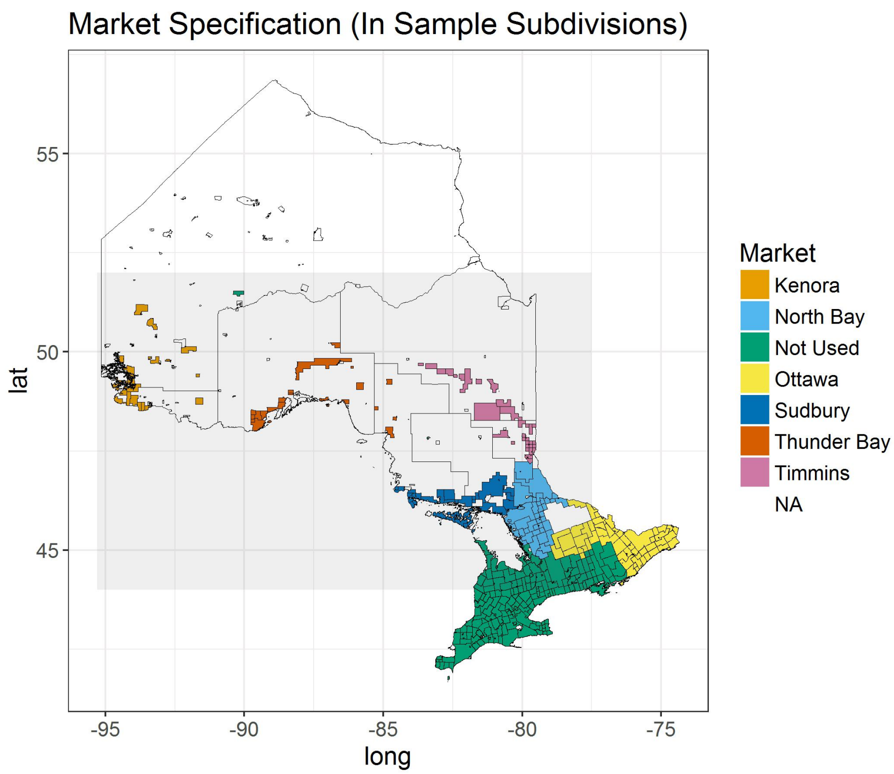
Table 2: Matching Summary Statistics

| | Summary Statistics |
|---|---|
| *Match Probabilities* | |
| N | 8,718,090 |
| Rurality (N (% unmatched)) | |
| Urban | 3,215,121 (26.50) |
| Semi-Urban | 2,482,297 (35.59) |
| Rural | 3,020,672 (37.85) |
| Age (N (% unmatched)) | |
| 0-14 | 1,404,284 (43.90) |
| 15-34 | 2,045,868 (37.00) |
| 35-50 | 1,787,272 (34.44) |
| 51-64 | 1,916,392 (27.57) |
| 65+ | 1,564,274 (20.51) |
| Comorbidities (N (% unmatched)) | |
| Without | 7,007,826 (36.40) |
| With | 1,710,264 (19.19) |
| *Match Characteristics* | |
| Minutes Drive Time (mean (sd)) | 17.23 (17.17) |
| Female Physician (% of matches) | |
| All Patients | 24.57 |
| Female Patients | 28.94 |
| Unit: Patient-Year; Panel: | 2004-2014 |

[17] Visit value is the service fee associated with the visit fee codes, not the amount paid to the doctor. This insures that physicians in fee-for-service payment models are not selected at higher frequency.

**Geographic Markets** Geographic markets are defined to ensure that there is limited interaction between physicians and patients outside of a geographic market. Population centers in Northern Ontario are fairly isolated. Five primary care markets with little inter-market matching are defined. These markets are the Kenora, Thunder Bay, Timmins, North Bay, and Sudbury regions. Figure 6 presents the regions. The Kenora, Thunder Bay, and Timmins regions are the most isolated, while the Sudbury and North Bay regions are less isolated. Table 3 describes how isolated each market is. Column 3 provides the share of matches that are between patients or physicians outside of the market.

Figure 6: Markets




Some census subdivisions are removed from the sample due to data issues. Figure 15 describes the reasons each were removed. The maps in the remainder of this paper correspond to the shaded region. The Ottawa market is used in some specifications to test the out-of-sample fit of the model. This map is built using 2006 CSD geographies. Geographical boundaries will differ slightly from year to year.

Table 3: Matchings Within/Across Markets

| | | % Connections | |
|---|---|---|---|
| Market Name | N CSDs | In Market | In Sample |
| Kenora | 18 | 100.00 | 100.00 |
| Thunder Bay | 19 | 99.84 | 99.98 |
| Timmins | 35 | 99.22 | 99.85 |
| Sudbury | 38 | 95.79 | 99.67 |
| North Bay/Parry Sound | 40 | 94.04 | 97.42 |
| Ottawa | 67 | 96.77 | 99.36 |

# 5 Specification

I specify the model to account for the primary determinants of matching patterns and to allow for tractable estimation. In the matching model, patient and physician preferences are flexibly specified as parametric functions of characteristics. Extreme value type I taste error terms and the Rationing-by-Waiting equilibrium concept provide an explicit formula for match probabilities.

Lastly, I specify a discrete choice model where physicians select payment models. Estimates from this model are used to account for unobserved heterogeneity in physician matching preferences.

## 5.1 Matching Model

**Patient Preferences Over Physicians** The latent utility that patient $i$ derives from matching with physician $j$ in market $t$ depends on a mean utility term, $\delta^{\boldsymbol{\beta}}_{\theta jt}$, and an extreme value type I taste shock. The mean utility can be categorized into three terms. First, patients derive utility from attaining medical care. Second, a patient of type $\theta$ has additional preferences for physician $j$, depending on $j$'s characteristics. Third, patients experience a disutility of traveling to physicians who are distant.

$$u_{i\theta jt} = \underbrace{\overbrace{\boldsymbol{\beta}^u_1 + \boldsymbol{x}'_{\theta t}\boldsymbol{\beta}^u_2}^{\text{Value of Attaining Care}} \quad + \quad \overbrace{\boldsymbol{x}'_{\theta jt}\boldsymbol{\beta}^u_3}^{\text{Net Match Value}} \quad - \overbrace{\boldsymbol{f_d}(d_{\theta jt})'\boldsymbol{\beta}^u_4}^{\text{Cost of Travel}}}_{\delta^{\boldsymbol{\beta}}_{\theta jt}} + \epsilon_{i\theta jt} \tag{6}$$

In the main specification, the variables that determine the value of attaining care, $\boldsymbol{x}_{\theta t}$, include an estimate of the expected number of visits a type $\theta$ patient will make and categorical descriptions of the patient type: 15-year age bins, sex, and whether the patients have comorbidities. Some interactions between patient characteristics are also included. The net match value variables, $\boldsymbol{x}_{\theta jt}$, include variables that describe physician $j$, such as payment model, gender, and whether she offers walk-in visits. Additionally, $\boldsymbol{x}_{\theta jt}$ includes interactions between physician $j$'s characteristics and patient type $\theta$'s characteristics. Namely, these are interactions between: physician payment model and the number of expected visits; patient and physician gender. The cost of distance is specified as a quadratic function of distance, interacted with the population density of the patient's census division.

**Physician Preferences Over Patients** Physician $j$'s mean utility for type $\theta$ patients in market $t$, $\gamma^{\boldsymbol{\beta}}_{\theta jt}$, consists of four terms. The first term is the revenue a physician expects to attain from a type $\theta$ patient. Second, the physician attains utility for treating a type $\theta$ patient. This non-revenue match value is a combination of the cost and the altruistic psychic benefit to the physician of supplying services to the patient. These cannot be separately identified. Third, physicians suffer a opportunity cost associated with adding a patient to their panel. Last, an additive random taste shock at the patient type-physician panel space-market level is assumed to be distributed extreme value type I.

$$v_{\theta jqt} = \underbrace{\overbrace{[\beta^v_1 + \beta^v_2 \alpha_j] R_{\theta jt}}^{\text{Revenue}} + \overbrace{\beta^v_3 V_{\theta t} + \boldsymbol{y}'_{\theta t} \boldsymbol{\beta}^v_4}^{\text{Net Match Value}} - \overbrace{[\beta^v_5 + \boldsymbol{y}'_{jt} \boldsymbol{\beta}^v_6]}^{\text{Value of Leaving Space Open}}}_{\gamma^{\boldsymbol{\beta}}_{\theta jt}} + \eta_{\theta jqt} \tag{7}$$

Expected revenue $R_{\theta jt}$ corresponds to the revenue of physician $j$'s payment model.[18] Recall that expected revenue is the revenue assuming fee-for-service utilization levels. That is, physicians select patients based on the revenue levels they would receive from those patients in the absence of practice style changes that are made in response to the alternative payment models.

Physicians vary in their responsiveness to revenue, $\beta^v_1 + \beta^v_2 \alpha_j$. This heterogeneity is specified as an unobserved taste for revenue, $\alpha_j$. Accounting for this heterogeneity is important for two reasons. First, the coefficient on revenue is identified by variation in revenue across physicians payment models. Physicians who choose the alternative payment models, however, are more likely to have high taste for revenue, generating a selection effect. Second, this selection effect also impacts the counterfactual analysis of the alternative payment model reforms. For accurate predictions of how a physician will select patients in a different payment model, I must account for physician selection into their current payment model.

The variables that enter the net match value term are the expected utilization (in visits) of a type $\theta$ patient, $V_{\theta t}$, and patient characteristics, $\boldsymbol{y}_{\theta t}$. The included patient characteristics are: 15-year age bins, sex, and whether the patient has a comorbidity. The value of leaving a space open depends on physician characteristics that change over time, $\boldsymbol{y}_{jt}$, including age, payment model, and years in practice.

Additional variables are included in $\boldsymbol{y}_{jt}$ that account for potential systematic error in the estimation of physician panel capacity $m_{jt}$. Since panel capacity is estimated as the maximum over all years, physicians who exist for less years in the panel are likely have less open panel spaces. To account for this, I include the proportion of years that the physician is in the sample as a variable in $\boldsymbol{y}_{jt}$.

[18] Note that expected revenue (visits) of a type $\theta$ patient is the average expected revenue (visits) of all patients in the patient type $R_{\theta jt} = \bar{R}_{i\theta jt}$ $(V_{\theta t} = \bar{V}_{i\theta t})$.

**Equilibrium** Matchings are assumed to be generated by a Rationing-by-Waiting equilibrium (equations 1, 2, 3, 4, and 5). Under the preference specifications above, the Rationing-by-Waiting equilibrium conditions for market $t$ can be written as explicit functions.

$$\text{Market Clearing}\begin{cases} \mu^{\boldsymbol{\beta}}_{\emptyset jt} + \sum_{\theta} \mu^{\boldsymbol{\beta}}_{\theta jt} = m_{jt} \ \forall j \\ \mu^{\boldsymbol{\beta}}_{\theta \emptyset t} + \sum_{j} \mu^{\boldsymbol{\beta}}_{\theta jt} = n_{\theta t} \ \forall \theta \\ \mu^{\boldsymbol{\beta}}_{\theta jt} = n_{\theta t}\frac{exp(\delta^{\boldsymbol{\beta}}_{\theta jt} - \tau^{u}_{\theta jt})}{1+\sum_{j'} exp(\delta^{\boldsymbol{\beta}}_{\theta jt} - \tau^{u}_{\theta jt})} = m_{jt}\frac{exp(\gamma^{\boldsymbol{\beta}}_{\theta jt} - \tau^{v}_{\theta jt})}{1+\sum_{\theta'} exp(\gamma^{\boldsymbol{\beta}}_{\theta jt} - \tau^{v}_{\theta jt})} \ \forall \theta \forall j \end{cases}$$

$$\text{One-Sided Effort}\Big\{ min\{\tau^{u}_{\theta jt}, \tau^{v}_{\theta jt}\} = 0 \ \forall \theta \forall j$$

Recall the interpretation of these equilibrium conditions. The market clearing conditions state that supply must equal demand. Given effort costs, $\tau^{u}_{\theta jt}, \tau^{v}_{\theta jt}$, and preferences $\boldsymbol{\beta}$, physician $j$ chooses to match with $\mu^{\boldsymbol{\beta}}_{\theta jt}$ type $\theta$ patients and $\mu^{\boldsymbol{\beta}}_{\theta jt}$ type $\theta$ patients choose to match with physician $j$. The last part of the equilibrium concept is the the one-sided effort condition. This is a constrained efficiency condition that guarantees that either type $\theta$ patients expend effort to match with physician $j$ or physician $j$ expends effort to match with type $\theta$ patients.

Under this specification, the equilibrium efforts $\tau^{u}_{\theta jt}, \tau^{v}_{\theta jt}$ can be eliminated. The odds ratio between the number of patients of type $\theta$ who match physician $j$ and the outside option shares for patient $\theta$ and physician $j$ are constructed from the equilibrium conditions.

$$\frac{\mu^{\boldsymbol{\beta}}_{\theta jt}}{\mu^{\boldsymbol{\beta}}_{\theta \emptyset t}} = exp(\delta^{\boldsymbol{\beta}}_{\theta jt} - \tau^{u}_{\theta jt})$$
$$\frac{\mu^{\boldsymbol{\beta}}_{\theta jt}}{\mu^{\boldsymbol{\beta}}_{\emptyset jt}} = exp(\gamma^{\boldsymbol{\beta}}_{\theta jt} - \tau^{v}_{\theta jt})$$

Solving for the equilibrium effort costs and plugging them into the One-Sided Effort condition implies that

$$0 = min\Big\{ \delta^{\boldsymbol{\beta}}_{\theta jt} - log(\frac{\mu^{\boldsymbol{\beta}}_{\theta jt}}{\mu^{\boldsymbol{\beta}}_{\emptyset jt}}) \ , \ \gamma^{\boldsymbol{\beta}}_{\theta jt} - log(\frac{\mu^{\boldsymbol{\beta}}_{\theta jt}}{\mu^{\boldsymbol{\beta}}_{\emptyset jt}}) \Big\}$$

Lastly, taking the monotonic transformation $\mu_{\theta jt} exp(\cdot)$ to both sides provides the matching condition provided by Galichon and Hsieh (2019):

$$\mu^{\boldsymbol{\beta}}_{\theta jt} = min\Big\{ \mu^{\boldsymbol{\beta}}_{\theta \emptyset t} exp(\delta^{\boldsymbol{\beta}}_{\theta jt}) \ , \ \mu^{\boldsymbol{\beta}}_{\emptyset jt} exp(\gamma^{\boldsymbol{\beta}}_{\theta jt}) \Big\}$$

The Rationing-by-Waiting Equilibrium conditions thus simplify to:

$$\sum_j \mu^{\boldsymbol{\beta}}_{\theta jt} + \mu^{\boldsymbol{\beta}}_{\theta\emptyset t} = n_{\theta t} \ \forall\theta \tag{8}$$

$$\sum_\theta \mu^{\boldsymbol{\beta}}_{\theta jt} + \mu^{\boldsymbol{\beta}}_{\emptyset jt} = m_{jt} \ \forall j \tag{9}$$

$$\mu^{\boldsymbol{\beta}}_{\theta jt} = min\left\{\mu^{\boldsymbol{\beta}}_{\theta\emptyset t} exp(\delta^{\boldsymbol{\beta}}_{\theta jt}) \ , \ \mu^{\boldsymbol{\beta}}_{\emptyset jt} exp(\gamma^{\boldsymbol{\beta}}_{\theta jt})\right\} \forall\theta\forall j \tag{10}$$

These equilibrium conditions show the logic of the rationing-by-waiting equilibrium. The number of matches between patient type $\theta$ and physician $j$ is driven by the preferences of the side of the market that is not expending effort to match.

## 5.2 Payment Model Selection

As previously discussed, accounting for physician unobserved taste for revenue $\alpha_j$ is important. I use additional identifying information from physicians' payment model choices to estimate the unobserved taste for revenue. Physicians with high taste for revenues are likely to choose the payment model that provides them the most revenues. A model of how physicians choose payment models is specified as follows.

**Physician Preferences over Payment Models** Physician $j$ in market $t$ derives utility in payment model $s$ from the matching market outcomes and the fixed costs of operating the payment model. Fixed costs include pecuniary and psychic costs, such as the distaste a physician may have for operating in an unfamiliar model.

$$w_{jst} = \overbrace{\phi_1 \varepsilon(v_{\theta jqt}, s)}^{\text{Matching Value}} - \overbrace{\phi_{3s} m_{jt} - \boldsymbol{\phi}'_{4s} \boldsymbol{z}^{FC}_{jt} - \boldsymbol{\phi}'_{5s} \boldsymbol{c}_{jy-1} + \phi_s}^{\text{Fixed/Switching Costs}} + \nu_{sjt} \tag{11}$$

The value of a matching is the physician's expectation of per-patient latent utility attained from the matching market $\varepsilon(v_{\theta jqt}, s)$. This expectation depends on the payment model chosen, $s$. Fixed and switching costs include the physician's panel capacity $m_{jt}$, physician-time varying variables, $\boldsymbol{z}^{FC}_{jt}$, indicators for lagged payment model $\boldsymbol{c}_{jy-1}$, and payment model fixed effects, $\phi_s$. Note that the subscript $t$ denotes a market, which are defined by year and geographic market. The subscript $y$ denotes only year.

The physician-specific characteristics in $\boldsymbol{z}^{FC}_{jt}$ are: age, squared age, sex, and indicators for rurality. Additionally, the average characteristics of physicians who are within a 30 minute drive from physician $j$ are also included in $\boldsymbol{z}^{FC}_{jt}$. To belong to an alternative payment model, physicians must generally join as a group of 3 physicians or more (see table 13). Therefore, if nearby physicians have a distaste for payment model $s$, it will be more difficult to do so. The average age and number of patients of close physicians are used.

**Myopic Choice Assumption** I assume that physicians are myopic when determining their expectation of the per-patient latent utility they would attain from the matching market. They are not looking forward at how their payment model choice will change the matching market equilibrium. Further, I assume that physicians make payment model choices before, not jointly with, the matching market equilibrium. These assumptions translate to simplifications in the payment model selection model. Specifically, physicians expect to keep the same patients irrespective of the payment model they choose. Loosening this assumption is a subject for further work, but is difficult for both conceptual and computational reasons. For example, it is not clear whether a unique equilibrium exists in a game where physicians first choose a payment model, then compete in a matching market.

Under the myopia assumption, $\varepsilon(v_{\theta jqt}, s)$ is equal to the expected average utility that physician $j$ choosing payment model $s$ attains from a patient on her panel.

$$\varepsilon(v_{\theta jqt}, s) = E[v_{\theta jqt}|c_{jy} = s] = \gamma^e + \frac{1}{m_{jt}} \sum_{\theta} \mu_{\theta jt} \gamma^{\boldsymbol{\beta}}_{\theta jst}$$

where $\gamma^e$ is Euler's constant and $\gamma^{\boldsymbol{\beta}}_{\theta jst}$ is the mean utility of physician $j$ matching with a type $\theta$ patient in market $t$ while in payment model $s$. $\mu_{\theta jt}$ is the observed matching in the data. I normalize such that choosing fee-for-service provides zero mean latent utility.

$$\begin{aligned} w_{jst} &= \overbrace{\phi_1 \Delta\varepsilon(v_{\theta jqt}, s)}^{\text{Matching Value}} - \overbrace{\phi_{3s} m_j - \boldsymbol{\phi}'_{4s} \boldsymbol{z}^{FC}_{jt} - \boldsymbol{\phi}'_{5s} \boldsymbol{c}_{jy-1} + \phi_s}^{\text{Fixed/Switching Costs}} + \nu_{sjt} \quad && s \in \{\text{Capitation}, \text{EFFS}\} \\ w_{jst} &= \nu_{sjt} && s = \text{FFS} \end{aligned} \tag{12}$$

where $\Delta\varepsilon(v_{\theta jqt}, s) = \varepsilon(v_{\theta jqt}, s) - \varepsilon(v_{\theta jqt}, \text{FFS})$ is the difference between the per-patient latent utility that physician $j$ attains in payment model $s$ and in the fee-for-service payment model.

Since only revenue and the value of leaving a panel space open vary by payment model in physician preferences over patients, the difference in the per-patient latent utility can be written simply.

$$\Delta\varepsilon(v_{\theta jqt}, s) = [\beta^v_1 + \beta^v_2 \alpha_j] \Delta_s R_{\theta jt} + \Delta_s \boldsymbol{y}'_{jt} \boldsymbol{\beta}^v_6$$

where $\Delta_s R_{jst} = \frac{1}{m_{jt}} \sum_\theta \mu_{\theta jt} [R_{\theta jst} - R_{\theta jFFSt}]$ is the difference in average revenues between payment model $s$ and FFS for physician $j$. $\Delta_s \boldsymbol{y}'_{jt}$ is the difference in physician characteristics between payment model $s$ and FFS. This becomes a constant for each payment model. Thus, physician latent utility can be simplified.

$$w_{jst} = \underbrace{\overbrace{[\tilde{\phi}_1 + \alpha_j] \Delta_s R_{jst}}^{\text{Matching Value}} - \overbrace{\tilde{\phi}_{3s} m_j - \tilde{\boldsymbol{\phi}}'_{4s} \boldsymbol{z}^{FC}_{jt} - \tilde{\boldsymbol{\phi}}'_{5s} \boldsymbol{c}_{jt-1} + \tilde{\phi}_s}^{\text{Fixed/Switching Costs}}}_{W_{jst}(\tilde{\boldsymbol{\phi}}, \epsilon^R_j)} + \nu_{sjt} \tag{13}$$

where $\tilde{\phi}_1 = \phi_1\beta_1^v$, $\tilde{\phi}_{3s} = \phi_{3s}$, $\tilde{\boldsymbol{\phi}}_{4s} = \boldsymbol{\phi}_{4s}$, $\tilde{\boldsymbol{\phi}}_{5s} = \boldsymbol{\phi}_{5s}$, and $\tilde{\phi}_s = \phi_s + \Delta_s \boldsymbol{y}'_{jt}\boldsymbol{\beta}_6^v$

Lastly, I allow unobserved taste for revenue to depend on physician observables $\boldsymbol{z}_{jt}^R$ and an error term $\epsilon_j^R$, which will be estimated. The vector $\boldsymbol{z}_{jt}^R$ includes physician variables that may affect sensitivity to revenue. These include age, squared age, and sex.

$$\alpha_j = \tilde{\boldsymbol{\phi}}_2 \boldsymbol{z}_{jt}^R + \epsilon_j^R$$

# 6 Estimation

Preferences are estimated in two stages. First, the unobserved physician taste for revenue $\alpha_j$ is estimated using data on observed payment model choices. In the second stage, patient and physician preferences are jointly estimated using matching data. I present the estimation procedure in reverse order.

## 6.1 Estimation of Matching Preferences

I estimate preferences using a nested fixed point maximum likelihood method. The likelihood function of the matching model can be written as a simple function of matching data and predicted matchings.

$$\mathcal{L}_t(\boldsymbol{\beta}) = 2\sum_j \sum_\theta \mu_{\theta jt} log(\mu_{\theta jt}^{\boldsymbol{\beta}}) + \sum_\theta \mu_{\theta\emptyset t} log(\mu_{\theta\emptyset t}^{\boldsymbol{\beta}}) + \sum_j \mu_{\emptyset jt} log(\mu_{\emptyset jt}^{\boldsymbol{\beta}})$$

where $\mu_{\theta jt}^{\boldsymbol{\beta}}$ is the number of patients of type $\theta$ who match with physician $j$ in the predicted Rationing-by-Waiting equilibrium when preferences are defined by $\boldsymbol{\beta}$. For brevity I use the notation $\mu_{\theta jt}^{\boldsymbol{\beta}} = \mu_{\theta jt}(\boldsymbol{\beta})$, $\mu_{\theta\emptyset t}^{\boldsymbol{\beta}} = \mu_{\theta\emptyset t}(\boldsymbol{\beta})$, and $\mu_{\emptyset jt}^{\boldsymbol{\beta}} = \mu_{\theta\emptyset t}(\boldsymbol{\beta})$.[19]

I predict matchings under the parametric specification laid out in section 5. Thus, we can reformulate the estimation problem as:

$$\max_{\boldsymbol{\beta}} 2\sum_t \sum_j \sum_\theta \mu_{\theta jt} log(\mu_{\theta jt}^{\boldsymbol{\beta}}) + \sum_t \sum_\theta \mu_{\theta\emptyset t} log(\mu_{\theta\emptyset t}^{\boldsymbol{\beta}}) + \sum_t \sum_j \mu_{\emptyset jt} log(\mu_{\emptyset jt}^{\boldsymbol{\beta}})$$

$$s.t. \ \sum_j \mu_{\theta jt}^{\boldsymbol{\beta}} + \mu_{\theta\emptyset t}^{\boldsymbol{\beta}} = n_{\theta t} \ \forall\theta\forall t$$

$$\sum_\theta \mu_{\theta jt}^{\boldsymbol{\beta}} + \mu_{\emptyset jt}^{\boldsymbol{\beta}} = m_{jt} \ \forall j\forall t$$

$$\mu_{\theta jt}^{\boldsymbol{\beta}} = min\left\{\mu_{\theta\emptyset t}^{\boldsymbol{\beta}} exp(\delta_{\theta jt}^{\boldsymbol{\beta}}) \ , \ \mu_{\emptyset jt}^{\boldsymbol{\beta}} exp(\gamma_{\theta jt}^{\boldsymbol{\beta}})\right\} \ \forall\theta\forall j\forall t$$

[19] The likelihood has an intuitive interpretation. Maximizing the likelihood is equivalent to minimizing the Kullback–Leibler divergence from the predicted matching $\mu_t^{\boldsymbol{\beta}}$ to the matching observed in the data $\mu_t$. The likelihood approach is therefore similar to the minimum distance approaches used in the existing literature (Hitsch et al., 2010; Boyd et al., 2013; Vissing, 2017; Agarwal, 2015; Matveyev, 2013). In both, a predicted matching is determined by an assumed equilibrium concept and parameterized preferences. Parameters are chosen to minimize a distance between the predicted matching and the matching observed in the data.

### 6.1.1 Nested Fixed Point Algorithm

The nested fixed point algorithm uses a non-linear optimization algorithm over $\boldsymbol{\beta}$ in an outer loop to maximize the likelihood. At each step of this outer loop, the predicted matching $\boldsymbol{\mu}^{\beta}$ must be calculated. To calculate $\boldsymbol{\mu}^{\beta}$, the solution to the Rationing-by-Waiting Equilibrium system of equations is found. This solution is found using an inner iterative loop that converges to a fixed point.

**Inner Loop** Given a parameter value $\boldsymbol{\beta}$, the inner loop solves the non-linear set of equations for $\boldsymbol{\mu}^{\beta}$:

$$\mu^{\beta}_{\theta\emptyset t} + \sum_j \mu^{\beta}_{\theta jt} = n_{\theta t} \ \forall\theta\forall t$$

$$\mu^{\beta}_{\emptyset jt} + \sum_\theta \mu^{\beta}_{\theta jt} = m_{jt} \ \forall j\forall t$$

$$\mu^{\beta}_{\theta jt} = min\Big\{\mu^{\beta}_{\theta\emptyset t}a^{\beta}_{\theta jt}\ ,\ \mu^{\beta}_{\emptyset jt}b^{\beta}_{\theta jt}\Big\} \ \forall\theta\forall j\forall t$$

Where $a^{\beta}_{\theta jt} = exp(\delta^{\beta}_{\theta jt})$ and $b^{\beta}_{\theta jt} = exp(\delta^{\beta}_{\theta jt})$. Given the parameter vector $\boldsymbol{\beta}$, $\boldsymbol{a}^{\boldsymbol{\beta}}$ and $\boldsymbol{b}^{\boldsymbol{\beta}}$ are known. Therefore, they are calculated at the beginning of each inner loop step and are treated as data thereafter.

This system of equations does not have a simple closed form solution due to non-linearities. I solve the system numerically using an iterative procedure similar to the iterative proportional fitting procedure used elsewhere in the empirical matching literature (Galichon and Salanié, 2010). Starting at an initial value of $\mu^{(0)}_{\emptyset jt}\forall j, t$, I iteratively solve the following two sets of equations until convergence. [20]

$$\mu^{\beta}_{\theta\emptyset t} + \sum_j \Big(\mathbf{1}\Big\{\mu_{\theta\emptyset t}a^{\beta}_{\theta jt} < \mu^{(k-1)}_{\emptyset jt}b^{\beta}_{\theta jt}\Big\}\mu^{\beta}_{\theta\emptyset t}a^{\beta}_{\theta jt} + \mathbf{1}\Big\{\mu_{\theta\emptyset t}a^{\beta}_{\theta jt} > \mu^{(k-1)}_{\emptyset jt}b^{\beta}_{\theta jt}\Big\}\mu^{\beta}_{\emptyset jt}b^{\beta}_{\theta jt}\Big) = n_{\theta t}\forall\theta\forall t \quad (14)$$

$$\mu^{\beta}_{\emptyset jt} + \sum_\theta \Big(\mathbf{1}\Big\{\mu^{(k)}_{\theta\emptyset t}a^{\beta}_{\theta jt} < \mu_{\emptyset jt}b^{\beta}_{\theta jt}\Big\}\mu^{\beta}_{\theta\emptyset t}a^{\beta}_{\theta jt} + \mathbf{1}\Big\{\mu^{(k)}_{\theta\emptyset t}a^{\beta}_{\theta jt} > \mu_{\emptyset jt}b^{\beta}_{\theta jt}\Big\}\mu^{\beta}_{\emptyset jt}b^{\beta}_{\theta jt}\Big) = Q_j\forall j\forall t \quad (15)$$

**Outer Loop** In the outer loop, I maximize the likelihood function.

$$\max_{\beta} 2\sum_t\sum_j\sum_\theta \mu_{\theta jt}log(\mu^{\beta}_{\theta jt}) + \sum_t\sum_\theta \mu_{\theta\emptyset t}log(\mu^{\beta}_{\theta\emptyset t}) + \sum_t\sum_j \mu_{\emptyset jt}log(\mu^{\beta}_{\emptyset jt})$$

using a quasi-newton non-linear optimization algorithm. To aid the algorithm, I derive the gradient of the likelihood. Section G.1 presents its derivation.

[20] An alternative iterative procedure was also used to test for consistency of results. This procedure involved iterating on the matrix form of the system. At each step, the procedure updated $R^{(k)}$, then $\boldsymbol{\mu}^{(k)}_{\emptyset} = [R^{(k)}]^{-1}\boldsymbol{v}$. This procedure returned identical results, but was less efficient.

### 6.2 Estimation of Unobserved Taste for Revenue

I estimate the physician choice of payment model using an L2 (Ridge) regularized maximum likelihood model. Recall that the object of interest of this exercise is the unobserved taste for revenue, $\alpha_j$, which is a function of observable characteristics and an error term $\epsilon_j^R$.

The regularized maximum likelihood method penalizes the squared magnitude of $\epsilon_j^R$ according to a smoothing parameter $\lambda$. Additionally, I assume that the idiosyncratic taste shock, $\nu_{sjt}$, is distributed Extreme Value type I. Thus, the regularized log likelihood problem has the form

$$\max_{\tilde{\boldsymbol{\phi}}, \boldsymbol{\epsilon^R}} \sum_j \sum_t \sum_s c_{jst} p_{jst}(\tilde{\boldsymbol{\phi}}, \epsilon_j^R) - \frac{\lambda}{2} \sum_j (\epsilon_j^R)^2$$

$$p_{jst}(\tilde{\boldsymbol{\phi}}, \epsilon_j^R) = log\left( \frac{exp(W_{jst}(\tilde{\boldsymbol{\phi}}, \epsilon_j^R))}{\sum_{s'} exp(W_{js't}(\tilde{\boldsymbol{\phi}}, \epsilon_j^R))} \right)$$

Recall that $W_{jst}(\tilde{\boldsymbol{\phi}}, \epsilon_j^R)$ is the mean utility of physician $j$ choosing payment model $s$ in market $t$. Thus, $p_{jst}(\tilde{\boldsymbol{\phi}}, \epsilon_j^R)$ is the probability that physician $j$ chooses payment model $s$ in market $t$. The smoothing parameter, $\lambda$, is chosen by cross-validation. I use a 12-fold cross validation technique and the log loss classification metric. Folds are created to accommodate the panel structure of the dataset. Each fold is a collection of randomly sampled observations. No fold includes more than one observation from the same physician. The $\lambda$ is chosen by the one-standard error rule. The one standard error rule accounts for error in the log loss function (Hastie et al., 2009). That is, define $LL(\lambda)$ as the log loss produced in cross validation using smoothing parameter $\lambda$. The $\lambda^*$ is chosen by the rule

$$\lambda_{min} = argmin_{\lambda \in \Lambda} LL(\lambda)$$
$$\lambda^* = min_{\lambda \in \Lambda} LL(\lambda)$$
$$s.t.\ LL(\lambda) > LL(\lambda_{min}) + SE(\lambda_{min})$$

where $SE(\lambda_{min})$ is the estimated standard error of the log loss function at the minimizing smoothing parameter. This is estimated using the variation of log loss among folds. $\Lambda$ is a grid of values. I use a grid of 200 points spread from .001 to .5.

The intuition behind this methodology is simple. This is a high dimension problem. There are 5,851 observations and 905 parameters. 879 of these parameters are random effects $\epsilon_j^R$. Thus, to avoid overfitting, the size of the random effects must be penalized.

Intuitively, the argument is that the estimators of the random effect $\epsilon_j^R$ suffer from the incidental parameters problem. However, it is reasonable to assume that physicians with similar observables will have similar revenue sensitivities. From a Bayesian perspective, I put a prior on the difference between the revenue sensitivity of each physician and the mean revenue sensitivity of physicians with identical characteristics. Specifically, the prior has the form $\epsilon_j^R \sim N(0, \frac{1}{\lambda})$.

**Alternatives** Due to the high dimensional nature of this model, $\epsilon_j^R$ are not point identified. Theoretically, the *distribution* of $\epsilon_j^R$ could be identified (Revelt and Train, 2000). However, due to the computational complexity of the empirical matching model, integrating the matching predictions over the distribution of $\epsilon_j^R$ is infeasible. Determining more efficient methods to predict Rationing-by-Waiting matchings with mixed coefficients is a potential avenue for further research.

### 6.3 Identification

Most importantly, identification rests on the assumption that patient preferences depend on distance, while physician preferences do not. This exclusion restriction allows the model to separately identify patient and physician preferences. Under this assumption, patients who are close to a physician are likely to have to wait for that physician. This implies that among those patients who close, match probabilities should reflect the choice probabilities of physicians. With physician preferences identified, patient preferences are then identified by the match probabilities conditional on physician preferences. Other exclusion restrictions may help in identification. Namely, revenue enters only into physician preferences. Revenue, however, is less predictive of matchings than distance.

Identification of preferences relies on three key features of the primary care market in Northern Ontario. First, the market is a many-to-one market. This is necessary for non-parametric identification of both the idiosyncratic taste shock distribution and preferences. Without a many-to-one market, any set of match probabilities could be explained by multiple taste shock distributions, therefore generating a reliance on the extreme value type I assumption for identification (Agarwal, 2013). Second, I observe multiple markets – both over years and geography. Additionally, even within a year and geographic market, the data structure is similar to many overlapping markets, as patients rarely match with physicians who are located more that 45 minutes away. This provides useful variation in the distribution of patient and physician types. Even in one-to-one settings, many markets can aid in identification (Hsieh, 2012).

## 7 Results: Patient and Physician Preferences

### 7.1 Physician Choice of Payment Model

Table 4 shows the results of the payment model choice model. The top panel presents the parameters that enter the unobserved taste for revenue $\alpha_j$. The bottom panel presents the fixed and switching cost parameters.

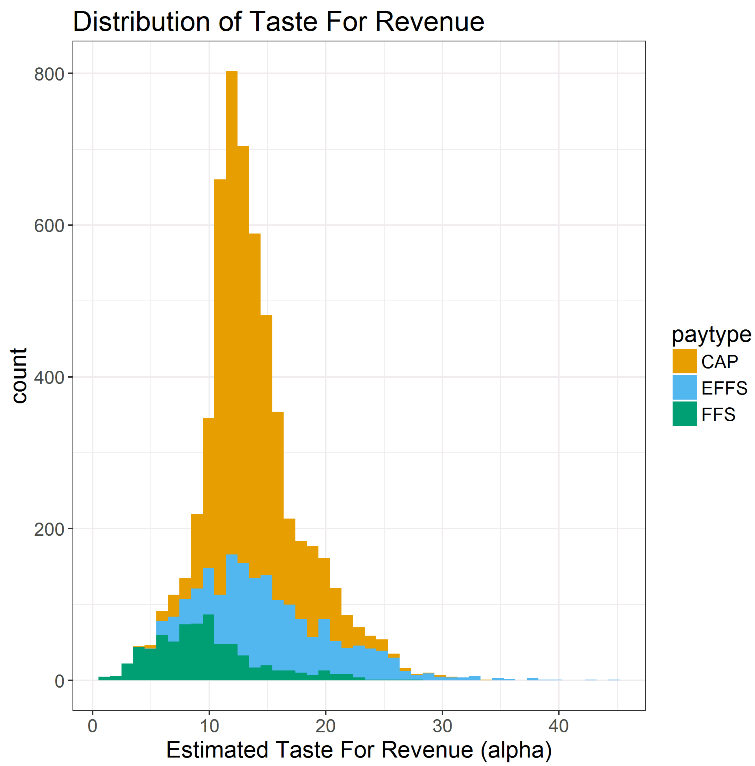
Fixed and switching cost parameters largely follow similar patterns to previous work by Rudoler et al. (2015b), who estimate payment model selection via mixed logit and allow for rich heterogeneity in physician panel characteristics. Conditional on revenue, characteristics that predict a higher probability of switching into an enhanced fee-for-service model are female, middle aged, larger patient panels and urban. Characteristics that predict a higher probability of switching to the capitation system are female, middle aged, rural, and smaller panel sizes.

Table 4: Physician Choice of Payment Model Estimates

| | All Payment Models | |
|---|---|---|
| Parameter | Estimate (Std. Dev.) | |
| **Revenue** | | |
| *Parameters in $\alpha_j$* | | |
| Constant | 7.449*** | |
| | (2.322) | |
| Num. of Patients | 6.573*** | |
| | (2.436) | |
| Female | -0.061 | |
| | (2.237) | |
| *Penalty* | | |
| $\lambda$ | 0.071 | |
| Payment Model: | EFFS | Capitation |
| Parameter | Estimate (Std. Dev.) | Estimate (Std. Dev.) |
| **Fixed/Switching Cost** | | |
| *Physician Characteristics* | | |
| Constant | -1.828 | -5.706*** |
| | (1.581) | (1.633) |
| Age | 6.463 | 5.600 |
| | (4.566) | (5.014) |
| Age$^2$ | -8.769* | -11.422** |
| | (4.557) | (5.143) |
| Num. of Patients | -0.402 | -2.039*** |
| | (0.362) | (0.639) |
| Semiurban | -0.244 | 0.667* |
| | (0.366) | (0.370) |
| Rural | -0.361 | 0.340 |
| | (0.487) | (0.523) |
| Female | -0.076 | -0.007 |
| | (0.378) | (0.669) |
| *Nearby Physicians* | | |
| Num. of Physicians | -0.101 | -0.528 |
| | (0.653) | (0.669) |
| Average Age | -2.198 | 7.620*** |
| | (2.451) | (1.800) |
| *Lagged Indicators* | | |
| EFFS | 6.445*** | 4.544*** |
| | (0.594) | (0.602) |
| Capitation | 1.023** | 5.318*** |
| | (0.436) | (0.359) |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 | |

There is substantial estimated unobserved taste for revenue. A penalty parameter of 0.071 is generated by the cross validation exercise – translating to a 3.75 hyperparameter standard deviation of the prior distribution of $\epsilon_j^R$. Figure 7 shows the estimated distribution of unobserved taste for revenue. The average estimated $\alpha_j$ is 13.75. Physicians who choose the capitation model are more likely to have high taste for revenue, while those in the fee-for-service model have a low taste for revenue.

Figure 7: Distribution of $\hat{\alpha}_j$




## 7.2 Empirical Matching Model Estimates

The results of the empirical matching model are presented in table 5 and table 6. Individual parameter magnitudes are difficult to interpret. The signs of the estimates are generally intuitive. As expected, I estimate that patients with more expected visits and female patients have higher preferences for attaining care. The estimates suggest that patients with more expected visits are more likely to prefer physicians in alternative payment models (non-fee-for-service models), perhaps due to the increased continuity of care these models provide. Evidence of homophily in sex is similarly strong.

Table 5: Patient Preferences

| Parameter | Estimate (Std. Dev.) | Parameter | Estimate (Std. Dev.) |
|---|---|---|---|
| *Care Value* | | *Net Match Value* | |
| Constant | -0.750*** | EFFS | -0.152*** |
| | (0.008) | | (0.006) |
| Expected Utilization (5 visits) | 0.842*** | EFFS × Expected Utiliz. | 0.080*** |
| | (0.011) | | (0.006)) |
| Comorbidity | -0.050*** | CAP | -0.608*** |
| | (0.004) | | (0.006) |
| Female | 0.326*** | CAP × Expected Utiliz. | 0.381*** |
| | (0.005) | | (0.006) |
| Under 14 | 0.017*** | Female doctor | -0.837*** |
| | (0.004) | | (0.002) |
| 35-49 | -0.044*** | Female doctor × Female | 0.851*** |
| | (0.003) | | (0.003) |
| 50-64 | -0.012 | Offers Walk-in Visits | -0.031*** |
| | (0.007) | | (0.001) |
| Over 65 | -0.168*** | *Travel Cost* | |
| | (0.010) | | |
| Over 50 × Expected Utiliz. | 0.190*** | Drivetime (hours) | 5.613*** |
| | (0.010) | | (0.004) |
| Female × Expected Utiliz. | -0.365*** | ... × log(Pop Density ($\frac{pop}{5km^2}$)) | 1.676*** |
| | (0.005) | | (0.009) |
| | | Drivetime$^2$ | -0.839*** |
| | | | (0.001) |
| | | ... × log(Pop Density) | -0.313*** |
| | | | (0.003) |
| *Included:* | | | Market Fixed Effects |
| *Note:* | | | $^*$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Physician preferences over patients are positively associated with revenue and expected number of visits. However, conditional on expected revenue and visits, the estimates suggest that physicians prefer male patients, younger patients, and patients without comorbidities. This could be explained by unobserved health that is observed by the physician but not the econometrician. For example, a young patient with high number of expected visits may have a serious unobserved health condition that physicians are particularly motivated to treat. The estimates also suggest that the cost in latent utility of adding a patient is negatively associated with alternative payment models.

Table 6: Physician Preferences

| Parameter | Estimate (Std. Dev.) |
|---|---|
| *Revenue* | |
| Revenue ($100) | 0.088*** |
| | (0.008) |
| $\alpha_j(Std.Dev.) \times$ Revenue($100) | 0.012 |
| | (0.011) |
| *Net Match Value* | |
| Expected Utilization (5 visits) | 3.321*** |
| | (0.004) |
| Comorbidity | -2.132*** |
| | (0.005) |
| Female | -0.289*** |
| | (0.004) |
| Under 14 | -0.216*** |
| | (0.003) |
| 35 to 49 | -0.744*** |
| | (0.007) |
| 50 to 64 | -1.220*** |
| | (0.010) |
| Over 65 | -2.318*** |
| | (0.010) |
| *Value of Leaving Panel Space Open* | |
| Constant | 2.394*** |
| | (0.005) |
| EFFS | -0.555*** |
| | (0.006) |
| CAP | -0.504*** |
| | (0.006) |
| Years in Practice | 0.109*** |
| | (0.006) |
| *Adjustments for* $\hat{m}_{jt}$ | |
| Proportion of years in sample | -0.161*** |
| | (0.006) |
| *Note:* | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

### 7.3 Exposition of Preferences

The estimated model suggests that physician preferences over patients are fairly responsive to expected utilization (visits) and are only mildly responsive to revenue. Figure 8 illustrates these findings. To do so, I construct an illustrative physician who is 50 years old, female, in a capitation payment model, and does not offer walk-in visits. For panel (a), I separate all patients in 2014 into 50 bins according to their expected number of visits. The figure shows the percent of patients in each bin who do not attain care, the probability that a patient chooses no care over choosing to go to the illustrative physician, and the probability that the illustrative physician would reject a patient. Panel (b) presents the results of the same procedure when conducted for expected capitation revenue.

Panel (a) suggests that physician preferences contribute to the low rates of attaining care among patients with low expected utilization. Both the probability of rejection and the patient's probability of attaining no care are downward sloping in expected utilization. Physicians discriminate in favor of patients with higher expected utilization. This suggests that physicians may be giving priority to patients with more need for healthcare.[21] However, there is no evidence from the matching model to separate this story from one where physicians prefer higher utilization due to cost efficiencies.

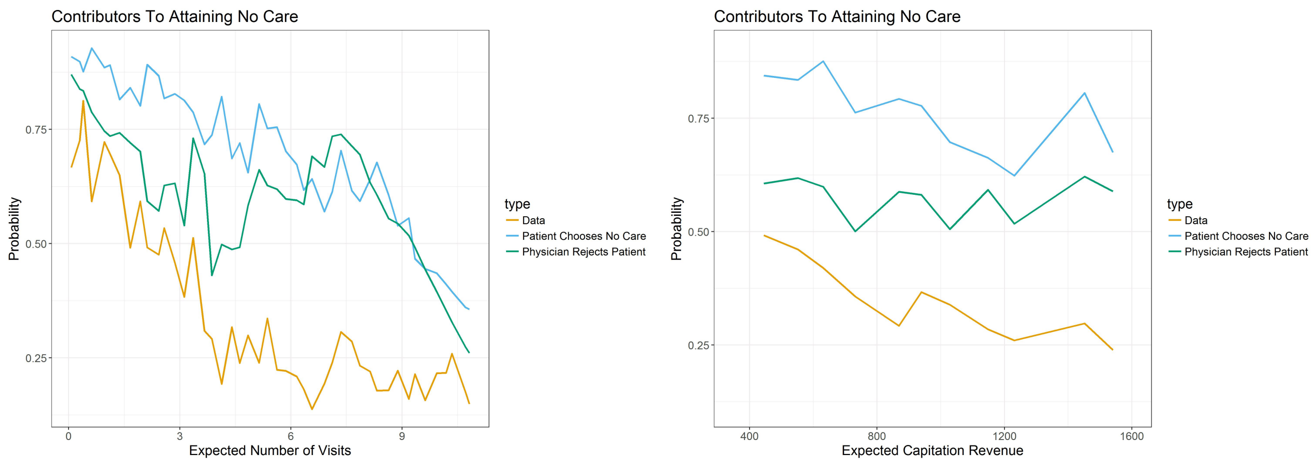
On the other hand, while the share of patients who attain care is decreasing in expected capitation revenue, the correlations do not suggest that physician responsiveness to revenue is driving those matching patterns. Other patient characteristics, such as age, exhibit similar patterns (see Appendix figure 18a).

This does not, however, imply that physicians are unresponsive to revenue on the margin, as matching patterns are a function of both demand and supply. To further explore the notion of how responsive physicians are to revenue, I calculate the elasticity of the number of type $\theta$ patients that physician $j$ chooses with respect to revenue. I find the average elasticity, weighted by observed matches, is 0.16. This is economically significant, but smaller than other estimates in the literature (Alexander and Schnell, 2018). The relatively low elasticity is consistent with the existing quasi-experimental literature, which finds little effect of revenue on selection by physicians in Ontario (Kantarevic and Kralj, 2014).

[21] Such preferential treatment is welfare improving under reasonable assumptions (Gravelle and Siciliani, 2008).

Figure 8: Preference Exposition

(a) Expected Number of Visits (b) Expected Capitation Revenue




### 7.4 Model Fit

The model predicts the data reasonably. Figure 9 compares the model's predictions to the data. The figure in panel (a) shows the density of the predicted share of patients without care compared to the share without care in the data. Most of the density falls along the 45-degree line, suggesting a reasonable fit. However, there is substantial density off of the 45-degree line. Panel (b) presents the relationship between the predicted and observed shares without care across three important cuts of the data. The table shows that the model underpredicts the outside option share in urban areas and overpredicts in rural areas.

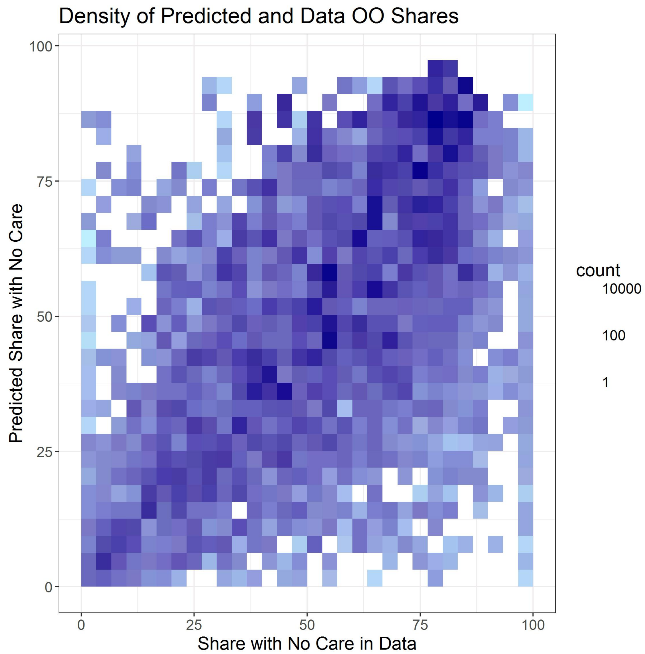
Unobserved heterogeneity in a patient's latent utility of attaining no care may cause the predicted share to differ from its empirical counterpart. For example, patients may experience shocks to their health which necessitate that they attain care that cannot be accounted for by the extreme value type 1 taste shock. Such unobserved heterogeneity could cause physician demand to be inelastic at both high levels of access and low levels of access. Unobserved heterogeneity could be easily added to the empirical specification, but at high computational costs. An avenue for future research is to determine an efficient methodology to implement unobserved heterogeneity in patient preferences.

## 8 Patterns And Determinants of Access to Care

I use the estimated model to study the distribution of access to care and what determinants contribute most to access loss. Before turning to theses results, I define access to care and access loss precisely.

Figure 9: Model Fit

(a) Comparison of Model to Data: Density




(b) Comparison of Model to Data: Across Cuts

| | | % Without Care | |
|---|---|---|---|
| Subset | N | Data | Model |
| *Area Type* | | | |
| Rural | 237,388 | 49.69 | 44.50 |
| Semiurban | 198,356 | 58.85 | 62.85 |
| Urban | 274,538 | 69.03 | 73.08 |
| *Age* | | | |
| 0-34 | 271,910 | 50.35 | 50.99 |
| 35-64 | 295,418 | 61.55 | 64.26 |
| 65+ | 142,954 | 73.76 | 71.69 |
| *Comorbidities* | | | |
| Yes | 572,588 | 55.81 | 57.29 |
| No | 137,694 | 75.99 | 74.73 |

## 8.1 Measures of Access to Care

I define access to care $A_{\theta t}$ for a type $\theta$ patient under choice conditions $(\mathcal{J}, \tau^u)$ as the share of patients that would have attained care under *full access choice conditions* that attain care under the conditions $(\mathcal{J}, \tau^u)$. Choice conditions include the set of all physicians in the patient's choice set $\mathcal{J}$, and a vector of effort costs to match with each physician $\tau^u$. The choice set $\mathcal{J}$ describes the characteristics and distances of each physician that the patient may match with.[22]

[22]I define a second measure of access to care as the share of patient surplus attained in a full access counterfactual that is attained in the present equilibrium.

$$A^2_{\theta t} = \frac{\text{logsum}_{\theta t}(\boldsymbol{\beta}, \boldsymbol{\tau}^u_\theta, \mathcal{J}_{\theta t})}{\text{logsum}_{\theta t}(\boldsymbol{\beta}, \mathbf{0}, \mathcal{J}_\theta{}^{FA})}$$

$$\text{logsum}_{\theta t}(\beta, \boldsymbol{\tau}_\theta, \mathcal{J}) = log(\sum_{j \in \{\mathcal{J}, \emptyset\}} exp(\delta_{\theta jt}(\beta) - \tau^u_{\theta jt})$$

Results are qualitatively similar when this alternative measure is used. However, since there is no normalizing coefficient in patient preferences, it is difficult to interpret comparisons of magnitudes of this measure of access across patient types. Section §E presents results for this measure, and others.

$$A_{\theta t} = \frac{\mathbb{P}_{\theta t}(\boldsymbol{\beta}, \tau^u_{\theta t}, \mathcal{J}_{\theta t})}{\mathbb{P}_{\theta t}(\boldsymbol{\beta}, \tau^{u,FA}, \mathcal{J}^{FA})}$$

$$\mathbb{P}_{\theta t}(\boldsymbol{\beta}, \tau^u_{\theta}, \mathcal{J}) = 1 - \Big[1 / \sum_{j \in \{\mathcal{J}_t, \emptyset\}} exp(\delta_{\theta jt}(\beta) - \tau^u_{\theta jt}(\beta))\Big]$$

where the probability that a patient attains no care, $\mathbb{P}_{\theta t}(\boldsymbol{\beta}, \tau^u, \mathcal{J})$, depends on preferences $\boldsymbol{\beta}$, effort costs $\tau^u$, and the choice set $\mathcal{J}$. $(\tau^u_{\theta t}, \mathcal{J}_{\theta t})$ defines the choice conditions of the estimated matching model in market $t$. $(\tau^{u,FA}, \mathcal{J}^{FA})$ defines the full access choice conditions.

The measure of access to care depends on how *full access* is defined. I assume that patients would have full access if they lived in Sudbury and every physician was willing to accept them. Sudbury is the largest city in my sample, with a population of 165,000 in 2014. Thus, in the main specification, I define the full access choice conditions to consist of the choice set of patients in Sudbury ($\mathcal{J}^{FA} = \mathcal{J}^{Sudbury}$) and an effort cost vector of zero ($\tau^{u,FA} = 0$).

For ease of exposition, I define access loss as the share of potential access that is not attained in the present equilibrium.

$$AL_{\theta t} = 1 - A_{\theta t}$$

## 8.2 Patterns in Access to Care

Access loss is large and unequally distributed. I find that 26.04% of patients who would attain care in a full access environment do not attain care in 2014. Figure 10 summarizes these results. For more detailed statistics, refer to table 16.

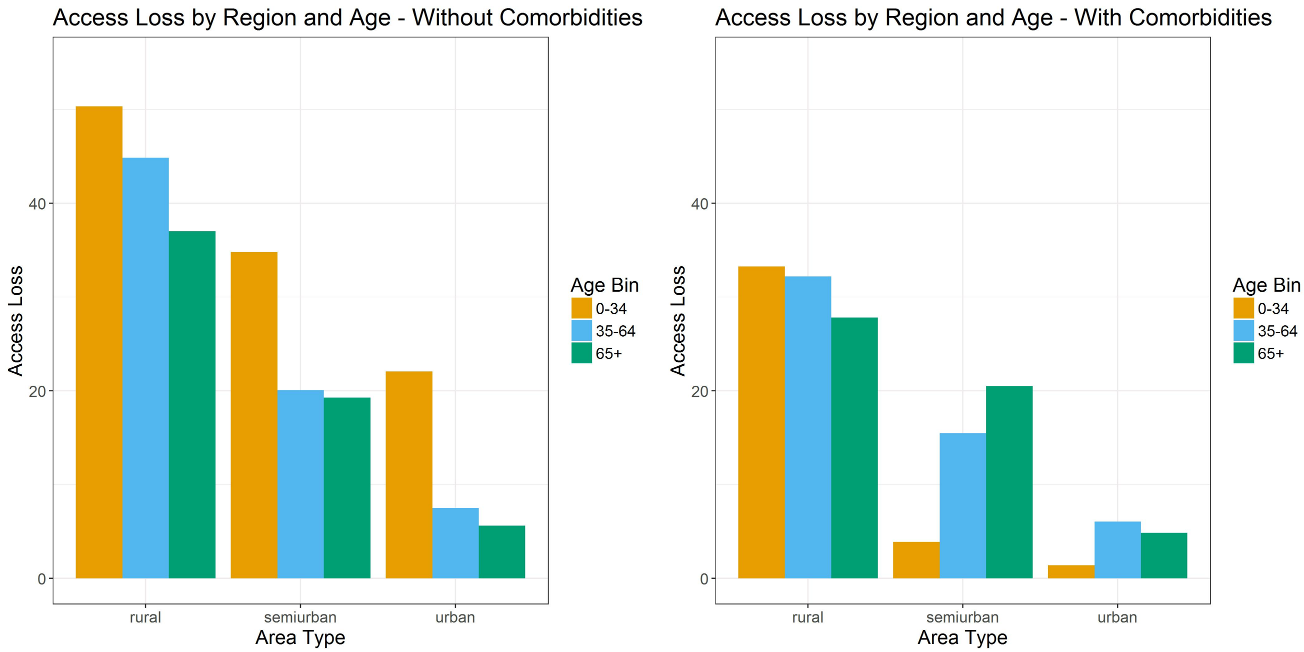
**Age and Comorbidities** Access loss is lower for older patients and patients with comorbidities. Access loss for patients without comorbidities and aged 0-34 is 34.65%. This compares to 17.17% for 65+ patients with comorbidities.

As discussed above, physicians discriminate in favor of patients with greater expected utilization. This discrimination decreases the effort that sicker and older patients must expend to attain care. In turn, this increases the effort that healthier and younger patients must expend. I find that patients without comorbidities face an average effort cost that is 11.07% higher than the average effort cost of patients with comorbidities. This discrimination results in lower access loss for sicker and older patients and higher access loss for healthier and younger patients.

Within the subset of patients who have comorbidities, older patients do not always have greater access to care than younger patients. This also derives from physician discrimination. Conditional on having comorbidities, younger patients still have high expected utilization. Additionally, physicians discriminate in favor of younger patients, all else equal.

Figure 10: Distribution of Access in 2014

(a) Healthier Patients (b) Sicker Patients




**Geography** Access loss is higher in more rural areas. Access loss is 43.04% in rural areas and 24.79% in semiurban areas, which are non-metropolitan agglomerations with a population of 10,000 or more. In comparison, urban areas have low levels of access loss at 12.25%. Pockets of extremely low access to care exist, where I estimate that access loss is over 90%. In such areas, the nearest primary care physician can be further than 60 km away.

## 8.3 Determinants of Access Loss

I decompose access loss for each patient type into potential determinants. I conduct two decompositions of access loss. First, access loss is decomposed into the amount of access lost due to capacity constraints and the amount of access lost due to an insufficient choice set. Second, access loss is decomposed into four aspects of the physician supply: the distribution of physician characteristics among existing physicians, the geographic distribution of physicians, the aggregate supply of physicians, and remaining access loss.

### 8.3.1 Decomposition One: Effort Costs and Insufficient Choice Sets

Access access loss is decomposed into the amount of access lost due to capacity constraints and the amount of access lost due to an insufficient choice set. The impact of capacity constraints on access loss can be interpreted as the access loss caused by patients expending effort to attain care. Since physicians are capacity constrained, patients must expend effort to match with a physician. The impact of capacity constraints on access loss is calculated as the difference in access loss between the present equilibrium and a counterfactual equilibrium with no effort.

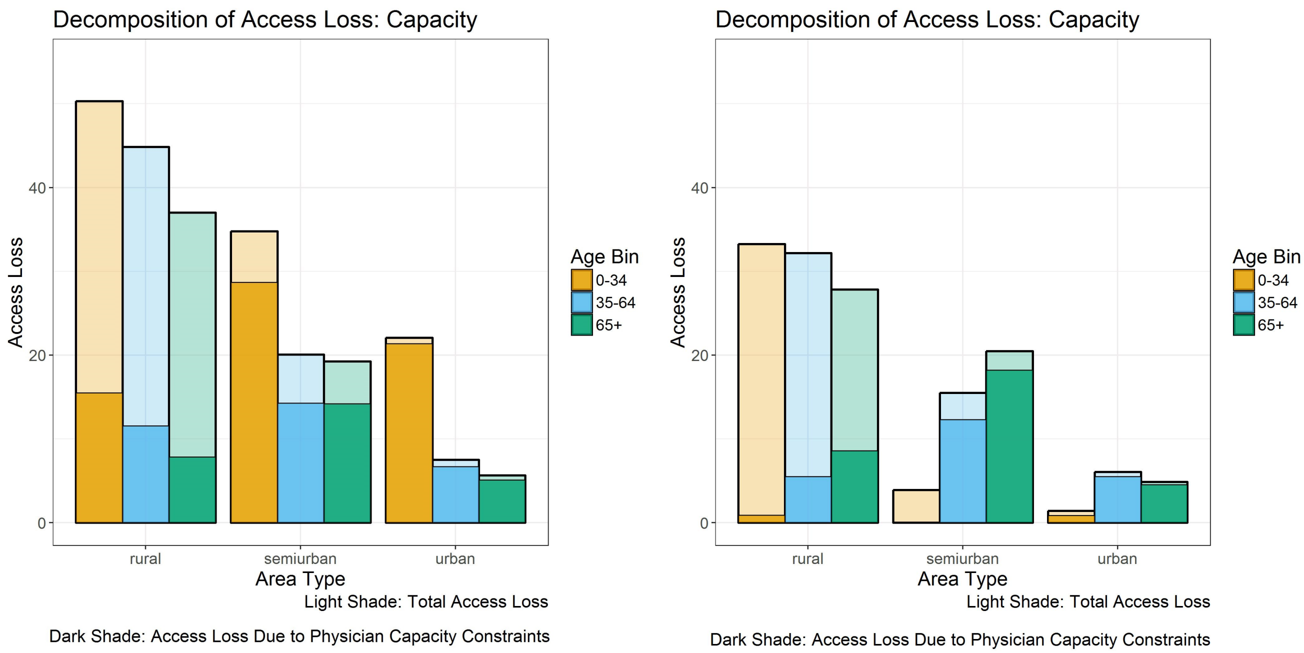
The remaining access loss is explained by the difference between the full access choice set and the current choice set. For patients living in Sudbury, this effect will be zero by construction. For others, it is the difference in access between their current choice set and the choice set in Sudbury, assuming that all physicians are accepting patients.

$$AL_{\theta t} = \underbrace{\frac{\mathbb{P}_{\theta t}(\boldsymbol{\beta},\mathbf{0},\mathcal{J}^{FA}) - \mathbb{P}_{\theta t}(\boldsymbol{\beta},\mathbf{0},\mathcal{J}_t)}{\mathbb{P}_{\theta t}(\boldsymbol{\beta},\mathbf{0},\mathcal{J}^{FA})}}_{\text{Choice Set}} + \underbrace{\frac{\mathbb{P}_{\theta t}(\boldsymbol{\beta},\mathbf{0},\mathcal{J}_t) - \mathbb{P}_{\theta t}(\boldsymbol{\beta},\boldsymbol{\tau}_{\theta t},\mathcal{J}_t)}{\mathbb{P}_{\theta t}(\boldsymbol{\beta},\mathbf{0},\mathcal{J}^{FA})}}_{\text{Capacity}}$$

Figure 11 summarizes the results of this decomposition. Table 16 provides more information. Physician capacity constraints contribute an average access loss of 13.80 percentage points. Determinants of access loss differ by patient type. Rural access loss is primarily driven by choice sets (31.47pp), rather than capacity constraints (11.57pp). Although patients must expend effort to match with physicians, the distance they must travel to the nearest doctor is a larger contributor to low access to care. In urban and semiurban areas, capacity constraints are the main drivers of access loss.

Figure 11: Share of Access Loss Attributed to Physician Capacities

(a) Healthier Patients (b) Sicker Patients




These results indicate that policy remedies should be calibrated to specific populations. Where capacity constraints are a main contributor, policies should aim to increase capacity. Either expanding the capacity of existing physicians or encouraging the entry of new physicians would suffice. When poor choice conditions is the primary determinant of access to care, policies should target entry of new physicians to areas with especially low access.

### 8.3.2 Decomposition Two: Physician Supply

In a second decomposition, I study how different aspects of physician supply across Northern Ontario contribute to access to care. To do so, I iteratively estimate counterfactual equilibria where one aspect of physician supply is changed at a time. First, I randomly redistribute physician characteristics, holding physician locations steady. Second, I randomly distribute physicians to make the ex-ante physician to population ratio equal across locations. Third, I add physicians such that there is one physician per 1000 patients in all locations. For each counterfactual, I report the average access loss over 10 simulated equilibria. I compare and discuss the change in access loss for each new counterfactual. Figure 12 presents the results and table 16 provides more detail.

I find that urban physicians benefit from the distribution of physician characteristics, mainly because of larger physician capacities in urban areas. In contrast, rural areas benefit from the distribution of physicians. Indeed, the physician to population ratio is larger in rural areas than urban areas. Although the aggregate supply of physicians contributes to the access loss of all patients, young urban and semiurban patients are the most affected. When supply is sufficient, physicians do not hit their capacity constraints and therefore have no need to discriminate against patients with low expected utilization. Thus, the low aggregate supply of physicians contributes especially to those discriminated-against patients: young and healthy urban patients.

**Distribution of Physician Characteristics** Physicians are heterogeneous in capacities, payment models, and other characteristics. This affects the spatial distribution of access loss. To determine the effect of the distribution of physician characteristics on access loss, I simulate counterfactual equilibria where the number of physicians in each location is kept the same, but physicians are randomly reallocated across locations. Comparing the average access loss across simulations to access loss in the present equilibrium provides an estimate of the impact of the distribution of physician characteristics on access loss.

I find that younger and more urban patients gain from the distribution of physician characteristics. The likely cause of this phenomenon is that physicians in urban areas have larger capacities than physicians in rural areas (mean panel size is 1,329 in urban areas; 1,246 in semi-urban, and 818 in rural areas). Interestingly, while the positive impact on urban patients is substantial, the negative impact on rural patients is small. In many rural areas, there are few potential patients. In these areas, physicians with large capacities would not be able to attract enough patients to hit their capacity constraint. Thus, adding high capacity physicians to rural areas does not significantly increase access. For this reason, the distribution of physician characteristics decreases aggregate access loss by 3.16pp.

**Distribution of Physician Locations** To estimate the impact of the distribution of physician locations on access loss, I simulate counterfactuals where physicians are randomly reassigned locations. The probability that a physician is assigned a location is proportional to the population of the location. If a physician is assigned to a location, the probability of the next physician being assigned that location decreases as if its population decreased by 1000. Simulations are averaged to attain the final estimates.

I find that the distribution of physician locations benefits rural patients. This result is unsurprising given the observed physician to population ratios. In the main sample, there are 0.96 physicians per 1000 patients in rural areas, 0.81 in semiurban areas, and 0.73 in urban areas. The effect of the physician location distribution is essentially zero-sum. Whereas rural access loss decreases, urban and semiurban access loss increases. In sum, the distribution of physician locations decreases aggregate access loss by 0.36pp.

**Aggregate Supply of Physicians** I assess the impact of the aggregate supply of physicians on access loss by simulating counterfactuals where each location is assigned a physician per 1000 residents.[23] Physicians are randomly sampled from the existing pool of physicians.

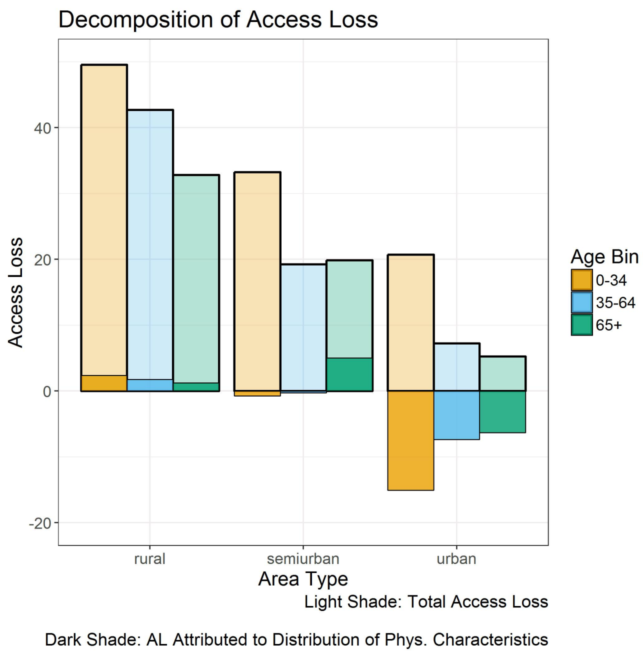
This effect is large for all patients. I estimate that if the supply increased to 1 physician per 1000 residents, aggregate access loss would decrease by 20.64pp (from the counterfactual with an equal distribution of physicians). Access to care would shrink to almost zero in urban and semiurban areas. Young urban and semiurban patients benefit the most from expanding supply. These patients are forced out of the market due to physician discrimination. When supply expands, physicians are no longer at capacity and therefore accept all patients.

**Remainder** After accounting for the distribution of physicians and the aggregate supply of physicians, some access loss remains. This remaining access loss is due to low levels of physician variety in sparsely populated areas. Even with a sufficient supply of physicians, rural patients are restricted to choose among the few physicians who are close to them. Those physicians may have low capacity, or they may not produce high match value with the patients. This remaining access loss is fairly large in rural areas, but is close to zero in urban areas.

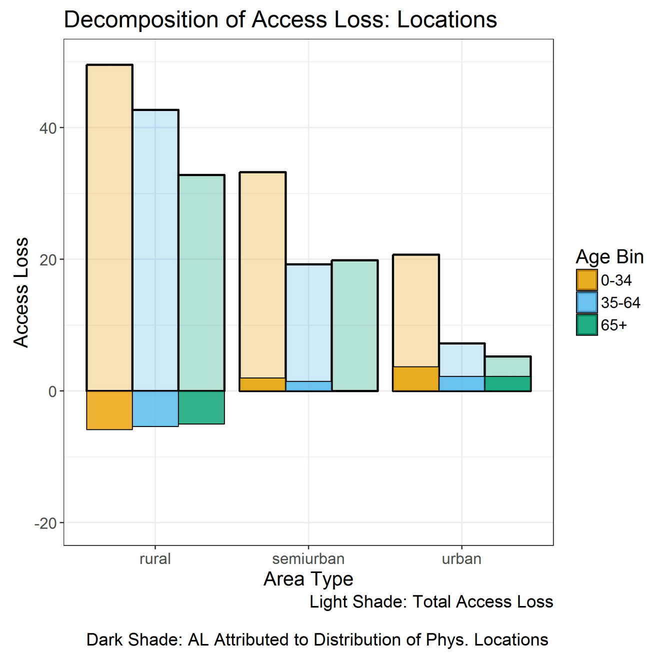
[23] An additional physician is assigned with probability $\frac{R}{1000}$ where R is the remainder after population is divided by 1000.

Figure 12: Access Loss Attributed to Characteristics of Physician Supply

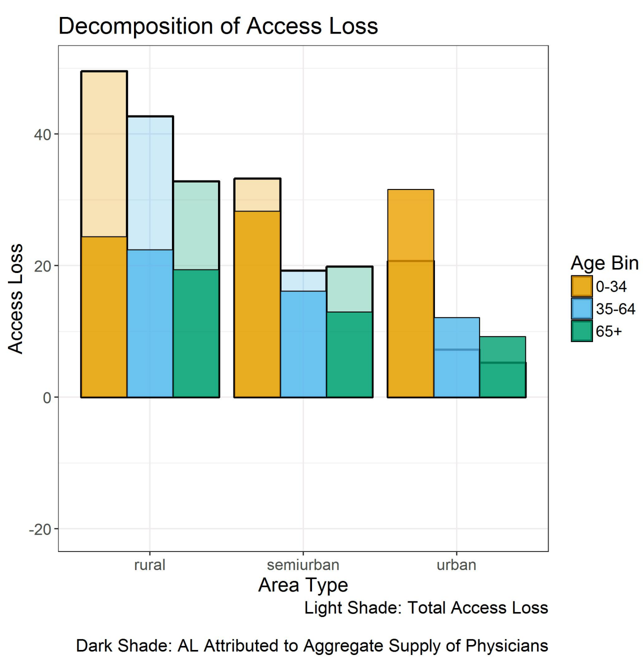
(a) Physician Characteristics Distribution



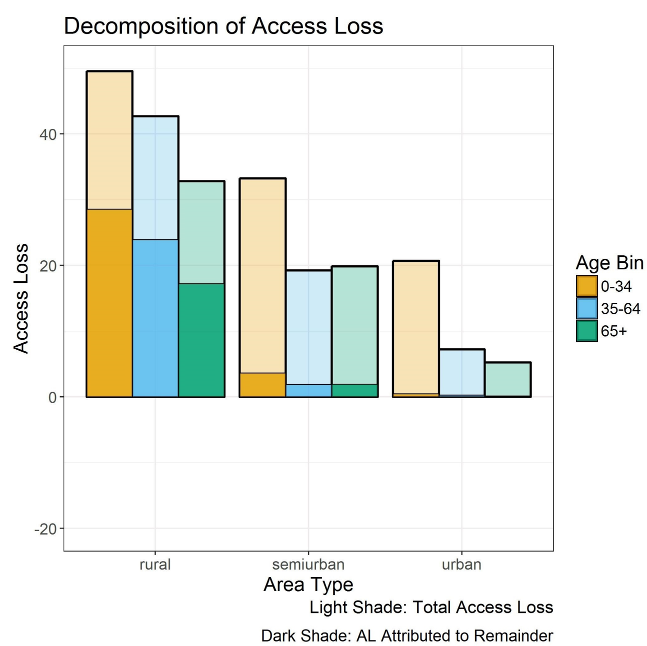

(b) Physician Location Distribution




(c) Aggregate Supply




(d) Remainder



# 9 Policy Implications

In this section, I use the estimated model to study two policies. First, I analyze grants to physicians who practice in low access locations. Second, I study the impact of alternative payment models on access to care.

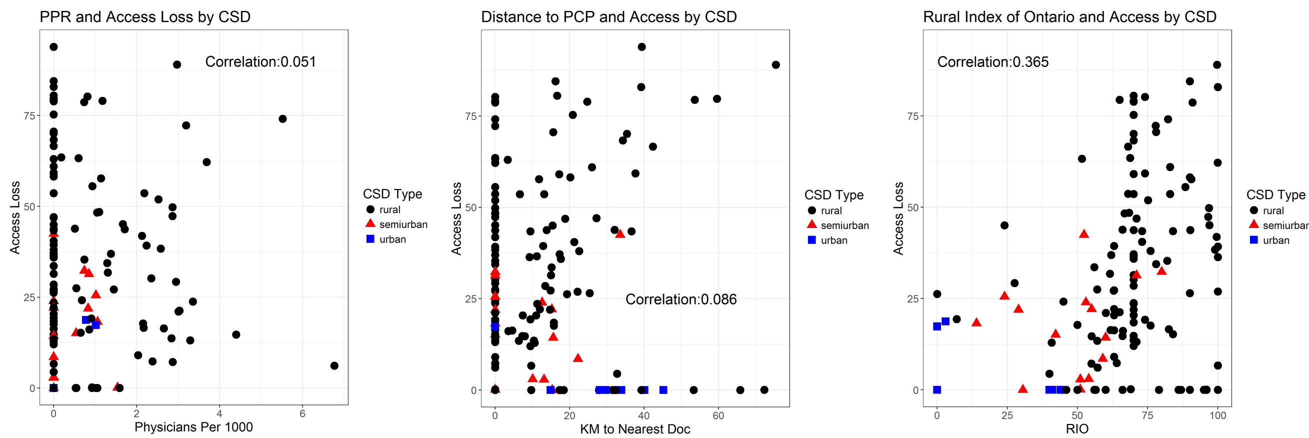
## 9.1 Location Incentive Grants

Ontario and other governments spend large sums to encourage physician entry into areas with low access to care. In Ontario, physicians can receive a grant of up to $117,600 to practice in low access areas (see section §C for details). Grant levels depend on a proxy of access, the Rural Index of Ontario (RIO). The Rural Index of Ontario is a weighted average of population, population density, and distance to the nearest physician. In the United States, an similar program that issues loan repayments and scholarships for health professionals who practice in low access areas, the National Health Services Corps (NHSC), has a budget of $310 million. The NHSC uses a proxy of access that is a weighted average of physician to population ratio ($\frac{2}{5}$ weight), a measure of poverty, an infant health index, and travel time to nearest doctor ($\frac{1}{5}$weight each).

First, I study how well proxies that governments use to distribute funds correlate with access to care. I find that physician to population ratios perform poorly, but the Rural Index of Ontario performs reasonably well. Figure 13 shows the relationship between average access loss and three measures: physician to population ratio, distance to nearest physician, and the Rural Index of Ontario (RIO). I find that physician to population ratios, when calculated at the census subdivision level, perform poorly at predicting access to care ($R^2$: 0.062, Correlation: .051). Physician to population ratios do not account for heterogeneity in physician capacity or the proximity of physicians outside of the census subdivision boundaries. Proxies for access to care that account for these factors perform better in predicting access to care. Distance to the nearest physician has a higher $R^2$ of .205, but still has a low correlation of .086. Encouragingly, the measure used by Ontario, RIO, seems to perform best ($R^2$ : 0.231, Correlation: 0.365). These results suggest that the use of physician to population ratios to distribute funds by the National Health Service Corps should be revisited. The Rural Index of Ontario is a good benchmark.

Figure 13: Relationship Between Estimated Access to Care and Common Measures

(a) Physician to Population Ratio (b) Distance To PCP (c) RIO




Second, I study whether these grant programs are justified. I find that the business stealing effect generates an inefficiently large incentive for physicians to enter into low access, high population areas. Ontario's grant program helps to counteract that effect.

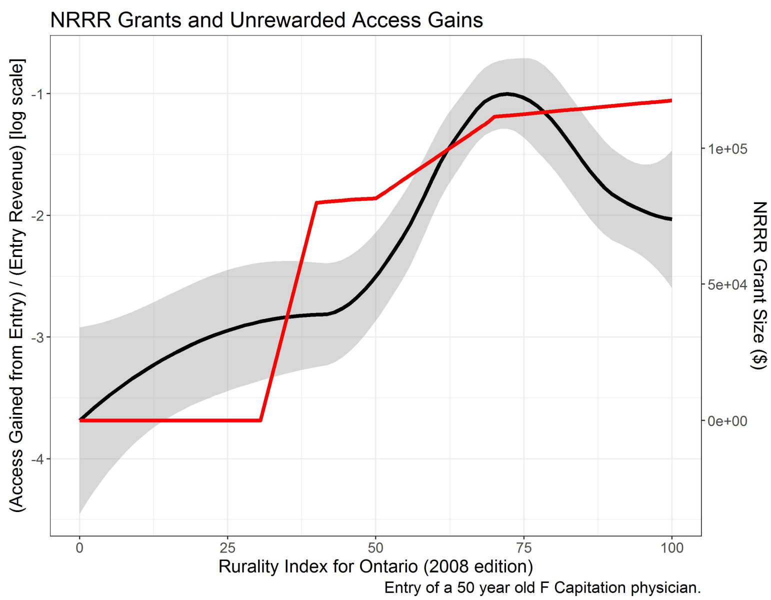
When physicians enter into a low access area, they are generally rewarded with high revenues. In those areas, demand is high and competition is low. In contrast, entry into high access areas generates low revenue. This relationship improves access under mild conditions. Physicians are attracted to areas where they can attain high revenues, thus attracting physicians to low access areas.

However, this relationship breaks down due to the business stealing effect. In areas with high access to care and high population density, physicians can attain high revenues at entry by stealing patients from incumbent physicians rather than attracting new patients who previously did not receive care. This effect works through the effort mechanism. After entry, incumbent physicians must expend more effort to attract patients. In response, individual physician supply decreases. The population is large enough, however, that the entrant physician can still attain high revenues. This may result in higher entry into high access urban areas than what is socially desirable (assuming access is a social objective). This suggests that there is scope for entry incentives to increase social welfare.

To show this, I estimate the impact of an illustrative physician entering into each census subdivision on access to care. Additionally, I estimate the revenue (in 2004 dollars) that the illustrative physician would attain if she entered in each census subdivision in 2014. As in previous analyses, the illustrative physician in is a 50 year old female physician in the capitation payment model who does not offer walk in visits and has a capacity of 1000 patients.

Figure 14 illustrates the phenomenon. The black curve represents the relationship between the Rural Index for Ontario and the ratio of the increase in access due to an entry and the revenue the entrant physician attains.[24] If access is the social objective and physicians are responsive to financial incentives at entry, it is clear that shifting entry incentives from low RIO areas to high RIO areas would be socially beneficial. I find that the grant provided by the Ontario government successfully incentivizes entry into areas with a large ratio of access gains to entry revenue. In figure 14, the grant size is represented in red.

Figure 14: Entry and Access




## 9.2 Alternative Payment Models

As described in 4.1, from 2002 to 2006, Ontario introduced voluntary capitation and enhanced fee-for-service models for primary care physicians to supplement the traditional fee-for-service payment model. Increasing access to care was a goal of these reforms. I find that the alternative payment models are successful at increasing access to care, primarily by incentivizing physicians to increase the number of patients they accept.

The alternative payment models affect access to care through physician incentives. First, alternative payment models incentivize physicians to increase the number patients they accept. Revenue per patient is higher in the alternative models than in the fee-for-service model. Physicians are therefore incentivized to increase patient panels either by working longer hours or changing practice styles. Second, alternative payment models incentivize physician to select patients on characteristics. Enhanced fee-for service physicians are incentivized to select patients who attain

[24]The line represents the local polynomial regression mean. The accompanying scatter plot, figure 18b, is in the appendix.

"in-basket" services, while capitation provides incentives to select physicians with low demand for services. The selection incentives for capitation physicians are large. Capitation payments are only risk adjusted by age in 5 year bins and sex.

The literature studying the effect of the Ontario payment reforms on Physician behavior is large and mixed. In an early analysis, Devlin and Sarma (2008) find the surprising result that physicians with high number of visits per week were attracted to the alternative payment models, but decreased the number of visits per week after adopting the new payment model. Kantarevic et al. (2011) find that physicians who switch to an enhanced fee-for-service model increase the number of patients they accept. Rudoler et al. (2016) find no selection of patients based on risk (expected utilization).

The estimated physician preferences (table 6) imply that physicians are more likely to accept patients when they are in an alternative payment model. Physicians are estimated to decrease the log odds of leaving a panel space open by .504 in the capitation model and by .555 in the enhanced fee-for-service model, relative to the fee-for-service model. Additionally, I find a small degree of selection of patients based on potential revenue.

Reduced form exercises confirm these patterns. Using a two-way fixed effect model with physician-year level panel data, I estimate that physicians in an alternative payment model increase the number of patients on their roster relative to when they are in the fee-for-service model. The magnitudes of these estimates (6.6% increase in capitation and 16.0% in enhanced fee-for-service) are significant. Additionally, fixed effect model results suggest that physicians have mildly different matching patterns in capitation models than in fee-for-service and enhanced fee-for-service models, perhaps due to physician selection of patients on expected revenue. Section §H describes this analysis in more detail.

Neither the previous literature nor the above reduced form work accounts for equilibrium effects. For example, when an individual physician changes to an alternative payment model, they may increase the number of patients that they accept. However, other physicians may respond by decreasing their own panel sizes. Therefore, to determine the effect of the alternative payment model on the market as a whole, I estimate a counterfactual equilibrium where the alternative payment models are never introduced. Counterfactual equilibrium outcomes are then compared with the current equilibrium outcomes.

Specifically, I estimate the impact of the alternative payment model on the primary care market in 2014 by simulating a counterfactual where all physicians are forced into the fee-for-service payment model. I keep patient preferences over physicians stable. I find that the alternative payment model increased access by 5.00 percentage points (pp) across the entire population.

The alternative payment models have heterogeneous affects on different types of patients. I find that the policy had a greater impact on the access of healthier and younger patients. Additionally, I find that the alternative payment models decrease urban and semiurban access loss (6.22pp, 5.02pp) more than rural area access loss (3.50pp).[25] As discussed in section 8.3, access loss in urban areas is primarily driven by capacity constraints, whereas access loss in rural areas are driven by limited choice sets. Therefore, a policy that primarily increases the number of patients that each physician

[25] In the appendix, figure 19 shows the estimated change in the distribution of access to care.

accepts will favor urban access. However, it should be noted that this analysis does not take into account changes in entry and exit. Alternative payment models could change entry and exit decisions in rural areas by changing the distribution of practice profitability across locations.

The alternative payment models primarily increase access to care by increasing physician propensity to accept patients. To determine the magnitude of this effect, I estimate an intermediate counterfactual where the expected revenue of patients $R_{\theta jt}$ is not adjusted from their fee-for-service levels. The latent utility a physician attains from leaving a panel space open, however, is adjusted to reflect the physician's participation in an alternative payment model. I find that access to care increases by 4.74pp in the intermediate counterfactual relative to the all fee-for-service counterfactual. Selection of patients accounts for the remaining 0.26pp of the access gains.

## 10 Conclusion

This paper studies access to care as an equilibrium output of a matching market between patients and physicians. In the model, the market is cleared by a non-price mechanism: the effort it takes for a patient (physician) to match with a physician (patient). Patient and physician preferences are estimated using data from the Northern Ontario primary care market. Using the estimated preferences, I study the distribution and determinants of access to care and the implications of two policies on access to care.

I find that access to care is low and unevenly distributed across types of patients. Further, I find that the determinants of low access to care differ for different patient types. In rural areas, low access to care is mostly driven by the high travel costs that patients must pay to attain care. In urban areas, low access to care is driven by physician capacities. Since physicians do not have enough capacity to accept all patients, patients with low willingness to expend effort and patients who are discriminated against by physicians are unable to attain care. These patients tend to be to be younger and healthier.

Differences in the determinants of access to care across regions have policy implications. In urban areas, policies should aim to increase physician capacity. In rural areas, policies should be more targeted. Entry of new physicians into areas that currently do not have physicians is the most effective way to increase access. Further, market incentives do not always align with increasing access. Physicians may be able to attain high revenues in densely populated areas without increasing access, via the market stealing effect. Thus, there is scope for grants that provides additional income to physicians who locate in areas with the highest potential to increase access to care.

Understanding the effectiveness of such grants in increasing access to care is important. Ontario, for example, provides a grant of up to \$117,600 for primary care physicians who begin practice in low access areas. A similar program in the United States has a budget of \$310 million. In this paper, I discuss the effectiveness of different proxies for inaccess to care that are used to distribute these funds. In future research, more can be done. To determine how effective the incentive grants are at increasing access to care, I must determine how physician entry and exit decisions are changed by the incentive grants. Once entry and exit decisions are understood, the impact of incentive grants in increasing access to care be estimated. Importantly, I would then be able to compare the effectiveness of incentive grants to other policy remedies, such as the introduction of alternative physician payment models.

In this paper, I estimate the impact of alternative payment models on access to care by estimating a counterfactual equilibrium where all physicians are returned to a fee-for-service model. I find that the alternative payment model regime increases access to care by 5 percentage points. I estimate that this impact is primarily driven by an incentive for physicians to increase the number of patients they accept. Physicians are estimated to be mildly responsive to revenue when selecting patients.

I conclude by identifying important avenues for future work. This paper is the first to my knowledge to measure access to care as an output of an equilibrium model. It does so under a purely positive framework. Using the estimated model, I am able to discuss the distribution, determinants, and the impact of policy remedies on a well-defined definition of access to care. However, I am unable to make judgments about which policies are *best.* For such normative statements to be made, I must determine the relative value to society of an additional unit of access for different types of patients.

Once a normative framework is established, the applications of the model expand. The model could then be used to design optimal policy and regulation. In Ontario, the model could be used to estimate the optimal spatial distribution of physician. Further afield, a model of access to care in equilibrium could inform the value to society of different network adequacy rules for insurance in the United States. These questions are beyond the scope of this paper. Yet, they are of primary importance for policymakers seeking to increase access to care and for the patients who would benefit.

# References

Hiroyuki Adachi. A search model of two-sided matching under nontransferable utility. *Journal of Economic Theory*, 113(2):182–198, 2003.

Nikhil Agarwal. *Essays in Empirical Matching: Chapter 3.* PhD thesis, Princeton, 2013.

Nikhil Agarwal. An empirical model of the medical match. *American Economic Review*, 105(7): 1939–78, 2015.

Diane Alexander. How do doctors respond to incentives? unintended consequences of paying doctors to reduce costs. *Journal of Political Economy*, 128(11), 2020.

Diane Alexander and Molly Schnell. Closing the gap: The impact of the medicaid primary care rate increase on access and health. Federal Reserve Bank of Chicago Working Paper 2017-10, 2018.

Eduardo M Azevedo and Jacob D Leshno. A supply and demand framework for two-sided matching markets. *Journal of Political Economy*, 124(5):1235–1268, 2016.

David Benson. *Lemon Dropping: Do Physicians Respond to Incentives?* PhD thesis, Northwestern University, 2018.

Kevin Bernstein. Family medicine in canada. Blog post, April 2013. URL http://futureoffamilymedicine.blogspot.com/2013/04/family-medicine-in-canada.html.

Donald Boyd, Hamilton Lankford, Susanna Loeb, and James Wyckoff. Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics*, 31(1):83–117, 2013.

Nele Brusselaers and Jesper Lagergren. The charlson comorbidity index in registry-based research: Which version to use? *Methods of Information in Medicine*, 56(5):401–406, 2017.

Gioia Buckley, Philip DeCicca, Jeremiah Hurley, and Jinhu Li. The response of ontario primary care physicians to pay-for-performance incentives. Working Paper 11-02, Centre for Health Economics and Policy Analysis, 2011.

Gioia Buckley, Philip DeCicca, Jeremiah Hurley, and Jinhu Li. Physician response to pay-for-performance: Evidence from a natural experiment. *Health Economics*, 23(8):962–978, 2014a.

Gioia Buckley, Philip DeCicca, Jeremiah Hurley, and Jinhu Li. Physician response to pay-for-performance: Evidence from a natural experiment. *Health Economics*, 23:8, 2014b.

Canadian Paediatric Society Public Education Subcommittee. Paediatricians in canada: Frequently asked questions. *Paediatric Child Health*, 9(6):431–432, 2004.

David C Chan and Michael J Dickstein. Industry Input in Policy Making: Evidence from Medicare*. *The Quarterly Journal of Economics*, 134(3):1299–1342, 04 2019.

Alice Chen. Do the poor benefit from more generous medicaid physician payments. Working paper, 2014.

Pierre-André Chiappori and Bernard Salanié. The econometrics of matching models. *Journal of Economic Literature*, 54(3):832–61, 2016.

Eugene Choo and Aloysius Siow. Who marries whom and why. *Journal of Political Economy*, 114 (1):175–201, 2006.

Jeffrey Clemens and Joshua D. Gottlieb. In the Shadow of a Giant: Medicare's Influence on Private Physician Payments. *Journal of Political Economy*, 125(1):1–39, 2017.

CMS. National health expenditures survey, table 7 physician and clinical services expenditures; aggregate and per capita amounts, percent distribution and annual percent change by source of funds: Calendar years 2012-2028. Data table, 2019.

College of Nurses of Ontario. Membership statistics report 2008. 2008.

Committee on GAFs. Evidence of geographic variation in access, quality, and workforce distribution. In *Geographic Adjustment in Medicare Payment: Phase II*. National Academies Press, 2011.

Christopher T Conlon and Julie Holland Mortimer. Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, 5(4):1–30, 2013.

Price Waterhouse Coopers. Evaluation of primary care reform pilots in ontario phase 2 interim report. Technical report, 2001.

André de Palma, Nathalie Picard, and Paul Waddell. Discrete choice models with capacity constraints: An empirical analysis of the housing market of the greater paris region. *Journal of Urban Economics*, 62(2):204–230, 2007.

Rose Anne Devlin and Sisira Sarma. Do physician remuneration schemes matter? the case of canadian family physicians. *Journal of Health Economics*, 27(5):1168–1181, 2008.

Christopher Frank and C. Ruth Wilson. Models of primary care for frail patients. *Canaidan Family Physician*, 61(7):601–606, 2015.

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

Alfred Galichon and Yu-Wei Hsieh. Aggregate stable matching wth money burning. (2887732), 2019.

Alfred Galichon and Bernard Salanié. Matching with trade-offs: Revealed preferences over competing characteristics. CEPR Discussion Paper 7858, 2010.

Gloria Galloway. The soul-destroying search for a family doctor. *The Globe and Mail*, Aug 2011.

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

Ana Gazmuri. School segregation in the presence of student sorting and cream-skimming: Evidence from a school voucher reform. Job market paper, 2019.

Richard Glazier, Michael Green, Eliot Frymire, Alex Kopp, William Hogg, Kamila Premji, and Tara Kiran. Do incentive payments reward the wrong providers? a study of primary care reform in ontario, canada. *Health Affairs*, 38(4):624–632, 2019.

Michelle Sovinsky Goeree. Limited information and advertising in the us personal computer industry. *Econometrica*, 76(5):1017–1074, 2008.

Hugh Gravelle and Luigi Siciliani. Is waiting-time prioritisation welfare improving? *Health Economics*, 17(2):167–184, 2008.

David Gray, William Hogg, Michael Green, and Yan Zhang. Did family physicians who opted into a new payment model receive an offer they should not refuse?: Experimental evidence from ontario. *Canadian Public Policy*, 41(2):151–165, June 2015.

M Green, P Gozdyra, E Frymire, and R Glazier. Geographic variation in the supply and distribution of comprehensive primary care physicians in ontario. Report, Institute for Clinical Evaluative Sciences, 2017.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY, 2009.

Roberta Heale and Marilyn Butcher. Canada's first nurse practitioner-led clinic: A case study in healthcare innovation. *Nursing Leadership (Toronto, Ont.)*, 23(3):21–29, 2010.

Health Quality Ontario. Measuring up 2018: A yearly report on how ontario's health system is performing. Technical report, Health Quality Ontario, 2018.

Heike Hennig-Schmidt, Reinhard Selten, and Daniel Wiesen. How payment systems affect physicians' provision behaviour: An experimental investigation. *Journal of Health Economics*, 30(11):637–646, 2011.

David Henry, Susan Schultz, Richard Glazier, Sacha Bharria, Irfan Dhalla, and Andreas Laupacis. Payments to ontario physicians from ministry of health and long-term care sources 1992/92 to 2009/10. Technical report, ICES Investigative Report, Toronto, 2012.

Gunter J Hitsch, Ali Hortaçsu, and Dan Ariely. Matching and sorting in online dating. *American Economic Review*, 100(1):130–63, 2010.

Yu-Wei Hsieh. *Econometric Analysis of Two-Sided Matching Markets*. PhD thesis, New York University, 2012.

Jeremiah Hurley, Jinhu Li, Philip DeCicca, and Gioia Buckley. Physician response to pay-for-performance: Evidence from a natural experiment. *Health Economics*, 23:8, 2013.

Brian Hutchison and Richard Glazier. Ontarios primary care reforms have transformed the local care landscape, but a plan is needed for ongoing improvement. *Health Affairs*, 32:4, 2013.

Brian Hutchison, Julia Abelson, and John Lavis. Primary care in canada: So much innovation, so little change. *Health Affairs*, 20:3, 2001.

ICES. Data dictionary. Data dictionary, 2020. URL https://www.ices.on.ca/Data-and-Privacy/ICES-data/Data-dictionary.

Tor Iversen and Luigi Siciliani. Non-price rationing and waiting times. In *The Oxford Handbook of Health Economics*. Oxford University Press, 2011.

Nanak Kakwani, Adam Wagstaff, and Eddy Van Doorslaer. Socioeconomic inequalities in health: measurement, computation, and statistical inference. *Journal of econometrics*, 77(1):87–103, 1997.

Jasmin Kantarevic and Boris Kralj. Risk selection and cost shifting in a prospective physician payment system: Evidence from ontario. *Healthy Policy*, 115(2-3):249–257, 2014.

Jasmin Kantarevic, Boris Kralj, and Darrel Weinkauf. Enhanced fee-for-service model and physician productivity: Evidence from family health groups in ontario. *Journal of health economics*, 30(1): 99–111, 2011.

John Kautter, Gregory C Pope, Melvin Ingber, Sara Freeman, Lindsey Patterson, Michael Cohen, and Patricia Keenan. The hhs-hcc risk adjustment model for individual and small group markets under the affordable care act. *Medicare & Medicaid research review*, 4(3), 2014.

Yeeun Kim, Young-Ji Byon, and Hwasoo Yeo. Enhancing healthcare accessibility measurements using gis: A case study in seoul, korea. *PLoS One*, 13(2), 2018.

Tara Kiran, Rahim Moineddin, Alexander Kopp, Eliot Frymire, and Richard H Glazier. Emergency department use and enrollment in a medical home providing after-hours care. *The Annals of Family Medicine*, 16(5):419–427, 2018.

Boris Kralj. Measuring rurality - rio2008 basic: Methodology and results. Technical report, OMA Economics Department, 2009.

Sanghoon Lee. Ability sorting and consumer city. *Journal of Urban Economics*, 68(1):20–33, 2010.

Wei Luo and Yi Qi. An enhanced two-step floating catchment area (e2sfca) method for measuring spatial accessibility to primary care physicians. *Health & place*, 15(4):1100–1107, 2009.

Wei Luo and Fahui Wang. Measures of spatial accessibility to health care in a gis environment: Synthesis and a case study in the chicago region. *Environment and Planning B: Planning and Design*, 30(6):865–884, 2003.

Gregory Marchildon. *Health Systems in Transition: Canada. 2*. University of Toronto Press, Toronto, 2013.

Egor Matveyev. How do firms and directors choose each other? evidence from a two-sided matching model of the director labor market. Job market paper, 2013.

Matthew McGrail and John Humphreys. Spatial access disparities to primary health care in rural and remote australia. *Geospatial Health*, 10:138–143, 2015.

Thomas McGuire. Physician agency. In *The Oxford Handbook of Health Economics*. Oxford University Press, 2000.

Natalie Mehra. Eroding public medicare: Lessons and consequences of for-profit health care across canada. Technical report, Ontario Health Coalition, 2008.

Natalie Mehra. Private clinics and the threat to public medicare in canada results of surveys with private clinics and patients. Technical report, Ontario Health Coalition, 2017.

Konrad Menzel. Large matching markets as two-sided demand systems. *Econometrica*, 83(3): 897–941, 2015.

Ministry of Health and Long-Term Care. Billing and payment information for family health group (fhg) signatory physicians. Billing guide, Ministry of Health and Long-Term Care, 2007a.

Ministry of Health and Long-Term Care. Family health organization (fho) fact sheet. Billing guide, Ministry of Health and Long-Term Care, 2007b.

Ministry of Health and Long-Term Care. Billing and payment guide for family health organization (fho) physicians. Billing guide, 2011.

Ministry of Health and Long-Term Care. Billing and payment guide for family health organization (fho) physicians opting for solo payment. Billing guide, Ministry of Health and Long-Term Care, 2014.

Ministry of Health and Long-Term Care. Healthforceontario northern and rural recruitment and retention initiative guidelines. Technical report, 2017. URL http://www.health.gov.on.ca/en/pro/programs/northernhealth/nrrr.aspx.

Tracy M. Mroz, Davis G. Patterson, and Bianca K. Frogner. The impact of medicare's rural add-on payments on supply of home health agencies serving rural counties. *Health Affairs*, 39(6):949–957, 2020.

Joseph P Newhouse, Albert P Williams, Bruce W Bennett, and William B Schwartz. Does the geographical distribution of physicians reflect market failure? *The Bell Journal of Economics*, pages 493–505, 1982.

News Staff. Only one in ten family doctors accepting new patients in ontario. *CityNews*, 2006.

NPAO. Npao faqs. Website, Nurse Practitioners' Association of Ontario, 2020. URL https://npao.org/about-npao/npao-faqs/.

OMA. Common billing codes 2015. Billing guide, Ontario Medical Association, 2015.

Ontario Legislature. Commitment to the future of medicare act, 2004, s.o. 2004, c. 5. Law, 2004.

Valérie Paris, Marion Devaux, and Lihan Wei. Health systems institutional characteristics: a survey of 29 oecd countries. Working paper, 2010.

Mohammad Habibullah Pulok, Kees van Gool, Mohammad Hajizadeh, Sara Allin, and Jane Hall. Measuring horizontal inequity in healthcare utilisation: A review of methodological developments and debates. *The European Journal of Health Economics*, 21(2):171–180, 2020.

David Revelt and Kenneth Train. Customer-specific taste parameters and mixed logit: Households' choice of electricity supplier. 2000.

D. Rudoler, A. Laporte, J. Barnsley, and R. H. Galzier. Paying for primary care: A cross-sectional analysis of cost and morbidity distributions across primary care payment models in ontario canada(article). *Social Science and Medicine*, 124:18–28, 2015a.

David Rudoler, Raisa Deber, Janet Barnsley, Richard H Glazier, Adrian Rohit Dass, and Audrey Laporte. Paying for primary care: the factors associated with physician self-selection into payment models. *Health Economics*, 24(9):1229–1242, 2015b.

David Rudoler, Raisa Deber, Adrian Rohit Dass, Janet Barnsley, Richard Glazier, and Audrey Laporte. Paying for primary care: The relationship between payment for primary care physicians and selection of patients based on case-mix. Working Paper 160007, Canadian Centre for Health Economics, 2016.

Susan Schultz and Richard Glazier. *CMAJ Open*, 5(4):E856–E863, 2017.

Yu-Chu Shen and Stephen Zuckerman. The effect of medicaid payment generosity on access and use among beneficiaries. *Health services research*, 40(3):723–744, 2005.

Barbara Starfield. Reinventing primary care: Lessons from canada for the united states. *Health Affairs*, 29(5), 2010.

Statistics Canada. Health fact sheets: Primary health care providers, 2016. Technical Report Catalogue no. 82-625-X, Statistics Canada, 2007.

Statistics Canada. Boundary files, reference guide. Data Dictionary 92-160-G, Statistics Canada, 2011.

U.S. Bureau of Labor Statistics. Occupational employment statistics, one occupation for multiple geographies, metropolitan or non metropolitan area, may 2019, family medicine physicians(soc code291215). Data table, 2019.

Ashley Vissing. One-to-many matching with complementary preferences: An empirical study of natural gas lease quality and market power. Job market paper, Duke University, 2017.

Adam Wagstaff, Eddy Van Doorslaer, and Pierella Paci. On the measurement of horizontal inequity in the delivery of health care. *Journal of health economics*, 10(2):169–205, 1991.

Jim Whaley. A longitudinal review of rural health policy in ontario. In *Healthcare Management Forum*, volume 33, pages 53–56. SAGE Publications Sage CA: Los Angeles, CA, 2020.

Xue Zhang and Arthur Sweetman. Blended capitation and incentives: Fee codes inside and outside the capitated basket. *Journal of Health Economics*, 60:16–29, 2018.

# A Dataset Construction

This section details the construction of the main analysis dataset.

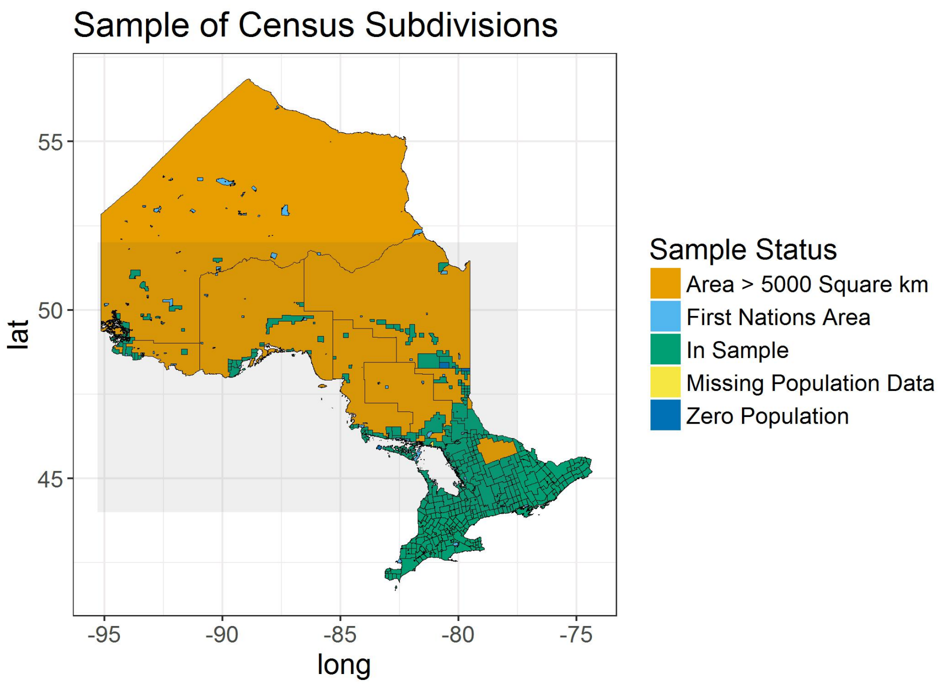
## A.1 Sample Subset

**Geographical** I geographically restrict the dataset to fit my empirical exercise. First, I remove all patients and physicians who do not reside in Ontario. Second, I remove all patients and physicians in census subdivisions that are affiliated with First Nations of Indian bands, as defined by Statistics Canada. These areas have data limitations and may not fully appear in the billings data. I also remove all patients and physicians in census subdivisions that are larger than 5000 square kilometers, or have populations that are lower than the reporting threshold (generally 40). These census subdivisions tend to be in frontier areas with extremely low population densities and have alternative healthcare environments. In each individual year, the sample is restricted further to those with population data in that year. All of these resulting CSDs had patients choosing physicians in them in every year.

One last census subdivision is removed from the sample: Unorganized Mainland Manitoulin (351091). This census subdivision was annexed by neighboring Killarney in 2006. It had a population of 5 and had 9 (mostly seasonally inhabited) dwellings in the 2006 census. This census subdivision was removed because it did not appear in the census population data in 2006.

Lastly, I restrict the census subdivisions to those in one of five markets in Northern Ontario. These markets are detailed in section 4.2.

Figure 15: Eligible Census Subdivisions




**Physicians** I restrict the physician dataset to comprehensive care primary care physicians with more than 300 patients. Schultz and Glazier (2017) define a comprehensive care physician as a primary care physician who saw patients more than 43 days per year, "more than half of their services were for core primary care and their services fell into at least 7 of 22 activity areas." I expand this definition to any physician who ever fell into this category and did not change their practice locations. This limits the influence of coding differences in different years and different payment models on the set of included physicians.

**Years** The original data span the years 2003-2015. The final sample includes only data from 2004 - 2014. 2003 and 2015 are excluded to allow lag and lead years to be used to construct patient characteristics and outside option shares.

## A.2 Comorbidities

Patient comorbidities are defined using the ICD-9 Royal College of Surgeons' Charlson Comorbidity Mapping (Brusselaers and Lagergren, 2017), with adjustments for the differences between the Ontario Health Insurance Plan (OHIP) diagnoses codes in the billings data and ICD-9 codes. Table 7 presents the differences between the ICD-9 codes and the OHIP diagnosis codes used in the comorbidity mapping. The primary difference is the omission of a number of ICD-9 codes from the set of possible OHIP diagnosis codes. Only one ICD-9 code was excluded because of a difference in definition (ICD-9 725). In the ICD-9 codebook, code 725 is Polymyalgia Rheumatica. In the OHIP diagnosis codebook, code 725 is Lumbar Disc Disease (degenerative) (ICES, 2020).

Comorbdities are defined for a patient based on their previous year's claims. If they have a diagnosis code associated a comorbidity in year $t-1$, they are labeled as having that comorbidity in year $t$.

Table 7: Charlson Comorbidity Codes

| Comorbidity | Code | ICD-9 Diagnosis Codes | OHIP Diagnosis Codes |
|---|---|---|---|
| Myocardial Infraction | MI | 410, 412 | 410, 412 |
| Congestive Heart Failure | CHF | 402, 425, 428, 429 | 402, 428, 429 |
| Peripheral Vascular Disease | PVD | 440-447, 785E, V43D | 440, 441, 443, 446, 447 |
| Cerebrovascular Disease | CD | 362C, 430-438 | 432, 435, 436, 437 |
| Dementia | Dem | 290, 294 | 290 |
| Chronic Pulmonary Disease | CPD | 416, 490-496, 500-505, 506D | 491-494, 496, 501, 502 |
| Rheumatic Disease | Rhe | 710-714, 725 | 710-712, 714 |
| Liver Disease | LD | 070, 456A-456C, 571-573 | 070, 571, 573 |
| Diabetes Mellitus | DM | 250 | 250 |
| Renal Disease | Ren | 403-404, 580-586, 588, V420, V451 | 403, 580, 581, 584, 585 |
| Metastatic Tumors | MT | 196-199 | 196-199 |
| Malignancy | Mal | 140-172, 174-195, 200-208 | 140-172, 174-195, 200-208 |

## A.3 Patient Characteristics

Age, sex, and location (census subdivision) are provided at the billing-level. Therefore, there is some variation in age, sex, and location for the same patient in each year. Simply, I define a patient's age and sex in a year to be the age and sex that are associated with the largest number of billings in that year. If the patient is not present in the data in year t, they are assigned age and sex from year t+1 data. If they are not present in year t or year t+1 data, they are assigned age and sex from year t-1 data.

The census subdivision (CSD) of a patient is determined by a more lengthy process. First, missing CSDs, non-Ontario CSDs, and CSDs that are more than 150km from the the patient's doctor are removed as possibilities, unless they are the only CSDs that are observed in the data in year t and year t+1. After these removals, if there is a unique CSD observed in the data, then that CSD is used. If there are still multiple CSDs that exist in that year, the CSD that is most associated with the next year is used. A CSD is most associated with the next year if: 1) it is the unique CSD of the patient in the next year; 2) it has the most billings the next year. Lastly, if no billings appear next year, then the CSD is assigned to the CSD with the most billings this year. In 2015, when there is no next year, the most associated CSD this year is used. If no billings exist in year t, then the CSD with the most associated claims in year t+1 is used as the CSD for this year. If no billings exist in year t or year t+1, then the CSD with the most associated claims in year t-1 is used.

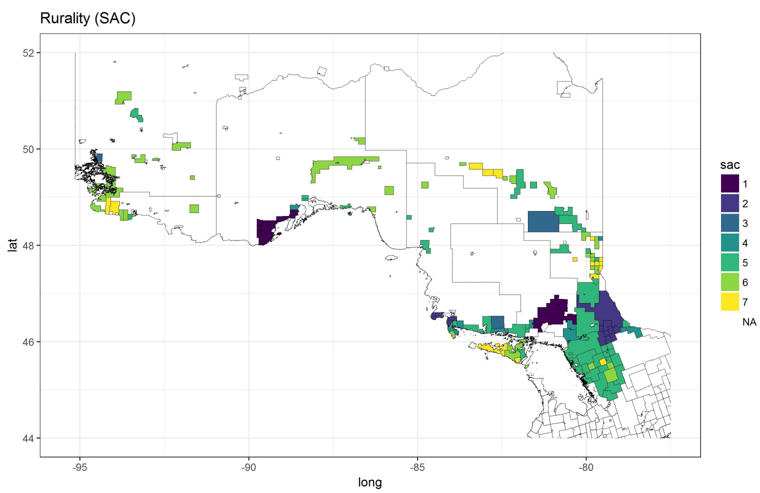
Five CSDs in the billings data did not exist in the census data due to changes in geographies over time. Three of these CSDs (3554012, 3554016, 3554018) were consolidated into a the new CSD of Temiskaming Shores in 2006. Two others (3554046, 3554048) were consolidated into a new CSD of Charlton and Dack. The coding of these CSDs were changed to reflect the consolidated CSDs.

I use the Canadian Census' Statistical Area Classification (SAC) as a proxy for rurality. CSDs with an SAC equal to 1 are labeled urban. CSDs with an SAC equal to 2 or 3 are labeled semi-urban. Semi-urban are non-metropolitan population agglomerations with more than 10,000 residents. Rural areas are CSDs with an SAC above 3. Table 8 details the SAC definitions. This table is replicated from from Statistics Canada (2011). Figure 16 is a map of SAC in the geographical sample.

Table 8: Statistical Area Classification (SAC) Definition

| SAC | Description |
|---|---|
| 1 | Census subdivision within census metropolitan area |
| 2 | Census subdivision within census agglomeration with at least one census tract |
| 3 | Census subdivision within census agglomeration having no census tracts |
| 4 | Census subdivision outside of census metropolitan area and census agglomeration area having strong metropolitan influence |
| 5 | Census subdivision outside of census metropolitan area and census agglomeration area having moderate metropolitan influence |
| 6 | Census subdivision outside of census metropolitan area and census agglomeration area having weak metropolitan influence |
| 7 | Census subdivision outside of census metropolitan area and census agglomeration area having no metropolitan influence |

Figure 16: Rurality (SAC)




## A.4 Physician Characteristics

Most physician characteristics are directly observed in the Corporate Provider Database or the ICES Physician Database. These variables include age, sex, specialty, practice type (specialist/CCPCP). Other variables are provided at the group level in the Corporate Provider Database and the Client Agency Program Enrolment Database. These include the location of the physician's office and payment model.

Some physicians are members of multiple groups. The distance/drive time between a physician and a patient is defined as the minimum distance/drive time between any of the physician's group locations and the patient. For the purposes of defining the physician's market, the physician's primary CSD is defined as the group location that is closest (in sum of squared distances) to the physician's patients.[26]

To determine whether a physician's payment model is fee-for-service, enhanced fee-for-service, or capitation, group payment model data from the Corporate Provider Dataset and revenue description data from the ICES Provider Database are used. 98.19% of physician group-years are in one of five payment models: CCM, FHG, FHN, FHO, and traditional fee-for-service. CCM and FHG payment models are labeled "enhanced Fee for Service" models and FHN and FHO payment models are labeled "capitation" models.

[26]For physicians who are not in a group, the office location is assumed to be equal to the maincsd variable from the ICES Provider Database, which is based on patient postcodes.

Most comprehensive care primary care physicians had only one payment model category each year (952,99 out of 107,724 physician-years for all of Ontario 2003-2015). However, the remaining physicians have multiple payment models in one year because they switch payment models halfway through the year or are in multiple groups. I must assign a unique payment model to each physician to estimate the model. To do so, I determine the main payment model for each physician by a series of tie-breaking rules. First, I break ties between fee-for-service and an alternative payment model in favor of the alternative payment model. Participation in an alternative payment model is an investment and in some cases comes with explicit patient minimums (e.g. 2,400 enrolled patients per group in the FHN model). Fee for service work, on the other hand, may be temporary work for another physician group or for a hospital. Second. I break ties between two payment models if only one remains as a payment model for that physician in the next year. In this case, physicians were likely to have switched payment models in the middle of the year. Physician and patients were likely to make choices knowing that this change was occurring. Third, if there are still multiple payment model, I label all physicians who make less than 50% of their revenue off of fee for service billings as being in a capitation model. I label physicians who make more that 50% of their revenue from fee for service billing as being in an enhanced fee for service model. I use this same rule to assign a physician when they are in a payment model that is not fee-for-service, CCM, FHG, FHN, or FHO.

Table 9 shows the number of physician payment models that are determined by each step. The first three steps appear to have been appropriate. Physicians are allocated to the enhanced fee-for-service payment model at a higher rate when they attain the majority of their revenue from fee-for-service payments.

Table 9: CCPCP Payment Model Data Cleaning

| | | % with most revenue coming from FFS payments | | |
|---|---|---|---|---|
| Step | # Determined | FFS | EFFS | CAP |
| Unique | 95,299 | 92.05% | 98.30% | 3.73% |
| First Step | 2,563 | N/A | 98.85% | 1.97% |
| Second Step | 2,337 | N/A | 71.43% | 40.96% |
| Third Step | 7,525 | N/A | 100% | 0% |

## A.5 Expected Revenue and Expected Number of Visits

I estimate expected number of visits and expected revenue using a risk adjustment-like methodology. Expected number of visits, $V_i$ is the expected number of visits that patient $i$ would make to their physician, if their physician was in a fee-for-service payment model. Expected revenue, $R_{is}$, is estimated separately for each payment model. Expected revenue is the revenue in 2004/2005 dollars that a physician in payment model $s \in$ {Capitation, EFFS, FFS} would expect to attain from patient $i$, assuming that patient attained care as if their physicians was in a fee-for-service model.

Care acquisition behavior is held constant in these estimations to remove a confounding factor in the identification of physician response to revenue. That is, physicians may change their practice styles to increase revenue after they adopt an alternative payment model, thus introducing variation into the revenue attained by a physician. This variation is correlated with an unobserved physician ability to change practice styles. My measure of expected revenue avoids this issue by holding patient utilization constant across the revenue calculations.

The following methodology is used to estimate expected revenue and expected number of visits. First, a dataset is constructed at the patient-year observation level. The dataset includes comorbidities, characteristics, number of visits, and the revenues the patient would provide for their physician in each of the three main payment models. Second, the dataset is restricted to a sample of patients with doctors in the fee-for-service model and who are similar to the sample used in the estimation of the empirical matching model. Third, regressions of revenues and visits onto patient comorbidities and characteristics are estimated. Fourth, predictions of revenues and visits are made for all patients in the main sample. The remainder of this section presents more detail on each of these steps.

Previous literature has shown that the alternative payment models increased the incomes of physicians significantly. Gray et al. (2015) find a 25% increase in incomes after a move from fee-for-service to capitation and a 12% increase after a move from fee-for-service to enhanced fee-for-service using a diff-in-diff strategy with survey data. This result is confirmed by estimates of per-patient revenue. I estimate that the average patient in Northern Ontario in 2015 would provide 46.63% more revenue in the capitation model and 23.30% more revenue in the enhanced fee-for-service model than in they would in the fee-for-service model, holding utilization fixed. Fixed and marginal costs may be larger in the alternative payment models.

**Dataset Construction** Patient comorbidities and characteristics are defined as in the matching model sample. Number of visits are the number of claims with a visit feecode. 83 feecodes are associated with a visit. 74 of these are the visit fee codes used by Buckley et al. (2014a). This list can be found in the appendices of the working paper version of this paper (Buckley et al., 2011). Nine additional feecodes were added to their list: P004, A002, A261, A268, K130, K131, K132, K267, K267, and K269. These feecodes were not in use during the time frame of the Buckley et al. (2011) sample.

Fee-for-service revenue is the sum of fees for services attained by the patient's matched physician. Each feecode is associated with a fee specified by the Ontario Ministry of Health's Schedule of Benefits and Fees (SBF). For each service a patient attains, I use the fee level specified by the SBF, rather than the physician paid variable. Thus, if physicians supply services under a non-fee-for-service payment model, the correct fee sum will still be tabulated.

Enhanced fee-for-service revenue is equal to the fee-for-service revenue, a comprehensive care capitation (CCC) payment, and an additional 10% bonus for fees associated with the comprehensive care premium eligible (CCPE) basket of services. The CCC payment is a monthly payment. However, I assume each patient is rostered for a full year. The base annual CCC payment used is $17.04 (Ministry of Health and Long-Term Care, 2007a). The CCC fee is adjusted according to age and sex, as specified below. Services that are in the CCPE basket are defined as the set of eligible fees for the FHG that existed in 2014 (33 fees) (OMA, 2015). 13 services were added to the FHG CCPE basket from 2007 to 2014 (Ministry of Health and Long-Term Care, 2007a). However, only 6.71% of the estimated comprehensive care premium revenue in 2004 is from those 13 services. In future work, variation in the size of the CCPE basket will be taken into account.

Capitation revenue is equal to the SBF fees associated with services outside of the capitation service basket (CSB) that are provided by the patient's matched physician, a comprehensive care capitation (CCC) payment, a capitation rate (CR) payment, an Access Bonus, and Shadow fees. See section §B for a clearer exposition of how capitation physicians are paid. The CCC payments are equivalent to the payments in the enhanced fee-for-service model. The base capitation rate payment used is \$126.04 (Ministry of Health and Long-Term Care, 2007b). The access bonus is estimated as the difference between .1895 times the CR payment and the sum of the SBF fees associated with physicians who are not the patient's matched physician. Lastly, shadow fees are estimated as $\frac{1}{10}$of the SBF fees of services inside the CSB.

The CSB used is the 2014 FHO basket. This basket consists of 157 fee codes (Ministry of Health and Long-Term Care, 2014). In 2007, the FHO basket consisted of 137 fee codes. Differences between these baskets are insignificant. Services associated with the 2014 basket but not with the 2007 basket are associated with .000037% of the estimated SBF fee payment totals for in-basket CSB services.

Physicians in the enhanced fee-for-service and capitation models also receive bonuses for quality of care (see section §B for details). However, these bonus payments are not included in the calculation of expected revenue. Bonuses are computed based on the percent of rostered patients who receive specific services. Therefore, the additional expected revenue a new patient would provide along this dimension is difficult to estimate, as it would be a function of the behavior of other rostered patients. Further, these bonuses contribute less than 2% of enhanced fee-for-service physician revenue (Henry et al., 2012).[27]

I restrict the sample to mirror the empirical matching sample and to minimize the effect of selection into payment models. The data used are at the physician-year observation level. The sample is restricted to patients in the geographical area of the 5 markets used to estimate the empirical matching model. The sample is further restricted to patients who are matched to physicians with more than 300 patients. Lastly, the sample is restricted to patients with a fee-for-service physician in 2004 and 2005.

Before regressions are estimated, I winsorize revenues at the 0.5th and 99.5th percentiles and visits from above at the 99.5th percentile.

Table 10 provides summary statistics of the dataset used in the regression analysis.

[27]This source only provides aggregate bonus payments for capitation physicians, which include quality bonuses and access bonuses. Quality bonus revenue for capitation physicians is likely to be similar.

Table 10: Revenue and Visit Estimation Data Summary Statistics

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| CD | 310,186 | 0.011 | 0.105 | 0 | 0 | 0 | 1 |
| CHF | 310,186 | 0.030 | 0.172 | 0 | 0 | 0 | 1 |
| CPD | 310,186 | 0.069 | 0.254 | 0 | 0 | 0 | 1 |
| Dem | 310,186 | 0.008 | 0.090 | 0 | 0 | 0 | 1 |
| DM | 310,186 | 0.065 | 0.246 | 0 | 0 | 0 | 1 |
| LD | 310,186 | 0.005 | 0.071 | 0 | 0 | 0 | 1 |
| Mal | 310,186 | 0.031 | 0.172 | 0 | 0 | 0 | 1 |
| MI | 310,186 | 0.044 | 0.205 | 0 | 0 | 0 | 1 |
| MT | 310,186 | 0.002 | 0.048 | 0 | 0 | 0 | 1 |
| PVD | 310,186 | 0.015 | 0.123 | 0 | 0 | 0 | 1 |
| Ren | 310,186 | 0.007 | 0.080 | 0 | 0 | 0 | 1 |
| Rhe | 310,186 | 0.014 | 0.119 | 0 | 0 | 0 | 1 |
| Has Comorbidity | 310,186 | 0.228 | 0.419 | 0 | 0 | 0 | 1 |
| Income | 310,186 | 35.240 | 6.084 | 16.763 | 30.929 | 37.098 | 61.089 |
| Female | 310,186 | 0.538 | 0.499 | 0 | 0 | 1 | 1 |
| FFS (Winsorized) | 310,186 | 123.743 | 138.103 | 8.500 | 33.530 | 153.250 | 1,003.895 |
| EFFS (Win.) | 310,186 | 152.257 | 148.306 | 8.500 | 56.511 | 188.877 | 1,062.889 |
| Capitation (Win.) | 310,186 | 177.072 | 114.509 | −168.026 | 110.682 | 234.343 | 827.820 |
| Visits (Win.) | 310,186 | 4.226 | 4.629 | 1 | 1 | 5 | 31 |

**Regression** I regress revenues and visits on comorbidity indicators, median income of a patient's census subdivision, age, sex, and interactions. Table 10 shows regression results.

Table 11: Revenue and Visits Estimation Regression Results

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Cap. Revenue | EFFS Rev | FFS Rev | Visits |
| | (1) | (2) | (3) | (4) |
| CD | 23.254*** | 67.885*** | 65.790*** | 2.155*** |
| | (1.453) | (2.370) | (2.247) | (0.074) |
| CHF | 4.779*** | 50.114*** | 47.788*** | 1.590*** |
| | (0.935) | (1.526) | (1.446) | (0.048) |
| CPD | 2.230*** | 57.613*** | 54.718*** | 1.561*** |
| | (0.838) | (1.368) | (1.297) | (0.043) |
| Dem | 106.596*** | 153.600*** | 161.666*** | 5.702*** |
| | (2.802) | (4.573) | (4.335) | (0.143) |
| DM | 4.533*** | 45.396*** | 42.088*** | 1.158*** |
| | (0.810) | (1.321) | (1.252) | (0.041) |
| LD | 5.052** | 53.084*** | 50.250*** | 1.014*** |
| | (2.110) | (3.443) | (3.264) | (0.107) |
| Mal | 7.787*** | 35.864*** | 34.067*** | 0.805*** |
| | (0.995) | (1.623) | (1.539) | (0.051) |
| MI | 9.793*** | 48.161*** | 44.881*** | 1.454*** |
| | (0.862) | (1.407) | (1.334) | (0.044) |
| MT | 3.066 | 54.067*** | 52.299*** | 0.780*** |
| | (3.057) | (4.988) | (4.728) | (0.155) |
| PVD | 1.724 | 34.355*** | 31.955*** | 0.981*** |
| | (1.242) | (2.027) | (1.922) | (0.063) |
| Ren | 7.359*** | 40.459*** | 39.369*** | 1.090*** |
| | (1.840) | (3.002) | (2.846) | (0.094) |
| Rhe | −3.964*** | 46.100*** | 43.163*** | 1.664*** |
| | (1.325) | (2.162) | (2.050) | (0.067) |
| Income(k) | −0.996*** | 4.100*** | 3.894*** | 0.138*** |
| | (0.169) | (0.276) | (0.261) | (0.009) |
| Income$(k)^2$ | 0.010*** | −0.052*** | −0.050*** | −0.002*** |
| | (0.002) | (0.003) | (0.003) | (0.0001) |
| hasclaims | −3.412*** | 32.740*** | 29.811*** | 1.331*** |
| | (0.371) | (0.606) | (0.574) | (0.019) |
| sac2 | 15.442*** | 20.146*** | 18.895*** | 0.428*** |
| | (0.464) | (0.757) | (0.717) | (0.024) |
| sac3 | 0.385 | 8.348*** | 7.716*** | 0.193*** |
| | (0.680) | (1.110) | (1.052) | (0.035) |
| sac4 | 8.651*** | 7.109*** | 6.385*** | 0.198*** |
| | (1.030) | (1.680) | (1.593) | (0.052) |
| sac5 | 1.540*** | 16.960*** | 16.347*** | 0.482*** |
| | (0.431) | (0.703) | (0.666) | (0.022) |
| sac6 | −4.494*** | −15.773*** | −13.570*** | −0.743*** |
| | (0.456) | (0.745) | (0.706) | (0.023) |
| sac7 | 0.935 | −17.492*** | −15.287*** | −0.760*** |
| | (1.172) | (1.913) | (1.814) | (0.060) |

Table 12: Revenue and Visits Estimation Regression Results (Continued)

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Cap. Revenue | EFFS Rev | FFS Rev | Visits |
| | (1) | (2) | (3) | (4) |
| age15-34 | −20.291*** | −3.034** | 0.148 | −0.341*** |
| | (0.741) | (1.210) | (1.147) | (0.038) |
| age35-49 | 33.400*** | 30.136*** | 26.029*** | 0.455*** |
| | (0.747) | (1.219) | (1.156) | (0.038) |
| age50-64 | 89.846*** | 54.544*** | 43.344*** | 1.061*** |
| | (0.786) | (1.283) | (1.216) | (0.040) |
| age65+ | 200.775*** | 100.260*** | 76.303*** | 2.455*** |
| | (0.963) | (1.571) | (1.489) | (0.049) |
| age0-14:female | 0.067 | −0.345 | −0.238 | 0.010 |
| | (0.801) | (1.307) | (1.239) | (0.041) |
| age15-34:female | 78.223*** | 36.178*** | 26.223*** | 0.746*** |
| | (0.638) | (1.041) | (0.987) | (0.032) |
| age35-49:female | 65.518*** | 26.736*** | 18.295*** | 0.635*** |
| | (0.670) | (1.093) | (1.036) | (0.034) |
| age50-64:female | 52.065*** | 22.524*** | 15.506*** | 0.568*** |
| | (0.753) | (1.229) | (1.165) | (0.038) |
| age65+:female | 19.747*** | 18.539*** | 15.602*** | 0.588*** |
| | (1.012) | (1.651) | (1.565) | (0.051) |
| Dem:female | 32.384*** | 40.598*** | 44.217*** | 1.697*** |
| | (3.483) | (5.684) | (5.389) | (0.177) |
| age0-14:hascomorbid:sexM | −15.450*** | −51.398*** | −48.969*** | −1.408*** |
| | (1.866) | (3.045) | (2.887) | (0.095) |
| age15-34:hascomorbid:sexM | −14.557*** | −39.266*** | −37.534*** | −1.011*** |
| | (1.839) | (3.001) | (2.845) | (0.094) |
| age35-49:hascomorbid:sexM | −9.659*** | −7.101*** | −7.524*** | 0.053 |
| | (1.433) | (2.338) | (2.216) | (0.073) |
| age50-64:hascomorbid:sexM | −6.715*** | −6.913*** | −7.479*** | 0.121** |
| | (1.188) | (1.939) | (1.838) | (0.060) |
| age65+:hascomorbid:sexM | −1.159 | −1.152 | −1.448 | 0.372*** |
| | (1.297) | (2.117) | (2.007) | (0.066) |
| age0-14:hascomorbid:sexF | −19.134*** | −53.077*** | −50.505*** | −1.495*** |
| | (2.056) | (3.356) | (3.181) | (0.105) |
| age15-34:hascomorbid:sexF | −21.029*** | −27.861*** | −27.193*** | −0.433*** |
| | (1.524) | (2.486) | (2.357) | (0.078) |
| age35-49:hascomorbid:sexF | −16.166*** | 0.026 | −1.052 | 0.490*** |
| | (1.319) | (2.152) | (2.041) | (0.067) |
| age50-64:hascomorbid:sexF | −11.762*** | −3.755* | −4.521** | 0.225*** |
| | (1.179) | (1.924) | (1.824) | (0.060) |
| age65+:hascomorbid:sexF | 2.827** | 4.852** | 4.948*** | 0.446*** |
| | (1.195) | (1.949) | (1.848) | (0.061) |
| Constant | 114.265*** | −18.189*** | −29.473*** | −0.797*** |
| | (3.441) | (5.614) | (5.322) | (0.175) |
| Mean of Dependent Var | 177.072 | 152.257 | 123.743 | 4.226 |
| Observations | 310,186 | 310,186 | 310,186 | 310,186 |
| $R^2$ | 0.500 | 0.206 | 0.177 | 0.208 |
| Adjusted $R^2$ | 0.500 | 0.206 | 0.177 | 0.208 |
| Residual Std. Error (df = 310144) | 80.988 | 132.152 | 125.281 | 4.120 |
| F Statistic (df = 41; 310144) | 7,560.055*** | 1,963.633*** | 1,628.893*** | 1,984.634*** |

*Note:* *p<0.1; **p<0.05; ***p<0.01

# B Details of the Physician Payment Models

There are five main payment models available for physicians to choose between 2003 and 2015. These are fee-for-service (FFS), Family Health Group (FHG), Comprehensive Care Model (CCM), Family Health Organization (FHO), and Family Health Network (FHN). The FHG and CCM models are enhanced fee for service models. The FHO and FHN are capitation models. Table 13 describes the main attributes of these models.[28]

Table 13: Characteristics of Payment Models

| Payment Model | Type | Introduced | % Revenue from fees | % Physicians in 2015 | Minimum Group Size |
|---|---|---|---|---|---|
| FFS | FFS | 1966 | 98 | 14 | 1 |
| FHN | Capitation | 2002 | 27 | 2 | 3 |
| FHG | EFFS | 2003 | 81 | 27 | 3 |
| CCM | EFFS | 2005 | 84 | 4 | 1 |
| FHO | Capitation | 2006 | 21 | 53 | 3 |

Under the fee-for-service model, physician $j$'s revenue is simple. The physician receives the sum of fees for services they provide as specified in the OHIP Schedule of Benefits and Fees (OSBF). Rostered patients and unrostered patients are treated identically under the FFS model. For clarity, the FFS revenue function is:

$$R_j^{FFS}(\mathbf{X}_R, \mathbf{N}_R, \mathbf{N}_0) = \sum_s p_s(N_{0s} + N_{Rs})$$

where $\mathbf{X}_R$ are the ages and sexes of rostered patients, $N_{Rs}$ are the number of services of type $s$ from rostered patients, $N_{0s}$ are the number of services of type $s$ from non-rostered patients, and $p_s$ is the OSBF fee for service $s$.

Under both enhanced fee-for-service models, EFFS physicians receive bonuses for hitting certain quality goals and small monthly payments based on their rostered patient's characteristics. These capitation payments are called "Comprehensive Care Capitation" payments. CCM physician $j$'s revenue is:

$$R_j^{CCM}(\mathbf{X}_R, \mathbf{N}_R, \mathbf{N}_0) = \sum_s p_s(N_{0s} + N_{Rs}) + QB(\mathbf{N}_R, \mathbf{X}_R) + CCC(\mathbf{X}_R)$$

Where $QB(\cdot)$ is the function from rostered patient age, patient sex, and services provided to quality bonus and $CCC(\cdot)$ is the function from rostered patient age and sex to comprehensive care capitation payments. In the FHG model, there are additional payments for a specific basket of services. This

[28]Source: Rudoler et al. (2015b), Hurley et al. (2013), author's calculations.

basket of services, $B_{FHG}$, is a set of services for which FHG physicians get an enhanced payment for their rostered patients. The enhanced payment is an additional 10% bonus for services in the basket. FHG physician $j$'s revenue can thus be written:

$$R_j^{FHG}(\mathbf{X}_R, \mathbf{N}_R, \mathbf{N}_0) = \sum_s p_s N_{0s} + \sum_s (1 + \frac{1}{10} 1\{s \in B_{FHG}\}) p_s N_{Rs} + QB(\mathbf{N}_R, \mathbf{X}_R) + CCC(\mathbf{X}_R)$$

Most physicians who chose an EFFS model chose the FHG model

The capitation models are structured differently than the EFFS and FFS models. For rostered patients, physicians receive only 10% (until October 2010, then 15% (Zhang and Sweetman, 2018)) of the OSBF fees for services in a basket of services, $B_{FHO}$and $B_{FHN}$. These payments are called "shadow fees," and exist to encourage physicians to report all services provided. Instead, physicians receive most of their revenue from yearly capitation payments that are a function of the age and sex of their rostered patients. Services provided to non-rostered patients and services outside of the basket are still paid for on a fee-for-service basis. There are some restrictions placed on these payments: payments for services provided to non-rostered patients cannot exceed $40,000. A certain percent of capitation fees called the "Access Bonus" is decreased dollar for dollar for every service a rostered patient receives from physicians outside of the physician group they are rostered with. Note that a patient is rostered to a physician, not a group. Lastly, physicians in the capitation models receive the same quality bonus and comprehensive care capitation payments as their enhanced fee-for-service colleagues.

FHO Physician $j$'s revenue is:

$$R_j^{FHO}(\mathbf{X}_R, \mathbf{N}_R, \mathbf{N}_R^{-g(j)}, \mathbf{N}_0) = \sum_s 1\{s \notin B_{FHO}\} p_s (N_{0s} + N_{Rs})$$

$$+ \frac{1}{10} \sum_s 1\{s \in B_{FHO}\} p_s N_{Rs} + max\{\sum_s 1\{s \in B_{FHO}\} p_s N_{0s}, \$40,000\}$$

$$+ Cap_O(\mathbf{X}_R) + max\{.1895 Cap_O(\mathbf{X}_R) - \sum p_s \mathbf{N}_{Rs}^{-g(j)}, 0\} + QB(\mathbf{N}_R, \mathbf{X}_R) + CCC(\mathbf{X}_R)$$

FHN Physician $j$'s revenue is:

$$R_j^{FHN}(\mathbf{X}_R, \mathbf{N}_R, \mathbf{N}_R^{-g(j)}, \mathbf{N}_0) = \sum_s 1\{s \notin B_{FHN}\} p_s (N_{0s} + N_{Rs})$$

$$+ \frac{1}{10} \sum_s 1\{s \in B_{FHN}\} p_s N_{Rs} + max\{\sum_s 1\{s \in B_{FHN}\} p_s N_{0s}, \$40,000\}$$

$$+ Cap_N(\mathbf{X}_R) + max\{.2065 Cap_N(\mathbf{X}_R) - \sum p_s \mathbf{N}_{Rs}^{-g(j)}, 0\} + QB(\mathbf{N}_R, \mathbf{X}_R) + CCC(\mathbf{X}_R)$$

Where $Cap(\cdot)$ is the function from rostered patient age and sex to capitation payments and $\mathbf{N}_{Rs}^{-g(j)}$ are the services provided to physician $j$'s rostered patients by other physicians who are not in $j$'s medical group.[29]

[29] Unless otherwise specified, this information was gathered from billings guides: Ministry of Health and Long-Term Care (2007a,b, 2014, 2011); OMA (2015).

## C Rural Practice Incentives

The province of Ontario has multiple policies that are designed to incentivize primary care physicians to locate in rural and Northern areas. The longest-standing and largest incentive policy is the Underserved Area Program/Northern and Rural Recruitment and Retention (UAP/NRRR) grant. This program was established in 1969 and provides grants that last up to 4 years for service in an underserved area. Underserved area designation before 2009 was conducted by a political process by the Ontario Ministry of Health and Long Term Care (MOHLTC). By 2009, the number of CSDs that were designated underserved were large and many areas were in the relatively densely populated South. In 2009, the ministry began to use a set formula for eligibility of CSDs for the UAP grant, the Rurality Index for Ontario (RIO) (Whaley, 2020). Grants are now valued at $40,000 to $120,000 and paid out over the first four years of practice. The grants are paid in installments: 40% in the first year, 15% in the second and third year, and 30% in the fourth year of practice. The value of the grant is a piece-wise linear function of the RIO (Ministry of Health and Long-Term Care, 2017):

$$\text{UAP grant} = \begin{cases} 0 & RIO < 0 \\ 72,0000 + 200RIO & 40 \leq RIO < 50 \\ 7,600 + 1480RIO & 50 \leq RIO < 70 \\ 97,6000 + 200RIO & 70 \leq RIO \end{cases}$$

The RIO itself is a function of CSD characteristics, such as population relative to the median CSD population, population density, drive time to a basic health care center, and drive time to an advanced health care center (Kralj, 2009).

$$RIO = .286P + .238T_a + .476T_b$$

where

$$P = 25\text{–}3.79(\text{Pop/Median Pop}) + 5\text{–}(\text{Pop Density}/22.6)$$
$$T_b = 10[(Time_{basic,i} - \text{Median Time}_{basic})/\text{Median Time}_{basic}]$$
$$T_a = 10[(Time_{adv,i} - \text{Median Time}_{adv})/\text{Median Time}_{adv}]$$

In addition to the UAP/NRRR, the MOHLTC provides a grant program named the Northern Physician Retention Initiative grants. This program is a straightforward flat payment of $7,000 per year for any physician that has been practicing in the North for more than 4 years. The North, as defined by the program is neither a subset nor a superset of those locations that qualify for the UAP/NRRR grant.[30]

[30] The North is defined as the districts of Algoma, Cochrane, Kenora, Manitoulin, Muskoka, Nipissing, Parry Sound, Rainy River, Sudbury, Thunder Bay and Timiskaming. Note that this includes the cities of Sudbury and Thunder Bay which are not eligible for UAP/NRRR grants.

Salary-based payment models for physicians who practice in the far North have existed since 1996 in two programs: the Community Sponsored Contracts (CSCs) and Northern Group Funding Plans (NGFPs). In 2004, these payment plans were consolidated into one new program: the Rural and Northern Physician Group Agreement (RNPGA). In this program, physicians are given a lump base salary that depends on the RIO of the community they serve (Buckley et al., 2011). This base salary starts at \$60,000 and increases according to the following function:

$$\text{RNPGA Base Salary} = \begin{cases} 0 & RIO < 45 \\ 12,000(-4 + \lfloor \frac{1}{5} RIO \rfloor) & 45 \leq RIO \end{cases}$$

Lastly, in 2001, Ontario established the Northern Ontario School of Medicine (NOSM). This medical school is the first in the province's North. Ontario has five more established medical schools, all in cities in the South: Queen's (Kingston), University of Ottawa (Ottawa), University of Toronto (Toronto), McMaster (Hamilton), and Western (London). The main premise for the NOSM was to train physicians that would be more likely to practice in Northern Ontario after graduation. It's curriculum is focused on rural and Northern health and includes clerkships in rural areas. The school began graduating students in 2009.

# D Figures and Tables

## D.1 Tables

Table 14: Patient Characteristics Summary Statistics

| | All Markets | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| N | 8,718,090 | 639,273 | 1,641,845 | 1,181,480 | 3,237,340 | 2,018,152 |
| Year (mean (sd)) | 2008.99 (3.16) | 2008.95 (3.17) | 2008.96 (3.17) | 2008.95 (3.16) | 2009.01 (3.16) | 2009.02 (3.16) |
| Male (%) | 4,263,135 (48.9) | 311,914 (48.8) | 801,872 (48.8) | 583,408 (49.4) | 1,574,031 (48.6) | 991,910 (49.1) |
| Age (%) | | | | | | |
| 0-14 | 1,404,284 (16.1) | 119,738 (18.7) | 267,075 (16.3) | 197,803 (16.7) | 512,524 (15.8) | 307,144 (15.2) |
| 15-34 | 2,045,868 (23.5) | 152,477 (23.9) | 406,312 (24.7) | 276,717 (23.4) | 758,775(23.4) | 451,587 (22.4) |
| 35-49 | 1,787,272 (20.5) | 131,529 (20.6) | 342,605 (20.9) | 249,566 (21.1) | 660,784 (20.4) | 402,788 (20.0) |
| 50-64 | 1,916,392 (22.0) | 133,097 (20.8) | 354,227 (21.6) | 261,262 (22.1) | 709,409 (21.9) | 458,397 (22.7) |
| 65+ | 1,564,274 (17.9) | 102,432 (16.0) | 271,626 (16.5) | 196,132 (16.6) | 595,848 (18.4) | 398,236 (19.7) |
| Has Comorbidity (mean (sd)) | 0.20 (0.40) | 0.15 (0.36) | 0.19 (0.40) | 0.18 (0.39) | 0.21 (0.41) | 0.20 (0.40) |
| Revenue (mean (sd)) | | | | | | |
| Capitation | 175.41 (81.99) | 163.77 (79.75) | 169.01 (80.99) | 168.58 (79.99) | 177.90 (82.39) | 184.29 (82.84) |
| EFFS | 147.50 (64.49) | 126.55 (60.53) | 141.81 (63.52) | 137.28 (63.56) | 151.98 (63.91) | 157.58 (65.19) |
| FFS | 119.63 (54.67) | 101.04 (50.94) | 114.47 (53.90) | 110.59 (53.91) | 123.49 (54.01) | 128.83 (55.31) |
| Visits (mean (sd)) | 4.07 (1.98) | 3.28 (1.86) | 3.94 (1.96) | 3.68 (1.95) | 4.25 (1.93) | 4.37 (2.00) |
| Area Type (%) | | | | | | |
| rural | 3,020,672 (34.6) | 459,452 (71.9) | 256,917 (15.6) | 694,848 (58.8) | 403,100 (12.5) | 1,206,355 (59.8) |
| semiurban | 2,482,297 (28.5) | 179,821 (28.1) | 0 (0.0) | 486,632 (41.2) | 1,004,047 (31.0) | 811,797 (40.2) |
| urban | 3,215,121 (36.9) | 0 (0.0) | 1,384,928 (84.4) | 0 (0.0) | 1,830,193 (56.5) | 0 (0.0) |
| Unmatched (mean (sd)) | 0.33 (0.18) | 0.42 (0.19) | 0.31 (0.15) | 0.45 (0.21) | 0.30 (0.16) | 0.30 (0.16) |

Table 15: Physician Characteristics Summary Statistics

| | All Markets | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| n | 5,374 | 467 | 970 | 652 | 1,896 | 1,389 |
| Male (%) | 3766 (70.1) | 302 (64.7) | 644 (66.4) | 498 (76.4) | 1314 (69.3) | 1008 (72.6) |
| Age (mean (sd)) | 49.82 (11.01) | 49.10 (10.54) | 49.31 (12.12) | 47.82 (10.53) | 50.75 (11.47) | 50.07 (9.71) |
| N Years (of 11)) | 10.31 (1.86) | 10.40 (1.67) | 10.21 (2.06) | 10.28 (1.83) | 10.22 (2.03) | 10.51 (1.51) |
| N Patients | 1,169.51 (700.86) | 717.08 (228.72) | 1,230.90 (769.40) | 946.56 (537.59) | 1,396.05 (818.19) | 1,074.16 (497.05) |
| Unfilled Capacity | 0.16 (0.16) | 0.17 (0.14) | 0.15 (0.14) | 0.15 (0.15) | 0.18 (0.18) | 0.15 (0.14) |
| Group Status (%) | | | | | | |
| Independent | 850 (15.8) | * (<20) | *(<30) | *(<30) | 314 (16.6) | 180 (13.0) |
| Multiple groups | 554 (10.3) | * (<20) | *(<30) | *(<30) | 155 (8.2) | 211 (15.2) |
| One Group | 3,970 (73.9) | 375 (80.3) | 713 (73.5) | 457 (70.1) | 1,427 (75.3) | 998 (71.9) |
| Payment Model (%) | | | | | | |
| Capitation | 3,369 (62.8) | 343 (73.4) | 461 (47.7) | 458 (70.2) | 1,120 (59.1) | 987 (71.3) |
| EFFS | 1,321 (24.6) | *(<30) | 352 (36.4) | *(<30) | 528 (27.8) | 245 (17.7) |
| FFS | 677 (12.6) | *(<30) | 154 (15.9) | *(<30) | 248 (13.1) | 153 (11.0) |
| Consultations (mean (sd)) | 23.89 (89.71) | 28.15 (68.92) | 9.85 (35.70) | 42.01 (126.97) | 20.64 (70.50) | 28.11 (120.99) |
| Visits (mean (sd)) | 4,893.79 (3,175.55) | 3,562.98 (1,623.85) | 4,821.09 (3,223.74) | 4,362.14 (2,429.72) | 5,451.92 (3,761.73) | 4,879.71 (2,765.68) |
| Consultations | 24.92 (97.38) | 38.24 (81.28) | 8.88 (36.15) | 45.16 (136.24) | 20.09 (73.93) | 28.76 (130.13) |
| FTE (mean (sd)) | 0.96 (0.35) | 0.84 (0.26) | 0.92 (0.39) | 0.92 (0.28) | 1.00 (0.38) | 0.99 (0.30) |
| Offers Walkins (%) | 1,589 (29.6) | 58 (12.4) | 361 (37.2) | 168 (25.8) | 572 (30.2) | 430 (31.0) |
| Area Type (%) | | | | | | |
| Rural | 2,159 (40.2) | 369 (79.0) | 232 (23.9) | 375 (57.5) | 274 (14.5) | 909 (65.4) |
| Semiurban | 1,487 (27.7) | 98 (21.0) | 0 (0.0) | 277 (42.5) | 632 (33.3) | 480 (34.6) |
| Urban | 1,728 (32.2) | 0 (0.0) | 738 (76.1) | 0 (0.0) | 990 (52.2) | 0 (0.0) |

*Excluded due to small bin sizes.

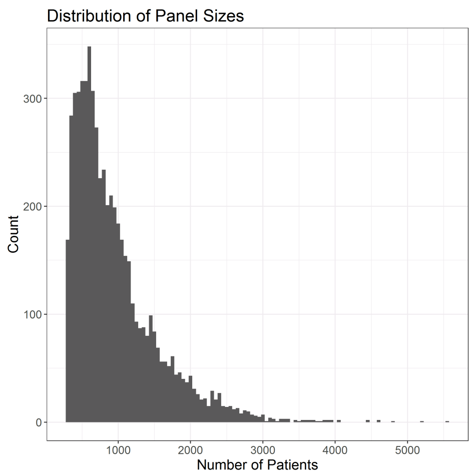
Table 16: Decomposition of Access Loss into Physician Distribution and Physician Supply Effects

| Patient Characteristics | | | | | | Decomposition One | Decomposition Two | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Area Type | Age | Has Comorbid | N | # Types | Access Loss | Capacities | Char Dist | Loc. Dist | Agg Supply | Remainder |
| Rural | 0-34 | No | 81,557 | 521 | 50.324 | 15.484 | 2.479 | -5.866 | 24.744 | 28.968 |
| Semiurban | 0-34 | No | 72,930 | 72 | 34.782 | 28.690 | -0.742 | 2.027 | 29.719 | 3.778 |
| Urban | 0-34 | No | 102,500 | 36 | 22.076 | 21.359 | -16.013 | 3.909 | 33.683 | 0.497 |
| Rural | 0-34 | Yes | 3,859 | 430 | 33.252 | 0.907 | 0.281 | -5.160 | 17.472 | 20.658 |
| Semiurban | 0-34 | Yes | 3,855 | 82 | 3.897 | 0.0140 | -0.356 | 1.791 | 1.289 | 1.173 |
| Urban | 0-34 | Yes | 7,209 | 47 | 1.407 | 0.872 | -1.445 | 0.464 | 2.186 | 0.202 |
| Rural | 35-64 | No | 83,785 | 713 | 44.845 | 11.556 | 1.925 | -5.422 | 23.120 | 25.221 |
| Semiurban | 35-64 | No | 66,459 | 103 | 20.084 | 14.271 | -0.694 | 1.690 | 17.057 | 2.031 |
| Urban | 35-64 | No | 91,189 | 63 | 7.518 | 6.696 | -6.838 | 2.104 | 11.954 | 0.298 |
| Rural | 35-64 | Yes | 16,875 | 521 | 32.198 | 5.517 | 0.818 | -5.076 | 18.994 | 17.461 |
| Semiurban | 35-64 | Yes | 14,767 | 61 | 15.490 | 12.292 | 1.714 | 0.520 | 12.013 | 1.242 |
| Urban | 35-64 | Yes | 22,343 | 41 | 6.053 | 5.515 | -9.577 | 2.790 | 12.657 | 0.183 |
| Rural | 65+ | No | 28,018 | 302 | 37.010 | 7.841 | 1.044 | -5.077 | 20.531 | 20.511 |
| Semiurban | 65+ | No | 20,396 | 30 | 19.270 | 14.208 | 3.274 | 0.847 | 12.666 | 2.483 |
| Urban | 65+ | No | 25,754 | 18 | 5.645 | 5.115 | -6.436 | 2.357 | 9.545 | 0.179 |
| Rural | 65+ | Yes | 23,294 | 243 | 27.822 | 8.580 | 1.526 | -4.956 | 18.039 | 13.213 |
| Semiurban | 65+ | Yes | 19,949 | 30 | 20.495 | 18.226 | 6.814 | -0.989 | 13.240 | 1.430 |
| Urban | 65+ | Yes | 25,543 | 18 | 4.860 | 4.546 | -6.217 | 2.067 | 8.913 | 0.097 |
| All Patients | | | 710,282 | 3,331 | 26.044 | 13.798 | -3.160 | -0.355 | 20.638 | 8.920 |

## D.2 Figures

Figure 17: Heterogeneity in Physician Outputs

(a) Number of Patients




(b) Number of Visits

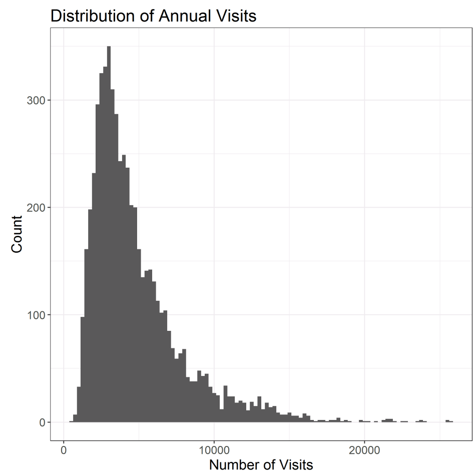
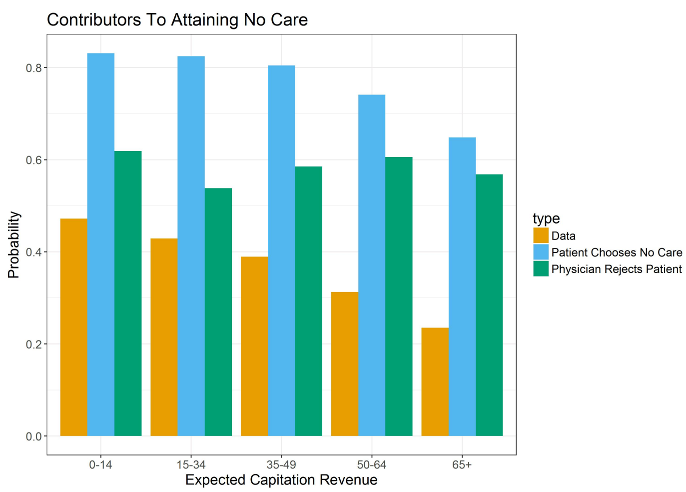



Figure 18: Results Tables

(a) Preferences: Age




(b) Entry and Access Scatterplot

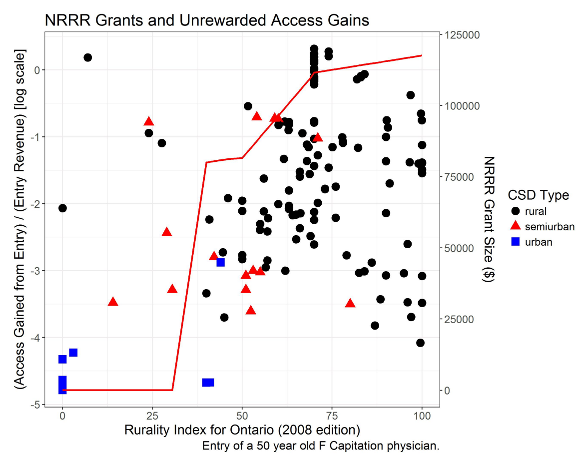
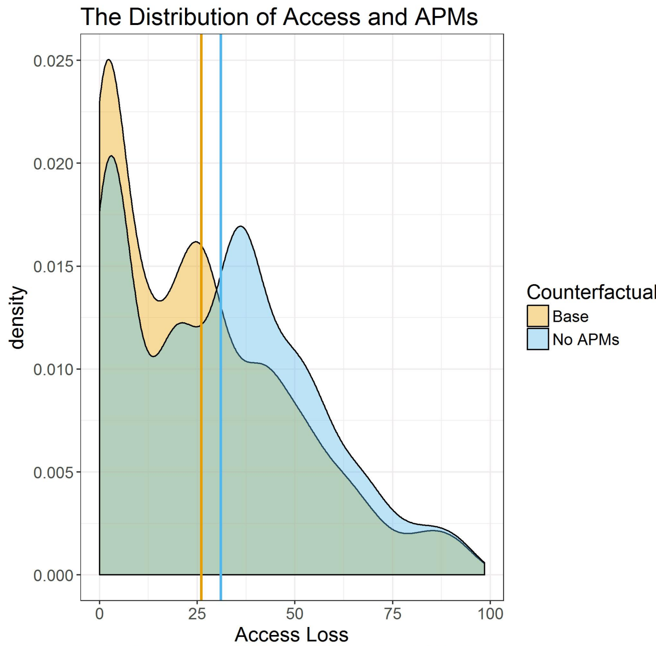


Figure 19: Effect of APMs on Access Distribution



Note: Vertical lines represent the means of the distributions.

# E Alternative Full Access Counterfactual Definitions

In the main specification, I define the full access environment as the choice environment that a patient would face in the city of Sudbury under the counterfactual assumption that all physicians are accepting all patients. This full access definition is subjective. Other definitions could be used. In this section, I present measures of access to care using the main specification and two alternative definitions of a full access environment. Additionally, I present measures of an alternative definition of access to care: the share of patient surplus that would be attained in the full access environment that is attained in the current equilibrium.

In a second full access environment definition, each patient chooses from 10 randomly selected physicians who are placed 0 kilometers from the patient. Additionally, each of these 10 physicians are always willing to accept the patient. This is equivalent to making all physician capacities infinite, shifting 10 to the same location as the patient, and shifting all other physicians to infinite distance from the patient.

In a third full access environment definition, I add physicians to the market until there is 1 physician per 1000 patients. Physicians are randomly sampled from the set of existing physicians. The reported results for each full access definition are the average over 10 simulated counterfactuals.

I define a second measure of access to care as the share of patient surplus attained in a full access counterfactual that is attained in the present equilibrium.

$$\text{Access}_{\theta t} = \frac{\text{logsum}_{\theta t}(\boldsymbol{\beta}, \boldsymbol{\tau}^u_{\theta}, \mathcal{J}_{\theta t})}{\text{logsum}_{\theta t}(\boldsymbol{\beta}, \tau^{u,FA}, \mathcal{J}_{\theta}{}^{FA})}$$

$$\text{logsum}_{\theta t}(\beta, \boldsymbol{\tau}_{\theta}, \mathcal{J}) = log( \sum_{j \in \{\mathcal{J}, \emptyset\}} exp(\delta_{\theta jt}(\beta) - \tau^u_{\theta jt})$$

where as before, individual patient surplus, $\text{logsum}_{\theta t}(\beta, \boldsymbol{\tau}_{\theta}, \mathcal{J})$, depends on preferences $\boldsymbol{\beta}$, effort costs $\tau^u$, and the choice set $\mathcal{J}$. $(\tau^u_{\theta t}, \mathcal{J}_{\theta t})$ defines the choice conditions of the estimated matching model in market $t$. $(\tau^{u,FA}, \mathcal{J}^{FA})$ defines the full access choice conditions.

Table 17 compares the different estimates of access loss for different patient types. Qualitatively, the results are similar for the first two full access environment definitions. The third full access environment definition produces significantly lower estimates of access for rural patients. Even in the choice environment where there is one physician per 1000 patients, rural patients still do not have a large variety of physicians in their choice set.

Table 17: Full Access Alternatives

| Patient Characteristics | | | | | | Share of Full Access With Doctor | | | Share of Full Access Surplus | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Area Type | Age | Comorbidities | N | # Types | % W/o Doctor | Sudbury | Ten Doctors | One Per 1000 | Sudbury | Ten Doctors | One Per 1000 |
| Rural | 0-34 | No | 81,557 | 521 | 62.244 | 50.324 | 55.999 | 30.627 | 63.495 | 71.426 | 37.084 |
| Semiurban | 0-34 | No | 72,930 | 72 | 53.381 | 34.782 | 38.187 | 35.131 | 47.870 | 53.144 | 50.717 |
| Urban | 0-34 | No | 102,500 | 36 | 39.291 | 22.076 | 22.758 | 27.864 | 36.149 | 36.108 | 46.779 |
| Rural | 0-34 | Yes | 3,859 | 430 | 45.063 | 33.252 | 41.276 | 20.123 | 45.702 | 58.954 | 27.137 |
| Semiurban | 0-34 | Yes | 3,855 | 82 | 37.951 | 3.897 | 5.600 | 8.727 | 5.777 | 9.171 | 21.245 |
| Urban | 0-34 | Yes | 7,209 | 47 | 25.399 | 1.407 | 0.648 | 6.380 | 3.064 | 1.018 | 19.483 |
| Rural | 35-64 | No | 83,785 | 713 | 52.822 | 44.845 | 49.918 | 27.528 | 60.004 | 67.985 | 35.046 |
| Semiurban | 35-64 | No | 66,459 | 103 | 42.498 | 20.084 | 22.72 | 23.212 | 30.789 | 35.536 | 40.493 |
| Urban | 35-64 | No | 91,189 | 63 | 31.222 | 7.518 | 6.605 | 13.047 | 14.173 | 10.933 | 30.509 |
| Rural | 35-64 | Yes | 16,875 | 521 | 31.775 | 32.198 | 36.912 | 19.829 | 50.482 | 59.809 | 29.318 |
| Semiurban | 35-64 | Yes | 14,767 | 61 | 23.282 | 15.490 | 17.243 | 18.337 | 30.513 | 34.836 | 40.708 |
| Urban | 35-64 | Yes | 22,343 | 41 | 17.048 | 6.053 | 6.420 | 10.945 | 14.738 | 14.697 | 31.086 |
| Rural | 65+ | No | 28,018 | 302 | 37.901 | 37.010 | 41.594 | 22.742 | 54.491 | 62.910 | 31.546 |
| Semiurban | 65+ | No | 20,396 | 30 | 26.736 | 19.270 | 20.869 | 22.405 | 34.502 | 38.463 | 45.684 |
| Urban | 65+ | No | 25,754 | 18 | 23.371 | 5.645 | 5.994 | 11.908 | 13.693 | 13.011 | 33.144 |
| Rural | 65+ | Yes | 23,294 | 243 | 28.737 | 27.822 | 30.766 | 18.101 | 49.875 | 57.405 | 29.854 |
| Semiurban | 65+ | Yes | 19,949 | 30 | 20.552 | 20.495 | 21.366 | 22.111 | 41.279 | 44.542 | 49.796 |
| Urban | 65+ | Yes | 25,543 | 18 | 18.107 | 4.860 | 4.598 | 7.933 | 13.744 | 11.293 | 28.422 |
| All Patients | | | 710,282 | 3,331 | 40.279 | 26.044 | 28.428 | 23.129 | 38.735 | 42.202 | 38.195 |

# F A Simple Proof of Stability

A matching is stable if

1. No patient prefers to be unmatched than be matched with their current physician.
2. No physician would rather have an empty space than have that space be filled with one of their current patients.
3. No patient and physician would both prefer to match with each other than keep their current matches.

Each of these conditions are met in the Rationing-by-Waiting equilibrium.

**Condition One** If a patient $i$ matches with physician $j$ then $u_{i\theta jt} - \tau^u_{\theta jt} > u_{i\theta\emptyset t}$. The one sided waiting condition implies that $\tau^u_{\theta jt} \geq 0$. Therefore when patient $i$ matches with physician $j$, they prefer $j$ to $\emptyset$: $u_{i\theta jt} > u_{i\theta\emptyset t}$.

**Condition Two** A symmetric argument applies to condition 2. If physician $j$ in panel space $q$ matches with patient type $\theta$ then $v_{\theta jqt} - \tau^v_{\theta jt} > v_{\theta jq\emptyset t}$. The one sided waiting condition implies that $\tau^v_{\theta jt} \geq 0$. Therefore, $v_{\theta jqt} > v_{\theta jq\emptyset t}$.

**Condition Three** Patient $i$ matches with physician $j'$ if $u_{i\theta j't} - \tau^u_{\theta j't} > u_{i\theta jt} - \tau^u_{\theta jt}$ for all $j \neq j'$. Physician $j$ in panel space $q$ matches with patient type $\theta'$ if $v_{\theta' jqt} - \tau^v_{\theta' jt} > v_{\theta jqt} - \tau^v_{\theta jt}$ for all $\theta \neq \theta'$. The one-sided waiting condition requires that $\tau^u_{\theta jt} = 0$ or $\tau^v_{\theta jt} = 0$. Therefore, either

$$u_{i\theta j't} - \tau^u_{\theta j't} > u_{i\theta jt}$$

or

$$v_{\theta' jqt} - \tau^v_{\theta' jt} > v_{\theta jqt}$$

Further, as both $\tau^u_{\theta j't} \geq 0$ and $\tau^v_{\theta' jt} \geq 0$, one of the following two conditions hold.

$$u_{i\theta j't} > u_{i\theta jt}$$
$$v_{\theta' jqt} > v_{\theta jqt}$$

Thus, if patient $i$ matches with physician $q'$ and physician space $q$ matches with a patient of type $\theta'$, then either patient $i$ prefers physician $j'$ to $j$ or physician $j$ prefers patient type $\theta'$to patient type $\theta$ for panel space $q$.

# G Gradient and Standard Errors

## G.1 Gradient

I calculate the gradient to aid the algorithm.

In order to do so, I rewrite the inner loop system in matrix form. I introduce the variable $W^{\boldsymbol{\beta}}_{\theta jt} = \mathbf{1}\Big\{\mu^{\boldsymbol{\beta}}_{\theta\emptyset t} a^{\boldsymbol{\beta}}_{\theta jt} < \mu^{\boldsymbol{\beta}}_{\emptyset jt} b^{\boldsymbol{\beta}}_{\theta jt}\Big\}$. This variable describes which side of the market is driving matching patterns for each patient type-physician pair. The likelihood function can be written in terms of $W^{\boldsymbol{\beta}}_{\theta jt}$ and the predicted outside option shares only.

$$\tilde{\mathcal{L}}(\boldsymbol{\beta}) = 2\sum_t \sum_j \sum_\theta \mu_{\theta jt}\Bigg(W^{\boldsymbol{\beta}}_{\theta jt} log(\mu^{\boldsymbol{\beta}}_{\theta\emptyset t} exp(\delta^{\boldsymbol{\beta}}_{\theta jt})) + (1 - W^{\boldsymbol{\beta}}_{\theta jt}) log(\mu^{\boldsymbol{\beta}}_{\emptyset jt} exp(\gamma^{\boldsymbol{\beta}}_{\theta jt}))\Bigg)$$

$$+ \sum_t \sum_\theta \mu_{\theta\emptyset t} log(\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}) + \sum_t \sum_j \mu_{\emptyset jt} log(\mu^{\boldsymbol{\beta}}_{\emptyset jt})$$

Now, take a parameter element that is in the patient utility function: $\beta \in \boldsymbol{\beta}^u$. The derivative of the likelihood function with respect to that parameter is:

$$\frac{\partial \tilde{\mathcal{L}}(\boldsymbol{\beta})}{\partial \beta} = 2\sum_t \sum_j \sum_\theta \mu_{\theta jt} W^{\boldsymbol{\beta}}_{\theta jt} \frac{\partial \delta^{\boldsymbol{\beta}}_{\theta jt}}{\partial \beta}$$

$$+ \sum_t \sum_\theta \frac{1}{\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}} \frac{\partial \mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}{\partial \beta} \Big(\mu_{\theta\emptyset t} + 2\sum_j W^{\boldsymbol{\beta}}_{\theta jt} \mu_{\theta jt}\Big) + \sum_t \sum_j \frac{1}{\mu^{\boldsymbol{\beta}}_{\emptyset jt}} \frac{\partial \mu^{\boldsymbol{\beta}}_{\emptyset jt}}{\partial \beta} \Big(\mu_{\emptyset jt} + 2\sum_\theta (1 - W^{\boldsymbol{\beta}}_{\theta jt}) \mu_{\theta jt}\Big)$$

Similarly, if the parameter is in the physician utility function ($\beta \in \boldsymbol{\beta}^v$), then

$$\frac{\partial \tilde{\mathcal{L}}(\boldsymbol{\beta})}{\partial \beta} = 2\sum_t \sum_j \sum_\theta \mu_{\theta jt} (1 - W^{\boldsymbol{\beta}}_{\theta jt}) \frac{\partial \gamma^{\boldsymbol{\beta}}_{\theta jt}}{\partial \beta}$$

$$+ \sum_t \sum_\theta \frac{1}{\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}} \frac{\partial \mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}{\partial \beta} \Big(\mu_{\theta\emptyset t} + 2\sum_j W^{\boldsymbol{\beta}}_{\theta jt} \mu_{\theta jt}\Big) + \sum_t \sum_j \frac{1}{\mu^{\boldsymbol{\beta}}_{\emptyset jt}} \frac{\partial \mu^{\boldsymbol{\beta}}_{\emptyset jt}}{\partial \beta} \Big(\mu_{\emptyset jt} + 2\sum_\theta (1 - W^{\boldsymbol{\beta}}_{\theta jt}) \mu_{\theta jt}\Big)$$

The derivatives have two phrases. The first phrase is the *direct effect* of the change in preferences on the choices made. The second is the *indirect effect* of changing preferences on the matching equilibrium. The direct effect can be easily computed, as mean preferences are linear in $\beta$. Computing the indirect effect involves finding the derivatives of the predicted outside options $\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}$ and $\mu^{\boldsymbol{\beta}}_{\emptyset jt}$.

To find the derivatives of the outside options with respect to $\beta$, I rewrite the equilibrium conditions in terms of $W^{\boldsymbol{\beta}}_{\theta jt}$.

$$\mu^{\beta}_{\theta\emptyset t} + \sum_j \left(W^{\beta}_{\theta jt}\mu^{\beta}_{\theta\emptyset t}a^{\beta}_{\theta jt} + (1 - W^{\beta}_{\theta jt})\mu^{\beta}_{\emptyset jt}b^{\beta}_{\theta jt}\right) = n_{\theta t}\forall\theta\forall t$$

$$\mu^{\beta}_{\emptyset jt} + \sum_\theta \left(W^{\beta}_{\theta jt}\mu^{\beta}_{\theta\emptyset t}a^{\beta}_{\theta jt} + (1 - W^{\beta}_{\theta jt})\mu^{\beta}_{\emptyset jt}b^{\beta}_{\theta jt}\right) = Q_j\forall j\forall t$$

This can be written in matrix form:

$$R^{\beta}\boldsymbol{\mu}^{\beta}_{\emptyset} = \boldsymbol{v}$$

where

$$\boldsymbol{\mu}^{\beta}_{\emptyset} = \begin{array}{c} \mu^{\beta}_{1\emptyset t} \\ \mu^{\beta}_{2\emptyset t} \\ \vdots \\ \hline \mu^{\beta}_{\emptyset 1t} \\ \mu^{\beta}_{\emptyset 2t} \\ \vdots \end{array}, \boldsymbol{v} = \begin{array}{c} n_{1t} \\ n_{2t} \\ \vdots \\ \hline Q_{1t} \\ Q_{2t} \\ \vdots \end{array}$$

$$R^{\beta} = \left.\begin{array}{ccc|ccc} 1 + \sum_j W^{\beta}_{1jt}a^{\beta}_{1jt} & 0 & \cdots & (1 - W^{\beta}_{11t})b^{\beta}_{11t} & (1 - W^{\beta}_{12t})b^{\beta}_{12t} & \cdots \\ 0 & 1 + \sum_j W^{\beta}_{2jt}a^{\beta}_{2jt} & \cdots & (1 - W^{\beta}_{21t})b^{\beta}_{21t} & (1 - W^{\beta}_{22t})b^{\beta}_{22t} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ \hline W^{\beta}_{11t}a^{\beta}_{11t} & W^{\beta}_{21t}a^{\beta}_{21t} & \cdots & \sum_\theta 1 + (1 - W^{\beta}_{\theta 1t})b^{\beta}_{\theta 1t} & 0 & \cdots \\ W^{\beta}_{12t}a^{\beta}_{12t} & W^{\beta}_{22t}a^{\beta}_{22t} & \cdots & 0 & \sum_\theta 1 + (1 - W^{\beta}_{\theta 2t})b^{\beta}_{\theta 2t} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{array}\right.$$

The derivative of $W^{\beta}_{\theta jt}$ with respect to any $\beta \in \boldsymbol{\beta}$ is zero with probability one, due to the existence of continuous characteristics in the utility specifications.

Recall that the equilibrium conditions can be written in matrix form.

$$R^{\beta}\boldsymbol{\mu}^{\beta}_{\emptyset} = \boldsymbol{v}$$

In this notation, I need to find the object $\frac{\partial \boldsymbol{\mu}^{\beta}_{\emptyset}}{\partial \boldsymbol{\beta}}$. I apply the implicit function theorem to derive this object:

Define the function $F(\boldsymbol{\beta}, \boldsymbol{\mu}) = R^{\beta}\boldsymbol{\mu}^{\beta}_{\emptyset} - \boldsymbol{v}$.

The implicit function theorem implies that

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} = -J^{-1}_{F,\mu}\frac{\partial F(\boldsymbol{\beta}, \boldsymbol{\mu})}{\partial \boldsymbol{\beta}}$$

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} = -R(\boldsymbol{\beta})^{-1}\frac{\partial R(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\boldsymbol{\mu}$$

### G.2 Standard Errors

There are three alternatives for the estimation of standard errors for parameter estimates in the main matching model. I present the methodology to calculate each in this section. Currently, the reported standard errors are the outer product of gradients standard errors.

**Outer Product of Gradients** In the main specification, the outer product of gradients estimator of the variance matrix is:

$$\hat{\Phi}(\hat{\boldsymbol{\beta}}) = \Big[\sum_t \Big(2\sum_j\sum_\theta \mu_{\theta jt}\hat{g}_{\theta jt}\hat{g}'_{\theta jt} + \sum_\theta \mu_{\theta\emptyset t}\hat{g}_{\theta\emptyset t}\hat{g}'_{\theta\emptyset t} + \sum_j \mu_{\emptyset jt}\hat{g}_{\emptyset jt}\hat{g}'_{\emptyset jt}\Big)\Big]^{-1}$$

where

$$\hat{g}_{\theta jt} = \frac{\partial ln(\mu^{\boldsymbol{\beta}}_{\theta jt})}{\partial\boldsymbol{\beta}} = \left(W^{\boldsymbol{\beta}}_{\theta jt}\left(\frac{1}{\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}\frac{\partial\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}{\partial\boldsymbol{\beta}} + \frac{\partial\delta^{\boldsymbol{\beta}}_{\theta jt}}{\partial\boldsymbol{\beta}}\right) + (1 - W^{\boldsymbol{\beta}}_{\theta jt})\left(\frac{1}{\mu^{\boldsymbol{\beta}}_{\emptyset jt}}\frac{\partial\mu^{\boldsymbol{\beta}}_{\emptyset jt}}{\partial\boldsymbol{\beta}} + \frac{\partial\gamma^{\boldsymbol{\beta}}_{\theta jt}}{\partial\boldsymbol{\beta}}\right)\right)$$

$$\hat{g}_{\theta\emptyset t} = \frac{\partial ln(\mu^{\boldsymbol{\beta}}_{\theta\emptyset t})}{\partial\boldsymbol{\beta}} = \frac{1}{\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}\frac{\partial\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}{\partial\boldsymbol{\beta}}$$

$$\hat{g}_{\emptyset jt} = \frac{\partial ln(\mu^{\boldsymbol{\beta}}_{\emptyset jt})}{\partial\boldsymbol{\beta}} = \frac{1}{\mu^{\boldsymbol{\beta}}_{\emptyset jt}}\frac{\partial\mu^{\boldsymbol{\beta}}_{\emptyset jt}}{\partial\boldsymbol{\beta}}$$

All objects have previously been derived in G.1.

**Empirical Hessian** The estimated hessian matrix of the log likelihood could also be used:

$$\bar{H}(\hat{\boldsymbol{\beta}}) = \Big[\sum_t \Big(2\sum_j\sum_\theta \mu_{\theta jt}\frac{\partial^2 ln(\mu^{\boldsymbol{\beta}}_{\theta jt})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta'}} + \sum_\theta \mu_{\theta\emptyset t}\frac{\partial^2 ln(\mu^{\boldsymbol{\beta}}_{\theta\emptyset t})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta'}} + \sum_j \mu_{\emptyset jt}\frac{\partial^2 ln(\mu^{\boldsymbol{\beta}}_{\emptyset jt})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta'}}\Big)\Big]^{-1}$$

For ease of exposition, I show the matrix-by-scaler derivations of the second derivatives. Take $(\beta_0, \beta_1) \in \boldsymbol{\beta}$,

$$\frac{\partial^2 ln(\mu^{\boldsymbol{\beta}}_{\theta jt})}{\partial\beta_0\partial\beta_1} = \left(W^{\boldsymbol{\beta}}_{\theta jt}\left(\frac{1}{\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}\frac{\partial^2\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}{\partial\beta_0\partial\beta_1} - \left(\frac{1}{\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}\right)^2\frac{\partial\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}{\partial\beta_0}\frac{\partial\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}{\partial\beta_1}\right) + (1 - W^{\boldsymbol{\beta}}_{\theta jt})\left(\frac{1}{\mu^{\boldsymbol{\beta}}_{\emptyset jt}}\frac{\partial^2\mu^{\boldsymbol{\beta}}_{\emptyset jt}}{\partial\beta_0\partial\beta_1} - \left(\frac{1}{\mu^{\boldsymbol{\beta}}_{\emptyset jt}}\right)^2\frac{\partial\mu^{\boldsymbol{\beta}}_{\emptyset jt}}{\partial\beta_0}\frac{\partial\mu^{\boldsymbol{\beta}}_{\emptyset jt}}{\partial\beta_1}\right)\right)$$

$$\frac{\partial^2 ln(\mu^{\boldsymbol{\beta}}_{\theta\emptyset t})}{\partial\beta_0\partial\beta_1} = \frac{1}{\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}\frac{\partial^2\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}{\partial\beta_0\partial\beta_1} - \left(\frac{1}{\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}\right)^2\frac{\partial\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}{\partial\beta_0}\frac{\partial\mu^{\boldsymbol{\beta}}_{\theta\emptyset t}}{\partial\beta_1}$$

$$\frac{\partial^2 ln(\mu^{\beta}_{\emptyset jt})}{\partial\beta_0\partial\beta_1} = \frac{1}{\mu^{\beta}_{\emptyset jt}}\frac{\partial^2\mu^{\beta}_{\emptyset jt}}{\partial\beta_0\partial\beta_1} - \left(\frac{1}{\mu^{\beta}_{\emptyset jt}}\right)^2\frac{\partial\mu^{\beta}_{\emptyset jt}}{\partial\beta_0}\frac{\partial\mu^{\beta}_{\emptyset jt}}{\partial\beta_1}$$

Therefore, the $i, j^{th}$ element of the matrix $\bar{H}(\boldsymbol{\beta})$ can be written as follows:

$$\bar{h}(\beta_i,\beta_j) = \sum_t\left(2\sum_j\sum_\theta \mu_{\theta jt}\left[W^{\beta}_{\theta jt}\left(\frac{1}{\mu^{\beta}_{\theta\emptyset t}}\frac{\partial\mu^{\beta}_{\theta\emptyset t}}{\partial\boldsymbol{\beta}}\right)+(1-W^{\beta}_{\theta jt})\left(\frac{1}{\mu^{\beta}_{\emptyset jt}}\frac{\partial\mu^{\beta}_{\emptyset jt}}{\partial\boldsymbol{\beta}}\right)\right]+\sum_\theta \mu_{\theta\emptyset t}\frac{1}{\mu^{\beta}_{\theta\emptyset t}}\frac{\partial\mu^{\beta}_{\theta\emptyset t}}{\partial\boldsymbol{\beta}}+\sum_j \mu_{\emptyset jt}\frac{1}{\mu^{\beta}_{\emptyset jt}}\frac{\partial\mu^{\beta}_{\emptyset jt}}{\partial\boldsymbol{\beta}}\right)$$

$$\bar{h}(\beta_i,\beta_j) = \sum_t\left(\sum_\theta\left(\mu_{\theta\emptyset t}+2\sum_j W^{\beta}_{\theta jt}\mu_{\theta jt}\right)\left(\frac{1}{\mu^{\beta}_{\theta\emptyset t}}\frac{\partial\mu^{\beta}_{\theta\emptyset t}}{\partial\boldsymbol{\beta}}\right)+\sum_j\left(\mu_{\emptyset jt}+2\sum_j(1-W^{\beta}_{\theta jt})\mu_{\theta jt}\right)\left(\frac{1}{\mu^{\beta}_{\emptyset jt}}\frac{\partial\mu^{\beta}_{\emptyset jt}}{\partial\boldsymbol{\beta}}\right)\right)$$

All objects have previously been derived except for the second derivatives. Recall that the derivative of $\boldsymbol{\mu}^{\beta}_{\emptyset}$ with respect to $\beta_0$was found to be:

$$\frac{\partial\boldsymbol{\mu}^{\beta}_{\emptyset}}{\partial\beta_0} = -[R^{\beta}]^{-1}\frac{\partial R^{\beta}}{\partial\beta_0}\boldsymbol{\mu}^{\beta}_{\emptyset}$$

where

$$\boldsymbol{\mu}^{\beta}_{\emptyset} = \frac{\begin{array}{c}\mu^{\beta}_{1\emptyset t}\\ \mu^{\beta}_{2\emptyset t}\\ \vdots\end{array}}{\begin{array}{c}\mu^{\beta}_{\emptyset 1t}\\ \mu^{\beta}_{\emptyset 2t}\\ \vdots\end{array}}, R^{\beta} = \left.\begin{array}{ccc|ccc} 1+\sum_j W^{\beta}_{1jt}a^{\beta}_{1jt} & 0 & \cdots & (1-W^{\beta}_{11t})b^{\beta}_{11t} & (1-W^{\beta}_{12t})b^{\beta}_{12t} & \cdots\\ 0 & 1+\sum_j W^{\beta}_{2jt}a^{\beta}_{2jt} & \cdots & (1-W^{\beta}_{21t})b^{\beta}_{21t} & (1-W^{\beta}_{22t})b^{\beta}_{22t} & \cdots\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots\\ \hline W^{\beta}_{11t}a^{\beta}_{11t} & W^{\beta}_{21t}a^{\beta}_{21t} & \cdots & \sum_\theta 1+(1-W^{\beta}_{\theta 1t})b^{\beta}_{\theta 1t} & 0 & \cdots\\ W^{\beta}_{12t}a^{\beta}_{12t} & W^{\beta}_{22t}a^{\beta}_{22t} & \cdots & 0 & \sum_\theta 1+(1-W^{\beta}_{\theta 2t})b^{\beta}_{\theta 2t} & \cdots\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{array}\right.$$

Taking the second derivative:

$$\frac{\partial^2\boldsymbol{\mu}^{\beta}_{\emptyset}}{\partial\beta_0\partial\beta_1} = [R^{\beta}]^{-1}\left(\frac{\partial R^{\beta}}{\partial\beta_1}[R^{\beta}]^{-1}\frac{\partial R^{\beta}}{\partial\beta_0}\boldsymbol{\mu}^{\beta}_{\emptyset} - \frac{\partial^2 R^{\beta}}{\partial\beta_0\partial\beta_1}\boldsymbol{\mu}^{\beta}_{\emptyset} - \frac{\partial R^{\beta}}{\partial\beta_0}\frac{\partial\boldsymbol{\mu}^{\beta}_{\emptyset}}{\partial\beta_1}\right)$$

**Sandwich Estimator** Given these derivations, the sandwich estimator of the variance matrix has the following form.

$$\hat{V}_{sandwich} = \frac{1}{\sum_t\left(\sum n_{\theta t}+\sum Q_{jt}+\sum n_{\theta t}Q_{jt}\right)}\bar{H}(\hat{\boldsymbol{\beta}})^{-1}\hat{\Phi}(\hat{\boldsymbol{\beta}})\bar{H}(\hat{\boldsymbol{\beta}})^{-1}$$

## H Reduced Form Evidence: Alternative Payment Models

In this section, I present reduced form evidence on the impact of the alternative payment models on physician outcomes. I exploit the quasi-experimental variation in the data caused by the staggered switching of physicians between payment models. I use a two-way fixed effects regression methodology to account for physician- and year- specific effects. I study physician-level outcomes, such as the total number of patients the physician matches with (panel size), and the characteristics of patients in the panel.

**Sample** This analysis uses a physician-year level dataset of physician characteristics and physician-patient matches. The dataset includes comprehensive care primary care physicians in Ontario from 2004 to 2014 with more than 300 patients. I restrict the sample further to only physicians who stay in the same location throughout the entire sample. The final sample includes 633 physicians and 4,496 physician-years.

**Model** The regression is specified as a two-way fixed effects model:

$$W_{jy} = \alpha_j + \lambda_y + \beta' \boldsymbol{c}_{jy} + \epsilon_{jy}$$

where $W_{jy}$ is an outcome variable of physician $j$ in year $y$. $\boldsymbol{c}_{jy}$ is a vector of indicators for physician $j$'s payment model.

**Results** Table 18 contains the key results. Physicians who switch from fee-for-service to an alternative payment model are estimated to increase the number of patients on their roster. I estimate that physicians increase their panel size by 6.6% in capitation and 16.0% in enhanced fee-for-service relative to when they are in fee-for-service. The results largely confirm the qualitative conclusions of the main specification. The magnitudes of these estimates, however, are larger than implied magnitudes of the main specification results.

Additionally, the fixed effect model results suggest that physicians have mildly different matching patterns in capitation models than in fee-for-service and enhanced fee-for-service models. I find a positive and significant association between capitation and a panel of patients who are older, sicker, provide more revenue.

Table 18: Regression Results

(a) Panel Observables

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | | Share of patients in panel | | |
| | $log(m_{jt} - \mu_{j\emptyset})$ | Over 50 | Female | Has Comorbidity |
| Capitation | 0.066*** | 0.041*** | −0.001 | 0.020*** |
| | (0.021) | (0.004) | (0.002) | (0.003) |
| EFFS | 0.160*** | 0.007* | −0.0003 | 0.012*** |
| | (0.021) | (0.004) | (0.002) | (0.003) |
| Mean of Dep. Var: | 6.788 | 0.472 | 0.557 | 0.243 |
| Observations | 4,492 | 4,492 | 4,492 | 4,492 |
| $R^2$ | 0.883 | 0.933 | 0.961 | 0.865 |
| Adjusted $R^2$ | 0.863 | 0.922 | 0.955 | 0.842 |
| Residual Std. Error | 0.205 | 0.041 | 0.020 | 0.030 |
| *Note:* | | | | *p<0.1; **p<0.05; ***p<0.01 |

(b) Expected Revenue and Visits

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Average of patients in panel | | | |
| | Cap. Revenue | EFFS Revenue | FFS Revenue | Visits |
| Capitation | 2.364*** | 1.714*** | 1.370*** | 0.057*** |
| | (0.660) | (0.541) | (0.459) | (0.017) |
| EFFS | −0.644 | −0.253 | −0.218 | −0.003 |
| | (0.651) | (0.533) | (0.453) | (0.017) |
| Mean of Dep. Var: | 181.808 | 156.123 | 126.916 | 4.323 |
| Observations | 4,492 | 4,492 | 4,492 | 4,492 |
| $R^2$ | 0.863 | 0.876 | 0.880 | 0.882 |
| Adjusted $R^2$ | 0.840 | 0.855 | 0.860 | 0.862 |
| Residual Std. Error | 6.409 | 5.248 | 4.458 | 0.165 |
| *Note:* | | | | *p<0.1; **p<0.05; ***p<0.01 |