Medical Testing as Optimal Information Acquisition*

Kyeongbae Kim†

August 13, 2021
Click here for the latest version.

Abstract

This paper formulates a model of optimal information acquisition, by developing a dynamic structural model with observation and adjustment costs. It then applies the empirical framework to medical testing in which a medical provider must balance the competing goals of making informed treatment decisions and saving testing costs for a diabetic patient's health. The novel feature of the model is that the state, a patient's blood sugar level, is not precisely observed by a provider. If she chooses not to pay an observation cost for medical testing, the true state remains unobserved and the following treatment adjustments may be inaccurate. I combine the dynamic structural model with confidential administrative data on patient health from universal health screening and testing decisions. I find that the higher cost of blood sugar testing leads to blood sugar levels that are more dispersed over time through ill-informed prescription adjustments. Counterfactual exercises show that performing an A1C test, the state-of-art method of blood sugar testing, every six months is the most cost-effective diabetes management considering its monetary costs and health benefits, with an additional cost of $11,018.66 for one extra quality-adjusted life-year.

*I thank Ali Hortaçsu, Fernando Alvarez, Stéphane Bonhomme, and Pietro Tebaldi for guidance and support. I thank Katherine Baicker, José Ignacio Cuesta, Angela Denis Pagliero, Elbert Huang, Joonhwi Joo, Chulhee Lee, Neale Mahoney, Karthik Nagarajan, Young Taeck Oh, Chad Syverson, James Traina, and seminar participants at the University of Chicago and Seoul National University for comments and suggestions. I acknowledge financial support from the Department of Economics at the University of Chicago and the IO Initiative of the Becker Friedman Institute. I am grateful to the National Health Insurance Service for help with accessing the data, specifically Jerim Cho and Seongyong Park. This paper uses data provided by the National Health Insurance Service (NHIS-2019-1-251). The views expressed herein are those of the author and do not reflect those of the National Health Insurance Service.

†Department of Economics, Sejong University. kimkb@sejong.ac.kr

1 Introduction

An economic agent often chooses to acquire information to resolve uncertainty. However, s/he may not always find it cost-efficient to obtain too much information. This paper formulates a model of optimal information acquisition, by developing a dynamic structural model with observation and adjustment costs. It then applies the empirical framework to medical testing in which a medical provider must balance the competing goals of making informed treatment decisions and saving testing costs. The novel feature of the model is that the state, patient health status, is not precisely observed by a provider. If she chooses not to pay an observation cost for medical testing, the true state remains unobserved and the following treatment adjustments may be inaccurate. It describes an economic mechanism of how a provider's endogenous testing decisions affect the patient's health status.

To empirically analyze the trade-off, I combine the dynamic structural model with confidential administrative data on diabetic patients' health and providers' testing decisions. Diabetes is an ideal example of information acquisition. A patient's blood sugar level is not only his/her health status a provider would like to control, but the information acquired through a blood test. Korean National Health Insurance beneficiaries' data consist of universal health screening data linked with medical claims data; blood sugar levels, and testing and prescription choices are observed. Diabetes is also one of the most widespread diseases requiring regular attention. Four hundred twenty-two million adults are estimated to live with diabetes, and 3.7 million annual deaths are caused by high blood glucose worldwide (World Health Organization, 2016). Since high blood sugar levels lead to many complications, including heart attacks, strokes, kidney failure, and blindness, it is important to monitor glucose levels properly through a blood test and adjust prescriptions whenever needed.

I first document substantial heterogeneity in blood glucose levels and testing frequency. Although an A1C test that monitors patients' average glucose levels represents "effective care," its frequency is highly heterogeneous across providers. Providers causing an additional dispersion of a patient's blood sugar level over time tend to under-test, compared with clinical guidelines.

Based on reduced-form evidence, I develop a dynamic structural model with observation and adjustment costs (Alvarez, Lippi and Paciello, 2011; Alvarez, Guiso and Lippi, 2012). The model features two choices—observation through medical testing and prescription adjustments. A medical provider would like to keep a patient's blood sugar level around the target over his/her lifetime by making observations and adjustment choices each period. A provider facing the trade-off optimally decides whether to perform the medical test. Observing the true state of the patient, the blood glucose level, is helpful with not only checking whether current prescriptions are working properly, but having a better prediction of the state in the future. However, a

fixed cost of observation must be incurred for the blood test. Without the test, she observes a noisy signal. Once the observation decision is made, she forms the best prediction using all test results and noisy signals she has ever observed. Based on the prediction of the state, she then optimally chooses whether to adjust existing prescriptions to keep the blood sugar levels under control. A fixed cost of adjustment must be paid whenever prescriptions become more or less aggressive than the previous one.

The model rationalizes interesting distributional and dynamic features of the data. First, prescription adjustments are not necessarily toward the target level if not accompanied by preceding observation choice. This is the economic mechanism in which a large observation cost leads to an additional dispersion of blood sugar levels over time, deteriorating patient health. Second, the belief on the state contains both state- and time-dependent components (Alvarez, Lippi and Passadore, 2017). The Kalman filter estimate describing the best prediction of the state reflects the true state, through signals she has received. However, under the normality assumption, its prediction error is purely a function of the agent's past observation choices, and it increases if not tested. The time- and state-dependent components rationalize the provider's tendency to perform a test regularly to keep the prediction error small and the situation in which a medical provider confirms her belief through medical testing whenever she believes the state is far from the target. I provide empirical evidence that supports the model implications.

I structurally estimate the model using data on blood sugar levels, testing, and prescription adjustments. Based on observation and adjustment frequencies, identifying two fixed cost components (Alvarez, Lippi and Paciello, 2011), I exploit patient moves in the matched provider-patient data to cluster patient and provider types simultaneously, to account for provider/patient heterogeneity. I then estimate structural parameters of interest for each combination of patient and provider types. Structural estimates suggest that heterogeneous frequencies are attributable to both provider and patient factors, and the fixed costs of observation substantially differ across both provider and patient types. Fixed costs of adjustment, however, are largely different across patient types They suggest a substantial portion of patients are reluctant to receive a blood test and take a new set of medications, and a large number of providers are unwilling to order the newly employed A1C test for diabetes management. I also find that prescription choices without medical testing are inaccurate across all combinations of patient-provider pairs. As a patient's blood sugar levels fluctuate highly over time, past test results quickly become uninformative.

Using the structural model, I conduct a counterfactual exercise to quantify the value of medical testing by making it mandatory. I find that performing an A1C test every six months is the most cost-effective diabetes care considering both health outcome improvements and monetary

cost of testing: with an incremental cost-effectiveness ratio of $11,018.66 per quality-adjusted life-year (QALY), far lower than the $100,000 threshold. It confirms current clinical guidelines that the A1C test must be performed at least twice a year from an economic perspective. It also indicates that current quality measures that count annual A1C testing rates are not sufficient; one may consider biannual testing rates for proper quality assessment instead.

This paper relates to two literature strands. First, it contributes to literature on costly information acquisition, including rational inattention (Sims, 2003; Alvarez, Lippi and Paciello, 2011; Alvarez, Guiso and Lippi, 2012; Matějka and McKay, 2015; Steiner, Stewart and Matějka, 2017; Joo, 2020), search and consideration set models (Hortaçsu and Syverson, 2004; Hong and Shum, 2006; De los Santos, Hortaçsu and Wildenbeest, 2012; Honka, 2014; Honka, Hortaçsu and Vitorino, 2017; Hortaçsu, Madanizadeh and Puller, 2017; Abaluck and Adams, 2018; Agarwal et al., 2020), and Bayesian learning (Erdem and Keane, 1996; Ackerberg, 2003; Crawford and Shum, 2005; Ching, 2010; Dickstein, 2018; Currie and MacLeod, 2020). Although it has been anecdotally known or argued that economic agents make choices considering the trade-off between benefits and costs of acquiring information, it is very rare for an information acquisition behavior to be directly and cleanly observed in observational data. The context, setup, and data I use in this paper overcome such limitations, allowing me to estimate a dynamic choice model of costly information acquisition using novel data on information acquisition behavior, subsequent choices, and a state variable that can be observed through the information-acquiring decisions. Combined with testing decisions which are clearly the action of information acquisition, as well as fasting blood sugar levels which serve as both a diabetic patient's health status and information acquired through a blood test, I empirically examine the role of information and its acquisition.

Second, this study contributes to extant literature on variations and efficiency of healthcare, including geographic variation (Zhang, Baicker and Newhouse, 2010*a*,*b*; Finkelstein, Gentzkow and Williams, 2016, 2018), provider practice style (Chandra and Staiger, 2007; Currie, MacLeod and Van Parys, 2016; Molitor, 2018; Cutler et al., 2019), and productivity (Baicker and Chandra, 2004; Chandra et al., 2016). I contribute to this strand of literature by proposing a unifying framework of testing, treatment, and clinical outcomes, and using an objective health measure to study the variations and their implications. Despite the importance of medical testing in not only accounting for healthcare utilization but affecting health outcomes, how medical providers decide to test has received little attention. Recent papers on over- and under-testing (Abaluck et al., 2016; Mullainathan and Obermeyer, 2020) abstract from corresponding treatment decisions by studying specific situations, where subsequent choices are completely determined

by test results.[1] In a more general setting, however, a medical provider is free to make any treatment decisions with or without raw test results. I explicitly disentangle testing and treatment decisions, by modeling a provider's belief on a patient's health status, which can be updated through medical testing. In addition, a fasting blood sugar level is an ideal healthcare performance measure. Because it is independently observed by universal health screening and its measurement does not reflect provider-specific measurement errors (Song et al., 2010), heterogeneity in blood sugar levels must reflect differential healthcare performance.

The remainder of the paper is organized as follows. In section 2, I describe institutional details and data. In section 3, I provide reduced-form evidence for the role of medical testing in healthcare performance. In section 4, I develop a dynamic model with observation and adjustment costs, and in Section 5, I provide structural estimation results. In section 6, I conduct counterfactual experiments. Section 7 concludes.

2 Background and Data

2.1 Diabetes

Diabetes is a chronic disease in which a patient experiences high blood sugar levels over a long period. It occurs either when the pancreas cannot produce insulin (Type 1 diabetes), or when the insulin does not lower the blood sugar level properly (Type 2 diabetes).[2] Increased blood sugar levels have cumulative health effects on the human body and lead to many complications, such as heart attacks, strokes, kidney failure, and blindness.[3] Four hundred twenty-two million adults worldwide were estimated to have diabetes in 2014.

It is important to keep glucose levels under control to reduce the health risks. A patient is diagnosed with diabetes when the fasting blood sugar level is greater than or equal to 126mg/dL. To eliminate direct symptoms and reduce long-term complications, a provider is recommended to set a target blood sugar level for a patient. The American Diabetes Association recommends 80-130mg/dL as the target preprandial glucose level.[4] There are a variety of anti-diabetic medications lowering blood glucose levels.[5] Multiple medications might be prescribed if monotherapy

[1] Oster, Shoulson and Dorsey (2013) provide a theoretical model of an individual who chooses to learn his true state through testing, although there is no subsequent treatment decision.

[2] I focus on patients with Type 2 diabetes, who are managed most commonly with oral medications.

[3] An individual diagnosed with diabetes at age 40 is estimated to lose 11.6 life-years for men and 14.3 for women (Narayan et al., 2003).

[4] This might be tightened or relaxed depending on a patient's condition. Treatment guidelines of the Korean Diabetes Association suggest the same criteria and target levels (Korean Diabetes Association, 2015).

[5] The Health Insurance Review and Assessment Service (HIRA) of Korea categorizes the medications into the following nine classes: Biguanides, Sulfonylureas, DPP-4 inhibitors, Insulin, Alpha-Glucosidase Inhibitors,

is insufficient to control blood glucose levels (Powers, 2015; World Health Organization, 2016; American Diabetes Association, 2019; Korean Diabetes Association, 2019).

The medical community advises a provider to perform regular blood tests to monitor a patient's blood glucose level. A variety of blood tests measure blood sugar levels with different levels of accuracy and convenience. To measure a *fasting blood sugar level,* the patient is required to fast at least eight hours before blood is drawn. *The random test* is done at any time of the day and its result can fluctuate due to food and physical activity. *The A1C test,* which measures the percentage of glucose attached to hemoglobin, represents an average blood sugar level over the past two to three months. Its reliability and convenience of not requiring fasting have made the medical community recommend the A1C test for diagnosis from 2009 (International Expert Committee, 2009). Current clinical guidelines advise a medical provider to "perform the A1C test at least two times a year" in usual diabetic patients and "quarterly in patients whose therapy has changed or who are not meeting glycemic goals." In 2019, the A1C test became the primary technique to assess diabetes management. It can be measured through either laboratory tests or point-of-care (POC) meters. Latter semi-quantitative methods tend to be less accurate (American Diabetes Association, 2019).[6][7]

2.2 Healthcare System in South Korea[8]

South Korea has a single-payer healthcare system. The National Health Insurance Act requires all Korean citizens to be covered by either National Health Insurance (NHI) or Medical Aid Program.[9] The National Health Insurance Service (NHIS) provides health insurance coverage to all NHI beneficiaries as a single health insurance agency. All NHI beneficiaries receive the same coverage, regardless of their enrollment status.[10]

Healthcare provision is heavily regulated in South Korea. Providers are paid based on a fee-for-service system; each service provided is paid for separately.[11] Fees and coinsurance vary only based on provider categories—clinics, hospitals, general hospitals, and tertiary hospitals.

Thiazolidinediones, Nonsulfonylurea secretagogues, GLP-1 agonists, and SGLT-2 inhibitors.

[6]Food and Drug Administration (2016) requires 95% and 99% of glucose readings from self-monitoring blood glucose test systems to be within 15% and 20% of the true value, respectively.

[7]Theranos, which falsely claimed to have developed finger-stick testing to detect health problems, is currently under regulatory and criminal review New York Times (2016).

[8]See Kwon, Lee and Kim (2015) for further details.

[9]2.9% of the population were Medical Aid beneficiaries in 2017.

[10]Insurance premiums are based on income and wealth, proportional to wage income for employees. For self-employed people, both income and property values, such as houses and vehicles, are counted. Non-working dependents are also covered without additional premiums in either case.

[11]The bundled payment system, also referred to as the Diagnosis Related Group (DRG)-based prospective payment system, was introduced in 2013 to seven disease categories that require surgical treatment, such as cataract procedures and cesarean sections.

Fees are negotiated between the government and provider associations annually. Providers have little financial incentive regarding prescribing behavior. The 2000 amendment to the Pharmaceutical Affairs Act enforces mandatory separation of prescribing and dispensing.

As universal health screening to NHI beneficiaries, the National Health Screening Program (NHSP) was introduced in 1995.[12] The NHSP offers biennial health screening to all NHI beneficiaries who pay the insurance premium, including their dependents aged 40 and older. Non-white-collar workers are offered health screenings each year, consisting of standardized tests and body measurements, such as blood and urine tests, blood pressure, eye and radiology examinations, height, and weight. It also includes a survey on family health history, exercise, cigarette smoking, and alcohol consumption. Participation was 78.5% as of 2017.

2.3 Data

I use population-level Korean National Health Insurance and Medical Aid beneficiaries' data from 2002-2017 and 2006-2017, respectively. For each beneficiary, I observe demographics, including sex, year of birth, enrollment and disability status, and insurance premiums, which proxy income. The county of residence is observed each month.[13] For each medical provider, I observe the county in which it is located, the category of the provider, its specialty, and the number of physicians.[14] These are linked with National Health Screening data and medical claims data. The National Health Screening dataset includes a fasting blood sugar level, a main outcome variable of the paper, in addition to the date the health screening was conducted. For each patient visit, the medical claims dataset provides both patient and provider identifiers, the date of visit, and a list of diagnosis codes. I observe all treatment codes, such as medical tests and prescriptions provided to the patient.

From population-level data, I draw 3 million diabetic patients, 5.8% of the population. For a complete treatment history of patients with Type 2 diabetes, I drop all patients treated as a diabetic in the first year in the data, because they may have been diagnosed and treated before they appeared in the data. I exclude all beneficiaries who did not have medical claims with the diagnosis code for diabetes for two consecutive years after first diagnosed and treated. I also drop patients with the diagnosis code for Type 1 diabetes over more than two-thirds of their treatment history. Among patients who had been prescribed anti-diabetic medications for more than a year, I retain patient-quarters whose medical or prescription records are observed each quarter.[15]

[12] It was expanded to Medical Aid beneficiaries in 2012.

[13] In this paper, the county stands for *si-gun-gu*-level.

[14] A physician identifier is not observed in data.

[15] To identify patients with diabetes from claims data and to avoid a situation in which a non-diabetic

A quarterly dataset is constructed for reliable treatment intervals. 97.73% of patient-quarters had at least one provider visit throughout the quarter.[16] A medical test is considered to have been conducted in the quarter if its record exists during the period. For prescription adjustment, I choose a prescription record with the longest prescription period of the quarter. After constructing nine indicators of anti-diabetic medications,[17] an indicator of prescription adjustment is defined as whether at least one of the nine indicatior variables changes in comparison to those of the last quarter.[18] Quarterly claims data are linked with a fasting blood sugar level measured through National Health Screening. Instead of actual testing results performed by a provider during a patient visit, the health screening dataset provides the health measure that would have been observed if a patient had been tested, regardless of whether s/he was tested.[19][20] Table 1 provides an overview of the datasets.

[Table 1 about here.]

3 Heterogeneity in Medical Testing

3.1 Heterogeneity in Healthcare Performance

[Figure 1 about here.]

The American Diabetic Association recommends 80-130mg/dL as the general preprandial glycemic target (American Diabetes Association, 2019). Figure 1 shows, however, a substantial portion of patients whose fasting blood sugar levels did not fall into the target level. Patient-level median glucose levels also show striking dispersion. 50.11% and 49.80% of patient-years and patients' glucose levels are not in the target range, respectively.

To investigate the source of the dispersion, I estimate regression model:

beneficiary was misclassified as diabetic, a patient with diabetes in claims data is commonly defined as those who are diagnosed during at least one inpatient stay or at least two outpatient visits in one year (Hux et al., 2002; Miller, Safford and Pogach, 2004; Rector et al., 2004; Asghari et al., 2009). The criterion imposed in the paper satisfies the common operational definition.

[16]Frequent provider visits in Korea are often at the expense of short consultation time, a so-called "three-minute consultation." See OECD (2010); Moon (2012).

[17]Following the Health Insurance Review and Assessment Service (HIRA)'s criteria, I consider nine classes of medications: Biguanides, Sulfonylureas, DPP-4 inhibitors, Insulin, Alpha-Glucosidase Inhibitors, Thiazolidinediones, Nonsulfonylurea secretagogues, GLP-1 agonists, and SGLT-2 inhibitors.

[18]It effectively rules out brand or dosage changes, as long as the classes of medications remain constant. During the first period, it is equal to 1 whenever a patient received at least one class of anti-diabetic drug.

[19]A patient must inform a provider directly about a health screening result if s/he wishes, because the result is not automatically passed to the provider.

[20]Panel A of Table A.1 shows that variables constructed from the claim data explain little about participation in health screenings, mitigating concern on potential sample selection due to its compliance.

$$y_{it} = \alpha_i + \psi_{j_{it}} + x_{it}\beta + \varepsilon_{it}, \tag{1}$$

where y_{it} is an outcome variable of interest for patient i at t. α_i and ψ_j represent additive effects for patient i and provider j, respectively. x_{it} indicates time-varying patient characteristics. j_{it} and p_{it} are indicators of the provider treating i at t and the place where i lives at t, respectively. ε_{it} is an unobserved time-varying error term. To measure health performance, I use the logarithm of blood sugar level as the outcome variable. The clinical outcome is measured from laboratory testing, which is unlikely affected by a diverse provider practice style. Health measures constructed from claim data are known to suffer from provider-specific measurement errors (Song et al., 2010). A patient's age and its square are included as controls.

The identifying variation of patient and provider heterogeneity, α_i and ψ_j, comes from patient moves between providers. Since Abowd, Kramarz and Margolis (1999)'s findings on decomposing worker and firm contributions to a wage distribution, it has been applied to a variety of settings, including health economics, industrial organization, and marketing (Bronnenberg, Dubé and Gentzkow, 2012; Skipper and Vejlin, 2015; Finkelstein, Gentzkow and Williams, 2016, 2018; Allcott et al., 2019). Identification relies on *network exogeneity.* Let D_{ijt} be an indicator of patient-provider network and $D_{ijt} = 1$ if and only if patient i is treated by j at t: $j_{it} = j$. Identification requires $E[\varepsilon_{it}|D, x, \alpha, \psi] = 0$.[21] The model allows patient-provider network D to correlate with patient and provider heterogeneity, α and ψ. It would be violated, however, if time-varying shocks affected by whom patients choose to be treated. For example, it permits the case in which an inherently sick patient (too low or high α_i) is treated by a certain provider, characterized by ψ_j. It does not allow for the situation in which a bad ε_{ijt} shock makes him/her switch providers.

The two samples (i.e., full and migrants) have their own advantages. The number of movers in the full sample is large, and as described below, it is less likely to suffer from limited mobility bias. In the migrants subsample, patient moves in the subsample are less likely attributable to the time-varying shock, ε_{ijt}, and moves are more likely exogenous. Survey data from the Korea Labor & Income Panel Study suggest that only 4.6% of households moved to current locations due to environmental or health issues.[22] It follows that the portion of household members who switched providers due to health concerns, not environmental issues, is even lower. Some household members might have moved to current locations due to a family member's health concerns, not their own. In both datasets, singleton observations are dropped for reliable estimations (Correia, 2015, 2016).

[21] Identification of Equation (1) requires no autocorrelation upon movers.

[22] Finkelstein, Gentzkow and Williams (2016, 2018) also use patient migration for causal interpretations.

When estimating the regression model (1) (AKM, hereafter), *limited mobility bias* arises if a network is not well-connected and the number of movers is small (Andrews et al., 2008, 2012; Bonhomme, 2020; Lamadon, Mogstad and Setzler, 2019). Table 1 shows that the average number of movers in the full sample and migrants subsample are 23.6503 and 2.2240, respectively. I consider Bonhomme, Lamadon and Manresa (2019*a*)'s classification method (BLM, hereafter) to address the concern. In the first step, I classify providers into K groups based on the empirical distribution of y_{it}, using the k-means clustering algorithm. In the second step, I estimate α_i and $\psi_{k(j)}$, where $k(j)$ is a clustering indicator. I consider $K = 10$ clusters for estimation.[23] I view discretized provider heterogeneity as an approximation of true provider heterogeneity Bonhomme, Lamadon and Manresa (2019*b*).

Equation (1) rationalizes the dispersion by allowing the conditional mean function to differ by patient and provider identifiers:

$$E[y_{it}|D, x, \alpha, \psi] = \alpha_i + x_{it}\beta + \psi_j.$$

The source of variation can be examined using the standard variance decomposition formula:

$$Var(y_{it}) = Var(\alpha_i + x_{it}\beta) + Var(\psi_{j_{it}}) + 2Cov(\alpha_i + x_{it}\beta, \psi_{j_{it}}) + Var(\varepsilon_{it}). \tag{2}$$

In addition to patient and provider heterogeneity contributing to the mean level of the outcome variable, I also consider potential heteroskedasticity in the error term by allowing its variance to differ by provider:[24]

$$Var(\varepsilon_{it}|j_{it} = j) = \sigma_j^2.$$

To identify providers' contribution to the dispersion of blood sugar levels from Equation (1), I adjust the variance of the error term so that each $\tilde{\varepsilon}_{it}$ satisfies the following equality:

$$\hat{Var}(\tilde{\varepsilon}_{it}|j_{it} = j) = \min_{j'} \hat{Var}(\varepsilon_{it}|j'), \tag{3}$$

for each j. This removes provider-level heteroskedasticity from contributing to the dispersion, $Var(\varepsilon_{it}|j_{it} = j)$, by adjusting the variance of the error term so it is not larger than $\min_{j'} Var(\varepsilon_{it}|j')$. Unlike the standard variance decomposition, it informs to what extent het-

[23] See Table A.2 and Table A.3 for robustness checks.

[24] Estimates of the variance of the error term, $\hat{\sigma}_j^2$, would be biased if patients and providers are sorted differentially. Table 2 shows, however, no evidence of provider-patient sorting, as provider and patient effects are uncorrelated. Figure A.1 provides an additional evidence of the no sorting. Except one provider class accounting for less than 1% of observations, there are no substantial differences of distributions of patients' blood sugar levels across provider classes. Richer specification of the variance allowing for potential sorting would be achieved by allowing it to differ by not only provider, but patient identity.

eroskedasticity explains overall dispersion. The minimum value of $\hat{Var}(\varepsilon_{it}|k')$ is chosen for the BLM estimates. For AKM estimates, I choose the 25th percentile of $\hat{Var}(\varepsilon_{it}|j')$ to account for potential sampling variability.[25]

[Table 2 about here.]

Table 2 shows that although provider heterogeneity is not particularly suited to explaining mean blood sugar levels, it contributes to the dispersion of a patient's blood sugar level over time. Panel A of Table 2, for the full sample, shows that patient fixed effects, α_i, explain substantial variation. Using the variance decomposition formula in Equation (2), 49.81% and 2.58% of patient and provider fixed effects explain the variance, respectively. Panel B of Table 2 shows estimates from the migrants subsample. Similar to Panel A of 2, 50.73% and 0.74% of the variance is explained by patient and provider fixed effects, respectively. When accounting for heteroskedasticity, however, provider heterogeneity explains larger portions of the variance. Table 2 shows that the variance decreases by 9.06% (Panel A) and 8.56% (Panel B), respectively, when I adjust heteroskedastic errors using Equation (3). Overall, provider heterogeneity explains 10.69% (Panel A) and 10.88% (Panel B) of the total variance, respectively.[26]

3.2 Testing and Prescription Adjustments

[Table 3 about here.]

I now document heterogeneity in testing and prescription adjustments, and their relationship with provider heterogeneity identified in Section 3.1. Table 3 documents heterogeneous and sub-optimal testing behavior. Panel A of Table 3 shows testing frequencies averaged by patient-year. Unlike the medical community's advice to perform an A1C test at least twice a year, A1C monitoring is conducted far below the guideline. 27.71% of patient-years did not get the test at all, and 27.08% received it once a year. Quantitative laboratory testing, which can usually be conducted to measure a patient's fasting blood sugar level, was performed much less. 48.86% did not take quantitative testing. Semi-quantitative tests, such as point-of-care meters, were more likely to be carried out. Panel B of Table 3 shows that the sub-optimal testing frequency is attributable to providers. Averaging them by provider-quarter, 21.87% performed the A1C test to 20% or less of their patients each quarter. They can thus be considered not to be treating their patients based on the guideline, since those who conducted the A1C test on each patient

[25]Since small estimated variance might reflect sample variability, I do not adjust it if it is smaller than the threshold value, $\hat{Var}(\varepsilon_{it}|j')$, to avoid artificially amplifying its dispersion. Table A.2 indicates that BLM estimates with the full sample show similar results.

[26]This is robust to alternative specifications. See Table A.2.

once a year were expected to conduct it to approximately 25% of their patients each quarter. This is consistent with medical literature on variations of testing frequencies (Yoo et al., 2017).

[Figure 2 about here.]

Other notable features are infrequent prescription adjustments. Often referred to as *clinical inertia* in medical literature (Phillips et al., 2001), providers are not willing to adjust prescriptions over time. Figure 2 shows survival analysis results for prescription adjustments. The median time of prescription adjustment is 5 quarters, suggesting that providers' reluctance to consider prescription adjustments when needed might cause suboptimal healthcare performance.

[Table 4 about here.]

Table 4 shows that low-performing providers, which set higher target levels and let patients' blood sugar levels fluctuate over time, tended to under-test. To investigate relationships between provider practice patterns and provider performance, I consider provider fixed effects, $\hat{\psi}_j$, and heteroskedastic standard deviation of the error term, $\hat{\sigma}_j$, as dependent variables. For regressors, I use the number of medical testing and prescription adjustments throughout a year, with a maximum of four. For testing, I consider the number of quarters the A1C test was performed. Panel A of Table 4 shows that providers who conducted an A1C test less frequently tended to set higher target levels (high $\hat{\psi}_j$). Panel B shows that they were also more like to let patients' blood sugar levels fluctuate over time, characterized by the large standard error of the error term ($\hat{\sigma}_j$). These results are robust to the addition of another regressor. Although poor diabetic management appears to associate with infrequent testing, its interpretation requires caution. Testing and prescription adjustments might reflect not only provider preference, but patient characteristics.

[Table 5 about here.]

Table 5 shows which clinics tended to under-test. I examine characteristics of testing clinics, since provider-level characteristics are more informative in clinics that consist of a small number of physicians. As clinic-level characteristics, I consider three types of regressors—year of entry, whether it specializes in internal medicine, and whether it has an A1C testing device. In South Korea, the qualification test for medical specialists is regulated by the government (Kwon, Lee and Kim, 2015). To measure the year of entry, I construct two variables—an indicator of whether it existed as of 2002, and if not, from which month it appears in data. These are proxies for when a physician was trained. The indicator of having the testing device is zero either if

50% or more A1C tests were outsourced or if it did not conduct the test throughout a year. Table 5 shows that those specializing in internal medicine tended to offer more frequent A1C tests, and those who entered recently were more likely to perform the test. These imply that being knowledgeable on diabetes care and its current trends is a predictor of testing providers. Those who can perform the A1C test without outsourcing is also a predictor of providers who test more frequently. The fact that it is equipped with the testing device, however, might reflect a clinic's preference for the A1C test.

3.3 Discussion

I document heterogeneity in healthcare performance and its potential mechanisms using a standardized health measure linked with matched provider-patient health data. Regression analysis suggests that dispersion of mean blood sugar levels is largely attributable to patient heterogeneity. There is, however, substantial provider heterogeneity when explaining dispersion in blood sugar levels over time. Dispersion is particularly large in those who performed medical tests infrequently, evidencing that providers' less-than-optimal testing behavior made the clinical outcome poorly controlled over time, leading to low healthcare performance.

Although regression results describe the mechanism of heterogeneous healthcare performance, the research questions of this paper have not yet been addressed. First, heterogeneity identified from Equation (1) themselves reveal nothing regarding economic mechanisms that drive dispersion of the clinical outcome. Examining predictors of heterogeneity provides only partial answers, since medical testing is an endogenous decision of a provider concerning patient health. Second and relatedly, they do not allow me to answer how an alternative allocation or policy affects healthcare performance. During counterfactual exercises, it is necessary to be explicit about an economic mechanism regarding how medical testing makes providers make more informed treatment choices. Third, they do not explain the distributional and dynamic features of data. The standard variance decomposition formula (2) does not include the heteroskedastic nature of healthcare performance. Equation (1) does not inherently permit potential state dependence of the outcome variable over time. To create an empirical framework that not only explains distributional and dynamic features but clarifies how economic mechanisms affect healthcare performance, I develop and estimate a dynamic choice model with observation and adjustment costs in the following sections.

4 A Dynamic Model with Observation and Adjustment Costs

I present a model that rationalizes a medical provider's testing and prescription adjustment behavior presented in the last section, serving three purposes. First, it proposes a unifying framework that relates medical testing, treatment decisions, and a clinical outcome. It explicitly models a provider's endogenous information acquisition, which affects consequent treatment choices and outcomes, thereby providing an empirical model regarding why testing behavior affects healthcare performance. Second, its structural estimates permit welfare calculations and policy-relevant counterfactual analysis in Section 6. Third, I combine the model with novel data on observations, adjustments, and states to estimate a dynamic model with observation and adjustment costs, which has drawn economists' attention to rationalizing information frictions. The model presented below is an extension of Alvarez, Lippi and Paciello (2011) and Alvarez, Guiso and Lippi (2012)'s model with observation and adjustment costs. In their model, an agent optimally chooses to observe the state by incurring a fixed observation cost, and adjusts a choice variable to minimize expected discounted losses. I apply this model by interpreting a provider's medical testing and prescription adjustments as observation and adjustment behaviors, respectively, where the provider's objective is to maintain the state, a blood sugar level, around the target level.

4.1 Setting

Let y_t be the true state, the logarithm of the patient's blood sugar level at the beginning of the period t. $y_t + a_t$ is the true state at the end of the period t, after prescription adjustment a_t is made. A medical provider's objective is to maximize a patient's expected discounted utility. Her per-period utility, $u(y_t + a_t; \theta)$, is a symmetric and concave function whose unique maximum is $y_t + a_t = y^*$, the target level. Described in Section 2.1, this represents her long-term goal because keeping a patient's blood sugar level close to the target is important to prevent complications from non-normal blood glucose levels. I assume that the utility function is quadratic:

$$u(y_t + a_t; \theta) = -\gamma(y_t + a_t - y^*)^2, \tag{4}$$

where γ is a scaling factor. It can be understood as a quadratic approximation to more general functional form of the per-period utility. She would optimally set $y_t + a_t = y^*$, or equivalently, $a_t = y^* - y_t$ each period by prescribing medication if she did not face any information frictions, discussed below.

Choice variable a_t denotes a prescription adjustment relative to the last period. If a_t is positive, y_{t+1} is expected to increase because treatment becomes less aggressive than that of the last one. This can be accomplished by adding or removing molecules from the last prescription.[27]The (log of the) blood sugar level follows a random walk:

$$y_{t+1} = y_t + a_t + w_t, \tag{5}$$

where $w_t \sim N(0, \sigma_w^2)$ denotes the i.i.d. error term. y_1 is also normally distributed with

$$y_1 \sim N(E[y_1], \sigma_{y1}^2). \tag{6}$$

Each period, a provider faces two decision problems: medical testing and adjustment choices during the first and second stages, respectively.Figure 3 describes a decision tree. Her value function is:

$$V(z^{t-1}, a^{t-1}) = \max\{\underbrace{-F_O + E[U(y_t, z^{t-1}, a^{t-1})|z^{t-1}, a^{t-1}]}_{\text{observe } y_t}, \underbrace{E[U(z_{0t}, z^{t-1}, a^{t-1})|z^{t-1}, a^{t-1}]}_{\text{do not observe} y_t}\},$$

where

$$U(z_t, z^{t-1}, a^{t-1}) = \max\{\underbrace{\max_{a_t}\{-F_A - \gamma E\left[(y_t + a_t - y^*)^2|z^t, a_t, a^{t-1}\right] + \beta V(z^t, a_t, a^{t-1})\}}_{\text{adjust } a_t},$$
$$\underbrace{-\gamma E\left[(y_t + 0 - y^*)^2|z^t, 0, a^{t-1}\right] + \beta V(z^t, 0, a^{t-1})}_{\text{do not adjust } a_t}\},$$

denotes the utility she faces during the second stage, and $V(z^t, a_t, a^{t-1}) \equiv V(z^t, a^t)$, F_O and F_A stand for fixed costs of observation and adjustment, respectively, and z_t denotes a signal about the true state y_t. The value function depends on all past signals $z^t = (z_1, ..., z_t)$ and prescription adjustments $a^t = (a_1, ..., a_t)$.[28]

During the first stage, she decides which signal $z_t \in \{y_t, z_{0t}\}$to observe; she chooses whether to perform a medical test to observe exact y_t and incurs fixed cost of observation F_O. Without the test, she observes a noisy signal potentially based on communication with the patient or simple finger-prick testing:

$$z_{0t} = y_t + v_{0t}, \tag{7}$$

[27]An advantage of using the log-transformed value of the blood sugar level is that it rationalizes progression of diabetes. Without any prescription adjustment, the mean of the raw blood sugar levels increase over time due to its log-normality: $E[\exp(y_t)] = E[\exp\left(y_1 + \sum_{s=1}^{t-1} w_s\right)] = \exp\left(E[y_1] + 0.5\left(\sigma_{y1}^2 + (t-1)\sigma_w^2\right)\right)$.

[28]In problems of imperfect state information, they are state variables from the agent's viewpoint (Bertsekas, 2017).

where $v_{0t} \sim N(0, \sigma_{v0}^2)$ is the i.i.d. noise term. After the observation decision is made, she forms beliefs on the current blood sugar level, $y_t|z^t = (z_1, ..., z_t)$ for subsequent treatment decision. Each z_s is either y_s or z_{0s}, depending on her testing choices at s. The belief at $t = 1$ coincides with Equation (6).

After she chooses to receive $z_t \in \{y_t, z_{0t}\}$ and forms beliefs, she decides whether to adjust existing prescriptions during the second stage. She chooses a prescription adjustment a_t. Due to symmetry and concavity of the per-period utility, the optimal adjustment is $a_t^* = y^* - E[y_t|z^t, a^{t-1}]$. She incurs fixed adjustment cost F_A if she makes an adjustment.[29] Her decision rule of the adjustment is to change the prescription if the utility of adjustment is greater than that of no adjustment:

$$\underbrace{-F_A - \gamma E\left[(y_t + a_t^* - y^*)^2|z^t, a_t^*, a^{t-1}\right] + \beta V(z^t, a_t^*, a^{t-1})}_{\text{adjust } a_t} > \underbrace{-\gamma E\left[(y_t + 0 - y^*)^2|z^t, 0, a^{t-1}\right] + \beta V(z^t, 0, a^{t-1})}_{\text{do not adjust } a_t},$$

and

$$\underbrace{-F_A + \beta V(z^t, a_t^*, a^{t-1})}_{\text{adjust } a_t} > \underbrace{-\gamma E\left[(y_t - y^*)^2|z^t, a^{t-1}\right] + \beta V(z^t, 0, a^{t-1})}_{\text{do not adjust } a_t}.$$

During the first stage, uncertain about the true state and its noisy signal she could receive, (y_t, z_{0t}), she performs a medical test if its expected utility is higher:

$$\underbrace{-F_O + E[U(y_t, z^{t-1}, a^{t-1})|z^{t-1}, a^{t-1}]}_{\text{observe } y_t} > \underbrace{E[U(z_{0t}, z^{t-1}, a^{t-1})|z^{t-1}, a^{t-1}]}_{\text{do not observe } y_t}.$$

4.1.1 Kalman filtering

Normality of y_1, w_t, and v_{0t}, together with the quadratic utility specification, make the problem more tractable. In addition to the fact that both $y_t|(z^t, a^{t-1})$ and $y_{t+1}|(z^t, a^t)$ are normal distributions, we have the recursive formula of the Kalman filter estimates and variances (Bertsekas, 2017):

$$\begin{aligned}
\hat{y}_{t|t} \equiv E[y_t|z^t, a^{t-1}] &= \frac{\hat{\sigma}_{t|t-1}^{-2}}{\hat{\sigma}_{t|t-1}^{-2}+\sigma_v^{-2}}\hat{y}_{t|t-1} + \frac{\sigma_v^{-2}}{\hat{\sigma}_{t|t-1}^{-2}+\sigma_v^{-2}} z_t \\
\hat{\sigma}_{t|t}^2 \equiv Var(y_t|z^t, a^{t-1}) &= \frac{1}{\hat{\sigma}_{t|t-1}^{-2}+\sigma_v^{-2}} \\
\hat{y}_{t+1|t} \equiv E[y_{t+1}|z^t, a^t] &= \hat{y}_{t|t} + a_t \\
\hat{\sigma}_{t+1|t}^2 \equiv Var(y_{t+1}|z^t, a^t) &= \hat{\sigma}_{t|t}^2 + \sigma_w^2,
\end{aligned} \tag{8}$$

[29] It rationalizes infrequent adjustment decisions, described in Section 3.2.

where $\sigma_v = 0$ if $z_t = y_t$ and $\sigma_v = \sigma_{v0}$ if $z_t = z_{0t}$. We also have a variance-bias decomposition of the expected values of the utility function:

$$E[\underbrace{u(y_t + a_t; \theta)}_{=-\gamma(y_t+a_t-y^*)^2} | z^t, a^t] = -\gamma\hat{\sigma}^2_{t|t} - \gamma(\hat{y}_{t|t} + a_t - y^*)^2,$$

where the first and second terms, $-\gamma\hat{\sigma}^2_{t|t}$ and $-\gamma(\hat{y}_{t|t}+a_t-y^*)^2$, are equal to 0 whenever observed and adjusted, respectively.

Several comments on these specifications are warranted. First, the updated formula for $\hat{y}_{t|t}$ is equal to the weighted average of the Kalman filter estimate of y_t at the end of the last period, $\hat{y}_{t|t-1}$, and a signal received in the current period, z_t. This places more weight on the signal whenever its variance is smaller, and especially $\hat{y}_{t|t} = y_t$ and $\hat{\sigma}^2_{t|t} = 0$ if $z_t = y_t$ because it is the exact information without noise. Second, due to normality of beliefs on y_t, which is its best prediction, $(\hat{y}_{t|t-1}, \hat{\sigma}^2_{t|t-1})$ contains all information for the best prediction of y_t at the end of period $t-1$. An immediate result is that the value function reduces to $V(z^{t-1}, a^{t-1}) = V(\hat{y}_{t|t-1}, \hat{\sigma}^2_{t|t-1})$. Third, the variance-bias decomposition result shows that each testing and adjustment decision corresponds to whether to remove variance $\hat{\sigma}^2_{t|t}$ or bias $(\hat{y}_{t|t} + a_t - y^*)^2$ terms of the expected utility function, at the expense of fixed observation and adjustment costs F_O and F_A, respectively. $\hat{\sigma}^2_{t+1|t} = \sigma^2_w$ if observed and $\hat{y}_{t+1|t} = y^*$if adjusted.

Combining these results, the agent solves a Bellman equation:

$$\begin{aligned}
V(\hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1}) &= \max\{\underbrace{E[-F_O + U(y_t, \hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1})|\hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1}]}_{\text{observe}}, \underbrace{E[U(z_{0t}, \hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1})|\hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1}]}_{\text{do not observe}}\} \\
&= \max\{E[\max\{\underbrace{-F_O - F_A + \beta V(y^*, \sigma^2_w)}_{\text{observe and adjust}}, \\
&\qquad \underbrace{-F_O - \gamma(y_t - y^*)^2 + \beta V(y_t, \sigma^2_w)}_{\text{observe but do not adjust}}\}|\hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1}], \\
&\qquad E[\max\{\underbrace{-F_A - \gamma\hat{\sigma}^2_{t|t} + \beta V(y^*, \hat{\sigma}^2_{t|t} + \sigma^2_w)}_{\text{do not observe but adjust}}, \\
&\qquad \underbrace{-\gamma\hat{\sigma}^2_{t|t} - \gamma(\hat{y}_{t|t} - y^*)^2 + \beta V(\hat{y}_{t|t}, \hat{\sigma}^2_{t|t} + \sigma^2_w)}_{\text{neither observe nor adjust}}\}|\hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1}]\}
\end{aligned}$$

where $\hat{y}_{t|t}$ and $\hat{\sigma}^2_{t|t}$ are obtained from (8) and $U(\cdot, \cdot, \cdot)$ is appropriately defined. Figure 3 depicts this two-stage decision problem.

[Figure 3 about here.]

4.2 Discussion

I discuss several implications of the model with observation and adjustment costs.

Comparison with Extant Literature. My model synthesizes two variants—adjustment cost models and models of information acquisition. Typical models with adjustment costs are a variant of an (s, S) inventory problem (Stokey, Lucas and Prescott, 1989; Stokey, 2008; Bertsekas, 2017). Endogenous information acquisition models include rational inattention models (Sims, 2003; Alvarez, Lippi and Paciello, 2011; Alvarez, Guiso and Lippi, 2012; Matějka and McKay, 2015; Steiner, Stewart and Matějka, 2017), search and consideration set models (Hortaçsu and Syverson, 2004; Hong and Shum, 2006; De los Santos, Hortaçsu and Wildenbeest, 2012; Honka, 2014; Honka, Hortaçsu and Vitorino, 2017; Hortaçsu, Madanizadeh and Puller, 2017; Abaluck and Adams, 2018; Agarwal et al., 2020), and Bayesian learning (Erdem and Keane, 1996; Ackerberg, 2003; Crawford and Shum, 2005; Ching, 2010; Dickstein, 2018; Currie and MacLeod, 2020). My two-stage model of optimal information acquisition shares a similarity with rational inattention models; an endogenous decision of information acquisition during the first stage affects beliefs for a standard choice problem during the second. However, unlike models of inattention, which rely on a specific parameterization of information costs, such as Shannon entropy, the cost of information is characterized by a fixed cost of observation. My parameterization of the costs of information and adjustment are closest to those of Alvarez, Lippi and Paciello (2011); Alvarez, Guiso and Lippi (2012)'s continuous time models with observation and adjustment costs, where an adjustment cannot be made unless the true state is observed. In addition to the two-stage nature of my discrete-time model, I also permit a free and noisy signal to identify inaccuracy of usual checkups over medical testing.[30]

Additional Dispersion from Information Frictions. Once the true state, a patient's blood glucose level, is observed through medical testing, a provider makes a perfect adjustment, in that the state at the end of period t is exactly equal to her target level: $\hat{y}_{t+1|t} = y_t + a_t = y^*$. The adjustment is thus always dispersion-reducing. If the test is not accompanied, however, the adjustment is made in that her belief on the true state at the end of the period equals the target level, $\hat{y}_{t+1|t} = \hat{y}_{t|t} + a_t = y^*$, and the true state after adjustment, $y_t + a_t$, is not necessarily equal to the target. If $y_t > y^* > \hat{y}_{t|t}$, she erroneously believes the patient needs less aggressive treatment $a_t > 0$ when his/her glucose level should be lowered. This gives rise to an additional dispersion of a patient's blood sugar level over time, identified in Section 3.1.

[30] Alvarez, Lippi and Paciello (2011) propose a continuous-time model with free signals in their appendix and provide analytical solutions for certain extreme cases.

In addition to the fixed costs that describe her tendency to observe and adjust, to what extent an adjustment without testing results in an extra dispersion is governed by two structural parameters—the standard deviation of a noisy signal, σ_{v0}, and the error term of the law of the motion, σ_w. Equation (8) shows that the accuracy of the belief when the adjustment is made, $\hat{\sigma}^2_{t|t}$, depends on them. The larger σ_{v0} is, the less informative the noisy signal is about the current glucose level. If σ_w is large, past signals become less informative when predicting the blood sugar level during the current period. In Section 5.3, I use structural estimates to show the extent the previous information becomes uninformative over time.

State- and Time-dependent Decision Rules. The model incorporates two important margins of medical testing and prescription adjustments called *state-* and *time-dependent* decision rules (Alvarez, Lippi and Passadore, 2017). Equation (8) shows that $\hat{y}_{t|t}$ and $\hat{y}_{t+1|t}$ depend on the state through $z_s = y_s + v_s$. A decision rule based on them is state-dependent in that sense. However, under the normality assumption $\hat{\sigma}^2_{t|t}$ and $\hat{\sigma}^2_{t+1|t}$ depend only on the agent's past choices and are unaffected by the true state. A cutoff decision rule solely based on them is, therefore, to wait until it reaches the cutoff level. As the variances grow deterministically until observed, the agent would perform it for every predetermined time interval.

[Figure 4 about here.]

The state vector of the model consists of $(\hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1})$, using both state- and time-dependent decision rules. Figure 4 shows net utility of medical testing and adjustment for each $(\hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1})$. Adjustment choice is largely state-dependent and she is willing to adjust prescriptions whenever beliefs about her patient's blood sugar level fall outside of the inaction region. Testing choice is, however, both state- and time-dependent. She is willing to observe by performing the test regularly (time-dependent) but at the same time, she would like to ensure her beliefs by observing the true state through medical testing if she believes the blood sugar level is poorly controlled (state-dependent). I provide empirical evidence on state- and time-dependent decision rules in Section 5.4.

Rational Expectations. The model assumes that an agent has rational expectations and recalls all signals she has ever received for the best prediction of the true state, the Kalman filter estimate. The rational expectations assumption accords with extant literature on consumer learning on medications when classes of drugs are considered (Erdem and Keane, 1996; Crawford and Shum, 2005; Ching, 2010; Dickstein, 2018). For the perfect recall, while this is reasonable to assume in this context since she observes a numeric value of a blood sugar reading whenever she performs a test and she can record all test results in a medical chart,

the normality assumption makes the prediction problem simpler. Instead of having to recall all past signals z^t and prescription adjustments a^t, Equation (8) shows the provider's updating formula of the best prediction is to compute the weighted average of the past Kalman filter estimate and a new signal. After the observation choice, the Kalman filter estimate equals y_t with zero prediction error and past information is no longer needed.

Choice of a_t. I assume that a perfect adjustment can be made; control variable a_t can take any real value and a blood sugar level during the next period changes exactly by that amount (5). The assumption is imposed mostly for convenience, but it appears in several observations of diabetes management. First, a provider can always choose more aggressive treatment by adding more classes of anti-diabetic medications (American Diabetes Association, 2019; Korean Diabetes Association, 2019).[31] Second, providers are aware of the effectiveness of their prescriptions, considering the large number of patients each provider treats and with help from clinical guidelines.[32] Third, a provider can confirm resulting changes to a patient's blood glucose level from another blood test and modify prescriptions if they do not reach the desired level. If dispersion of $y_t + a_t$ remains after the adjustment, it might bias some structural parameters, such as the standard deviation of the error term of the law of motion (σ_w), since its identification argument relies on the fact that dispersion after both and adjustment reflect variation in the error term.

Patient-Provider Relationship. A simple interpretation of the model is for a provider to behave in a patient's best interest and their interests align, which can be far from realistic. My view is to interpret key parameters, including fixed costs of medical testing and adjustment, as a result of provider-patient relationship and negotiation. A patient might express a preference or physical condition during a provider visit. A provider considers the preferences of both sides, and physical condition, when making medical decisions.

From this viewpoint, a provider might face heterogeneous fixed costs and target levels across patients. To dissociate provider and patient contributions to structural parameters, I simultaneously identify both provider and patient heterogeneity using matched provider-patient data. As in Section 3, the identifying variation comes from patient moves. I revisit the bi-clustering technique for classifying provider-patient types in Section 5.3.

[31] See Section 2.1.

[32] Table 1 shows that each provider treated 77.28 patients, on average, during the sample period 2002 to 2017. Clinical guidelines contain A1C reduction from each class of agents, ranging from A1C 0.5% to 2% (Powers, 2015; Korean Diabetes Association, 2019). They can be translated into pre-breakfast plasma glucose 14-54mg/dL, but they should be interpreted with caution (Rohlfing et al., 2002).

5 Estimation

5.1 Specification

The structural parameters of interest are fixed costs of observation and adjustment, F_O and F_A, mean and standard deviation of y_1, $E[y_1]$ and σ_{y1}, standard deviation of the error term of the law of motion (5) and a free signal (7), σ_w and σ_{v0}, and the target value and a scaling factor in the per-period utility function (4), y^* and γ. For the first- and second-stage decision problems, I add logit errors to rationalize idiosyncratic choices. The probabilities of observation and adjustment are:

$$\begin{aligned} Pr(obs_t = 1|\hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1}) &= \frac{\exp\left(-F_O + E[U(y_t, \hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1})|\hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1}]\right)}{\exp\left(-F_O + E[U(y_t, \hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1})|\hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1}]\right) + \exp\left(E[U(z_{0t}, \hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1})|\hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1}]\right)} \\ Pr(adj_t = 1|\hat{y}_{t|t}, \hat{\sigma}_{t|t}) &= \frac{\exp\left(-F_A + \beta V(y^*, \hat{\sigma}^2_{t|t} + \sigma^2_w)\right)}{\exp\left(-F_A + \beta V(y^*, \hat{\sigma}^2_{t|t} + \sigma^2_w)\right) + \exp\left(-\gamma(\hat{y}_{t|t} - y^*)^2 + \beta V(\hat{y}_{t|t}, \hat{\sigma}^2_{t|t} + \sigma^2_w)\right)}, \end{aligned} \tag{9}$$

where $U(z_t, \hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1})$ is now the inclusive utility after $z_t \in \{y_t, z_{0t}\}$ is observed:

$$\begin{aligned} U(z_t, \hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1}) &= \log(\exp(-F_A - \gamma\hat{\sigma}^2_{t|t} + \beta V(y^*, \hat{\sigma}^2_{t|t} + \sigma^2_w)) \\ &\quad + \exp(-\gamma\hat{\sigma}^2_{t|t} - \gamma(\hat{y}_{t|t} - y^*)^2 + \beta V(\hat{y}_{t|t}, \hat{\sigma}^2_{t|t} + \sigma^2_w))). \end{aligned}$$

An agent faces a maximization problem under uncertainty on not only $\{y_t, z_{0t}\}$, but the logit error during the second stage.

Estimation of the model requires data on the sequence of medical testing and adjustment indicators and the true state, $(obs_t, adj_t, y_t)_{t=1}^T$, where patient i's data are observed from $1 \leq t \leq T$ and i is omitted for notational convenience.[33] True state variable $(y_t)_{t=1}^T$ is observed potentially with some missing periods. Equation (8) shows that prediction errors $\hat{\sigma}_{t|t-1}$ and $\hat{\sigma}_{t|t-1}$ depend only on the agent's past testing choices. Kalman filter estimates $\hat{y}_{t|t-1}$ and $\hat{y}_{t|t}$ require information on the true state and signals $(y_s, v_s)_{s=1}^t$, and adjustment indicators. For $z_t \in \{y_t, z_{0t}\}$, the free signals v_{0s} are not observed by the econometrician and must be integrated over. The state variable y_t must also be integrated over whenever it is not observed. From an agent's viewpoint, $z_t|(\hat{y}_{t|t-1}, \hat{\sigma}^2_{t|t-1})$ is normally distributed with:

$$z_t|(\hat{y}_{t|t-1}, \hat{\sigma}_{t|t-1}) \sim N\left(\hat{y}_{t|t-1}, \hat{\sigma}^2_{t|t-1} - \hat{\sigma}^2_{t|t}\right). \tag{10}$$

The variance reflects a reduction in uncertainty by observing $z_t \in \{y_t, z_{0t}\}$ and the prediction

[33] Although I observe patient i's medical claims record up to T in data, the agent solves an infinite horizon problem.

error decreasing from $\hat{\sigma}^2_{t|t-1}$ to $\hat{\sigma}^2_{t|t}$. Expectations in (9) are taken with respect to (10).

Combining (9) and the distribution of the law of motion (5), the integrated likelihood for each patient is:

$$\begin{aligned} l &= \int f(y_1)^{d_1} \prod_{t=2}^{T} f(y_t|y_{t-1}+a_{t-1})^{d_t} \\ &\quad \cdot \prod_{t=1}^{T} Pr(obs_t=1|\hat{y}_{t|t-1},\hat{\sigma}_{t|t-1}) Pr(adj_t=1|\hat{y}_{t|t},\hat{\sigma}_{t|t}) dF(\hat{y}_{t|t-1},\hat{y}_{t|t},a_t)_{t=1}^{T}, \end{aligned} \tag{11}$$

where $a_0=0$, $d_t=1$ whenever y_t is observed by the econometrician. The likelihood is integrated over potentially unobserved beliefs $(\hat{y}_{t|t-1},\hat{y}_{t|t})_{t=1}^{T}$, which can be approximated by the simulated likelihood, discussed in Section 5.3.[34]

The value function in the likelihood (11) solves fixed point equation:

$$\begin{aligned} V(\hat{y}_{t|t-1},\hat{\sigma}_{t|t-1}) &= \log(\exp(\underbrace{-F_O + E[U(y_t,\hat{y}_{t|t-1},\hat{\sigma}_{t|t-1})|\hat{y}_{t|t-1},\hat{\sigma}_{t|t-1}]}_{\text{observe}}) \\ &\quad + \exp(\underbrace{E[U(z_{0t},\hat{y}_{t|t-1},\hat{\sigma}_{t|t-1})|\hat{y}_{t|t-1},\hat{\sigma}_{t|t-1}]}_{\text{do not observe}})) \\ &= \log(\exp(\log(\exp(\underbrace{-F_O - F_A + \beta E[V(y^*,\sigma_w^2)|\hat{y}_{t|t-1},\hat{\sigma}_{t|t-1}]}_{\text{observe and adjust}}) \\ &\quad + \exp(\underbrace{-F_O - E[\gamma(y_t-y^*)^2 + \beta V(y_t,\sigma_w^2)|\hat{y}_{t|t-1},\hat{\sigma}_{t|t-1}]}_{\text{observe but do not adjust}})) \\ &\quad + \exp(\log(\exp(\underbrace{-F_A - \gamma\hat{\sigma}^2_{t|t} + \beta E[V(y^*,\hat{\sigma}^2_{t|t}+\sigma_w^2)|\hat{y}_{t|t-1},\hat{\sigma}_{t|t-1}]}_{\text{do not observe but adjust}}) \\ &\quad + \exp(\underbrace{-\gamma\hat{\sigma}^2_{t|t} - E[\gamma(\hat{y}_{t|t}-y^*)^2 + \beta V(\hat{y}_{t|t},\hat{\sigma}^2_{t|t}+\sigma_w^2)|\hat{y}_{t|t-1},\hat{\sigma}_{t|t-1}]}_{\text{neither observe nor adjust}})), \end{aligned} \tag{12}$$

where expectation is taken with respect to (10). For each set of structural parameters, it can be computed through value function iterations.

5.2 Identification

I briefly discuss how data variation identifies structural parameters. Fixed costs are identified from observation and adjustment frequencies, up to scale. Alvarez, Lippi and Paciello (2011) show that F_A/F_O is identified from the ratio of the number of adjustments and observations, and F_A and F_O are functions of the frequencies of adjustments and observations multiplied by σ_w^2. The intuition is that an agent's observation and adjustment behaviors depend on how fast the true state changes over time (σ_w^2), and her costs from observation and adjustment decisions.

[34] Both $\hat{y}_{t|t}$ and $\hat{y}_{t+1|t}$ are observed by the econometrician when the agent chooses to observe y_t ($obs_t=1$) and it is not missing in the data. When the agent adjusts, $\hat{y}_{t+1|t}=y^*$ is observed.

This implies that fixed costs are identified up to scale without the state variable. γ is identified from idiosyncrasies of observation and adjustment decisions, since it is an inverse of the scale factor of the logit error.

Data on the true state identify other parameters. σ_w is identified from consecutive observations of y_t and y_{t+1}, where no adjustment is made during t. Dispersion of blood sugar level after observation and adjustment gives additional information on σ_w, since blood sugar levels reach the target from choices.[35] The distribution of y_{t_i}, in which both observation and adjustment are performed during $t_i - 1$, is centered at y^*, and therefore it identifies target level y^*. Once y^* and σ_w are identified, we can find σ_{v0} by examining the distribution of y_{s_i}, in which the only adjustment made was at $s_i - 1$, and the state is observed before $s_i - 1$. Its dispersion reflects both the per-period contribution (σ_w) to the law of motion (5) and the agent's best prediction not being accurate (σ_{v0}). How much a provider is willing to adjust when y_t is far from y^* also helps identify γ. $E[y_1]$ and σ_{y1} are identified from the distribution of y_{i1}.

5.3 Estimation

To estimate parameters in the model, I use the nested fixed point algorithm (NFXP) (Rust, 1987, 2000). Iskhakov et al. (2016) demonstrate that when a more efficient version of the value function iteration algorithm is adopted, its performance is comparable to the mathematical programming with equilibrium constraints (MPEC) method that Su and Judd (2012) suggested. As Rust (2000) recommended, I combine conventional value function iterations and Newton-Kantorovich iterations to solve the inner fixed point problem (12). I use cubic spline interpolation for two-dimensional state variables (Adda and Cooper, 2003).[36] $\beta = 0.985$ is used for estimation.

Data. I use the quarterly-level dataset described in Section 2.3. An indicator of observation, obs_{it}, is equal to 1 whenever patient i receives the A1C test during t. An adjustment indicator, adj_{it}, equals 1 if and only if the combination of anti-diabetic medication classes changes in comparison to that of the last period. y_{it} is the logarithm of patient i's fasting blood sugar level during t.

Grouped Fixed-Effects to Consider Discrete Heterogeneity. To permit heterogeneity

[35] We would identify σ_w from a_t were it observed from the data. In their continuous-time model without free signal z_{0t}, Alvarez, Le Bihan and Lippi (2016) show that σ_w^2 is identified from the variance of a_t and frequency of prescription adjustments. Their identification argument applies to conventional models with only adjustment costs. I use an indicator of adjustment, $adj_t = 1\{a_t \neq 0\}$, for estimation.

[36] $15 \times 15 = 225$ grid points are used.

documented in Section 3 and to address the issue of patient-provider relationships in Section 4.2, I consider a two-step grouped fixed-effects estimator proposed by Bonhomme, Lamadon and Manresa (2019*b*); Bonhomme (2020). They are fixed-effects in that they allow for arbitrary correlation with observables, but they are grouped fixed-effects because those in the same group share the same fixed effect. For estimtation, in the first step, providers and patients are simultaneously classified based on moments, and in the second, I estimate the model for each combination of provider and patient groups, identified from the clustering algorithm in the first step.

A bi-clustering method is used to exploit matched provider-patient data and identify provider and patient groups, which I call types. For provider type $k = 1, ..., K$ and patient type $l = 1, ..., L$, I simultaneously find grouping indicators $\hat{k}_j \in \{1, ..., K\}$ and $\hat{l}_i \in \{1, ..., L\}$, respectively, which minimize objective:

$$\min_{\left(\tilde{h}, \{k_j\}, \{l_i\}\right)} \sum_{i,t} \left\| h_{it} - \tilde{h}(k_{j_{it}}) - \tilde{h}(j_i) \right\|, \tag{13}$$

where h_{it} denote the moments for the clustering, and $\|\cdot\|$ stands for the Euclidean norm I consider $K = L = 2$ and then estimate the model for each subsample out of $K \times L = 4$. Only non-movers across provider types are used for structural estimation.[37]

[Table 6 about here.]

I use observation and adjustment frequencies as provider-patient specific moments h_{it}. The frequencies identify fixed costs of testing and adjustment (Section 5.2) and they are observed each period. Table 6 shows clustering results. Those with high testing and adjustment frequencies are called a high type, suggesting substantial variation of adjustment frequency across both patient and provider types.

Simulated Likelihood. I consider up to $2T + 2$ dimensional simulation draws for estimation.[38] For value function iterations, I use two-dimensional draws to approximate two expectations (12). To approximate the integrated likelihood (11), I simulate up to T-dimensional noisy signals, $(z_{0t})_{t=1}^{T}$. Missing y_t must also be simulated, since state variables depend on the true state (8) and potentially observed y_{t+1} centers on $y_t + a_t$. I also use another set of up to T

[37] The full dataset is used for bi-clustering, but for structural estimation, I use randomly drawn subsamples to reduce computational burden.

[38] $N = 500$ leaped Halton sequences are used to approximate the high-dimensional integral (Kocis and Whiten, 1997). My structural estimates are not sensitive to the number of simulation draws.

dimensional draws to simulate the true state. The simulation draws are drawn from Equation (5), assuming the model is correctly specified. The simulated likelihood for estimation is:

$$\begin{aligned} l(obs_i, adj_i, y_i) &= \tfrac{1}{S}\textstyle\sum_{s=1}^{S} f(y_1)^{d_1} \prod_{t=2}^{T} f(y_t|y_{t-1}^{(s)} + a_{t-1}^{(s)})^{d_t} \\ &\quad \cdot \textstyle\prod_{t=1}^{T} Pr(test_t|\hat{y}_{t|t-1}^{(s)}, \hat{\sigma}_{t|t-1}) Pr(adj_t|\hat{y}_{t|t}^{(s)}, \hat{\sigma}_{t|t}), \end{aligned}$$

where subscript (s) denotes a simulated variable. $y_{t-1}^{(s)} = y_{t-1}$ if not missing.

5.4 Additional Empirical Evidence

[Table 7 about here.]

Data exhibit distributional and dynamic relationships among testing, adjustment, and the true state. Table 7 shows that testing and adjustment choices are time- and state-dependent, respectively, discussed in Section 4.2. Panel A of Table 7 shows that past testing choice is a strong predictor of the current choice of testing across all bi-clustered subsamples. Due to a potential correlation between past choice and provider and patient heterogeneity, the full sample estimate suggests a positive association between past and current choices. Conditional on provider and patient heterogeneity, however, past and current choices associate negatively across all subsamples. The fact that a provider is unwilling to test for two consecutive periods is evidence that she considers past testing choices when making a testing choice and it is time-dependent. Panel B of Table 7 shows a U-shaped relationship between blood sugar levels and adjustment choices. If the blood sugar level is either too high or low, a provider is more willing to adjust, evidencing that adjustment choice is affected by the state and therefore is state-dependent.

5.5 Estimation Results

[Table 8 about here.]

Table 8 shows estimation results. Both F_O/γ and F_A/γ describe fixed costs of testing and adjustment, normalized into the logarithm of the blood sugar level. Since I use testing and adjustment frequencies for bi-clustering, high patient and provider types correspond to low fixed costs. Consistent with 6, the fixed costs of testing, F_O/γ, has substantial heterogeneity across both patient and provider types, demonstrating that both patients' reluctance to receive medical testing and providers' preference for the relatively new A1C test contribute to differential testing frequencies. Table 8 also shows that heterogeneity in adjustment costs, F_A/γ, is mostly attributable to patient types.

Target levels across patient and provider types, $\exp(y^*)$, are around the upper bound of the target range of the clinical guideline (80-130mg/dL), but they do not exhibit substantial variation across types. This result accords with the variance decomposition in Table 2; provider fixed effects ψ_j, denoting the target levels, explain little of the total variance of blood sugar levels.

Testing technology is informative across all samples. Since σ_{v0} represents the standard deviation of a noisy signal, it describes the extent to which standard communication with a patient without medical testing is less informative than the medical test. Conversely, it characterizes the relative accuracy of testing over verbal communication, and it is significant and positive across all samples during estimation.

[Figure 5 about here.]

Discussed in Section 4.2, σ_{v0} and the standard deviation of the error term of the law of motion, σ_w, affect the degree to which no testing is uninformative in comparison to a medical test. I plot $\hat{\sigma}^2_{t|t}$ over time to assess the extent medical testing is informative when predicting future states. I assume testing is performed only at $t = 10$. Figure 5 uses the full sample structural estimates to show how quickly $\hat{\sigma}^2_{t|t}$ returns to the long-term prediction error when a patient remains untested, which solves $\hat{\sigma}^2_{\cdot|\cdot} = \left(\hat{\sigma}^{-2}_{\cdot|\cdot} + \sigma^{-2}_{v0}\right)^{-1} + \sigma^2_w$ in Equation (5). At $t = 11$ and 12, $\hat{\sigma}^2_{t|t}$ is 74.17% and 94.63% of the long-term prediction error, respectively. It reaches 98.94% and 99.79% of the long-term variance at $t = 13$ and 14, respectively. It has large σ_{v0}, and σ_w makes the previous testing result quickly uninformative over time, underscoring the importance of regular testing. However, past signals are still informative. The long-term prediction is 55.88% of $\hat{\sigma}^2_v$, which is the prediction error were only the current signal used. I formally investigate these effects in a counterfactual exercise in Chapter 6.

5.5.1 Model Performance

I examine model performance by comparing observed and predicted outcomes, conditional on observed time periods, T_i. For each simulated patient $i = 1, ..., N$ with type $l = 1, ..., L$, I assign a type-k provider using estimates of the mobility model during $t = 1$. At the beginning of period $t \geq 2$, the mobility decision is made using mobility model estimates during $t \geq 2$. If patient i does not switch, he/she is treated by the same provider. T_i is drawn randomly from its empirical distribution.

[Table 9 about here.]

Table 9 shows that both observation and adjustment frequencies match well in real and simulated data. Conditional on patient and provider types, differences in observation and adjustment frequencies are approximately 0.2% to 0.8% and 0.8% to 2.1%, respectively. Incorporating the mobility model, differences are approximately 0.5% to 2.1%.

[Figure 6 about here.]

Figure 6 shows striking fit between real and simulated fasting blood sugar levels,across patient types.

6 Counterfactual Experiments

I quantify the effects of mandatory medical testing on blood sugar levels by considering two counterfactuals. First, medical testing must be performed every s periods, and second, it must be given before a prescription adjustment. This serves three purposes. First, such counterfactual exercises are policy-oriented. Identifying the health effects of A1C testing informs to what extent a policy must consider direct or indirect incentives for desired health outcomes. In the United States, the Healthcare Effectiveness Data and Information Set (HEDIS), maintained by the National Committee for Quality Assurance (NCQA), includes an annual A1C testing rate as a performance measure of diabetes care across providers. HEDIS measures are used to reward quality Medicare Advantage (MA) plans and providers.[39] In South Korea, the Health Insurance Review & Assessment Service (HIRA) uses annual testing rates for diabetes quality assessment, and subsidizes providers that offer quality diabetes management (Kim et al., 2017). Second, it permits assessment of the appropriateness of the current clinical guideline—perform A1C tests twice a year and each quarter in usual patients and those whose blood sugar levels appear to fluctuate (American Diabetes Association, 2019). Third, from a model perspective, they isolate the effects of imperfect information on health outcomes from other influences, such as heterogeneous target levels or adjustment costs across providers.

During the first mandatory testing counterfactual, testing is performed regardless of utility. Starting from $t = 1$, when a test was conducted, it is performed for predetermined time intervals. For example, for an exercise in which the test was given every six months, it is performed every odd period. I conduct the second experiment by prohibiting a provider from adjusting if no testing was performed during the first stage; once the provider decides not to test during the first stage, she cannot adjust during the second. I compare simulated blood sugar levels with

[39]Source: "Federal Affairs," NCQA, 2019, accessed at https://www.ncqa.org/public-policy/federal-work/ on 11/13/2019.

those from simulated data in which a provider chooses the medical testing optimally. As in Section **??**, I consider $N = 10000$ and $T = 40$ for each bi-clustered subsample to quantify its influence on QALE. I choose \$10.19 as A1C testing cost.[40]

[Table 10 about here.]

Table 10 shows counterfactual outcomes. QALE increases by 0.01 year, if he/she received a medical test each quarter or every six months (Panel A and B). Performing the test less frequently leads to a negligible effect in QALE (Panel C). Discussed in Section 5.5, this result might be attributed to the fact that prediction error quickly returns to its long-term limit. Prediction errors are 0% during the period the test is given, and 74.17% during the next period. If the test is not given for three or four consecutive periods, prediction errors are 98.94% and 99.79% of its long-term counterpart, respectively. Panel D of Table 10 shows that mandatory testing before an adjustment is bad for health. There is a small increase in testing frequency from 36.88% to 37.66%, but testing frequency decreases significantly from 12.80% to 7.89%. Thus, testing requirements before adjustment discourage an adjustment choice, thus leading to poor diabetes management.

To assess their cost-effectiveness, I compute incremental cost-effectiveness ratios (ICER): that of incremental cost (\$) to incremental effectiveness (quality-adjusted life-year). It is considered cost-effective if the cost per QALY gained is lower than a certain threshold (Neumann, Cohen and Weinstein, 2014). Table 10 shows that, except mandatory testing before adjustment (Panel D), all ICERs of regular testing (Panel A-C) are lower than the \$100,000 threshold, thus indicating they are cost-effective diabetes management. Testing every quarter (Panel A), however, does not meet the conventional \$50,000 threshold. Because performing the test every year (Panel C) not only improves health outcomes but saves testing costs, its ICER is negative. Comparing their incremental benefits, however, show performing the test every six months is the most cost-effective diabetes care, with a net benefit of \$457.19 over ten years.[41] This result corroborates medical literature that suggests that performing an A1C test at least every six months is good practice for proper diabetes care (American Diabetes Association, 2019). It also suggests that biannual testing rates may be considered for quality measures instead.[42]

[40]It is from the Medicare clinical laboratory fee schedule (CLFS) for 2019, which is based on weighted median rates. Source: "Clinical Laboratory Fee Schedule," Centers for Medicare & Medicaid Services, 2019, accessed at https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ClinicalLabFeeSched/index.html on 11/13/2019.

[41]It is robust to alternative per-year QALE thresholds: \$50,000 to \$250,000.

[42]It is not uncommon that interventions give modest gains in health outcomes; UK Prospective Diabetes Study Group (1998) and Clarke et al. (2004) show that intensive blood-glucose control with oral medications and insulin injections lowered an 11% reduction in an A1C level and increased QALE by 0.10-year, over ten

7 Conclusions

I study the role of medical testing on healthcare performance. I develop a dynamic model in which a provider optimally performs a test for treatment choices. Test results help her make informed treatment decisions, but a fixed testing cost must be incurred, rationalizing that infrequent testing leads to additional dispersion of health measures over time, and explaining a provider's propensity to perform a test regularly. I combine the empirical framework with matched provider-patient data on testing prescription decisions, and a standardized health measure. I find that provider fixed effects and provider-specific heteroskedasticity explain 11% of the dispersion in blood sugar levels, and low-performing providers, in terms of large variance of the error term, tend to under-test. Structural estimates have substantial heterogeneity across provider and patient types. I also find that without medical testing, prescription choices are not necessarily inaccurate, underscoring the role of information acquisition through medical testing. Counterfactual exercises show that with the requirement that an A1C test must be performed every six months, patients' blood sugar levels would be well controlled: its incremental cost-effectiveness ratio is \$11,018.66 per quality-adjusted life-year (QALY), far lower than the \$100,000 threshold.

years, indicating even aggressive treatments cause modest QALE improvement. QALE gains from screening are around 0.01-0.02 years in different settings (The CDC Diabetes Cost-Effectiveness Study Group, 1998; Hoerger et al., 2004; Nicholson et al., 2005; Shono et al., 2018). As their cost-effectiveness is also dependent on costs, however, low costs of testing or screening still make the interventions very cost-effective. See Li et al. (2010) for review.

References

Abaluck, Jason, and Abi Adams. 2018. "What Do Consumers Consider Before They Choose?: Identification from Asymmetric Demand Responses." *Working Paper*.

Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh. 2016. "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care." *American Economic Review*, 106(12): 3730–3764.

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica*, 67(2): 251–333.

Ackerberg, Daniel A. 2003. "Advertising, learning, and consumer choice in experience good markets: an empirical examination." *International Economic Review*, 44(3): 1007–1040.

Adda, Jérôme, and Russell W. Cooper. 2003. *Dynamic economics: quantitative methods and applications.* MIT Press.

Agarwal, Sumit, John Grigsby, Ali Hortaçsu, Gregor Matvos, Amit Seru, and Vincent Yao. 2020. "Searching for Approval." National Bureau of Economic Research, Cambridge, MA.

Allcott, Hunt, Rebecca Diamond, Jean-Pierre Dubé, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell. 2019. "Food Deserts and the Causes of Nutritional Inequality." *The Quarterly Journal of Economics*, 134(4): 1793–1844.

Alvarez, Fernando E., Francesco Lippi, and Luigi Paciello. 2011. "Optimal Price Setting With Observation and Menu Costs." *The Quarterly Journal of Economics*, 126(4): 1909–1960.

Alvarez, Fernando, Francesco Lippi, and Juan Passadore. 2017. "Are State- and Time-Dependent Models Really Different?" *NBER Macroeconomics Annual*, 31(1): 379–457.

Alvarez, Fernando, Hervé Le Bihan, and Francesco Lippi. 2016. "The Real Effects of Monetary Shocks in Sticky Price Models: A Sufficient Statistic Approach." *American Economic Review*, 106(10): 2817–2851.

Alvarez, Fernando, Luigi Guiso, and Francesco Lippi. 2012. "Durable Consumption and Asset Management with Transaction and Observation Costs." *American Economic Review*, 102(5): 2272–2300.

American Diabetes Association. 2019. "Standards of Medical Care in Diabetes-2019." *Diabetes Care*, 42(Supplement 1).

Andrews, M. J., L. Gill, T. Schank, and R. Upward. 2008. "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3): 673–697.

Andrews, M.J., L. Gill, T. Schank, and R. Upward. 2012. "High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias." *Economics Letters*, 117(3): 824–827.

Asghari, Shabnam, Josiane Courteau, André C Carpentier, and Alain Vanasse. 2009. "Optimal strategy to identify incidence of diagnostic of diabetes using administrative data." *BMC Medical Research Methodology*, 9(1): 62.

Baicker, Katherine, and Amitabh Chandra. 2004. "The Productivity of Physician Specialization: Evidence from the Medicare Program." *American Economic Review*, 94(2): 357–361.

Bertsekas, Dimitri P. 2017. *Dynamic Programming and Optimal Control Volume 1.* . 4 ed., Athena Scientific.

Bonhomme, Stéphane. 2020. "Econometric analysis of bipartite networks." In *The Econometric Analysis of Network Data.* , ed. Bryan S. Graham and Áureo de Paula, Chapter 5, 83–121. Elsevier.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019*a*. "A Distributional Framework for Matched Employer Employee Data." *Econometrica*, 87(3): 699–739.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019*b*. "Discretizing Unobserved Heterogeneity." *Working Paper*.

Bronnenberg, Bart J, Jean-Pierre H Dubé, and Matthew Gentzkow. 2012. "The Evolution of Brand Preferences: Evidence from Consumer Migration." *American Economic Review*, 102(6): 2472–2508.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector." *American Economic Review*, 106(8): 2110–2144.

Chandra, Amitabh, and Douglas O. Staiger. 2007. "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy*, 115(1): 103–140.

Ching, Andrew T. 2010. "Consumer learning and heterogeneity: Dynamics of demand for prescription drugs after patent expiration." *International Journal of Industrial Organization*, 28(6): 619–638.

Clarke, P. M., A. M. Gray, A. Briggs, A. J. Farmer, P. Fenn, R. J. Stevens, D. R. Matthews, I. M. Stratton, and R. R. Holman. 2004. "A model to estimate the lifetime health outcomes of patients with Type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68)." *Diabetologia*, 47(10): 1747–1759.

Correia, Sergio. 2015. "Singletons, Cluster-Robust Standard Errors and Fixed Effects: A Bad Mix." *Working Paper*.

Correia, Sergio. 2016. "A Feasible Estimator for Linear Models with Multi-Way Fixed Effects." *Working Paper*.

Crawford, Gregory S., and Matthew Shum. 2005. "Uncertainty and Learning in Pharmaceutical Demand." *Econometrica*, 73(4): 1137–1173.

Currie, Janet M., and W. Bentley MacLeod. 2020. "Understanding Doctor Decision Making: The Case of Depression Treatment." *Econometrica*, 88(3): 847–878.

Currie, Janet, W. Bentley MacLeod, and Jessica Van Parys. 2016. "Provider practice style and patient health outcomes: The case of heart attacks." *Journal of Health Economics*, 47: 64–80.

Cutler, David, Jonathan S. Skinner, Ariel Dora Stern, and David Wennberg. 2019. "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending." *American Economic Journal: Economic Policy*, 11(1): 192–221.

De los Santos, Babur, Ali Hortaçsu, and Matthijs R Wildenbeest. 2012. "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior." *American Economic Review*, 102(6): 2955–2980.

Dickstein, Michael. 2018. "Efficient provision of experience goods: Evidence from antidepressant choice." *Working Paper*.

Erdem, Tülin, and Michael P. Keane. 1996. "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science*, 15(1): 1–20.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of Geographic Variation in Health Care: Evidence From Patient Migration." *The Quarterly Journal of Economics*, 131(4): 1681–1726.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2018. "What Drives Prescription Opioid Abuse?: Evidence from Migration." *Working Paper*.

Food and Drug Administration. 2016. "Self-Monitoring Blood Glucose Test Systems for Over-the-Counter Use."

Hoerger, Thomas J, Russell Harris, Katherine A Hicks, Katrina Donahue, Stephen Sorensen, and Michael Engelgau. 2004. "Screening for Type 2 Diabetes Mellitus: A Cost-Effectiveness Analysis." *Annals of Internal Medicine*, 140(9): 689.

Hong, Han, and Matthew Shum. 2006. "Using price distributions to estimate search costs." *The RAND Journal of Economics*, 37(2): 257–275.

Honka, Elisabeth. 2014. "Quantifying search and switching costs in the US auto insurance industry." *The RAND Journal of Economics*, 45(4): 847–884.

Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino. 2017. "Advertising, consumer awareness, and choice: evidence from the U.S. banking industry." *The RAND Journal of Economics*, 48(3): 611–646.

Hortaçsu, Ali, and Chad Syverson. 2004. "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds." *The Quarterly Journal of Economics*, 119(2): 403–456.

Hortaçsu, Ali, Seyed Ali Madanizadeh, and Steven L. Puller. 2017. "Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market." *American Economic Journal: Economic Policy*, 9(4): 192–226.

Hux, Janet E, Frank Ivis, Virginia Flintoft, and Adina Bica. 2002. "Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm." *Diabetes Care*, 25(3): 512–6.

International Expert Committee. 2009. "International Expert Committee Report on the Role of the A1C Assay in the Diagnosis of Diabetes." *Diabetes Care*, 32(7): 1327–1334.

Iskhakov, Fedor, Jinhyuk Lee, John Rust, Bertel Schjerning, and Kyoungwon Seo. 2016. "Comment on "Constrained Optimization Approaches to Estimation of Structural Models"." *Econometrica*, 84(1): 365–370.

Joo, Joonhwi. 2020. "Rational Inattention as an Empirical Framework: Application to the Welfare Effects of New Product Introduction."

Kim, Yu Jin, Suk Chon, Seungjoon Oh, Jeong-taek Woo, Sung Woon Kim, and Sang Youl Rhee. 2017. "Analysis of diabetes quality assessment findings and future directions for the appropriate management of diabetes in Korea." *The Korean Journal of Internal Medicine*.

Kocis, Ladislav, and William J. Whiten. 1997. "Computational investigations of low-discrepancy sequences." *ACM Transactions on Mathematical Software*, 23(2): 266–294.

Korean Diabetes Association. 2015. *Treatment Guidelines for Diabetes.* . 5 ed.

Korean Diabetes Association. 2019. *Treatment Guidelines for Diabetes.* . 6 ed.

Kwon, Soonman, Tae-jin Lee, and Chang-yup Kim. 2015. "Republic of Korea Health System Review." *Health Systems in Transition*, 5(4).

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler. 2019. "Imperfect Competition and Rent Sharing in the U.S. Labor Market." *Working Paper*.

Li, Rui, P. Zhang, Lawrence E. Barker, Farah M. Chowdhury, and Xuanping Zhang. 2010. "Cost-Effectiveness of Interventions to Prevent and Control Diabetes Mellitus: A Systematic Review." *Diabetes Care*, 33(8): 1872–1894.

Matějka, Filip, and Alisdair McKay. 2015. "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review*, 105(1): 272–298.

Miller, Donald R, Monika M Safford, and Leonard M Pogach. 2004. "Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data." *Diabetes care*, 27(Supplement 2): B10–21.

Molitor, David. 2018. "The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration." *American Economic Journal: Economic Policy*, 10(1): 326–356.

Moon, Tai Joon. 2012. "Light and Shadows of the Korean Healthcare System." *Journal of Korean Medical Science*, 27(Suppl): S3.

Mullainathan, Sendhil, and Ziad Obermeyer. 2020. "A Machine Learning Approach to Low-Value Health Care: Wasted Tests, Missed Heart Attacks and Mis-Predictions."

Narayan, K. M. Venkat, James P. Boyle, Theodore J. Thompson, Stephen W. Sorensen, and David F. Williamson. 2003. "Lifetime Risk for Diabetes Mellitus in the United States." *JAMA*, 290(14): 1884–90.

Neumann, Peter J., Joshua T. Cohen, and Milton C. Weinstein. 2014. "Updating cost-effectiveness - The curious resilience of the $50,000-per-QALY threshold." *New England Journal of Medicine*, 371(9): 796–797.

New York Times. 2016. "A Theranos Timeline."

Nicholson, W. K., L. A. Fleisher, H. E. Fox, and N. R. Powe. 2005. "Screening for Gestational Diabetes Mellitus: A decision and cost-effectiveness analysis of four screening strategies." *Diabetes Care*, 28(6): 1482–1484.

OECD. 2010. *OECD Economic Surveys: Korea 2010.* Vol. 2010 of *OECD Economic Surveys: Korea*, OECD.

Oster, Emily, Ira Shoulson, and E. Ray Dorsey. 2013. "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease." *American Economic Review*, 103(2): 804–830.

Phillips, Lawrence S., William T. Branch, Curtiss B. Cook, Joyce P. Doyle, Imad M. El-Kebbi, Daniel L. Gallina, Christopher D. Miller, David C. Ziemer, and Catherine S. Barnes. 2001. "Clinical Inertia." *Annals of Internal Medicine*, 135(9): 825.

Powers, Alvin C. 2015. "Diabetes Mellitus." In *Harrison's Principles of Internal Medicine.* . 19 ed., , ed. Dennis L. Kasper, Anthony S. Fauci, Stephen L. Hauser, Dan L. Longo, J. Larry Jameson and Joseph Loscalzo, Chapter 417-419, 2399–2430. New York, NY:McGraw-Hill Education.

Rector, Thomas S., Steven L. Wickstrom, Mona Shah, N. Thomas Greeenlee, Paula Rheault, Jeannette Rogowski, Vicki Freedman, John Adams, and José J. Escarce. 2004. "Specificity and Sensitivity of Claims-Based Algorithms for Identifying Members of

Medicare+Choice Health Plans That Have Chronic Medical Conditions." *Health Services Research*, 39(6p1): 1839–1858.

Rohlfing, C. L., H.-M. Wiedmeyer, R. R. Little, J. D. England, A. Tennill, and D. E. Goldstein. 2002. "Defining the Relationship Between Plasma Glucose and HbA1c: Analysis of glucose profiles and HbA1c in the Diabetes Control and Complications Trial." *Diabetes Care*, 25(2): 275–278.

Rust, John. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica*, 55(5): 999–1033.

Rust, John. 2000. "Nested fixed point algorithm documentation manual."

Shono, Aiko, Masahide Kondo, Shuling Hoshi, Reiko Okubo, and Naoya Yahagi. 2018. "Cost-effectiveness of a New Opportunistic Screening Strategy for Walk-in Fingertip HbA 1c Testing at Community Pharmacies in Japan." *Diabetes Care*, 41(6): 1218–1226.

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of Monetary Economics*, 50(3): 665–690.

Skipper, Niels, and Rune Vejlin. 2015. "Determinants of generic vs. brand drug choice: Evidence from population-wide Danish data." *Social Science & Medicine*, 130: 204–215.

Song, Yunjie, Jonathan Skinner, Julie Bynum, Jason Sutherland, John E. Wennberg, and Elliott S. Fisher. 2010. "Regional Variations in Diagnostic Practices." *New England Journal of Medicine*, 363(1): 45–53.

Steiner, Jakub, Colin Stewart, and Filip Matějka. 2017. "Rational Inattention Dynamics: Inertia and Delay in Decision-Making." *Econometrica*, 85(2): 521–553.

Stokey, Nancy L. 2008. *The Economics of Inaction: Stochastic Control Models with Fixed Costs.* NJ: Princeton:Princeton University Press.

Stokey, Nancy L., Robert E. Jr. Lucas, and Edward C. Prescott. 1989. *Recursive Methods in Economic Dynamics.* Harvard University Press.

Su, Che-Lin, and Kenneth L Judd. 2012. "Constrained Optimization Approaches to Estimation of Structural Models." *Econometrica*, 80(5): 2213–2230.

The CDC Diabetes Cost-Effectiveness Study Group. 1998. "The Cost-effectiveness of Screening for Type 2 Diabetes." *JAMA: The Journal of the American Medical Association*, 280(20): 1757–1763.

UK Prospective Diabetes Study Group. 1998. "Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33)." *The Lancet*, 352(9131): 837–853.

World Health Organization. 2016. *Global Report on Diabetes.*

Yoo, Kyoung-Hun, Dong-Wook Shin, Mi-Hee Cho, Sang-Hyuck Kim, Hyun-Jung Bahk, Shin-Hye Kim, Su-Min Jeong, Jae-Moon Yun, Jin-Ho Park, Heesun Kim, and BeLong Cho. 2017. "Regional variations in frequency of glycosylated hemoglobin (HbA1c) monitoring in Korea: A multilevel analysis of nationwide data." *Diabetes Research and Clinical Practice*, 131: 61–69.

Zhang, Yuting, Katherine Baicker, and Joseph P. Newhouse. 2010*a*. "Geographic Variation in Medicare Drug Spending." *New England Journal of Medicine*, 363(5): 405–409.

Zhang, Yuting, Katherine Baicker, and Joseph P. Newhouse. 2010*b*. "Geographic Variation in the Quality of Prescribing." *New England Journal of Medicine*, 363(21): 1985–1988.

Table 1: Overview of the Datasets

Health Screening Observations	Claims Observations	Patients	Providers	Movers	Movers/Providers
A. Full Sample, Quarterly Dataset					
6,383,849	77,429,081	2,994,766	38,754	1,814,274	46.8151
B. Full Sample, Yearly Dataset					
5,863,047	8,823,970	2,582,646	31,681	902,123	28.4752
C. Migrants Subsample, Yearly Dataset					
2,480,559	3,681,833	596,315	21,408	65,487	3.0560

Notes: The quarterly dataset includes all observations that satisfy criteria in Section 2.3. Patient-quarters with no medical records but with remaining prescription drugs during at least one day of the quarter are also included. Yearly datasets do not include patient-years in which at least one patient-quarter observation is missing during a year. Patient-years that the National Health Screening is not offered are not included in the yearly dataset for estimation of the AKM model (1).

Table 2: Contributions of Patient and Provider Heterogeneity in Blood Sugar Levels

		Value	Share
A. Full Sample, AKM			
Overall	$Var(y_{it})$	0.0791	100.00%
Patient FE	$Var(\alpha_i + x_{it}\beta)$	0.0413	52.16%
Provider FE	$Var(\psi_j)$	0.0020	2.58%
Sorting	$2Cov(\alpha_i + x_{it}\beta, \psi_j)$	-0.0008	-1.00%
Residual	$Var(\varepsilon_{it})$	0.0366	46.26%
No Heteroskedasticity	$Var(\tilde{\varepsilon}_{it})$	0.0294	37.20%
B. Migrants Subsample, BLM			
Overall	$Var(y_{it})$	0.0689	100.00%
Patient FE	$Var(\alpha_i + x_{it}\beta)$	0.0374	54.34%
Provider FE	$Var(\psi_j)$	0.0059	0.85%
Sorting	$2Cov(\alpha_i + x_{it}\beta, \psi_j)$	0.0001	1.45%
Residual	$Var(\varepsilon_{it})$	0.0299	43.35%
No Heteroskedasticity	$Var(\tilde{\varepsilon}_{it})$	0.0240	34.79%

Notes: In Panel A, estimates of the AKM (1) model using the full sample are shown. In Panel B, a classification method (Bonhomme, Lamadon and Manresa, 2019*a*) is applied to the migrants subsample. $K = 10$ clusters are used. For classification, I evaluate provider-level empirical cumulative distributions at 20 grid points, from Q_5 to Q_{95} of the unconditional empirical distribution of log blood sugar levels. Once providers are clustered using the k-means clustering algorithm, the AKM model (1), with full patient fixed effects and $K = 10$ provider-class fixed effects, is estimated. For both datasets, singleton observations are dropped for reliable estimation (Correia, 2015). Provider level and provider-class level variances of $\hat{\varepsilon}_{it}$ are computed to measure provider-level heteroskedasticity. $\hat{Var}(\tilde{\varepsilon}_{it})$ are computed to satisfy Equation (3).

Table 3: Patient- and Provider-Level Testing Frequencies

A. Patient-Years

	0Q	1Q	2Q	3Q	4Q
Semi-Quantitative	15.68%	7.84%	7.31%	11.50%	57.66%
A1C	27.79%	27.08%	22.19%	14.35%	8.58%
Quantitative	48.79%	23.15%	12.72%	7.54%	7.80%
A1C, Quantitative	23.99%	26.04%	22.36%	15.31%	12.31%

B. Provider-Quarters

	$\leq 20\%$	$\leq 40\%$	$\leq 60\%$	$\leq 80\%$	$\leq 100\%$
Semi-Quantitative	9.64%	7.00%	6.22%	17.63%	59.52%
A1C	21.44%	38.10%	22.15%	13.69%	4.62%
Quantitative	51.07%	26.01%	9.57%	7.38%	5.96%
A1C, Quantitative	16.61%	34.35%	26.01%	13.28%	9.76%

Notes: Semi-quantitative tests include point-of-care (POC) blood glucose monitoring conducted by a provider. Panel A uses the yearly dataset for patient-level testing frequencies. Patient-years that the National Health Screening is not offered are also included. Panel B uses the quarterly dataset in Table 1. For each provider, I compute portions of all patient-quarters that receive a medical test. Each column in Panel B is the raw frequency of providers performing the test.

Table 4: Predictors of Provider Heterogeneity

	Full Sample			Migrants Subsample		
A. Provider Fixed Effects ($\hat{\psi}_j$)						
A1C Test	-0.0090		-0.0091	-0.0016		-0.0016
	(0.0000)		(0.0000)	(0.0000)		(0.0000)
Adjustments		-0.0016	0.0003		0.0000	0.0004
		(0.0000)	(0.0000)		(0.0000)	(0.0000)
R^2	0.0626	0.0007	0.0626	0.0066	0.0000	0.0068
N		5,192,761			2,078,763	
B. Heteroskedastic Errors ($\hat{\sigma}_j$)						
A1C Test	-0.0032		-0.0033	-0.0003		-0.0003
	(0.0000)		(0.0000)	(0.0000)		(0.0000)
Adjustments		0.0009	0.0016		0.0000	0.0001
		(0.0000)	(0.0000)		(0.0000)	(0.0000)
R^2	0.0251	0.0007	0.0274	0.0034	0.0000	0.0036
N		5,192,761			2,078,763	

Notes: Estimates of provider fixed effects ($\hat{\psi}_j$) and heteroskedastic errors ($\hat{\sigma}_j$) in Table 2 are used as dependent variables. Two regressors presented above are patient-level count variables, with a maximum of four. Standard errors appear in parentheses.

Table 5: Characteristics of Testing Clinics

Provider-level Mean Number of A1C Tests	
Year of Entry ($\geq$ 2002)	0.0358
	(0.0001)
Year of Entry $<$ 2002	-0.1032
	(0.0008)
Specialist in	0.4390
Internal Medicine	(0.0006)
Has the A1C Device	0.3314
	(0.0007)
R^2	0.1728
N	5,362,386

Notes: Provider-level average number of A1C tests are used as an outcome variable. Observations are weighted by the number of patient-years. The year of entry variable is the month a clinic appears for the first time in the dataset, divided by 12. The 'Year of Entry $<$ 2002' variable is an indicator of whether the clinic is observed in January 2002, the first month in the dataset. The 'Specialist in Internal Medicine' is an indicator of whether there is at least one specialist in internal medicine working at the clinic. The 'Has the A1C Device' is one if and only if at least one A1C test is performed and 50% of all A1C tests are not out-sourced during a year. Standard errors appear in parentheses.

Table 6: Overview of Bi-clustered Subsamples

Patient	Full	Low		High	
Provider	Sample	Low	High	Low	High
Observation	37.7816%	15.4093%	42.7184%	42.6859%	77.6453%
Adjustment	15.1313%	12.4947%	14.6085%	17.9929%	17.9879%
Patient-Quarter	77,429,081	32,374,809	12,852,126	17,816,238	14,385,908

Notes: Raw frequencies of observation and adjustment for each patient-provider type pair are shown. Patient and provider types are clustered, minimizing Equation (13). I apply the k-means algorithms iteratively. I classify providers, minimizing objective function (13), net of patient-level centroids. Given provider types, I find patient types, minimizing the objective. Iterative k-means clustering is repeated until convergence. Fifty starting values of provider types are used. Indicators of an A1C test and prescription adjustments are used as moments.

Table 7: Predictors of Testing and Adjustment

Patient	Full	Low		High	
Provider	Sample	Low	High	Low	High
A. A1C Test					
Test Conducted	0.7609	-0.1284	-0.2717	-0.5078	-0.3067
Last Period	(0.0005)	(0.0016)	(0.0012)	(0.0011)	(0.0017)
N	69,276,161	29,157,573	11,539,152	15,640,261	12,939,175
B. Adjustment					
log(Blood Sugar Level)	-5.9995	-4.5516	-5.0601	-7.0111	-10.6458
	(0.0738)	(0.1176)	(0.2014)	(0.1412)	(0.1776)
log(Blood Sugar Level)^2	0.7215	0.5565	0.7344	0.8330	1.2123
	(0.0073)	(0.0116)	(0.0199)	(0.0139)	(0.0177)
N	6,383,848	2,593,459	1,012,914	1,198,807	1,578,668

Notes: The table shows estimates from logistic regression. In Panel A, all patient-quarters staying at the provider treated during the last quarter are used for estimation. In Panel B, I use all patient-quarters with observed blood sugar levels. Standard errors appear in parentheses.

Table 8: Structural Estimates

Patient	Full	Low		High	
Provider	Sample	Low	High	Low	High
$E[y_1]$	4.8504	4.8911	4.8460	4.8933	4.9017
	(0.0025)	(0.0036)	(0.0050)	(0.0049)	(0.0166)
σ_{y1}	1.0798	1.0273	1.1453	0.9250	1.1186
	(0.0181)	(0.0178)	(0.0354)	(0.0222)	(0.0340)
σ_w	0.1230	0.1117	0.1244	0.1228	0.1266
	(0.0007)	(0.0008)	(0.0011)	(0.0012)	(0.0014)
σ_{v0}	0.1462	0.1922	0.1873	0.1895	0.2212
	(0.0073)	(0.0070)	(0.0161)	(0.0110)	(0.0232)
γ	1.6530	1.7225	1.4815	2.6780	1.8847
	(0.0683)	(0.0748)	(0.1095)	(0.1430)	(0.1264)
y^*	4.8650	4.8760	4.8600	4.8669	4.8523
	(0.0020)	(0.0025)	(0.0038)	(0.0029)	(0.0033)
F_O	0.5577	1.8225	0.3688	0.4390	-1.1971
	(0.0034)	(0.0087)	(0.0087)	(0.0097)	(0.0097)
F_A	2.5848	3.0028	2.6622	2.4444	2.2162
	(0.0159)	(0.0207)	(0.0276)	(0.0236)	(0.0240)
$\exp(y^*)$	129.6746	131.1059	129.0306	129.9212	128.0393
	(0.2658)	(0.3333)	(0.4918)	(0.3799)	(0.4170)
F_O/γ	0.3374	1.0581	0.2489	0.1639	-0.6351
	(0.0140)	(0.0458)	(0.0191)	(0.0095)	(0.0428)
F_A/γ	1.5637	1.7433	1.7970	0.9127	1.1759
	(0.0611)	(0.0713)	(0.1225)	(0.0486)	(0.0766)
N	7986	6338	2578	4369	2922

Notes: N denotes all patient-provider pairs in the sample. I use 0.1% (full sample) and 0.2% (bi-clustered subsamples) subsamples for structural estimation to reduce computational burden. 500 and 300 Halton draws for value function iterations and simulated maximum likelihood, respectively, are used. Once the value function is evaluated at 30×29 grid points, I use cubic spline interpolation and extrapolation to evaluate it at other points. Symmetry of the value function around y^* is imposed. Outlier glucose levels have been censored. Standard errors appear in parentheses. For functions of structural estimates, delta method standard errors are shown.

Table 9: Model Performance: Medical Testing and Adjustments

	Patient	Provider	Actual	Simulated
A. Conditional on Patient and Provider Types				
Observation (%)	Low	Low	15.4093	16.5662
		High	42.7184	44.3057
	High	Low	42.6859	41.5194
		High	77.6453	74.4861
Adjustment (%)	Low	Low	12.4947	13.8691
		High	14.6085	15.9127
	High	Low	17.9929	18.6924
		High	17.9879	20.0759
B. Incorporate Mobility				
Observation (%)	Low	-	23.1698	24.6557
	High	-	58.3036	54.7746
Adjustment (%)	Low	-	13.0954	14.4651
	High	-	17.9907	19.2487

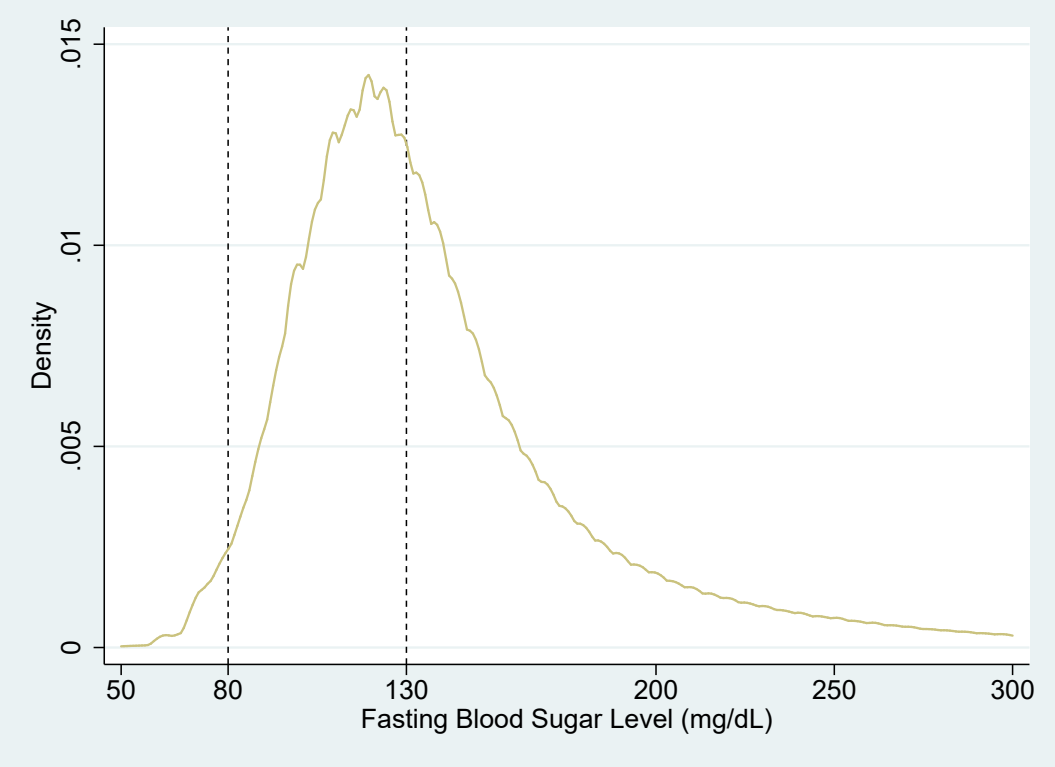
Notes: Panel A shows observation and adjustment frequencies for each patient and provider type. Panel B shows frequencies for each patient type only. Panel B combines the mobility model in Equation (**??**) and the structural model. For each patient type, at the beginning of each period, a mobility decision is made before a provider makes a testing choice.

Table 10: Mandatory Testing

Increase in QALE (years)				Incremental	Incremental	Incremental
Q_{50}	Q_{75}	Q_{90}	Mean	Cost ($)	Cost-effectiveness Ratio ($/QALE)	Benefit ($)
A. Performed Every Quarter						
0.00	0.01	0.02	0.01	272.41	53140.23	237.63
B. Performed Every 6 Months						
0.00	0.01	0.01	0.01	56.61	11018.66	457.19
C. Performed Every Year						
0.00	0.00	0.00	0.00	-51.29	-33340.78	205.11
D. Mandatory Testing before Adjustment						
0.00	-0.03	-0.07	-0.01	3.35	-344.72	976.11

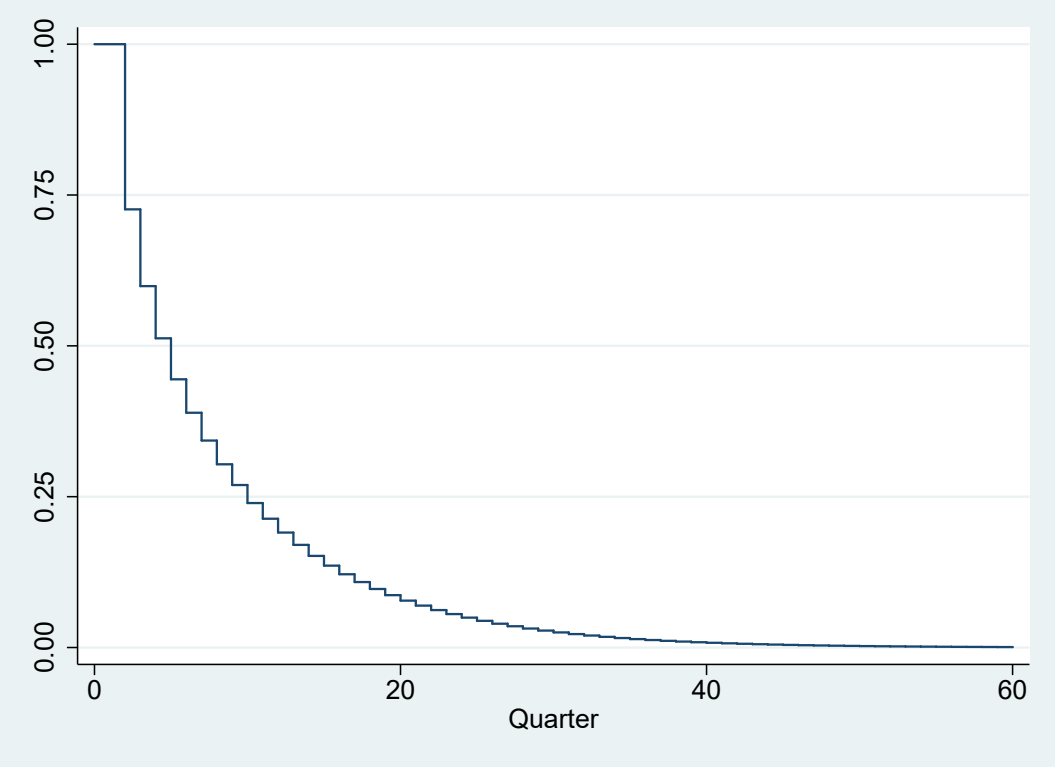
Notes: All outcomes except testing choices in Panel A-C are simulated from the structural model. In Panels A-C, predetermined testing choices are implemented regardless of the utility of medical testing. In Panel D, a provider optimally makes a testing choice each period. However, she cannot adjust if no testing choice is made. For each patient, I compute an average blood sugar level over 40 quarters. To translate the distribution of average blood sugar levels over 10 years to quality-adjusted life expectancy (QALE), I use Clarke et al. (2004)'s QALE computation that those who were randomly assigned intensive policy that resulted in an 11% reduction in an A1C level over 10 years experience 0.10-year QALE increase. $100,000-per-year threshold is used to monetize QALE and compute incremental benefits (Neumann, Cohen and Weinstein, 2014). I choose $10.19 as A1C testing cost, which is from the Medicare clinical laboratory fee schedule (CLFS) rates for 2019.

Figure 1: Distribution of Blood Sugar Levels Among Patients with Diabetes



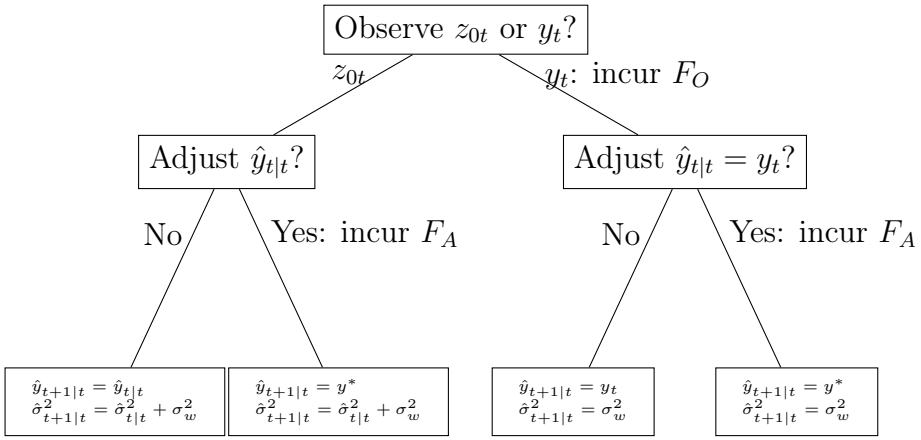
Notes: Observations are weighted by the inverse of the number of observations for each patient. Dotted lines represent the target level (i.e., 80-130mg/dL) that clinical guidelines suggest (American Diabetes Association, 2019).

Figure 2: Survival Probability of No Prescription Adjustments



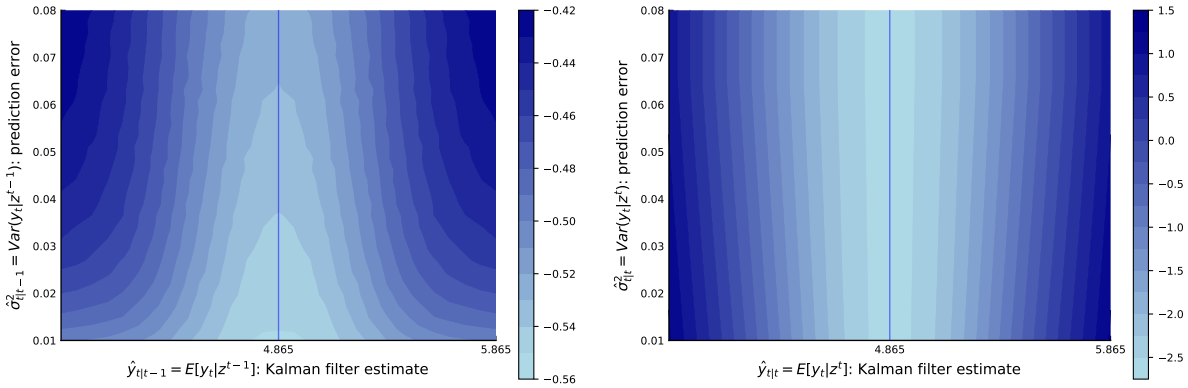
Notes: An indicator of prescription adjustment is one if and only if a composition of prescription drugs changes in comparison to that during the last quarter. The composition consists of 9 classes of anti-diabetic medications: Biguanides, Sulfonylureas, DPP-4 inhibitors, Insulin, Alpha-Glucosidase Inhibitors, Thiazolidinediones, Nonsulfonylurea secretagogues, GLP-1 agonists, and SGLT-2 inhibitors. A patient's entire treatment history is used during survival analysis. Prescription adjustment at $t = 1$ is excluded.

Figure 3: Two-Stage Decision Problem



- $\hat{y}_{t|t} = E[y_t|z^t, a^{t-1}]$: best prediction of y_t based on $z^t = (z_1, ..., z_t)$ and $a^{t-1} = (a_1, ..., a_{t-1})$
- $\hat{\sigma}^2_{t|t} = Var(y_t|z^t, a^{t-1})$: prediction error of $\hat{y}_{t|t}$
- $(\hat{y}_{t+1|t}, \hat{\sigma}^2_{t+1|t})$: state variables from a provider's viewpoint

Figure 4: Net Utility of Medical Testing and Adjustment

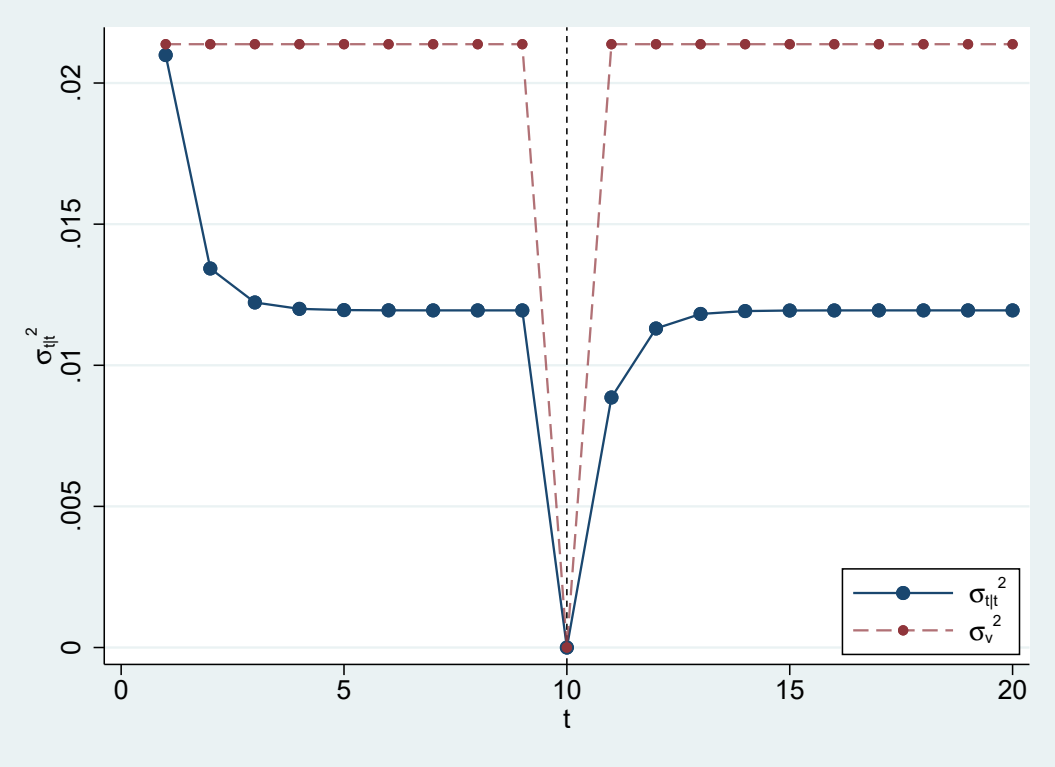


A. Net Utility of Medical Testing

B. Net Utility of Adjustment

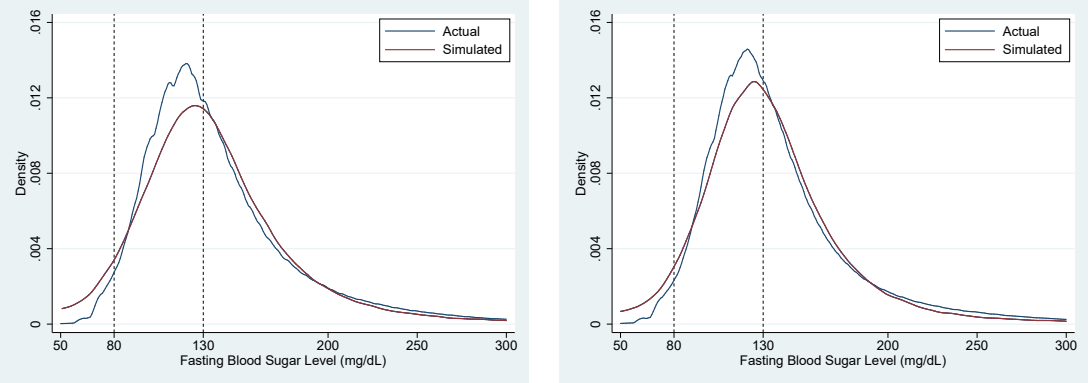
Notes: Panels A and B are plotted using full sample estimates from Table 8. Logit errors are added to rationalize idiosyncratic choices. 30×29 grid points and cubic spline interpolation are used to solve Equation (12). Symmetry of the value function around y^* is imposed.

Figure 5: Prediction Errors over Time



Notes: The blue line denotes the prediction error after a testing choice is made, $\sigma^2_{t|t}$. The red dashed line represents the variance of a signal received after the testing choice. No adjustment and testing choices are made, except for the test conducted at $t = 10$. Full sample estimates in Table 8 are used.

Figure 6: Model Performance: Blood Sugar Levels



A. Low Patient Type B. High Patient Type

Notes: The blue and red line denote a distribution of blood sugar levels in actual and simulated data, respectively. The simulated data combine the mobility model in Equation (**??**) and the structural model. For each patient type, at the beginning of each period, a mobility decision is made before a provider makes a testing choice.

[Table 11 about here.]

[Table 12 about here.]

[Table 13 about here.]

[Figure 7 about here.]

Table A.1: Selection Issues in Health Screening Data

	A. Takes the Health Screening			
A1C Tests	0.0017		0.0019	
	(0.0002)		(0.0002)	
Adjustments		-0.0010	-0.0014	
		(0.0002)	(0.0002)	
R^2	0.5667	0.5667	0.5667	
N	8,823,970	8,823,970	8,823,970	
Patient FE	Y	Y	Y	
Year FE	Y	Y	Y	
	B. Performs the A1C Test		C. Adjusts Prescriptions	
Takes the Health Screening	0.0523	0.6794	0.0178	2.1236
	(0.0002)	(0.0306)	(0.0001)	(0.0247)
Takes the Health Screening		-0.3139		-0.9970
$\times$ log(Blood Sugar Level)		(0.0123)		(0.0099)
Takes the Health Screening		0.0378		0.1154
$\times$ log(Blood Sugar Level)^2		(0.0012)		(0.0010)
R^2	0.2538	0.2539	0.1083	0.1097
N	77,429,081	77,429,081	77,429,081	77,429,081
Patient FE	Y	Y	Y	Y
Year-Quarter FE	Y	Y	Y	Y

Notes: The table shows estimates of fixed effects linear regression models. In Panel A, a dependent variable of the regression using the yearly dataset is an indicator of whether a patient takes the National Health Screening. All patient-years that the health screening is not offered are not included. In Panel B, dependent variables are indicators of the A1C test and prescription adjustment during the quarter, respectively. The interacted variables are zero whenever the health screening is not conducted and therefore blood sugar levels are not observed. Standard errors appear in parentheses.

Table A.2: Contributions of Patient and Provider Heterogeneity in Blood Sugar Levels, Full Sample

		Value	Share
A1. Full sample, $K = 20$ clusters			
Overall	$Var(y_{it})$	0.0791	100.00%
Patient FE	$Var(\alpha_i + x_{it}\beta)$	0.0406	51.28%
Provider FE	$Var(\psi_j)$	0.0006	0.74%
Sorting	$2Cov(\alpha_i + x_{it}\beta, \psi_j)$	0.0010	1.21%
Residual	$Var(\varepsilon_{it})$	0.0370	46.77%
No Heteroskedasticity	$Var(\tilde{\varepsilon}_{it})$	0.0292	36.91%
A2. Full sample, $K = 10$ clusters, 10 model-based iterations			
Overall	$Var(y_{it})$	0.0791	100.00%
Patient FE	$Var(\alpha_i + x_{it}\beta)$	0.0407	51.47%
Provider FE	$Var(\psi_j)$	0.0017	2.15%
Sorting	$2Cov(\alpha_i + x_{it}\beta, \psi_j)$	0.0000	-0.01%
Residual	$Var(\epsilon_{it})$	0.0367	46.39%
No Heteroskedasticity	$Var(\tilde{\varepsilon}_{it})$	0.0340	42.96%
A3. Full sample, $K = 10$ clusters			
Overall	$Var(y_{it})$	0.0791	100.00%
Patient FE	$Var(\alpha_i + x_{it}\beta)$	0.0409	51.70%
Provider FE	$Var(\psi_j)$	0.0004	0.56%
Sorting	$2Cov(\alpha_i + x_{it}\beta, \psi_j)$	0.0007	0.92%
Residual	$Var(\epsilon_{it})$	0.0370	46.82%
No Heteroskedasticity	$Var(\tilde{\varepsilon}_{it})$	0.0316	39.99%

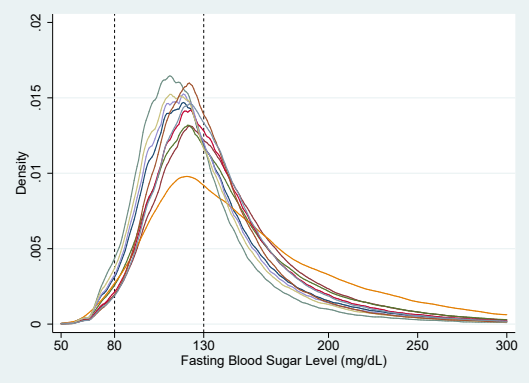
Notes: For a robustness check, a classification method (Bonhomme, Lamadon and Manresa, 2019*a*) is applied to both the full sample and the migrants subsample. For classification, I evaluate provider-level empirical cumulative distributions at 20 grid points, from Q_5 to Q_{95} of the unconditional empirical distribution of log blood sugar levels. Once providers are clustered using the k-means clustering algorithm, the AKM model (1), with full patient fixed effects and K provider-class fixed effects, is estimated. For estimates with model-based iterations, for given estimates, I find an alternative classification, minimizing the mean squared error. I then estimate the AKM model based on the new classification. For both datasets, singleton observations are dropped for reliable estimation (Correia, 2015). Provider and provider-class level variances of $\hat{\varepsilon}_{it}$ are computed to measure provider-level heteroskedasticity. $\hat{Var}(\tilde{\varepsilon}_{it})$ are computed to satisfy Equation (3).

Table A.3: Contributions of Patient and Provider Heterogeneity in Blood Sugar Levels, Migrants Subsample

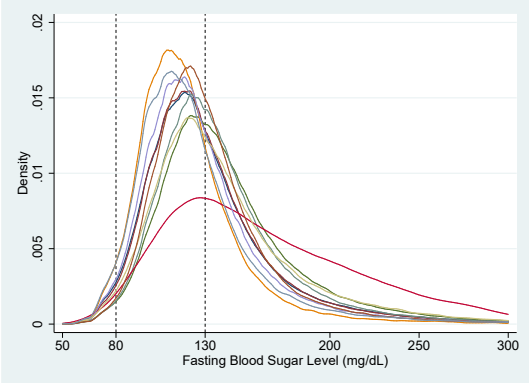
		Value	Share
B1. Migrants sample, $K = 20$ clusters			
Overall	$Var(y_{it})$	0.0689	100.00%
Patient FE	$Var(\alpha_i + x_{it}\beta)$	0.0371	53.91%
Provider FE	$Var(\psi_j)$	0.0007	1.01%
Sorting	$2Cov(\alpha_i + x_{it}\beta, \psi_j)$	0.0012	1.74%
Residual	$Var(\epsilon_{it})$	0.0299	43.34%
No Heteroskedasticity	$Var(\tilde{\varepsilon}_{it})$	0.0224	32.57%
B2. Migrants Subsample, $K = 10$ clusters, 10 model-based iterations			
Overall	$Var(y_{it})$	0.0689	100.00%
Patient FE	$Var(\alpha_i + x_{it}\beta)$	0.0410	59.48%
Provider FE	$Var(\psi_j)$	0.0056	8.09%
Sorting	$2Cov(\alpha_i + x_{it}\beta, \psi_j)$	-0.0073	-10.62%
Residual	$Var(\epsilon_{it})$	0.0297	43.05%
No Heteroskedasticity	$Var(\tilde{\varepsilon}_{it})$	0.0252	36.56%

Notes: For a robustness check, a classification method (Bonhomme, Lamadon and Manresa, 2019*a*) is applied to both the full sample and the migrants subsample. For classification, I evaluate provider-level empirical cumulative distributions at 20 grid points, from Q_5 to Q_{95} of the unconditional empirical distribution of log blood sugar levels. Once providers are clustered using the k-means clustering algorithm, the AKM model (1), with full patient fixed effects and K provider-class fixed effects, is estimated. For estimates with model-based iterations, for given estimates, I find an alternative classification, minimizing the mean squared error. I then estimate the AKM model based on the new classification. For both datasets, singleton observations are dropped for reliable estimation (Correia, 2015). Provider and provider-class level variances of $\hat{\varepsilon}_{it}$ are computed to measure provider-level heteroskedasticity. $\hat{Var}(\tilde{\varepsilon}_{it})$ are computed to satisfy Equation (3).

Figure A.1: Evidence of No Sorting



A. Full Sample, $K = 10$ clusters



B. Migrants Subsample, $K = 10$ clusters

Notes: Panels A and B show distributions of patients' blood sugar levels by provider-class. Clusters are those of Panel A3 of Table A.2 and B of Table 2, respectively.