Estimating Nursing Home Quality with Selection

Andrew Olenski Szymon Sacher*

Department of Economics
Columbia University

March 10, 2022

Abstract

We estimate a Bayesian model of nursing home quality using variational inference. We then conduct three exercises. First, we examine the correlates of quality, finding that public report cards have near-zero correlation. Second, we show that higher quality nursing homes fared better during the pandemic: a one standard deviation increase in quality corresponds to 2.4% fewer Covid-19 cases. Finally, we show that a 10% increase in the Medicaid reimbursement rate raises quality, leading to a 1.85 percentage point increase in 90-day survival. Such a reform would be highly cost-effective even under conservative estimates of the quality-adjusted statistical value of life.

*We are very grateful to Charles Angelucci, Matt Backus, Bo Cowgill, Gautam Gowrisankaran, Stephen Hansen, Suresh Naidu, Adam Sacarny, Ashley Swanson, and Pietro Tebaldi for their helpful comments and suggestions. We thank Mohan Ramanujan, Elizabeth Adams, and Maurice Dalton for their help in obtaining and managing the data. Olenski gratefully acknowledges support from the National Science Foundation Graduate Research Fellowship.

1 Introduction

In order for health care markets to function well, it is critical for patients and insurers to be able to observe provider performance. Lacking information, consumers will not be able to identify and select better providers, which in turn under-incentivizes investment in quality (Dranove and Satterthwaite 1992). Yet, quality measurement is notoriously challenging, and this complexity can diminish the returns to efforts to improve information. The nursing home industry may be the poster child of this problem. Despite significant investments in improving measurement, quality remains infamously low.[1] Meanwhile, multiple investigations have found that the public report cards that serve as the primary system of quality measurement are subject to rampant manipulation, as facilities systematically misreport information so as to maximize their ratings (Thomas 2014; Silver-Greenberg and Gebeloff 2021).

In this paper, we propose a new method to obtain consistent estimates of the causal effect of admission to a given nursing home on 90-day survival, while controlling for patient selection. Though averting mortality is not the exclusive aim of nursing home care, this causal effect is a relevant dimension of facility quality, and may be used to evaluate the validity of existing measures. We account for endogenous patient selection using the distance between the facility and each patient's home zip code as an instrumental variable, a common approach in this literature (Grabowski et al. 2013; Einav, Finkelstein, and Mahoney 2022).

To link this quasi-experimental variation in facility choice with our binary health outcome measure, we propose a structural Bayesian model of nursing home quality, in the style of Hull (2020) and Geweke, Gowrisankaran, and Town (2003). Due to the large number of parameters, traditional estimation approaches are computationally infeasible, and so we adapt a technique, variational inference, from machine learning which replaces the objective function with an approximation. We show in a simulation exercise that this procedure performs extremely well. Using the facility-level quality estimates, we then answer three questions.

First, we evaluate the performance of the public report card system. These scores, meant to serve as easy reference points for consumers, are organized into "star ratings" and include measures of staffing, facility inspections, and patient-level outcome measures. Prior work studying similar report cards in other health care settings has found mixed results with respect to quality (Abaluck, Caceres Bravo, Hull, and Starc 2021; Doyle, Graves, and Gruber 2019). We find that that there is near-zero correlation between the report cards and our survival-based quality estimates. Even conditioning on narrow measures of geography, very

1. For instance, one report found that more than one in five Medicare patients who stayed in a nursing facility for 35 days or fewer experienced harm as a result of their medical care (Office of Inspector General 2014).

little of the variation in nursing home quality can be explained by the variation in the star ratings.

Second, we apply the estimates to a question that has emerged in the wake of the Covid-19 pandemic's devastation of the nursing home sector: did higher quality facilities fare better? A flurry of recent papers have found no relationship between Covid-19 and star ratings, with one recent systematic review concluding that providers therefore had few levers available to combat Covid-19 spread (Konetzka et al. 2021). We re-visit this question using our estimates of quality, and find that contrary to the consensus, higher quality nursing homes have indeed fared better during the Covid-19 pandemic: facilities with one standard deviation higher quality had 2.4% fewer Covid-19 cases per bed, suggesting that there may be some scope to limit adverse outcomes.

Finally, in light of recent commitments by the federal government to improve nursing home quality, we use our estimates to evaluate the effect of one policy: increasing the reimbursement rate (The White House 2022). The regulated price paid for most skilled nursing care – the Medicaid reimbursement rate – is commonly cited as a chief cause of low quality (Grabowski 2001; Harrington et al. 2008). Firm operators have lamented that rates have been too slow to rise over time, even prompting a recent lawsuit in one state in a bid to raise reimbursement rates (Rucinski 2018). We ask whether an increase to the Medicaid rate would actually increase quality, using an instrumental variable strategy mirroring Hackmann (2019). We find that a 10% increase in the Medicaid rate would raise quality by 0.28 standard deviations, which corresponds to a 1.85 percentage point increase in 90-day survival. We calculate that even under conservative estimates of the value of a quality-adjusted life year, this policy easily passes a cost-benefit analysis.

This paper primarily contributes to a considerable literature on quality estimation in the health care sector. The majority of work in this area studies hospital quality: for instance, Gowrisankaran and Town (1999), Doyle, Graves, and Gruber (2019), Geweke, Gowrisankaran, and Town (2003) and Hull (2020) all estimate models of hospital quality using mortality. The latter two derive models from which we borrow heavily. For these models, we show the power of variational inference as a general purpose estimation algorithm, particularly in cases where traditional maximum likelihood methods can be computationally intractable. Despite its popularity in computer science, VI has seen very limited adoption in economics; however, now that desirable theoretical properties have been established, we anticipate interest in the method to grow substantially (Wang and Blei 2019; Medina, Olea, Rush, and Velez 2021).

Our work is closely related to contemporaneous research by Einav, Finkelstein, and Mahoney (2022). To study nursing home value-added, they construct a unidimensional health index that measures how fit a patient is to return to the community, and study the average change in this index over 30 days. Doing so requires estimating a model of "selection into" nursing homes – similar to ours – and "selection out," where discharge decisions are a function of the health index. This health index arguably comprises a richer measure of value than the binary 90-day survival outcome that we use here.[2]

In terms of measuring quality and value-added (two tightly intertwined concepts), our findings are complementary with those of Einav, Finkelstein, and Mahoney (2022). We both find that public report cards are poor predictors of the causal effect of a given nursing home admission on health outcomes. Moreover, despite using a simpler outcome measure, we also show that our causal measures of quality would have been able to predict out-of-sample heterogeneity in nursing home outcomes during the Covid-19 pandemic, and we demonstrate that increasing Medicaid payments would lead to higher facility quality and lower mortality.

2 Institutional Details and Data

2.1 Specialization and Selection

This paper considers quality estimation of skilled nursing facilities (SNFs), commonly referred to as nursing homes. SNFs are health care facilities that provide a broad range of services, including skilled nursing care, specialized rehabilitative services such as physical therapy, occupational therapy, and speech therapy, in addition to treatment for the mentally ill and developmentally disabled.

Given the diversity of treatment requirements across patient types, it is natural that SNFs involve some degree of specialization and differentiation (Mor, Banaszak-Holl, and Zinn 1995). Facilities invest in specialty care units, such as Alzheimer's wards or rehabilitative care, which may affect the types of patients who select these facilities. That is, in the presence of comparative advantage across SNFs, there may be non-random selection across patients. In addition to patient-side selection, a pair of recent papers present evidence that nursing homes themselves may worsen this selection problem by strategically admitting and discharging less profitable patients as the opportunity cost of maintaining those patients rises (Gandhi 2020; Hackmann, Pohl, and Ziebarth 2021). These papers therefore raise an additional concern. To the extent that some nursing homes are better able to select their patient censuses on unobserved risk, quality estimation based purely on risk-adjustment would

2. However, this added richness comes at the cost of additional assumptions, as the authors must account for selection out of the facility. Because we observe mortality for all residents, we only have to grapple with selection in. Additionally, our estimation method allows us to consider larger markets, and this results in a small share of patients choosing a nursing home outside their local market. See Section 3.2.

be biased. Specifically, one would systematically overestimate quality for more selective facilities, and underestimate quality for less selective facilities.

2.2 Nursing Home Compare

The most prominent of the policy reforms to raise quality is the introduction of the Nursing Home Compare website. NHC provides public report cards on each SNF, aggregating information on patient outcomes and process measures, results from recent facility inspections, and data on the levels of staffing. Facilities are assigned a star rating from one to five for each of the quality factors, in addition to an overall composite rating. The logic of NHC is to reduce information asymmetry; by giving consumers more information, they can choose higher quality providers.

While the NHC star ratings are widely used measures of quality by both consumers and researchers, there are substantial concerns regarding their accuracy. Two investigations found that even 5-Star facilities can have horrific living conditions, and consumers often feel misled by the system (Thomas 2014; Silver-Greenberg and Gebeloff 2021). In a sweeping overview of the literature on the NHC system, Konetzka, Yan, and Werner (2021) provide two potential explanations for the system's shortcomings. The first is a standard multitasking moral hazard problem: because NHC includes relatively narrow quality measures, facilities may target these measures while shirking on excluded measures (Holmstrom and Milgrom 1991). For instance, NHC initially reported the use of physical restraints, but not antipsychotic use – a class of drugs commonly misused as sedatives for the nursing home population, against clinical guidance. Konetzka, Brauner, Shega, and Werner (2014) find that public reporting of only physical restraints led to an increased use of antipsychotics among residents with severe cognitive impairments.

The second issue arises from the fact that two components of the NHC star ratings (the staffing and quality measures) come from records the facilities themselves submit, and so there is substantial bandwidth for data manipulation. By nature, research on this front must be somewhat indirect, but nonetheless the findings largely support claims of manipulation. For instance, despite large reported increases in staffing ratios under NHC, there was little evidence of increases in staffing costs (Sharma, Konetzka, and Smieliauskas 2019).

In light of these shortcomings, our first exercise after estimation is to evaluate the correlation between the NHC star ratings and the survival-based quality estimates.

2.3 Data

We combine a variety of administrative and public data sources from the Centers for Medicare & Medicaid Services (CMS). The base of our analytic file is resident-level assessment data from the Minimum Data Set (MDS) spanning 2000-2017. All CMS-certified SNFs are required to complete regular assessments of each resident, beginning at admission. These assessments include a high-dimensional array of clinical and demographic information that are reported to CMS to develop quality metrics and determine Medicare reimbursement rates. As we are interested in mortality irrespective of where the beneficiary was at time of death,[3] we merge the MDS data with the annual Medicare enrollment files, which contain each beneficiary's home zip code as well as their date of death, through 2017.

The patient data are supplemented by several public datasets that contain information on the SNFs themselves. These include the LTCFocus files, which collect facility-level characteristics (such as location, size, and ownership status) from administrative sources, as well as the quarterly NHC 5-Star Ratings.[4] To examine the relationship between SNF quality and Covid-19 outcomes, we collect the cumulative number of confirmed cases and deaths through December 19th, 2021 in addition to the resident and staff vaccination rates, available from CMS.[5] In our final exercise, we consider the role of Medicaid reimbursement in determining quality, for which we draw upon the California Medi-Cal cost reports.

2.4 Sample Construction

As we study survival to 90 days after admissions, when constructing our analytic sample we make use only of information available on the initial admission MDS assessment. We identify all new nursing home admissions, defined as no prior admission assessment at any SNF in the prior 365 days. Further, we restrict the assessments to patients who are enrolled in the Medicare program, which is required to track mortality and residence prior to admission. We define each patient's choice set as the 100 nearest within-state SNFs to their home zip code centroid. We exclude any admissions to nursing homes outside the choice set, including travel to any out-of-state SNFs. This restriction drops 4.5% of the sample. To increase statistical power while allowing time-varying quality estimates, we estimate the model over 4-year bins. This procedure leaves us with facility-level estimates for the year-bins 2001-2004, 2005-2008, 2009-2012, and 2013-2016, across each state separately.

3. For instance, a resident may be discharged to the hospital or the community where their death may not be recorded in the MDS assessments.

4. The LTCFocus data are provided by the Shaping Long Term Care in America Project at Brown University and funded in part by the National Institute on Aging (1P01AG027296).

5. The latest available data at the time of download from https://data.cms.gov/covid-19/covid-19-nursing-home-data.

All together, these restrictions leave us with a sample of 20,514,758 new nursing home admissions across all state-year-bins. Summary statistics are given in Appendix Table E.1. Reflecting the overall demographics of the nursing home population, our sample is disproportionately white (86.1%), older (average age 79.3), traveled 7.8 miles to their admitting SNF, and have an average 90 day survival rate of 85.8%.

3 Quality Estimation

To accommodate the potential for endogenous selection, we specify the following two-index model that is similar in spirit to Geweke, Gowrisankaran, and Town (2003) and Hull (2020).

3.1 Econometric Framework

A patient i who chooses SNF j has a latent health index, h_{ij}, that depends on j's quality β_j, patient demographics and comorbidities X_i and a health shock ε_{ij}. The potential outcome (90-day survival) of patient i at SNF j is Y_{ij}. These are expressed as:

$$h_{ij} = \beta_j + \gamma X_i^T + \varepsilon_{ij} \tag{1}$$

$$Y_{ij} = \mathbb{1}\left[h_{ij} > 0\right] \tag{2}$$

Patients do not choose SNFs at random. Instead, they choose the SNF that yields the highest utility, u_{ij}, defined as:

$$u_{ij} = \underbrace{\xi_j + \pi Z_{ij}^T}_{\delta_{ij}} + \eta_{ij} \tag{3}$$

Where ξ_j parametrizes constant SNF popularity and Z_{ij} is a vector of utility shifters. Most important among these is the geodetic distance between patient i and SNF j.[6] Maximization of this latent utility index over choice of j yields the patient's choice indicator, equal to one if patient i chooses SNF j, zero otherwise:

$$D_{ij} = \mathbb{1}\left[u_{ij} > u_{ik}, \forall k \neq j\right] \tag{4}$$

We model selection as a facility-specific correlation between the preference shock η_{ij} and the health shock ε_{ij}.[7] We assume that the shocks are jointly normally distributed

6. Here we draw upon the vast literature in health economics documenting a strong preference for proximity when selecting providers.

7. Notice that the scale of the latent indices h_{ij} and u_{ij} is unidentified. Therefore, without loss of generality we normalize the scale of the shocks to one: $\sigma_\varepsilon = \sigma_\eta = 1$, to mirror a conventional probit model.

with mean zero, unit variance and covariance α_j. Specifically, define the health shock as $\varepsilon_{ij} = \alpha_j \eta_{ij} + \sqrt{1-\alpha_j^2}\tilde{\varepsilon}_{ij}$ where η_{ij} and $\tilde{\varepsilon}_{ij}$ are standard normal random variables. Notice that because the probability of selecting SNF j is increasing in $\eta_{i,j}$, facilities with higher draws of α_j will have more favorable 'unobserved'[8] selection.

Estimation requires evaluating the log-probability of both the observed choices and survival outcomes. Typically, estimation of discrete choice models relies on the existence of closed-form probabilities, which our model does not produce. Instead, our approach starts by observing that, conditional on the structural parameters and, crucially, the realized preference shock for the selected SNF $\eta_{i,j(i)}$, the log-probabilites take a simple form:

$$\log P(D_{ij} = 1 \mid \theta_i) = \Sigma_{j' \neq j(i)} \log \Phi(\delta_{ij(i)} - \delta_{ij'} + \eta_{ij(i)}) \tag{5}$$

$$\log P(Y_{i,j(i)} = 1 \mid \theta_i) = \log\left(1 - \Phi\left(\frac{-(\beta_{j(i)} + \gamma X_i^T + \alpha_{j(i)}\eta_{i,j(i)})}{\sqrt{1-\alpha_{j(i)}^2}}\right)\right) \tag{6}$$

where $\theta_i := (\beta, \alpha, \gamma, \xi, \pi, \eta_{i,j(i)})$ is a vector that stacks all of the parameters we estimate and Φ is the standard normal cumulative distribution function. Appendix B contains the derivations and further details.

This approach bears resemblance to a Heckman correction, as we model the selection problem instead as an omitted variable problem (Heckman 1979). Moreover, we reduce the high-dimensional selection problem (a preference shock for each SNF in the choice set) to a single, sufficient, estimable parameter for each individual (the value of the preference shock for only the chosen SNF). The quantities in equations (5) and (6) are easy to compute. By conditioning on the value of $\eta_{i,j(i)}$ we only need to evaluate the CDF of the univariate standard normal, for which highly efficient, differentiable approximations exist. While this comes at the cost of estimating $\eta_{i,j(i)}$ for each individual, with variational inference we are able to easily estimate a model in which the number of parameters grows linearly with the sample size.

A common issue in quality estimation is accounting for few observations in some facilities, which may increase statistical noise in the quality estimates. To accommodate this, we use Bayesian inference. We specify the priors in Appendix Section B.2. Importantly, we use hierarchical priors for the quality parameters β_j, which implicitly regularizes the quality estimates towards the mean, increasing reliability especially for small SNFs. Empirical Bayes methods frequently used in quality estimation (e.g. Chetty, Friedman, and Rockoff 2014; Chandra, Finkelstein, Sacarny, and Syverson 2016) can be viewed as an approximation to

8. Specifically, unobserved in patient characteristics X_i.

the hierarchical model.

3.2 Estimation with Variational Inference

Estimating a choice model such as ours over a large number of admissions (more than 20 million) is computationally challenging. One common solution to relieve the computational burden is to estimate separately across narrowly defined geographic markets (Hull 2020; Einav, Finkelstein, and Mahoney 2022). This solution imposes strict assumptions on patients' choice sets, resulting in limited substitution patterns between nursing homes. In addition to the statistical efficiency concerns from losing observations, in our setting it is possible that patients who travel farther into other markets may not be representative of all patients who select a given facility, giving rise to potentially biased estimates. We relax assumptions on substitution patterns between facilities by defining large person-specific choice sets, as we do not need to impose narrow geographic markets to ensure computational tractability.

Defining such rich choice sets allows us to include many more observations in our estimation; however, doing so is computationally taxing. Our solution is to adopt a technique from the machine learning literature, variational inference, which is tailored for estimating complex Bayesian models efficiently. VI typically converges much faster than traditional, sampling-based approaches as it instead uses optimization to approximate the posterior distributions of each parameter in θ_i. This means that VI is able to take advantage of recent advances in optimization, such as automatic differentiation and massive parallelization, significantly improving the computation time. To implement VI, the researcher specifies a family of distributions (such as a factorizable Gaussian) and then searches over possible values of the parameters of those distributions (such as the means and scales of the Gaussian) so as to minimize the Kullback-Leibler divergence between the model-implied posteriors and the specified approximating distributions.

We provide a more in-depth overview of VI, along with further implementation details in Appendix A. We also detail the results from a simulation exercise, where we compare the VI estimates with both true (simulated) values as well as estimates recovered with an MCMC-based approach (NUTS, Hoffman and Gelman 2014). We find that VI recovers the true values accurately; the R^2 from a regression of the true β_j on the estimated values is equal to 0.98. Moreover, we find that VI estimates are indistinguishable from exact MCMC estimates, while run-time is improved by approximately 50-fold.

4 Correlates of Quality

We estimate the model separately across state-year-bins, and so we are limited in our ability to generate statements about heterogeneity in SNF quality across states, as the levels of β_j are unidentified. To resolve this issue for our subsequent exercises, we standardize all quality estimates within-market and denote the standardized estimates as $\hat{\beta}_{jt}$, such that a SNF with $\hat{\beta}_{jt} = 1$ has quality that is one standard deviation above the mean for that state-year-bin.

Assessing the correlation between SNF quality $\hat{\beta}_{jt}$ and unobserved selection $\hat{\alpha}_{jt}$, we find a sharp negative relationship ($\rho = -0.32$). Appendix Figure E.1 contains a binned scatterplot of $\hat{\alpha}_{jt}$ by $\hat{\beta}_{jt}$. This negative relationship indicates a traditional adverse selection mechanism: the sickest patients (lowest h_{ij}) have stronger preferences for higher quality SNFs. Because we model travel distances, rather than prices, this result implies that sicker patients travel farther for high quality SNFs.

4.1 Performance of Nursing Home Compare

Table 1 summarizes the findings on the relationship between SNF quality and the Nursing Home Compare star ratings. Each column represents the coefficients of a regression of quality $\hat{\beta}_{jt}$ on the overall rating, as well as the component ratings for facility inspections ('Survey'), MDS-derived patient-based measures ('Quality'), and overall levels of nurse and nurse-aid staffing ('Staffing'). Partial correlations between the component ratings and $\hat{\beta}_{jt}$, adjusted for geographic fixed effects, are also included.

Our baseline specification includes state-by-year-bin fixed effects. To further facilitate comparisons between SNFs that plausibly treat similar groups of patients, we also include fixed effects for hospital referral regions.[9] HRRs represent regional markets for tertiary medical care, and so provide natural markets to adjust for patient populations. Note that because the NHC data are quarterly and our parameter estimates correspond to 4-year bins, we assign the median NHC star rating for each period. The star ratings were introduced in 2009, and so we include estimates only for the last two year-bins (2009-2012, 2013-2016).

There is startlingly little relationship between the estimates of quality $\hat{\beta}_{jt}$ and the NHC measures. We find a within-state correlation of only 0.033. The estimates in Table 1 suggest that moving from a 1-star to 5-star facility corresponds to an increase in quality of only 0.06 standard deviations. The only NHC rating component across which $\hat{\beta}_{jt}$ is monotonically increasing is the 'Quality' rating, which aggregates MDS-derived patient-based quality metrics, but even in this best-performing component, the partial correlation within-state

9. More precisely, because some HRRs cross state boundaries and $\hat{\beta}_{jt}$ is standardized within-state, we instead construct HRR-by-state indicators.

is only 0.147. These results are unsurprising in light of the pitfalls of the Nursing Home Compare rating system outlined in Section 2.2. Such poor performance of CMS quality scores with outside measures is not unheard of. For instance, Abaluck, Caceres Bravo, Hull, and Starc (2021) study Medicare Advantage plans and find very low correlation between a mortality-based quality measure and analogous CMS scores.

4.2 Facility Characteristics

Table 2 reports univariate regressions of $\hat{\beta}_{jt}$ on several facility characteristics, similarly adjusted for state-year and HRR-year fixed effects. After conditioning on HRR, we find that for-profit and chain facilities perform no worse than their counterparts, though hospital-based SNFs do have lower quality. We also find that each measure of nurse or nurse-aide hours is positively correlated with $\hat{\beta}_{jt}$. Overall, we find that very little of the variation in $\hat{\beta}_{jt}$ can be explained by the characteristics of patients or facilities. A regression of $\hat{\beta}_{jt}$ on all of the facility and patient characteristics in Table 2 has an R^2 of only 0.10, after conditioning on HRR and year-bin.

5 Quality and Covid-19

The Covid-19 pandemic has been particularly devastating for nursing home residents, as SNFs have been centers of outbreak and excess mortality since the start of the crisis. In the wake of the pandemic, a flurry of recent research has sought to examine the determinants of Covid-19 spread in nursing homes, with a particular focus on whether higher quality SNFs performed better in preventing adverse outcomes. Konetzka et al. (2021) provide a systematic review of the literature, and find that there is no relationship between the NHC star ratings and various Covid-19 outcomes. The authors conclude that there is little that SNFs could have done to avert severe outcomes.

Given the lack of correlation between SNF quality $\hat{\beta}_{jt}$ and the star ratings documented in Section 4.1, it is possible that the relatively poor performance of higher rated SNFs is instead due to shortcomings in conventional metrics. We ask whether higher quality SNFs, as measured by $\hat{\beta}_{jt}$, fared relatively better during the first 21 months of the pandemic. Considering the 3-year lag between last of our year-bins (ending in 2016) and the beginning of the Covid-19 pandemic, we first verify that the estimates of $\hat{\beta}_{jt}$ are highly stable across time: the AR(1) regression coefficient of $\hat{\beta}_{jt}$ across year-bins is 0.805 (Appendix Table E.2 reports the autocorrelations.)

Figure 1 presents binned scatter plots of Covid-19 resident cases and deaths by SNF quality $\hat{\beta}_{jt}$ from the latest available year-bin, 2013-2016. Due to the geographic variation in Covid-19 spread, we adjust the figures for HRR fixed effects.[10] Given the mechanical relationship between size and counts of Covid-19 cases, we follow McGarry, Barnett, Grabowski, and Gandhi (2022) and calculate the number of cases and deaths per bed. Higher quality SNFs experienced lower rates of both Covid-19 cases and deaths, even conditional on HRR or county fixed effects.

Given the negative relationship between $\hat{\beta}_{jt}$ and Covid-19 outcomes observed in Figure 1, we ask whether any omitted variables may explain these results. To assess the extent of this possibility, we run a series of regressions of the following form:

$$y_j = \lambda_1 \hat{\beta}_{jt} + \lambda X_j^T + \epsilon_j \tag{7}$$

where y_j is one of several Covid-19 outcomes, X_j^T is a vector of facility characteristics and $\hat{\beta}_{jt}$ is the quality estimate from the last available year-bin, 2013-2016. We include the NHC Overall Rating as it appeared on the website in December 2019. We also include measures of size, the mean household income for the zip code, for-profit ownership and chain membership, the presence of an Alzheimer's unit, and whether the SNF is located in a hospital. We consider several Covid-19 outcomes, including the cumulative number of confirmed resident cases/deaths per bed, as well as staff and resident vaccination rates.

Table 3 presents the results from these regressions for several Covid-19 outcomes. We consider four specifications. In column (1), we replicate the regressions underlying the binscatters in Figure 1, and adjust only for HRR fixed effects. In column (2), we include the NHC Overall Rating. Column (3) includes each of the additional controls contained in X_j. Column (4) replaces the HRR fixed effects with county fixed effects.

Consistent with prior literature, we find that the 2019 NHC Overall Rating is not predictive of resident Covid cases or deaths, though vaccination rates at those facilities are higher. In contrast, we find that $\hat{\beta}_{jt}$ significantly predicts resident cases and deaths, and is stable across specifications. Facilities with a one standard deviation higher value of $\hat{\beta}_j$ had 2.4% fewer cases and 3.3% fewer deaths due to Covid-19. We find evidence that higher quality SNFs also vaccinated their staff and residents at higher rates, though these estimates are more sensitive to the choice of controls. While these results explain relatively little of the overall variation in Covid-19 outcomes, they suggest that the reported non-correlation between quality and Covid-19 outcomes is partially due to inadequate measures of quality.

10. In a robustness exercise, we also consider narrower geographic fixed effects (county-level) and find similar results (Appendix Figure E.2). However, this approach drops counties that contain only a single SNF.

6 Role of Medicaid Reimbursement Rates

Finally, we examine how an increase in the average Medicaid reimbursement rate would affect nursing home quality. Low Medicaid reimbursements are frequently cited as a chief cause of low quality (Harrington et al. 2008; Grabowski 2001). The daily rate can often fall below average cost, while more generous Medicare and private pay rates (partially) offset the losses.

Although we do not restrict our sample to nursing home residents whose stays are paid by Medicaid, nursing home quality has been found to be a common good across payer types, and facilities are legally required to provide equal quality to all residents, regardless of payer status (Grabowski, Gruber, and Angelelli 2008). Moreover, due to the institutional features governing long-term care reimbursement policy, nursing homes are highly dependent on Medicaid rates. Medicare provides coverage only for short-term, post-acute care provided at SNFs, whereas Medicaid coverage tends to be unlimited. Accordingly, while most patients begin their stays covered by Medicare, many who require long-term, custodial care eventually transition to Medicaid coverage. As such, on average Medicaid tends to account for approximately 50% of facility revenue (Spanko 2019).

6.1 Setting: California

To examine the effect of increasing Medicaid reimbursement rates on our nursing home quality estimates $\hat{\beta}_{jt}$, our aim is to estimate an equation of the form:

$$\hat{\beta}_{jt} = \delta_1 \log(R_{jt}^{mcaid}) + \delta X_{jt} + \nu_{jt} \tag{8}$$

However, we face two immediate challenges to doing so. The first is that Medicaid policies vary by state, and so national data on rates R_{jt}^{mcaid} are not readily available. We resolve this by narrowing our focus to one large state (California) where we have rich data on reimbursements.

The second challenge regards identification of the coefficient of interest, δ_1. We handle identification of this parameter using an instrumental variables approach in the style of Hackmann (2019). However, before understanding the potential endogeneity problem, it is important to consider how R_{jt}^{mcaid} is determined. In 2005, California reformed its reimbursement methodology so that facility j's per diem rate is a function of its own lagged average costs as well as the lagged average costs of facilities in a geographic peer group p, defined as groups of counties c. That is:

$$R_{jt}^{mcaid} = g(AC_{j,t-2}, AC_{c(j),t-2}^{p(j)}, AC_{-c(j),t-2}^{p(j)}) \tag{9}$$

where $AC_{c(j),t}^{p(j)}$ denotes average costs in county $c(j)$ and peer group $p(j)$. The empirical challenge to estimating δ_1 is the potential correlation between $\log(R_{jt}^{mcaid})$ and ν_{jt}. This is concerning, as we see that j's lagged costs enter directly into R_{jt}^{mcaid} from equation (9), and we find that quality is highly persistent across time (Appendix Table E.2). The direction of the bias is theoretically ambiguous. For instance, higher quality facilities may attract sicker patients, prompting facilities to hire more staff thereby raising average costs, which generates a positive correlation. Alternatively, an unobserved negative supply shock (such as an increase in wages) would raise average costs but may reduce quality, generating a negative relationship.

6.2 Hackmann Instruments

Hackmann (2019), who studies the relationship between average Medicaid rates and the level of nurse staffing in Pennsylvania (which uses a similar reimbursement methodology), notices that under the assumption that nursing homes compete in locally segmented markets (counties), one could use variation in the lagged average costs of distant facilities in the peer group of j to estimate δ_1: that is, by assuming independence of $AC_{-c(j),t-2}^{p(j)} \perp \nu_{jt}$. Under this assumption, one can use these lagged average costs of distant 'peer' facilities to instrument for the reimbursement rate of a given SNF.

Following this approach, for each county-year we construct average costs across facilities in the same peer group, but operating in a different county. California uses four primary cost categories for determining reimbursements: direct care labor costs (such as nurses), indirect care labor costs (such as housekeeping), non-labor costs, and administrative costs.[11] For each facility, we calculate the per diem Medicaid rate in year t as the amount of Medicaid revenue from skilled nursing divided by the number of Medicaid-covered skilled nursing days. We use the log average of each cost category separately in our first stage specification.

Appendix Figure E.3 presents a map of California's seven peer regions.[12] The median peer group contains 9 counties, and all contain at least 5 counties. The lone exception is Los Angeles county, which serves as its own peer group due to its greater size and relatively lower average costs. To construct the instruments for Los Angeles county, we instead use

11. The vast majority of costs come from direct care, via nurses and nurse aides.

12. These peer regions were generated by a consultant hired by the state using cost report data during the implementation process of the new reimbursement methdology. The peer groups are the result of a clustering analysis performed on 2003 nursing facility direct care costs.

the average costs of all other facilities in LA. In a robustness exercise, we find similar results when we exclude facilities in Los Angeles entirely (Appendix Table E.3).

Because our quality estimates are estimated at the four-year bin level, we calculate the averages of R_{jt}^{mcaid} and $AC_{-c(j),t-2}^{p(j)}$ across years within a bin. Using data from the American Community Survey 1% annual files, we include county-level controls for the population over age 65, the share with any physical or cognitive difficulty, and the log of household income. Following Hackmann, we also include facility-level controls for the presence of an Alzheimer's ward, for-profit ownership, chain membership, and the number of beds. We also include year-bin fixed effects. Finally, given the peer group definitions, we include an indicator for Los Angeles county.

6.3 Results

Figures 2a and 2b visualize the first stage and second stage results, respectively. The first panel presents a binned scatterplot of the average Medicaid reimbursement rate by the lagged average direct care labor costs[13] of distant peer facilities, as this represents the largest cost component. Similarly, the second panel plots the quality estimates $\hat{\beta}_{jt}$ by the predicted (second-stage) Medicaid reimbursement rate. Both figures are in log-scale, reflecting the underlying regressions. Table 2c reports the corresponding results of the 2SLS estimation of equation (8) using each of the lagged average cost components of distant facilities as the excluded instruments. We find that the lag of distant facilities' costs, in particular direct care costs are predictive of facility j's Medicaid rate. The first-stage F-stat is 19.68.

Turning to our primary outcome in column (2), nursing home quality $\hat{\beta}_{jt}$, we find that a 10% increase in average Medicaid reimbursement over a 4-year bin corresponds to an increase in estimated quality of 0.28 standard deviations. For an average patient in California, this corresponds to an increase in predicted 90-day survival probability of 1.85 percentage points, though we cannot rule out an effect size as low as 0.34 percentage points. Using a Cox proportional hazards model to project the implied number of life-years saved, we estimate this effect corresponds to approximately 67,883 life-years from 2009-2012. Considering the costs to the taxpayer (approximately \$363 million annually), this policy is cost-effective for quality-adjusted life-years valued more than \$21,377, which is substantially below the existing estimates (Ganz, Simmons, and Schnelle 2005). Further details of this analysis are provided in Appendix D.

To examine the mechanism by which this quality improvement occurs, in columns (3)-(5) we replicate Hackmann (2019) and examine the roles of Medicaid reimbursement on the level

13. This includes nurse and nurse aide hours, but excludes non-direct care labor costs, such as housekeeping.

of nurse staffing. We find that in response to increases in the Medicaid rate, SNFs increase the number of registered nurse hours per resident day, with a corresponding decrease in the percent of licensed practical nurses, suggesting that firms use the greater reimbursement rates to substitute their medium-skilled labor with high-skilled labor. We find no effect of increasing the mean Medicaid rate on the number of certified nursing assistants, suggesting no change in the amount of low-skilled labor.

7 Conclusion

The low quality of care provided at nursing homes has long posed a challenge to policymakers. In this project, we propose a new method of estimating SNF quality. Applying the method to the universe of SNF admissions, we find that conventional quality measures have near-zero correlation with our survival-based approach. We also find that higher quality SNFs have fared better during the Covid-19 pandemic, in contrast to the medical consensus. Finally, we conduct an evaluation of the effects of raising the Medicaid reimbursement rate, and find that such a reform would improve survival rates, and that the increase is cost-effective under conservative assumptions of the value of a quality-adjusted life-year.

References

Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc. 2021. "Mortality Effects and Choice Across Private Health Insurance Plans." *The Quarterly Journal of Economics* 136, no. 3 (June): 1557–1610.

Bingham, Eli, Jonathan P. Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D. Goodman. 2018. "Pyro: Deep Universal Probabilistic Programming," October 18, 2018. Accessed 1/4/2022. arXiv: 1810.09538 `[cs, stat]`. http://arxiv.org/abs/1810.09538.

Blei, David M. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.

Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe. 2017. "Variational Inference: A Review for Statisticians." *Journal of the American Statistical Association* 112, no. 518 (April 3, 2017): 859–877.

Bonhomme, Stéphane. 2021. *Teams: Heterogeneity, Sorting, and Complementarity.* SSRN Scholarly Paper ID 3778749. Rochester, NY: Social Science Research Network, February 3, 2021.

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, no. 8 (August): 2110–2144.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104, no. 9 (September): 2593–2632.

Department of Tranportation. 2013. "Guidance on Treatment of the Economic Value of a Statistical Life in U.S. Department of Transportation Analyses — US Department of Transportation." Accessed 2/26/2022. https://www.transportation.gov/regulations/guidance-treatment-economic-value-statistical-life-us-department-transportation-analyses.

Doyle, Joseph, John Graves, and Jonathan Gruber. 2019. "Evaluating Measures of Hospital Quality: Evidence from Ambulance Referral Patterns." *The Review of Economics and Statistics* 101, no. 5 (December): 841–852.

Dranove, David, and Mark A. Satterthwaite. 1992. "Monopolistic Competition When Price and Quality Are Imperfectly Observable." *The RAND Journal of Economics* 23 (4): 518–534.

Einav, Liran, Amy Finkelstein, and Neale Mahoney. 2022. *Producing Health: Measuring Value Added of Nursing Homes.* Working Paper. Unpublished.

Gandhi, Ashvin. 2020. *Picking Your Patients: Selective Admissions in the Nursing Home Industry.* SSRN Scholarly Paper ID 3613950. Rochester, NY: Social Science Research Network, May 29, 2020.

Ganz, David A., Sandra F. Simmons, and John F. Schnelle. 2005. "Cost-Effectiveness of Recommended Nurse Staffing Levels for Short-Stay Skilled Nursing Facility Patients." *BMC Health Services Research* 5, no. 1 (May): 35.

Geweke, John, Gautam Gowrisankaran, and Robert J. Town. 2003. "Bayesian Inference for Hospital Quality in a Selection Model." *Econometrica* 71 (4): 1215–1238.

Gowrisankaran, G., and R. J. Town. 1999. "Estimating the Quality of Care in Hospitals Using Instrumental Variables." *Journal of Health Economics* 18, no. 6 (December): 747–767.

Grabowski, David C. 2001. "Medicaid Reimbursement and the Quality of Nursing Home Care." *Journal of Health Economics* 20, no. 4 (July): 549–569.

Grabowski, David C., Zhanlian Feng, Richard Hirth, Momotazur Rahman, and Vincent Mor. 2013. "Effect of Nursing Home Ownership on the Quality of Post-Acute Care: An Instrumental Variables Approach." *Journal of health economics* 32, no. 1 (January): 12–21.

Grabowski, David C., Jonathan Gruber, and Joseph J. Angelelli. 2008. "Nursing Home Quality as a Common Good." *The Review of Economics and Statistics* 90, no. 4 (November): 754–764.

Guarino, Cassandra M., Michelle Maxfield, Mark D. Reckase, Paul N. Thompson, and Jeffrey M. Wooldridge. 2015. "An Evaluation of Empirical Bayes's Estimation of Value-Added Teacher Performance Measures." *Journal of Educational and Behavioral Statistics* 40, no. 2 (April 1, 2015): 190–222.

Hackmann, Martin B, R Vincent Pohl, and Nicolas R Ziebarth. 2021. *Patient Versus Provider Incentives in Long-Term Care.* Working Paper. Unpublished, May.

Hackmann, Martin B. 2019. "Incentivizing Better Quality of Care: The Role of Medicaid and Competition in the Nursing Home Industry." *American Economic Review* 109, no. 5 (May): 1684–1716.

Harrington, Charlene, Janis O'Meara, Eric Collier, Taewoon Kang, Caroline Stephens, and Jamie Chang. 2008. *Impact of California's Medi-Cal Long Term Care Reimbursement Act on Access, Quality and Cost.* University of California San Francisco.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–161.

Hoffman, Matthew D., and Andrew Gelman. 2014. "The No-U-turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *The Journal of Machine Learning Research* 15, no. 1 (January 1, 2014): 1593–1623.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7:24–52.

Hull, Peter. 2020. *Estimating Hospital Quality with Quasi-Experimental Data.* Working Paper. Unpublished.

Jordan, Michael I., Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. 1998. "An Introduction to Variational Methods for Graphical Models." In *Learning in Graphical Models,* edited by Michael I. Jordan, 105–161. Dordrecht: Springer Netherlands.

Kingma, Diederik P., and Jimmy Ba. 2017. "Adam: A Method for Stochastic Optimization," January 29, 2017. Accessed 1/6/2022. arXiv: 1412.6980 `[cs]`. http://arxiv.org/abs/1412.6980.

Kingma, Diederik P., and Max Welling. 2014. "Auto-Encoding Variational Bayes," May 1, 2014. Accessed 1/6/2022. arXiv: 1312.6114 `[cs, stat]`. http://arxiv.org/abs/1312.6114.

Konetzka, R. Tamara, Daniel J. Brauner, Joseph Shega, and Rachel M. Werner. 2014. "The Effects of Public Reporting on Physical Restraints and Antipsychotic Use in Nursing Home Residents with Severe Cognitive Impairment." *Journal of the American Geriatrics Society* 62, no. 3 (March): 454–461.

Konetzka, R. Tamara, Elizabeth M. White, Alexander Pralea, David C. Grabowski, and Vincent Mor. 2021. "A Systematic Review of Long-Term Care Facility Characteristics Associated with COVID-19 Outcomes." *Journal of the American Geriatrics Society* 69 (10): 2766–2777.

Konetzka, Tamara, Kevin Yan, and Rachel M. Werner. 2021. "Two Decades of Nursing Home Compare: What Have We Learned?" *Medical Care Research and Review* 78, no. 4 (August 1, 2021): 295–310.

McGarry, Brian E., Michael L. Barnett, David C. Grabowski, and Ashvin D. Gandhi. 2022. "Nursing Home Staff Vaccination and Covid-19 Outcomes." *New England Journal of Medicine* 386, no. 4 (January 27, 2022): 397–398.

Medina, Marco Avella, José Luis Montiel Olea, Cynthia Rush, and Amilcar Velez. 2021. "On the Robustness to Misspecification of Alpha-Posteriors and Their Variational Approximations," April 16, 2021. Accessed 12/21/2021. arXiv: 2104.08324 `[cs, math, stat]`. http://arxiv.org/abs/2104.08324.

Mele, Angelo, and Lingjiong Zhu. 2021. "Approximate Variational Estimation for a Model of Network Formation." *The Review of Economics and Statistics* (Forthcoming 2021).

Mor, Vincent, Jane Banaszak-Holl, and Jacqueline S. Zinn. 1995. "The Trend Towward Specialization in Nursing Care Facilities." *Generations: Journal of the American Society on Aging* 19 (4): 24–29.

Office of Inspector General. 2014. *Adverse Events in Skilled Nursing Facilities: National Incidence Among Medicare Beneficiaries.* Department of Health and Human Services, February.

Phan, Du, Neeraj Pradhan, and Martin Jankowiak. 2019. "Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro," December 24, 2019. Accessed 1/4/2022. arXiv: 1912.11554 `[cs, stat]`. http://arxiv.org/abs/1912.11554.

Robbins, Herbert, and Sutton Monro. 1951. "A Stochastic Approximation Method." *The Annals of Mathematical Statistics* 22, no. 3 (September): 400–407.

Rucinski, Tracy. 2018. "Illinois Nursing Homes Sue State over Low Medicaid Rates." Reuters, January 13, 2018. Accessed 2/26/2022. https://www.reuters.com/article/us-illinois-medicaid-idUSKBN1F201Y.

Ruiz, Francisco J. R., Susan Athey, and David M. Blei. 2020. "SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements." *The Annals of Applied Statistics* 14, no. 1 (March 1, 2020): 1–27.

Sharma, Hari, R. Tamara Konetzka, and Fabrice Smieliauskas. 2019. "The Relationship Between Reported Staffing and Expenditures in Nursing Homes." *Medical Care Research and Review* 76, no. 6 (December 1, 2019): 758–783.

Silver-Greenberg, Jessica, and Robert Gebeloff. 2021. "How U.S. Ratings of Nursing Homes Mislead the Public." The New York Times, March 13, 2021. Accessed 1/11/2022. https://www.nytimes.com/2021/03/13/business/nursing-homes-ratings-medicare-covid.html.

Spanko, Alex. 2019. "Medicaid's Share of Nursing Home Revenue, Resident Days Hits Record High as Medicare Drops to Historic Low." Skilled Nursing News, December 11, 2019. Accessed 2/26/2022. https://skillednursingnews.com/2019/12/medicaids-share-of-nursing-home-revenue-resident-days-hits-record-high-as-medicare-drops-to-historic-low/.

The White House. 2022. "FACT SHEET: Protecting Seniors and People with Disabilities by Improving Safety and Quality of Care in the Nation's Nursing Homes." The White House, February 28, 2022. Accessed 2/28/2022. https://www.whitehouse.gov/briefing-room/statements-releases/2022/02/28/fact-sheet-protecting-seniors-and-people-with-disabilities-by-improving-safety-and-quality-of-care-in-the-nations-nursing-homes/.

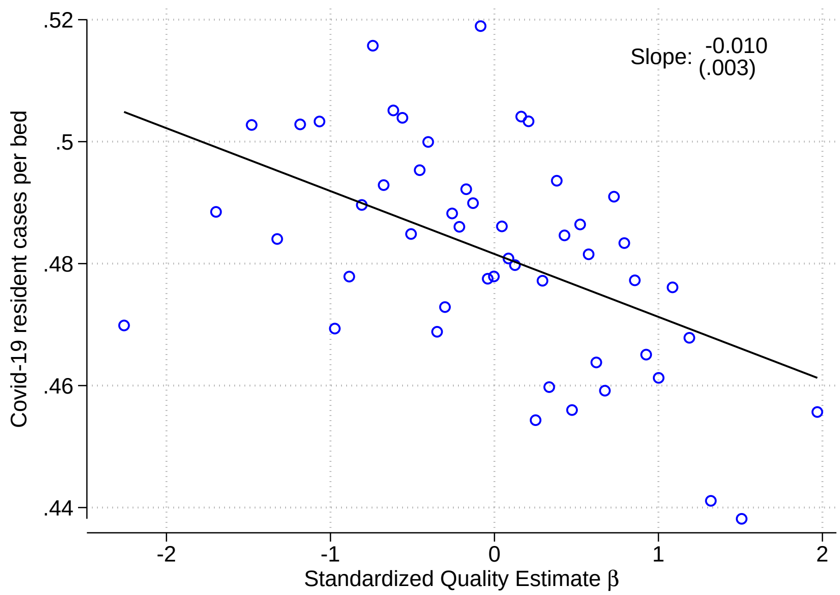
Thomas, Katie. 2014. "Medicare Star Ratings Allow Nursing Homes to Game the System." The New York Times, August 24, 2014. Accessed 11/22/2019. https://www.nytimes.com/2014/08/25/business/medicare-star-ratings-allow-nursing-homes-to-game-the-system.html.

Vafa, Keyon, Suresh Naidu, and David M. Blei. 2020. "Text-Based Ideal Points," July 21, 2020. Accessed 12/21/2021. arXiv: 1312.6114 `[cs, stat]`. http://arxiv.org/abs/2005.04232.

Wang, Yixin, and David M. Blei. 2019. "Frequentist Consistency of Variational Bayes." *Journal of the American Statistical Association* 114, no. 527 (July 3, 2019): 1147–1161.

8 Tables and Figures

(a) Cumulative Covid-19 resident cases per bed



(b) Cumulative Covid-19 resident deaths per bed

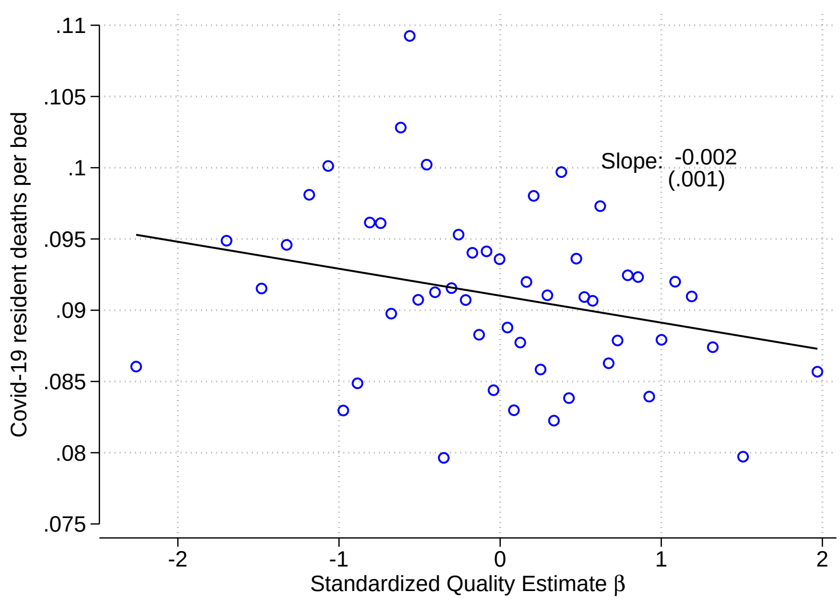


Figure 1: Covid-19 Outcomes by Nursing Home Quality Estimates β
Notes: Figures establish the negative relationship between nursing home quality in 2013-2016 and Covid-19 outcomes through 2021. Top panel presents a binned scatterplot of cumulative confirmed Covid-19 cases per bed through December 19, 2021 by nursing home quality $\hat{\beta}_j$ estimated in the last available year-bin, 2013-2016. Bottom panel presents Covid-19 deaths per bed by nursing home quality. Both are adjusted for hospital referral region fixed effects.

	(a) State x Year Fixed Effects				(b) HRR x Year Fixed Effects			
	(1) Overall	(2) Survey	(3) Quality	(4) Staffing	(1) Overall	(2) Survey	(3) Quality	(4) Staffing
2-Star	0.0203	-0.0406	0.132	-0.0632	-0.0246	-0.0866	0.0819	-0.0243
	(0.0219)	(0.0187)	(0.0285)	(0.0247)	(0.0195)	(0.0164)	(0.0254)	(0.0216)
3-Star	-0.0316	-0.0908	0.243	-0.101	-0.0925	-0.138	0.127	-0.0337
	(0.0225)	(0.0197)	(0.0277)	(0.0247)	(0.0199)	(0.0174)	(0.0250)	(0.0218)
4-Star	-0.0108	-0.0484	0.385	-0.0806	-0.0743	-0.111	0.207	-0.00109
	(0.0228)	(0.0207)	(0.0279)	(0.0256)	(0.0204)	(0.0186)	(0.0252)	(0.0225)
5-Star	0.180	0.0445	0.554	0.0957	0.0631	-0.0235	0.289	0.132
	(0.0262)	(0.0284)	(0.0307)	(0.0351)	(0.0235)	(0.0257)	(0.0278)	(0.0310)
Observations	28813	28813	28813	28743	28738	28738	28738	28668
Partial R^2	.0045	.0015	.0217	.0027	.0036	.0033	.0072	.0022
Partial ρ	.033	-.004	.147	.005	.005	-.026	.084	.024

Table 1: Regressions of Quality Estimates $\hat{\beta}_{jt}$ on Nursing Home Compare 5-Star Ratings

Notes: Table establishes the weak correlation between nursing home quality and public report cards. Table reports regressions of standardized SNF quality estimates $\hat{\beta}_{jt}$ on each component of the Nursing Home Compare 5-Star ratings. The mean of $\hat{\beta}_{jt}$ is zero. Partial R^2 reports the fraction of the variance in $\hat{\beta}_{jt}$ explained by the star ratings, conditional on the fixed effects. Partial ρ is the partial correlation between the star ratings (as continuous variables) and quality $\hat{\beta}_{jt}$, conditional on the fixed effects. We include estimates only for the 2009-2012 and 2013-2016 year-bins, due to availability of the star ratings. Standard errors in parentheses are clustered at the facility-level.

	(a) State x Year Fixed Effects		(b) HRR x Year Fixed Effects	
For-Profit	0.0443	(0.0162)	-0.00739	(0.0141)
Chain	-0.00752	(0.0137)	0.00648	(0.0119)
Alzheimer's Unit	-0.0219	(0.0176)	0.0000246	(0.0153)
Hospital-Based	-0.392	(0.0280)	-0.314	(0.0243)
log(Total Beds)	0.0661	(0.0132)	-0.0459	(0.0114)
% Occupancy	0.00363	(0.000511)	0.00204	(0.000453)
log(RNs/Resident)	0.0667	(0.00883)	0.0178	(0.00785)
log(LPNs/Resident)	0.00518	(0.0151)	0.0400	(0.0131)
log(CNAs/Resident)	0.173	(0.0225)	0.202	(0.0196)
% White	-0.0147	(0.000353)	-0.0100	(0.000373)
% Medicare	-0.00217	(0.000425)	-0.00400	(0.000375)
% Medicaid	-0.000765	(0.000317)	0.000678	(0.000279)
Mean ADL Score	0.0490	(0.00318)	0.0258	(0.00274)
Mean Age	-0.0284	(0.00119)	-0.0193	(0.00108)

Table 2: Univariate Regressions of Nursing Home Quality $\hat{\beta}_{jt}$ on Facility Characteristics

Notes: Table reports a series of univariate regressions of estimated nursing home quality $\hat{\beta}_{jt}$ on several facility characteristics, collected from the Long-Term Care Focus files. The dependent variable is the market-standardized estimate of $\hat{\beta}_{jt}$, so coefficients may be interpreted in terms of within-state year-bin standard deviations of quality. Column (a) reports regressions that include market (state-by-year-bin) fixed effects. Column (b) includes HRR-by-year-bin fixed effects. For-profit, chain, Alzheimer's unit, and hospital-based are each indicator variables. % Occupancy, White, Medicare, and Medicaid are each scaled from 0 to 100. log(RNs/Resident), log(LPNs/Resident), and log(CNAs/Resident) measure the log number of hours worked per resident-days for registered nurses, licensed practical nurses, and certified nursing aides, respectively. Mean ADL Score is an index of patient severity (higher values indicate that patients require more help with activities of daily living). Standard errors in parentheses are clustered at the facility-level.

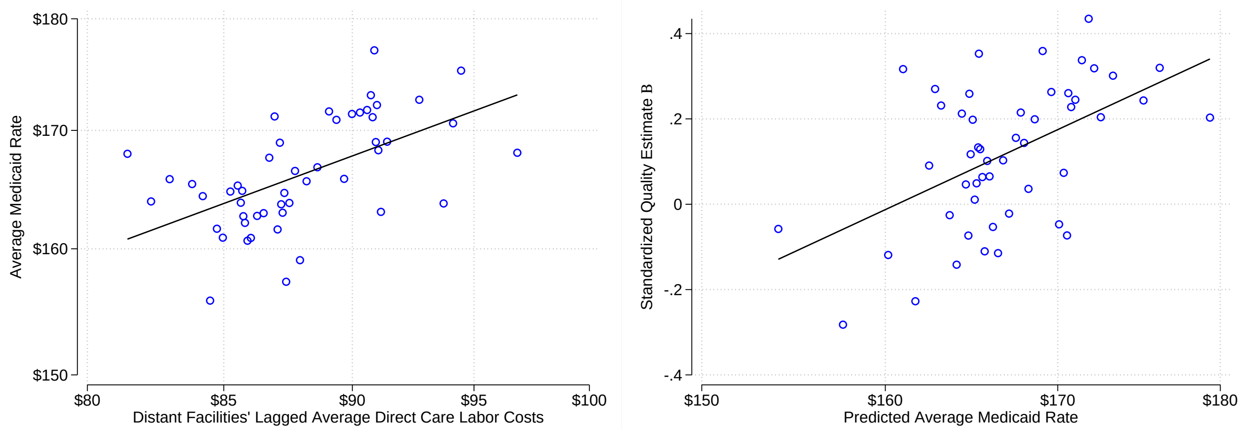
		(1)	(2)	(3)	(4)
(a) Resident Cases Per Bed	SNF Quality β	-0.0103	-0.0107	-0.0101	-0.0117
		(0.00275)	(0.00276)	(0.00278)	(0.00396)
	2019 NHC Rating		-0.00667	0.00175	0.00367
			(0.00178)	(0.00188)	(0.00210)
	Mean	0.482	0.482	0.482	0.480
	Observations	13160	13087	13074	12245
(b) Resident Deaths Per Bed	SNF Quality β	-0.00189	-0.00188	-0.00241	-0.00304
		(0.000859)	(0.000862)	(0.000884)	(0.00123)
	2019 NHC Rating		0.000132	0.000851	0.00117
			(0.000565)	(0.000606)	(0.000695)
	Mean	0.0911	0.0912	0.0912	0.0910
	Observations	13160	13087	13074	12245
(c) Staff Cases Per Bed	SNF Quality β	-0.00863	-0.00804	-0.0109	-0.0132
		(0.00227)	(0.00221)	(0.00218)	(0.00301)
	2019 NHC Rating		0.0343	0.0209	0.0223
			(0.00148)	(0.00143)	(0.00159)
	Mean	0.467	0.467	0.467	0.465
	Observations	13160	13087	13074	12245
(d) Vaccinated Residents, %	SNF Quality β	0.263	0.308	0.0567	0.434
		(0.103)	(0.101)	(0.101)	(0.145)
	2019 NHC Rating		1.509	1.219	1.295
			(0.0723)	(0.0745)	(0.0850)
	Mean	87.25	87.26	87.26	87.22
	Observations	13158	13085	13072	12247
(e) Vaccinated Staff, %	SNF Quality β	1.402	1.453	1.289	-0.143
		(0.141)	(0.139)	(0.140)	(0.186)
	2019 NHC Rating		1.513	1.288	1.382
			(0.0919)	(0.0955)	(0.103)
	Mean	78.23	78.25	78.25	78.56
	Observations	13163	13090	13077	12251
	Additional Controls	No	No	Yes	Yes
	HRR FEs	Yes	Yes	Yes	No
	County FEs	No	No	No	Yes

Table 3: Regressions of Covid-19 Outcomes on Nursing Home Quality Estimates

Notes: Table reports regressions of several Covid-19 outcomes on several facility characteristics. SNF quality estimates $\hat{\beta}_j$ are from the last available year-bin, 2013-2016. The Nursing Home Compare Overall Rating is as it appeared on the website in December 2019. Additional controls include the log of mean household income for the SNF's zip code, log number of total beds, and indicators for the for-profit status, chain membership, the presence of an Alzheimer's unit, and whether the SNF is hospital-based. Standard errors in parentheses are all clustered at the facility-level.

(a) Distant Facilities' Costs and Medicaid Rate

(b) Predicted Medicaid Rate and Quality



(c) Medicaid Rate and Nursing Home Quality Regressions

	(1) First Stage	(2) Quality β	(3) log(RN^{res})	(4) log(LPN^{res})	(5) log(CNA^{res})
log(Direct Care)	0.426				
	(0.101)				
log(Indirect Care)	-0.0432				
	(0.154)				
log(Non-Labor)	-0.0772				
	(0.161)				
log(Admin)	0.103				
	(0.0918)				
log(Medicaid Rate)		2.831	3.573	-1.519	-0.0620
		(1.190)	(1.073)	(0.436)	(0.132)
Observations	2533	2533	2533	2533	2533
F-Statistic	19.68				

Figure 2: Effects of the Average Medicaid Rate on Quality and Staffing

Notes: Results of the two-stage least squares Medicaid reimbursement exercise. Panel (a) presents the first-stage, a binned scatterplot of each SNF's Medicaid rate and lagged average direct labor costs of distant facilities in the peer group, controlling for the other instruments and the variables listed below. Panel (b) presents the second-stage, a binned scatterplot of facility quality $\hat{\beta}_{jt}$ and its predicted Medicaid rate from distant facilities' costs, controlling for the variables listed below. Both figures are in log-scale, reflecting the underlying regressions. Panel (c) presents the corresponding regressions. log(RN^{res}), log(LPN^{res}), log(CNA^{res}) respectively abbreviate the log number of registered nurses, licensed practical nurses, and certified nursing aide hours per resident day. All specifications include controls for for-profit ownership, the presence of an Alzheimer's unit, the log number of beds, and local demographics. Standard errors in parentheses clustered at the county-level.

A Details on Variational Inference

A.1 Overview of Variational Inference

There are very few applications of variational inference in estimating an economic model, and so many readers may be unfamiliar with the approach. As such, we provide a quick overview of the method. Readers interested in a more in depth treatment should refer to Blei, Kucukelbir, and McAuliffe (2017).

The goal of Bayesian computation is to estimate $p(\theta|x)$, the posterior density of the parameters of the model, θ, given the data, x. By Bayes' rule this is proportional to the joint density of data and parameters,

$$p(\theta|x) = \frac{p(\theta, x)}{p(x)} \tag{10}$$

For a given prior, $p(\theta)$ the numerator in (10) is easily computed since $p(\theta, x) = p(x|\theta)p(\theta)$, where $p(\theta)$ is the prior density. However, the denominator is generally computationally intractable.

Variational inference starts by postulating a parametrized family of distributions:

$$Q_\Psi = \{q(\theta|\psi) \quad \forall \psi \in \Psi\}$$

It then approximates the posterior density in (10) by finding a member of Q_Ψ that minimizes the distance between $p(\theta|x)$ and $q(\theta|\psi)$. Given a family of distributions the goal is then to find $\psi^\star$ that minimizes the distance, measured by the Kullback-Leibler divergence, between the posterior and the approximating distribution:

$$\psi^* = \arg\min_{\psi \in \Psi} KL(q(\theta|\psi)||p(\theta|x)), \tag{11}$$

where KL divergence is defined as

$$\begin{aligned} \mathrm{KL}(q(\theta|\psi) \mid\mid p(\theta|x)) &= \int q(\theta|\psi) \log \frac{q(\theta|\psi)}{p(\theta|x)} d\theta \\ &= \mathbb{E}_q[\log q(\theta|\psi) - \log p(\theta|x)]. \end{aligned} \tag{12}$$

The quantity in (12) is not computable as it still involves the constant, $p(x)$. However, for the purpose of optimization, the constant can be ignored as its value does not depend on ψ. The resulting objective function, known as the Evidence Lower Bound (ELBO), is obtained by subtracting the log evidence, $p(x)$ from the KL divergence and switching the sign.

$$\begin{aligned} \mathrm{ELBO}(\psi) &:= \log p(x) - \mathrm{KL}(q(\theta|\psi) \mid\mid p(\theta|x)) \\ &= \log p(x) + \mathbb{E}_q[\log p(\theta|x) - \log q(\theta|\psi)] \\ &= \log p(x) + \mathbb{E}_q[\log p(\theta, x) - \log p(x) - \log q(\theta|\psi)] \\ &= \mathbb{E}_q[\log p(\theta, x)] - \mathbb{E}_q[\log q(\theta|\psi)] \end{aligned} \tag{13}$$

Notice that maximizing the ELBO is equivalent to minimizing the KL divergence. That is,

$$\psi^* = \arg\min_{\psi\in\Psi} KL(q(\theta|\psi)||p(\theta|x)) \iff \psi^* = \arg\max_{\psi\in\Psi} \text{ELBO}(\psi). \tag{14}$$

A typical tradeoff in Bayesian computation is that (a) the parameters with a high posterior density should be good at fitting the data, and (b) the posterior distribution of parameters should be close to the prior. This tradeoff can be easily seen by rewriting (13) as

$$\text{ELBO}(\psi) = \mathbb{E}_q[\log p(\theta, x)] - \mathbb{E}_q[\log q(\theta|\psi)] = \mathbb{E}_q[\log p(x|\theta)] - \text{KL}(q(\theta|\psi) \mid p(\theta)).$$

The first term in the equation above ensures that the parameters fit the data, while the second ensures that the approximate posterior $q(\theta|\psi)$ is as close as possible to the prior $p(\theta)$.

A.2 Use of Variational Inference in Economics

Variational inference and related approaches have long been popular in computer science (see Jordan, Ghahramani, Jaakkola, and Saul 1998). It has been used, for instance, to estimate the Latent Dirichlet Allocation model (Blei 2003) which is now popular among economists who use text data. Despite this popularity, there are few instances of the use of VI for estimating economic models. Notable exceptions include Bonhomme (2021) (team production) and Mele and Zhu (2021) (network formation). There are also several examples at the intersection of economics and computer science including Ruiz, Athey, and Blei (2020) (demand) and Vafa, Naidu, and Blei (2020) (polarization in text). One reason for this limited adoption may be that, until recently, there have been few theoretical results on the statistical properties of the approach. However, recent work (Wang and Blei 2019; Medina, Olea, Rush, and Velez 2021) established appealing properties of the approach regarding asymptotic behavior and robustness to model mis-specification. It is worth noting that VI has a great potential in terms of scalability (though massive parallelization and data-subsampling) and can benefit from modern specialized hardware (GPUs/TPUs). We anticipate that VI may see widespread adoption by economists in the near future.

A.3 Implementation

The parameters whose posterior distributions we estimate are: (a) the parameters of the choice model $((\xi_j)_{j\in\mathcal{J}},\ \pi)$, (b) the parameters of the survival model $((\beta_j)_{j\in\mathcal{J}},\ \gamma)$, (c) the selection parameters $(\alpha_j)_{j\in\mathcal{J}}$ and (d) the preference shocks for the selected SNFs $((\eta_{i,j(i)})_{i\in\mathcal{I}})$. Each of these parameters receives a variational family that we describe in the subsequent section. All the parameters of the variational families are collected into a single vector ψ. The joint log-probability of the model given the parameters is given by

$$\log p(\theta, x) = \sum_i (\log p(x_i|\theta)) + \log p(\theta) = \sum_i \left(\log P(Y_{i,j(i)}|\theta) + \log P(D_{i,j(i)}|\theta)\right) + \log p(\theta), \tag{15}$$

where $p(\theta)$ is the prior density of the parameters. We use independent priors for each parameter so $\log p(\theta) = \sum_i \log p(\eta_{i,j(i)}) + \sum_j \log p(\beta_j) + \sum_j \log p(\xi_j) + \sum_j \log p(\alpha_j) + \log p(\gamma) + \log p(\pi)$. Likewise, we use a factorizable variational family so the approximate posterior density of the parameters is given by

$$\begin{aligned}\log q(\theta|\psi) = &\sum_i \log q(\eta_{i,j(i)}|\psi_{\eta_{i,j(i)}}) + \sum_j \log q(\beta_j|\psi_{\beta_j}) + \sum_j \log q(\xi_j|\psi_{\xi_j}) \\ &+ \sum_j \log q(\alpha_j|\psi_{\alpha_j}) + \log q(\gamma|\psi_\gamma) + \log q(\pi|\psi_\pi)\end{aligned} \tag{16}$$

We specify the ELBO using Numpyro (Phan, Pradhan, and Jankowiak 2019; Bingham et al. 2018), a Python library for probabilistic programming. Full implementation details are beyond the scope of this paper, though we emphasize a few key points. First, as Equations (15) and (16) show, the ELBO is an expectation over a sum of components related to individual observations which makes it easily parallelizable. Second, we are able to leverage automatic differentiation to compute an unbiased estimate of the gradient of the ELBO. This is achieved using the so-called 'reparametrization trick' of Kingma and Welling (2014). This trick enables expressing the gradient of the ELBO as an expectation of the gradient with respect to a noise distribution, τ where the density does not depend on the parameters ψ.

$$\nabla_\psi ELBO(\psi) = \nabla_\psi \mathbb{E}_q(\log p(\theta, x) - \log q(\theta|\psi)) = \mathbb{E}_\tau \left(\nabla_\psi (\log p(f(\psi,\tau)) - \log q(f(\psi,\tau)|\psi))\right), \tag{17}$$

where f is the reparametrizing function.[14]

Consequently, we can obtain an unbiased estimate of $\nabla_\psi \, \text{ELBO}(\psi)$ by taking draws from the noise distribution. This is critical given the dimension of the parameter space (which in our model scales linearly with the sample size), so gradient-free optimization methods are infeasible. On the other hand, the analytic form of $\nabla_\psi \, \text{ELBO}(\psi)$ may not exist. Fortunately, this unbiased estimate of the gradient is sufficient to specify an algorithm that is guaranteed to converge to a local optimum (Robbins and Monro 1951).[15]

A.4 Variational Families

Several factors contribute to our choice of the variational family. First, due to the number of parameters we estimate (including the value of the preference for the selected SNF $\eta_{i,j(i)}$), and the fact that in the subsequent analysis we focus on posterior means of the parameters (and not the variance), we selected factorizable families. Second, for each parameter we chose a family that includes the prior. For example, for each parameter whose prior is Gaussian, we specified a Gaussian variational approximation. For α_j, whose prior is Uniform, we specify

14. As an example, if $\theta \sim N(\mu, \sigma)$ then $\theta = \mu + \sigma\tau = f((\mu,\sigma),\tau)$ where τ is the standard normal noise distribution. The reparametrization trick is not available for some distributions, for instance categorical distributions.

15. This approach differs significantly from Bonhomme (2021) who instead uses a (gradient-free) variational expectation maximization algorithm to maximize an objective function closely related to the ELBO. By exploiting a gradient-based optimizer, our approach performs well in significantly larger parameter spaces.

a Beta variational approximation. Finally, for the posterior scale of the hierarchical prior we specify a Gamma variational approximation. Specifying the variational family in such a way is not necessary, but it facilitates analytically computing the KL divergence between the prior and the approximate posterior, which reduces the variance in the estimation of the gradient of the ELBO and may speed up convergence.

A.5 Optimization Details

We minimize the ELBO with the Adam optimizer (Kingma and Ba 2017) using clipped gradients (to ensure numerical stability) and geometrically decreasing step size. For a typical state-year bin, a single Adam update step takes less than a second on a single core CPU[16] and our model converged after approximately 30 minutes to one hour, depending on the state-year bin. The optimization of the VI objective function can further benefit from specialized hardware (such as a GPU), but we did not implement this due to hardware limitations on the secure server where our patient data reside.

A.6 Simulation Exercise

To evaluate the performance of variational inference in our setting, we conduct a brief simulation exercise. We generate two datasets according to the data-generating process specified in Section 3: one smaller, with $I = 5,000$ patients, and one larger, with $I = 50,000$ patients. For each facility $j \in \{1, \dots, 100\}$, we sample a quality parameter $\beta_j \sim N(0, 0.8)$ and a selection parameter $\alpha_j \sim \text{Uniform}(-1, 1)$. For each (i, j) pair we simulate the distance from a triangular distribution, bounded between 0 and 30, with a mode at 30. We specify the distance and square of distance coefficients, π, to be $(-1, 0.001)$. For simplicity, we ignore the observable risk-shifters. In addition to estimating the model with VI, we also use a Markov Chain Monte Carlo approach, the No-U-Turn-Sampler (Hoffman and Gelman 2014) implemented in Numpyro.

Appendix Figure E.4 presents the results obtained with VI. We plot the estimated mean posterior values of β_j and α_j against their true values for both datasets. In the smaller dataset, the estimates of β_j and α_j are understandably noisy. However, with the bigger dataset, which is comparable in magnitude to the real data, the precision improves significantly. In that case the R-squared values from regressions of the true β_j and true α_j on their estimated equivalents are 0.984 and 0.868, respectively.

To evaluate whether the lack of precision in the smaller dataset is caused by the approximating algorithm we use, we compare the VI estimates against MCMC-based equivalents. Appendix Figure E.5 presents the results. In both the smaller and the larger datasets, the estimates are virtually identical between the two algorithms for both β_j and α_j. This should reassure readers that in our setting, despite being an approximate method, VI performs extremely well and matches a state-of-the-art MCMC approach in precision. We do not report runtimes, but in all of our experiments we observed at least a 10-fold improvement using VI compared with NUTS, and often much larger.

16. The biggest state-year bin, California 2013-2016, contains 500k+ admissions. A single Adam update there takes approximately 2.5 seconds and objective function converges in less than 2 hours.

B Further Model Details

B.1 Deriving Likelihoods

A necessary element in computing the VI objective, the ELBO, is the log-joint likelihood of the data and parameters. As mentioned in the main text, we explicitly condition on and estimate the value of the preference shock for each person for their selected facility, $\eta_{ij(i)}$. Adding in the assumption that the shocks for different facilities are independent, we can then express the probability that i selects SNF j as the product of the probabilities that the utility of SNF j exceeds that of $\tilde{j}$, for each other facility $\tilde{j}$ in i's choice set. The remaining randomness then comes the utility shocks for those non-chosen facilities $\tilde{j}$, which we do not estimate. Consequentely, the log-probability of the observed choice can be derived as:

$$\begin{aligned}\log P(D_{ij} = 1 \mid \theta_i) &= \log P(u_{ij(i)} > u_{ij'},\ \ \forall j' \neq j(i) \mid \theta_i) \\ &= \log P(\eta_{ij'} < \delta_{ij(i)} - \delta_{ij'} + \eta_{ij(i)},\ \ \forall j' \neq j(i) \mid \theta_i) \\ &= \log \Pi_{j' \neq j(i)} P\left(\eta_{ij'} < \delta_{ij(i)} - \delta_{ij'} + \eta_{ij(i)} \mid \theta_i\right) \\ &= \Sigma_{j' \neq j(i)} \log \Phi(\delta_{ij(i)} - \delta_{ij'} + \eta_{ij(i)})\end{aligned} \tag{18}$$

where Φ is the standard normal cumulative distribution function.

Similarly, by plugging in for our definition of the health shock ε_{ij}, we can express the log-probability of the observed survival indicator as:

$$\begin{aligned}\log P(Y_{i,j(i)} = 1 \mid \theta_i) &= \log P\left(\beta_{j(i)} + \gamma X_i^T + \alpha_{j(i)}\eta_{i,j(i)} + \sqrt{1 - \alpha_{j(i)}^2}\tilde{\varepsilon}_{i,j(i)} > 0 \mid \theta_i\right) \\ &= \log P\left(\tilde{\varepsilon}_{i,j(i)} > \frac{-(\beta_{j(i)} + \gamma X_i^T + \alpha_{j(i)}\eta_{i,j(i)})}{\sqrt{1 - \alpha_{j(i)}^2}} \mid \theta_i\right) \\ &= \log\left(1 - \Phi\left(\frac{-(\beta_{j(i)} + \gamma X_i^T + \alpha_{j(i)}\eta_{i,j(i)})}{\sqrt{1 - \alpha_{j(i)}^2}}\right)\right)\end{aligned} \tag{19}$$

B.2 Prior Distributions

Our variational inference estimation approach does not restrict us to conjugate priors – it is only required that the joint log-likelihood is differentiable. This restriction is mild. While it excludes models with explicit discrete latent variables, these models can often be parametrized to remove these latent variables, a process known as marginalization.

Starting with the quality parameters, β_j, we specify a hierarchical prior $\beta_j \sim N(\mu_\beta, \sigma_\beta)$. The mean and scale of this prior are given the following prior: $\mu_\beta \sim N(0, 5)$, $\sigma_\beta \sim \text{Gamma}(3, \frac{1}{3})$. This specification implicitly regularizes (or shrinks) the quality estimates towards the state-year bin mean, increasing reliability especially for small SNFs that have relatively few patients. Note that the empirical Bayes methods frequently used in quality estimation to reduce noise (e.g. Chetty, Friedman, and Rockoff 2014; Guarino et al. 2015; Chandra, Finkelstein, Sacarny, and Syverson 2016) can be viewed as an approximation to the hierarchical model.

For the remaining parameters we use uninformative or weakly informative priors. Specifically, we use a uniform prior for the selection parameter, α_j. For the coefficients on the health (γ) and preference (π) shifters, we employ dispersed normal priors with mean zero and scale 8. For the observed SNF popularity parameters ξ_j, we use a standard normal prior. It is well-known that utility parameters are identified only up to a constant. We do not normalize any of the ξ_j, instead we rely on the prior to fix the mean of these parameters.

C Sample Construction

We begin with all MDS admissions assessments for Medicare-enrolled residents during the period 2000-2017. We further restrict these admission assessments to only those assessments that occur within 30 days of entry. In the event that an admission spans multiple assessments, we collect the non-missing covariates from the last assessment conducted, and we exclude any assessments with missing information on diagnoses or the activities of daily living.

We combine the cleaned sample of admissions assessments with the Long-Term Care Focus files; a small number of assessments are excluded at this stage if their provider numbers do not match any active SNFs from the LTCFocus panel. We then impose several sample restrictions. First, we require no prior nursing home admission assessments in the prior 365 days. Second, for this reason we exclude any assessments for admissions that begin prior to 2001. Third, we exclude assessments for whom we have no home zip code or for whom the home zip code does not align with the state code[17] provided in the beneficiary summary files. Fourth, we exclude any patients residing in Alaska, as the LTCFocus files do not contain any Alaskan SNFs. Finally, after we pool admissions across 4-year bins, we exclude all assessments at SNFs with fewer than 50 patients.

To construct the patient choice sets, we calculate the distance to each nursing home in the state, for all beneficiaries in the new admissions sample. To do so we assign each beneficiary a home zip code, coming from the Medicare beneficiary summary files.[18] Distance is calculated between the beneficiary's home zip code centroid[19] and the coordinates of the nursing home, which we geocode from the address available in the LTCFocus files.

A natural concern with using zip code centroids is that our distances may not accurately represent the true distance for each beneficiary to nursing home. This may reflect both that a beneficiary's residence may not be close to the centroid, as well the distinction between geodetic and driving distance. However, the extent to which we are misattributing distance will only bias the disutility of distance parameters in our selection equation towards zero. The two-stage least squares equivalent is that measurement error will lead our instruments to be weaker than they otherwise would be.

17. We collect the zip code ranges available in each state from the IRS: https://www.irs.gov/pub/irs-utl/zip_code_and_state_abbreviations.pdf

18. For long-term stays at nursing homes, it is common for a resident to update her address on file with Social Security to that of the nursing home. To account for this, for admissions in year t, we assign the zip code in year $t-1$ if the zip code matches that of the nursing home; otherwise, we use the zip code from year t.

19. Available at https://www.nber.org/research/data/zip-code-distance-database.

D Cost-Benefit Analysis of Raising Medicaid Rates

To compute the cost-benefit break-even point, we draw upon several sources. To calculate the cost to the taxpayer, we use the California Medicaid costs reports to compute the implied change in total Medicaid costs resulting from a 10% increase in the Medicaid rate. We rely on the observed number of Medicaid-paid skilled nursing days for each facility and the computed Medicaid rate, R_{jt}^{mcaid}, as described in Section 6.

To calculate the benefit of this Medicaid rate increase, we use our causal estimates of the effect of the Medicaid rate on quality, $\hat{\beta}_{jt}$, as reported in Table 2c and the formula derived in equation (6) to compute the difference in the implied 90-day survival probability between the counterfactual scenario and the status quo for all patients in California for the 2009-2012 year bin. We limit to this year bin to allow for a sufficient follow-up to estimate a model of remaining life-years, as discussed below. We only model the 90-day survival probability, and so need to project forward the number of saved life-years assuming a patient survives past this endpoint. To calculate the expected number of additional life-years, we estimate a Cox proportional hazards model for the subset of patients who survived beyond the first 90 days, and use this model to predict the number of additional life-years for each patient. We use the same set of risk-adjusters as in equation (6).

The implied break-even point is the ratio of the total cost increase for the taxpayer to the expected number of life-years saved. Estimates of the value of a statistical life-year vary, but are typically much higher then our estimate of the break-even point, even adjusting for the ill health of the SNF patient population. For instance, Department of Tranportation (2013) places the value of statistical life at $9.1m for 2013, or approximately $115,000 per year given the life expectancy in 2013. To account for the quality of life among nursing home patients, Ganz, Simmons, and Schnelle (2005) uses a multiplier of 0.79. With this multiplier, the resulting quality-of-life-adjusted value of statistical life-year would be almost $91,000, or more than four times our estimated break-even point.

E Additional Tables and Figures

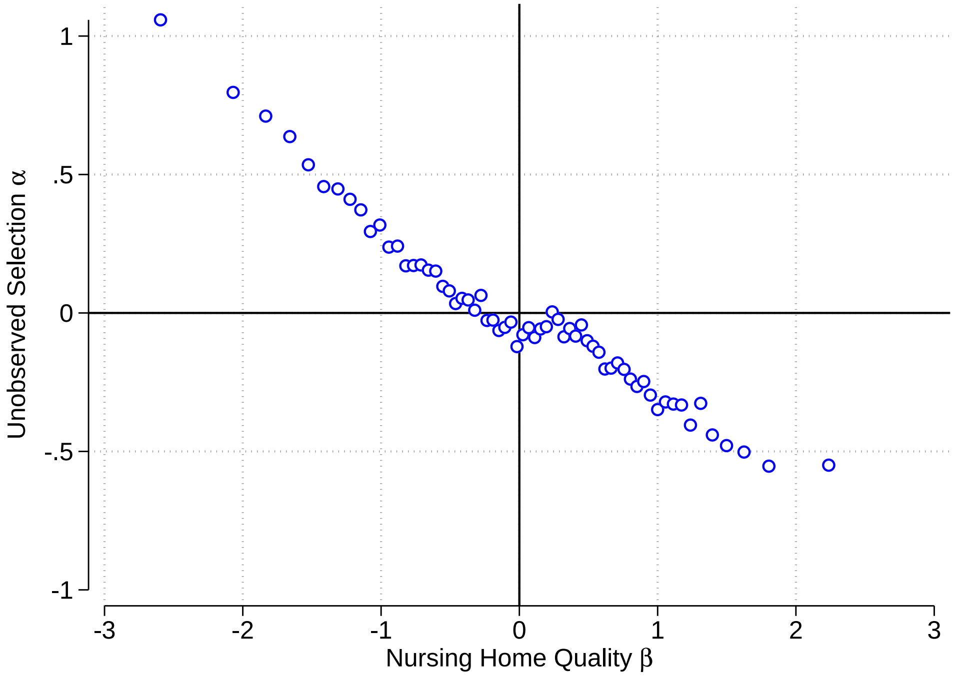
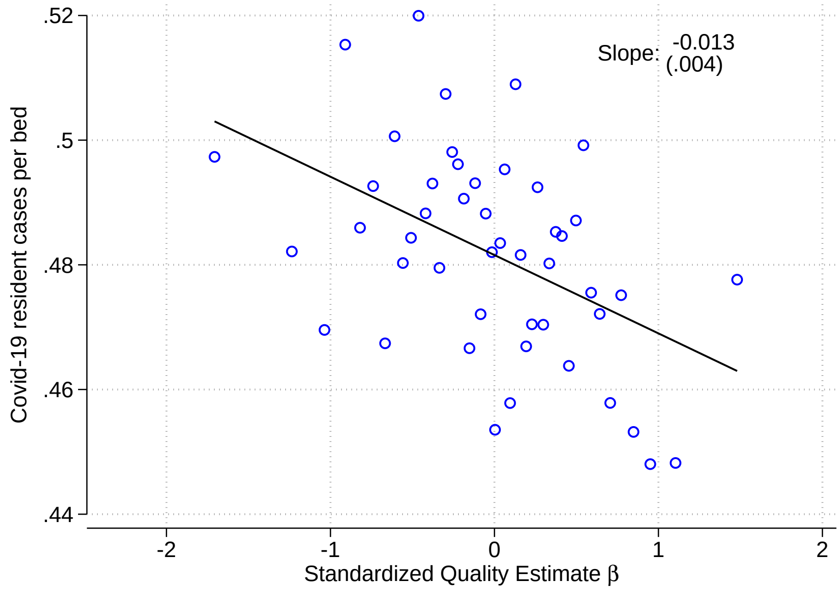
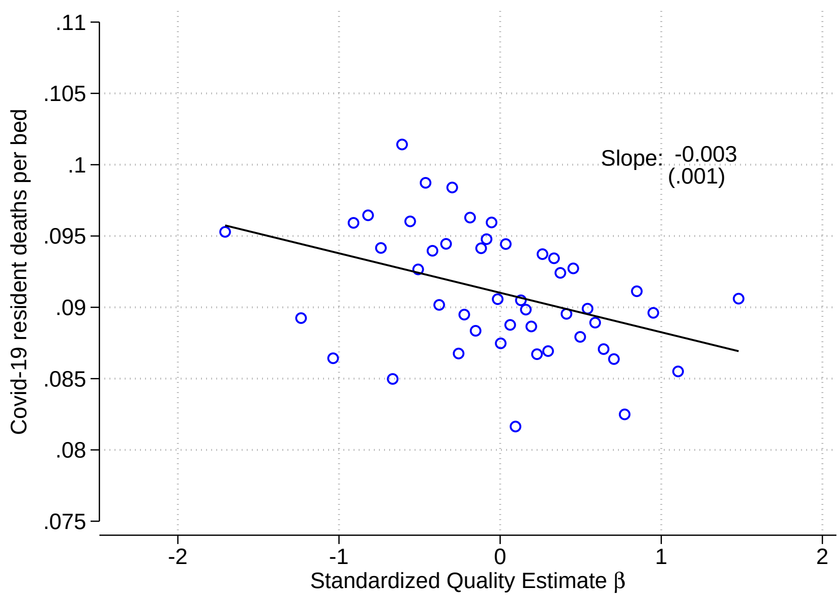


Figure E.1: Binscatter of Unobserved Selection $\hat{\alpha}_{jt}$ by Nursing Home Quality $\hat{\beta}_{jt}$

Notes: Figure presents a binscatter plot of the market-standardized unobserved selection parameter $\hat{\alpha}_{jt}$ by the market-standardized quality parameter $\hat{\beta}_{jt}$. The negative relationship indicates the presence of adverse selection.



(a) Cumulative Covid-19 resident cases per bed



(b) Cumulative Covid-19 resident deaths per bed

Figure E.2: Binscatters of Covid-19 Outcomes by Nursing Home Quality Estimates β with County Fixed Effects

Notes: Top panel presents a binned scatterplot of cumulative confirmed Covid-19 cases per bed through December 19, 2021 by nursing home quality β estimated in the last available year-bin, 2013-2016. Bottom panel presents Covid-19 deaths per bed by nursing home quality. Both are adjusted for county fixed effects.

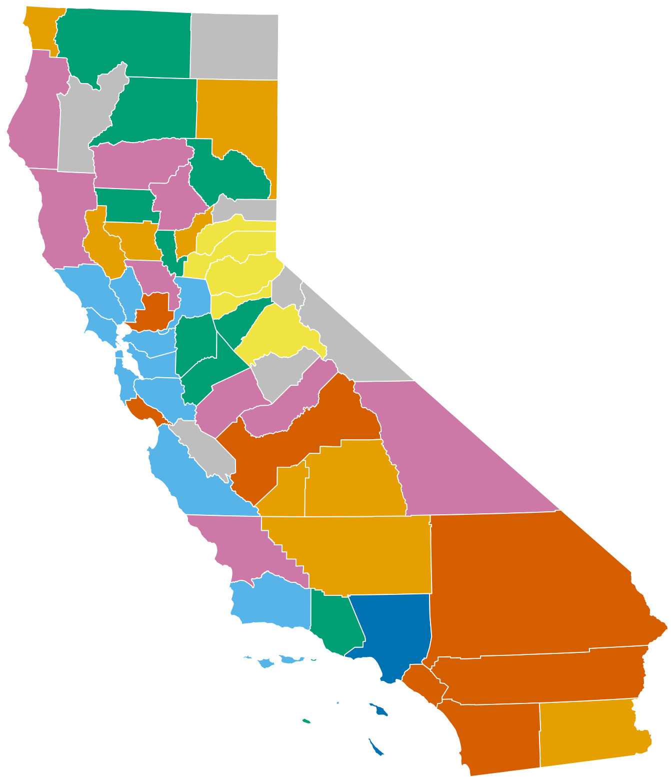
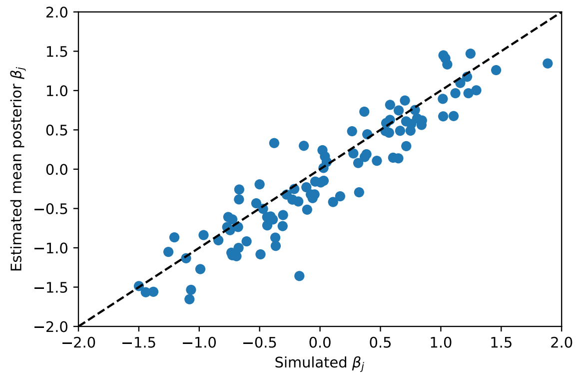
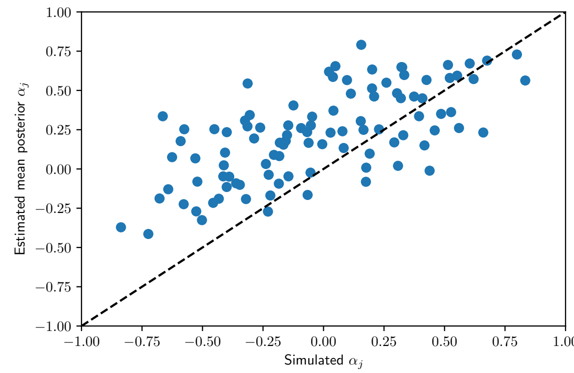


Figure E.3: Map of California skilled nursing facility reimbursement peer group regions.

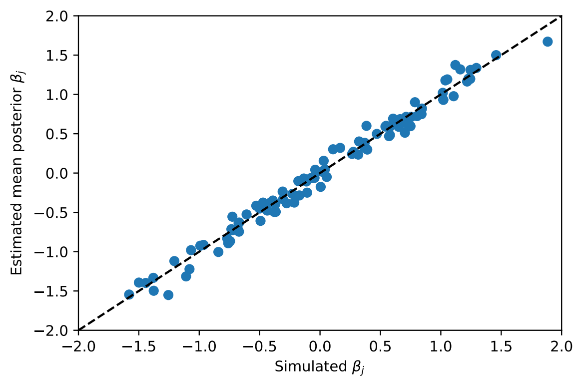
Notes: Map of California's SNF peer group regions used to calculate Medicaid per diem reimbursement rate for free-standing SNFs. Gray areas indicate counties without a free-standing SNF during the sample period.



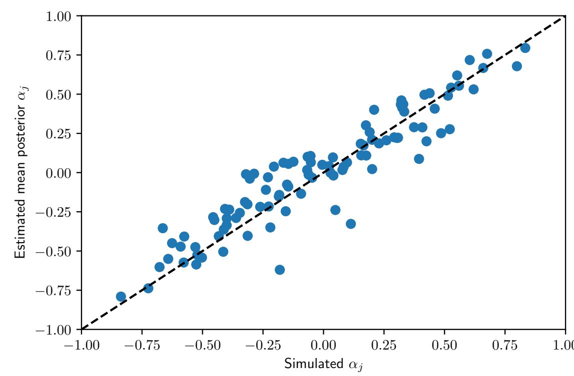
(a) Simulated and Estimated Values of β_j, $I = 5,000$



(b) Simulated and Estimated Values of α_j, $I = 5,000$



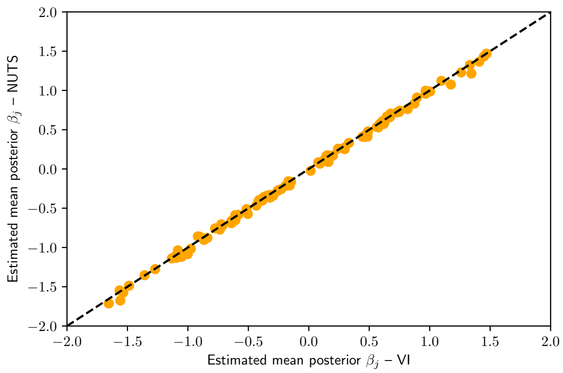
(c) Simulated and Estimated Values of β_j, $I = 50,000$



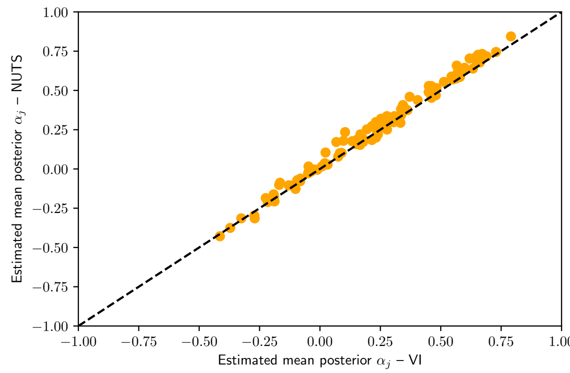
(d) Simulated and Estimated Values of α_j, $I = 50,000$

Figure E.4: Simulation Study: VI Estimates vs True Values

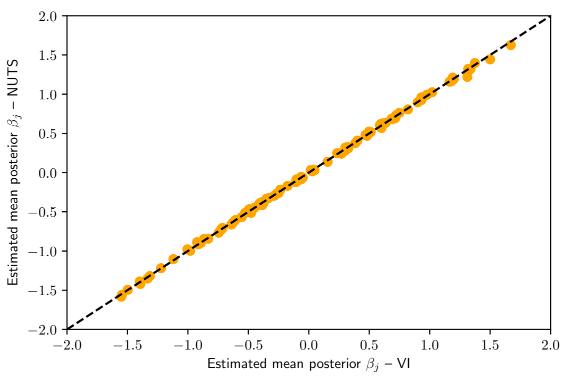
Notes: Figure presents scatter plots of the true parameters (on the horizontal axes) and the means of the respective posterior distributions estimated using variational inference. Panels (a) and (b) present the estimates using a small dataset, with approximately 50 observations per facility. Panels (c) and (d) present the same estimates when the number of observations is increased to 50,000 or about 500 per facility.



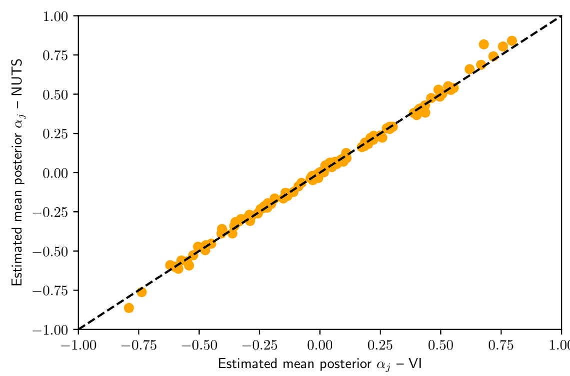
(a) Simulated and Estimated Values of β_j, $I = 5,000$



(b) Simulated and Estimated Values of α_j, $I = 5,000$



(c) Simulated and Estimated Values of β_j, $I = 50,000$



(d) Simulated and Estimated Values of α_j, $I = 50,000$

Figure E.5: Simulation Study: VI Estimates vs NUTS Estimates

Notes: Figure presents scatter plots of the means of the posterior distributions estimated with variational inference (on the horizontal axes) against their equivalents obtained with an MCMC approach, NUTS. Panels (a) and (b) present the estimates using a small dataset, with approximately 50 observations per facility. Panels (c) and (d) present the same estimates when the number of observations is increased to $50,000$ or about 500 per facility.

	Mean	P10	P50	P90
Female	0.64	0	1	1
Age	79.34	66	81	91
White	0.86	0	1	1
Black	0.10	0	0	0
Number of ADLs	5.19	1	6	8
Number of Diagnoses	1.66	0	1	3
Disability	0.19	0	0	1
ESRD	0.01	0	0	0
BMI<18.5	0.08	0	0	0
BMI 18.5-24.9	0.39	0	0	1
BMI 25-29.9	0.26	0	0	1
BMI >30	0.26	0	0	1
Mortality, 30-day	0.05	0	0	0
Mortality, 90-day	0.14	0	0	1
Mortality, 180-day	0.21	0	0	1
Mortality, 365-day	0.30	0	0	1
Distance to NH (mi)	7.83	1	4	17
Observations	20,514,758			

Table E.1: Individual summary statistics

Notes: Table reports summary statistics of the estimating sample of new nursing home admissions. Demographics and comorbidities are derived from the admission assessment for each stay. Disability, ESRD status, and mortality are all derived from the Beneficiary Summary Files.

	2001	2005	2009	2013
2001	1	0.803	0.746	0.713
2005	0.803	1	0.816	0.766
2009	0.746	0.816	1	0.821
2013	0.713	0.766	0.821	1

Table E.2: Autocorrelation of $\hat{\beta}_{jt}$

Notes: Table reports correlations of $\hat{\beta}_{jt}$ across 4-year bins. The estimates suggest that the quality is highly stable across time.

	(1) First Stage	(2) Quality β	(3) log(RN^{res})	(4) log(LPN^{res})	(5) log(CNA^{res})
log(Direct Care)	0.420				
	(0.160)				
log(Indirect Care)	-0.0424				
	(0.183)				
log(Non-Labor)	-0.0678				
	(0.171)				
log(Admin)	0.103				
	(0.0919)				
log(Medicaid Rate)		4.385	3.750	-1.203	-0.152
		(1.310)	(1.420)	(0.491)	(0.149)
Observations	1722	1722	1722	1722	1722
F-Statistic	12.64				

Table E.3: Regressions of Covid-19 Outcomes on Nursing Home Quality Estimates Excluding Los Angeles County

Notes: log(RN^{res}), log(LPN^{res}), log(CNA^{res}) respectively abbreviate the log number of registered nurses, licensed practical nurses, and certified nursing aide hours per resident day. β is the quality estimate for each nursing home. All specifications include controls for for-profit ownership, the presence of an Alzheimer's unit, the log number of beds, and local demographics. Standard errors clustered at the county-level.

Variable	Mean	% Positive	% Negative	% Positive Significant	% Negative Significant
(a) Choice Model					
Distance	-1.66	0.00	1.00	0.00	1.00
Distance Sq	0.65	1.00	0.00	1.00	0.00
(b) Survival Model					
Female	0.11	1.00	0.00	1.00	0.00
Age at Entry	-0.08	0.04	0.96	0.01	0.76
Age at Entry Sq	-0.05	0.17	0.83	0.10	0.64
ADL: Bath	-0.01	0.38	0.62	0.15	0.28
ADL: Bed	-0.07	0.00	1.00	0.00	0.92
ADL: Dressing	-0.02	0.18	0.82	0.02	0.18
ADL: Eating	-0.15	0.00	1.00	0.00	1.00
ADL: Hygiene	-0.09	0.01	0.99	0.00	0.95
ADL: Walking	0.04	0.98	0.02	0.87	0.00
ADL: Locomotion	-0.05	0.04	0.96	0.01	0.71
ADL: Toileting	-0.03	0.03	0.97	0.00	0.28
ADL: Transferring	0.00	0.45	0.55	0.04	0.03
Disability	0.01	0.87	0.13	0.23	0.01
ESRD	-3.12	0.01	0.99	0.00	0.80
Fall: 1 month	0.01	0.84	0.16	0.42	0.03
Fall: Last 6 months	0.00	0.32	0.68	0.08	0.27
Fracture: last 6 months	0.08	1.00	0.00	1.00	0.00
Race: Black	-0.81	0.80	0.20	0.74	0.18
Race: Other	-1.10	0.67	0.33	0.51	0.28
Weight: Underweight	-0.07	0.00	1.00	0.00	1.00
Weight: Overweight	0.06	1.00	0.00	0.99	0.00
Weight: Obese	0.08	1.00	0.00	0.98	0.00

Table E.4: Parameter Estimates of the Choice and Survival Models

Notes: Table reports the parameter estimates of the (a) choice and (b) survival models. Additionally, we report the percentage of markets in which the estimates are positive or negative, as well as the share in which they are statistically significant at the 5% level. For brevity we omit the coefficients from the 19 diagnosis codes that enter the survival equation.