Producing Health: Measuring Value Added of Nursing Homes*

Liran Einav† Amy Finkelstein‡ Neale Mahoney§

July 16, 2022

Abstract

We develop a stylized model that allows us to estimate a value-added measure for nursing homes ("SNFs") which accounts for patient selection both into and out of a SNF. We use the model, together with detailed data on the physical and mental health of about 6 million Medicare SNF patients between 2011 and 2016, to estimate the value added for about 14,000 distinct SNFs. We document substantial heterogeneity in value added. Nationwide, compared to a 10th percentile SNF, a 90th percentile SNF is able to discharge a patient at the same health level about a week sooner, which is about one third of the median length of stay. Heterogeneity in value added within a market is almost as large as it is nationwide. Our results point to the potential for substantial gains through policies that encourage reallocation of patients to higher-quality SNFs within their market.

*We thank Alden Cheng, Eric Zwick, and seminar participants at Chicago Booth, Duke/UNC, Tinos IO conference, UBC, UPenn, USC, Rice, and Washington University for comments and suggestions. We are grateful to Sean Gao, Sam Wang, Alicia Weng, Xuyang Xia, Connie Xu, Yuci Zhou, and especially James Okun for superb research assistance, and to the Laura and John Arnold Foundation, and the National Institute on Aging (Finkelstein R01-AG032449) for financial support.

†Stanford University and NBER, leinav@stanford.edu

‡MIT and NBER, afink@mit.edu

§Stanford University and NBER, nmahoney@stanford.edu. Mahoney is currently at the White House, on leave from Stanford; all his work on this paper was done while he was still at Stanford.

1 Introduction

The production of health lies at the core of health economics (e.g., Grossman 1972), yet with the exception of mortality, the study of health has been hampered by a lack of consistently and comprehensively measured data on health outcomes. As a result, instead of focusing on health outputs, most research in health economics has focused on inputs into the production of health – such as health-care utilization and health behavior – where extensive data is available.

Nursing homes represent an important exception to the general paucity of data on health. For decades, virtually all nursing homes in the United States have been required to administer detailed health assessments to all of their patients at the time of admission, and at regular intervals during their stay. These assessments capture multiple different measures of each patient's physical health, mental health, daily functioning, and cognitive capacity. They present a rare opportunity to study the health production process across health-care providers, and to estimate potential gains from reallocation of patients.

In this paper, we use this rich data on health outcomes at nursing homes to construct a holistic measure of patient health and to estimate nursing home value added in improving this health measure for about 14,000 nursing homes. Our estimation accounts for potential bias from endogenous selection of patients into nursing homes and endogenous timing of when patients are discharged. We use the estimates to examine heterogeneity in nursing home quality across and within markets.

The nursing home sector is an important setting for assessing the production of health, because of both its size and widespread concerns about the quality of nursing home care. In 2016, there were over 15,000 certified nursing homes in the U.S., providing care to over 1.3 million patients at an annual cost of $160 billion, or roughly 5% of national health expenditures (Harrington et al. 2018; CMS 2019b). For decades, policymakers and the public have expressed concerns about the quality of nursing home care (e.g., Rau 2018; Rau and Lucas 2018; Goldstein and Gebeloff 2019; Jacobs and Richtel 2019) and about the ability of patients (or their caregivers) to assess this quality (Wunderlich et al. 1996; GAO 2016, 2019). To help patients make choices, the Centers for Medicare and Medicaid Services (CMS) construct "star ratings" of each nursing home that are widely used (CMS 2008a; GAO 2016) but also widely criticized (Thomas 2014; Rau 2018; Ryskina et al. 2018; Silver-Greenberg and Gebeloff 2021).

We focus on Medicare patients, for whom nursing homes are intended to provide short-

term care that is designed to aid in the recovery from a hospitalization or medical condition.[1] In 2016, about one fifth of hospitalized Medicare patients were discharged to nursing homes, at an annual cost to Medicare of nearly $30 billion, or about 8% of overall Medicare spending (MedPAC 2018). We combine the consistently and comprehensively measured health assessment data into a one-dimensional "health index" that measures how fit the patient is to return to the community; specifically, it measures the probability the individual will be discharged to the community within a week of the assessment. A nursing home's value added reflects the average improvement in a patient's health index between the assessment at admission and the one taken 30 days later. It is straightforward to use our approach to estimate value added for other health measures, as we briefly illustrate below.

Estimating a nursing home's value added requires overcoming two econometric challenges. First, there is potentially non-random sorting of patients to nursing homes, even conditional on the rich set of observable patient characteristics at admission. This problem of "selection in" is a standard one in the education literature on value added, and we follow the standard approach of constructing a control function to address it (Dubin and McFadden 1984; Abdulkadiroğlu et al. 2020). We use distance between the patient's residence and the nursing homes in their market as the excluded instrument in their nursing home demand model.

A second challenge in our setting is that patients may leave the nursing home before their 30-day assessment, either because they recover and return to the community (which we refer to as "discharged downstream") or because their health worsens and they are sent back to the hospital, to a hospice, or die ("discharged upstream"). We model this "selection out" with an explicit model of the nursing home's discharge decision that is similar in spirit to Heckman (1979).

We estimate the model by maximum likelihood using data from 2011 to 2016 on health assessments for about 6 million Medicare patients at about 14,000 different nursing homes.[2] Our results indicate substantial heterogeneity in value added. While the (patient-weighted) average nursing home increases the health index – i.e., the weekly probability of community discharge – by 4.5 percentage points between admission and 30 days, there is a 6.3 percentage point difference in value added between the 90th and 10th percentile nursing home. To put that in perspective, consider that in the cross section of admitted patients, a 1 percentage point higher health index is associated with approximately 1 fewer day in the nursing home.

[1] Specifically, we focus on the approximately 70% of Medicare patients with coverage from Traditional Medicare (henceforth "Medicare"), as opposed to coverage from a private Medicare Advantage plan.

[2] The value added measures for each nursing home are available to researchers upon request; we will post them publicly once the paper is published.

This means that the 6.3 percentage point difference in value added between the 90th and 10th percentile nursing home is roughly equivalent to a nursing home discharging a patient at a given health level an entire week sooner, a one-third reduction compared to the median length of stay in our sample. Given that a nursing home costs Medicare about $470 per day (MedPAC 2018), this means that a 90th percentile value-added nursing home can save Medicare about $3,000 per patient relative to the 10th percentile nursing home. Interestingly, our estimates of nursing homes' value added are only weakly correlated with their CMS star rating (correlation of 0.12).

We find considerable variation in (patient-weighted) nursing home value added both across and within geographic markets. Value added is on average 3.2 percentage points higher in the 90th percentile market than in the 10th percentile market; nursing homes in markets in the South tend to have substantially lower value added. Within markets, the average difference between the 90th and 10th percentile nursing home is 5.6 percentage points, almost as large as the corresponding nationwide difference (of 6.3). Even in markets in the top decile of average value added, close to 25% of patients are in nursing homes that perform below the national median. This points to the potential for substantial gains from within-market reallocation of patients.

Our focus on geographic variation in health production complements an influential body of research associated with the Dartmouth Atlas, which has documented substantial regional variation in Medicare spending and health-care utilization without corresponding variation in mortality (e.g., Congressional Budget Office 2008; Gawande 2009; Skinner 2011; Institute of Medicine 2013). Variation in spending for post-acute care (mainly nursing homes) explains a substantial share of the spending variation.[3] Relative to this existing literature, we are able to use rich, consistently measured data on (non-mortality) health outcomes to estimate and examine heterogeneity in health-care quality. Our finding of substantial heterogeneity in quality across markets suggests that there may be scope to improve the quality of care in low-value-added regions. Moreover, our findings of substantial variation in value added *within* markets is particularly encouraging, since there are likely more policy levers for encouraging reallocating patients within markets than across them.

In addition to the Dartmouth Atlas literature, our paper relates to several other distinct literatures. It complements the literature on value added for hospitals (e.g., Geweke et al. 2003; Chandra et al. 2016; Doyle et al. 2019; Hull 2020) and nursing homes (e.g., Olenski and Sacher 2022) which, like most of the prior health literature, has focused on mortality as

[3]Specifically, despite accounting for only 16% of Medicare spending, post-acute care accounts for 73% of the geographic variation in Medicare spending (Institute of Medicine 2013).

the single outcome. Our paper also contributes to a growing literature in health economics on nursing home behavior (e.g., Grabowski et al. 2008; Hackmann 2019; Gandhi 2020; Gandhi et al. 2021; Gupta et al. 2021; Hackmann et al. 2021) and a literature in health services research that examines cross-sectional variation in nursing home process measures and outcomes (e.g., Castle and Ferguson 2010; Morris et al. 2018). Finally, our methodological approach connects to a well-developed literature on teacher and school value added (e.g., Todd and Wolpin 2003; Chetty et al. 2014a, 2014b; Koedel et al. 2015; Abdulkadiroğlu et al. 2020). This education literature has mainly focused on bias from "selection in," while in our setting both "selection in" and "selection out" are prominent concerns.

The rest of the paper proceeds as follows. Section 2 provides background on our setting and data. In Section 3 we present our econometric model of value added. Section 4 discusses implementation and estimation, and Section 5 presents the results. In Section 6 we briefly compare our estimates to other measures of nursing home quality (including the widely-used CMS "star ratings"), and explore how our estimates change with other health measures or econometric specifications. Section 7 concludes.

2 Setting and Data

2.1 Institutional Setting

Nursing homes provide both short-term care to patients recovering from a hospitalization or illness, and long-term care to patients in need of ongoing assistance with their daily living. Most short-term patients are covered by Medicare, which pays for short-term nursing and rehabilitation services for Medicare patients recovering from a surgical procedure (e.g., hip replacement) or a health event (e.g., stroke). Crucially, Medicare coverage is predicated on the expectation that the patient is on a path to recovery and return to the community (CMS 2019a), which is why we focus on discharge to the community as a key marker of "success."

By contrast, patients who need ongoing assistance typically receive nursing home coverage from Medicaid. Some patients are initially admitted with Medicaid coverage, but many others transition to Medicaid during their stay. Therefore, while over 70% of nursing home patients are covered by Medicare at the time of admission,[4] only 14% of nursing home patients have Medicare as their primary payer at a given point in time. Because Medicaid patients have much longer stays than Medicare patients, at a given point in time 60% of

[4]Authors' calculation using data described in Section 2.2.

nursing home patients have Medicaid as their primary payer. The remainder are covered by private insurance or pay out of pocket (Harrington et al. 2018).

Almost all nursing homes (96% of beds) are certified to care for both Medicare and Medicaid patients (Harrington et al. 2018). Nursing homes that are certified to treat Medicare patients are referred to by Medicare as Skilled Nursing Facilities (SNFs, pronounced "sniffs"), a terminology we will adopt for the remainder of the paper.

During our 2011-2016 study period, Medicare reimbursed SNFs at a prospective daily rate, with the rate depending on both the SNF's geographic location and a measure of the patient's health at admission derived from health assessments (MedPAC 2021). Starting at the 21st day in the SNF, patients must pay daily co-pays (either directly or via supplemental coverage), and after 100 days Medicare coverage ends.

To be eligible for Medicare or Medicaid reimbursement, a SNF must comply with federal requirements regarding residents' civil rights and ethical treatment of patients, as well as the health care services they provide (*42 C.F.R. §483* 2016). States conduct an initial certification of a SNF, after which the SNF must be re-certified annually. During (re)-certification, a state inspector collects (facility-reported) data on staffing levels and nursing home characteristics (such as number of beds, occupancy rates, and ownership type) and records any quality deficiencies that are identified during inspection (CMS 2018; Harrington et al. 2018; LTCFocus, no date). As part of the re-certification process, certified SNFs are also required to assess their Medicare patients' health at regularly defined intervals during their stay, producing the health assessment data we use below (CMS 2017).

The information collected during re-certification is aggregated by CMS into a "five-star" rating system. These ratings, which have been constructed since 2008, are then publicly posted on the Nursing Home Compare website. Patients and their families are encouraged to use the five-star ratings in choosing a facility, and many do (CMS 2008a; GAO 2016); the ratings are also used by insurers for determining their provider networks, and by federal regulators deciding on loans to facilities (*12 U.S.C. §1715w* 2014; CMS, no date). Extremely poorly-performing SNFs – less than 0.1% of facilities – are subject to twice as many inspections, as well as financial penalties and potential termination (CMS 2022).

2.2 Data Sources

Our primary data source is the Long-Term Care "Minimum Data Set" for Resident Assessment and Care Screening (hereafter, MDS). These data contain a series of federally mandated, standardized patient assessments that track the health status of all patients in Medicare

and Medicaid certified nursing homes. The assessments were developed by a consortium of professionals in the late 1980s in response to a congressional mandate to create a national resident assessment system for nursing homes (Morris et al. 1990). These data have since been widely used by researchers studying the economics of nursing homes (e.g., Grabowski et al. 2008; Cornell et al. 2019; Hackmann 2019; Gandhi 2020; Gupta et al. 2021; Hackmann et al. 2021). The assessment form has been revised several times since it was first created. We use version 3.0, which was in effect during our entire 2011-2016 study period.

The MDS covers all nursing home patients, not just the Medicare patients who are our focus. Assessments are required for all patients at admission and at discharge. In addition, during our study period, assessments were also required for Medicare-covered patients at days 14, 30, 60, and 90.[5] The rich, longitudinal information that the assessments provide on patient physical health, mental health, and cognition will be the backbone of our measurement of value added. The MDS also provides basic demographics (age, race, gender, and marital status), length of stay, and discharge destination.

We supplement the MDS with additional data on Medicare patients and on nursing homes. We merge the MDS with Medicare data to identify the patient's 5-digit zip code of residence, whether the patient is dually eligible for Medicaid (a proxy for low income), whether the patient has each of 27 different chronic conditions at the start of the year of their SNF admission, whether the patient was discharged to the SNF from an acute care hospital, and the patient's diagnoses for the prior hospital admission if she was discharged directly from a hospital.[6] We use facility-level data from the OSCAR/CASPER system, created during the annual re-certification process, to measure the number of beds in the SNF, its occupancy rate, for-profit status, and whether the SNF is based within a hospital.

2.3 Sample Construction and Summary Statistics

We start with the full sample of all patients stays at SNFs during the five fiscal years that begin on October 1, 2011 and end on September 30, 2016. The start of the period corresponds to the start of a new schedule for required assessments. The data contain 24.1 million patient-

[5]Specifically, initial assessments must be conducted by day 8, day-14 assessments must be conducted during days 13-18, day-30 assessments during days 27-33, day-60 assessments during days 57-63, and day-90 assessments during days 87-93.

[6]The MDS does not include information on whether the patient is covered by Medicare. We use a crosswalk between the internal resident ID in the MDS and the Medicare beneficiary ID to identify patients covered by Medicare at admission to the SNF. We use the Master Beneficiary Summary File to provide additional demographics, the Chronic Conditions file for chronic conditions diagnosed in the year of SNF admission, and the MedPAR file to identify whether the patient had a prior hospital admission and which of the 241 Clinical Classifications Software (CCS) diagnoses applied for that admission.

stays, admitted to 16,355 distinct SNFs. We restrict attention to patients who, at the time of their SNF admission, are at least 65 years old and are covered by Medicare. This leaves 10.7 million patient-stays at 16,307 distinct SNFs.

We then make several additional sample restrictions, which bring our final analysis sample to 6.2 million patient-stays at 13,996 distinct SNFs. Appendix Table A1 provides more detail about the number of observations that are dropped due to each of these restrictions. First, we restrict attention to a patient's first stay within an episode of nursing home care; this accounts for the vast majority of the sample decline.[7] Second, we require that the patient enters the SNF following an acute-care hospital stay with a diagnosis code from that hospital stay (which about 91% of the remaining sample does). Third, we require that the patient has a 5-day health assessment (as defined in Appendix A). Finally, to ensure an adequate sample size for estimating value added, we require that the SNF has at least 50 stays that meet our sample requirements over the five-year period.

Table 1 presents summary statistics on the patients in the sample (column (1)), as well as the standard deviation of the SNF-level average across SNFs, with each SNF weighted by the number of patient-stays (column (2)). The average patient is 81 years old, almost two thirds are female, one third are married, and 86% are white. Almost one fifth of patients are dually eligible for Medicaid (a marker of low income), and the average hospital stay prior to admission lasts 7.6 days. The variation across patients in race and Medicaid status has a considerable SNF-specific component (see column (2)).

Panel B presents statistics on length of stay. Median length of stay is 22 days (not shown), but some patients stay for much longer; the average length of stay is 45 days. Almost 60% of the patients in our sample are discharged from the SNF prior to their 30-day assessment, while a non-trivial share (8%) remain in the SNF for more than 90 days.

Panel C shows discharge destinations for the approximately three fifths of our sample who are discharged prior to their 30-day assessment. Nearly two thirds are discharged "downstream" (i.e. to the community), which is the good outcome, and about one third are discharged "upstream" (most of whom are sent back to the hospital).

[7] A stay is defined as a set of contiguous days within a facility between admission and discharge or death. To account for patients with brief interruptions in their nursing home stay, CMS defines an episode of nursing home care as a period of time that spans one or more stays and ends either when the patient (i) is discharged without a return anticipated, or (ii) is discharged and does not return within 30 days, or (iii) dies. About 85% of Medicare patients admitted to a SNF have an episode consisting of a single stay. The stay is the more natural unit of analysis when measuring health changes, since time between stays can be of varying length and Medicare's requirements for patient assessment are based on days within a stay (although Medicare cost-sharing rules are based on the cumulative number of days to date across stays within an episode) (CMS 2016).

Figure 1 shows the movement of patients through our environment. Every patient in our sample receives an initial health assessment. Our value-added measure, described below, is based on the change in outcomes between the initial assessment and the assessment conducted at 30 days; specifically, it measures the change in the probability that a patient is discharged to the community within a week of their assessment. We chose the 30-day assessment as the endpoint because, relative to the 14-, 60- and 90-day assessments, it seemed to strike a balance between being long enough after admission for SNFs to have had the opportunity to have an effect and short enough so that there is still a significant share of patients at the SNF. Before the 30-day assessment, 36.8% of patients are discharged downstream, 19.8% are discharged upstream, and 3.1% are discharged elsewhere, leaving 40.3% of patients in the SNF for the 30-day assessment. Within a week of the 30-day assessment, 7.4% of patients are discharged downstream, 2.0% are discharged upstream, and 0.6% are discharged elsewhere, leaving 30.3% of all patients in the SNF one week after the 30-day assessment.

Table 2 reports summary statistics on our baseline sample of 13,996 SNFs. The average SNF has just over 110 beds; the 5th percentile SNF has 40 beds and the 95th percentile one has 210 beds. For most SNFs, occupancy rates hover around 80-90%, although there is a small fraction with occupancy rates lower than 60%. More than two thirds of SNFs are for-profit, which is similar to the rate elsewhere in the post-acute care sector (for example, almost three quarters of long-term care hospitals are for-profit) but much higher than the rate for acute-care hospitals (where less than one fifth are for-profit) (MedPAC 2020). About 4% of SNFs are located on the premises of a hospital. On average, our baseline sample includes 446 patient admissions per SNF over our five-year sample period; the 5th percentile SNF has 65 admissions and the 95th percentile has more than a thousand.[8]

To define the market or "choice set" of SNFs for each patient, we use the standard market definition for hospital choice, namely the 306 Hospital Referral Regions (HRRs).[9] Less than 12% of stays occur outside of the patient's HRR. The average HRR has about 45 SNFs and 20,000 stays, 70% of them have more than 20 SNFs, and 30% of them have more than 50 SNFs.

[8]Recall that we require a minimum of 50 patient admissions as part of our sample restrictions, and that our admission count only includes Medicare patients 65 and over, among other sample restrictions (see Appendix Table A1).

[9]HRRs are designed to approximate health-care markets for tertiary medical care, each containing at least one hospital that performs operations on the heart and brain (Wennberg et al. 1998).

2.4 Health Measures

The rich data from patient health assessments are a key feature of our setting. These assessments are conducted by a nurse or an assessment team (*42 C.F.R. §483* 2016). We draw on 109 health measures that are consistently recorded over our study period. Appendix A explains how we selected these measures and describes them in more detail; Appendix Table A2 provides the full set of measures and their complete definitions.

The 109 measures include a wide range of health outcomes. Many are physical health measures, such as vomiting or falls; or physical limitations to activities of daily living – such as walking, dressing, or toileting – that are measured on a 5-point scale ranging from fully independent (requiring no help or staff oversight at any time) to total dependence (requiring full staff help every time during a 7-day period). Other categories of health measures include those that deal with mental health, use of restraints, pain, use of treatment and equipment, and interactions with others.

Table 3 lists a selected group of the 109 measures and provides summary statistics for them at admission. It shows the mean and standard deviation across patients, and the standard deviation across SNFs (weighted by patients).[10]

There is a great deal of variation across measures in the frequency of health problems. For example, about two thirds of patients use a walker, about one third are taking antidepressants, 9% are on anti-psychotics, and 2% are experiencing vomiting. The median patient scores about 3 out of 4 in terms of activities of daily living for walking, dressing, and toileting, meaning that for each of these activities she requires extensive assistance but is not totally dependent on the staff. Naturally, there is considerable heterogeneity across patients in these measures (column (2)). Although not as heterogeneous (by design), there is also a non-trivial variation in the averages of these measures across SNFs (column (3)). This heterogeneity underscores the importance of grappling with potential unobservable selection on initial health.

3 An Econometric Model of Nursing Home Value Added

3.1 Defining Value Added

To fix ideas, consider a population of patients, each denoted by i, who are randomly assigned to a set of SNFs, each denoted by j. Patients arrive at SNFs with baseline health described

[10]The fourth column discusses how each measure is used in constructing our baseline health index; we defer a discussion of this to Section 4.

by the index h_{i1} at the start of period 1 (initial assessment). We then observe health h_{i2} at period 2 (30-day assessment) for all patients.

Given this setup, we define the value added of SNF j as α_j in the equation:

$$h_{i2} = \alpha_j + \theta h_{i1} + \varepsilon_i, \tag{1}$$

where ε_i is a mean-zero iid error term and h_{i1} controls for the effect of baseline health. Thus, SNF value added, α_j, can be interpreted as the average improvement in health at SNF j between period 1 and period 2, conditional on baseline health.[11] This conceptualization of value added is very similar to that used in the education literature. For instance, teacher value added in test scores is often based on a regression of test scores at the end of a year on teacher fixed effects and lagged test scores.

While the specification may be familiar, it imposes several restrictions that are worth noting. First, it assumes that α_j is homogeneous across patients. This allows us to estimate a single value added for each SNF; it would be straightforward to estimate the model separately for different groups of patients and thus allow for heterogeneity in SNF value-added based on observables. Second, it assumes that the serial correlation in health (captured by θ) does not vary across SNFs. This assumption is needed for the cross-SNF differences in α_j to be interpretable; if we allowed θ to vary by SNF, we would not be able to characterize the health improvement across SNFs with α_j alone.

A natural concern is that SNFs may strategically misreport health measures. The fact that health measures are reported by nursing home staff adds to this concern. A similar issue arises in the education setting, where test scores are the outcome and teachers or principals may have incentives to "teach to the test" (Hoffman et al. 2001; Lazear 2006); indeed, there is evidence that some teachers fraudulently change students' answers (Jacob and Levitt 2003).

In our setting, however, incentives to shade assessments operate in both directions. On the one hand, better patient health will improve a facility's quality ratings. On the other hand, worse patient health at admission will increase the daily reimbursement rate received for the patient, which has created concerns about SNF "upcoding" (Bowblis and Brunt 2014; Levinson 2015). Importantly for our purposes, since our value added estimates are based on within-patient changes in health assessment, they should not be affected by any systematic differences in coding intensity across SNFs.

[11] In practice, we will estimate θ to be very close to 1 (0.99). As a result, our measure of SNF value-added effectively measures the average improvement in patient health between period 1 and period 2, and we will use the terms "health improvement" and "value added" interchangeably in what follows.

3.2 Key Challenges to Estimation

Estimating equation (1) for a given health measure would be straightforward if patients were randomly assigned to SNFs and we observed health for all patients in periods 1 and 2. In practice, neither of these conditions holds, and we explicitly address this in our estimation.

First, patients are not randomly assigned to SNFs. If there is a correlation between patient health improvements and SNF quality, then estimates of α_j may be biased, although the direction of this bias is unclear. For example, if more savvy patients are more likely to choose higher-quality SNFs and are more likely to improve, this would bias upward the estimates of α_j for these high-quality SNFs, stretching the distribution of α_j. Alternatively, if SNF quality is particularly important for those who would not improve otherwise, estimates of α_j for high-quality SNFs would be biased downwards, compressing the distribution. We call this issue "selection in," and address it by using the patient's distance to different SNFs as instrumental variables that shift SNF choice.

Second, some patients leave the SNF before receiving a follow-up health assessment. Specifically, three fifths of patients are discharged prior to their period 2 (30-day) health assessment, with about two thirds of them discharged "downstream" to the community, and one third discharged "upstream," meaning they either are sent to a hospital or hospice or they die at the SNF (recall Figure 1).[12] If there is a correlation between patient health improvements and SNF discharge propensities, estimates of α_j using only those patients who remain in the SNF until period 2 may be biased. This bias will most likely generate a compression of estimates of α_j around the mean. SNFs with higher value added will have a sicker pool of patients at day 30 than they would without discharge, since they are likely to discharge to the community the patients that improve the most, understating the health improvements. SNFs with lower value added will also have a sicker pool of patients at day 30, but because the sickest are more likely to die or be transferred to a hospital before the 30-day assessment, the observed pool of patients may be healthier, overstating health improvements. We call this issue "selection out," and address it with an explicit model of the SNF's discharge decision.

[12]There are 58 (smaller) SNFs in our baseline sample that discharge all patients prior to the 30-day assessment. We cannot estimate value added for these SNFs so our discussion of value-added estimates below is focused on the 13,938 SNFs that have some patients who are still in the SNF at 30 days.

Selection Out of the Nursing Home

Selection of patients out of the SNF can occur both "downstream" when they are discharged to the community, presumably once their health is sufficiently good, and "upstream" when their health deteriorates and they have to be discharged to a hospital or (less frequently) die or are sent to hospice.

To account for this selection in an econometrically-tractable fashion, we model it as a two-step sequential process. First, SNFs actively make downstream discharge decisions, $d_i^D \in \{0,1\}$, based on a SNF-specific health discharge threshold. Then, upstream discharge decisions are realized for the remaining patients, $d_i^U \in \{0,1\}$, as a result of stochastic health deterioration. We choose to model the two discharge decisions sequentially and not simultaneously (for example via an ordered probit) in order to allow for different amounts of noise in each direction; this makes the model more flexible and allows us to fit the data significantly better.

We assume that downstream discharge is a result of a stylized discharge model, in which SNFs decide whether to discharge each patient to the community (that is, downstream) using the following rule:

$$d_i^D = 1 \iff h_{i2} \geq \lambda_j + \nu_i, \tag{2}$$

where λ_j is a SNF-specific discharge threshold and ν_i is an iid error term, drawn from $N(0, \sigma_\nu)$. That is, SNFs are more likely to discharge healthier (higher h_{i2}) patients, with a discharge threshold that varies across SNFs and is affected by unobservables, such as the availability of family members to assist in the discharge and the opportunity cost of SNF beds.[13]

We model discharge upstream in a more statistical way, assuming that upstream discharges are the result of a probit stochastic process, which is a (declining) function of h_{i2}. That is, we assume

$$\mathbb{P}\left(d_i^U = 1 | d_i^D = 0\right) = \Phi(\gamma_0 + \gamma_1 h_{i2}), \tag{3}$$

where $\Phi(\cdot)$ is the standard normal cumulative density function. We expect γ_1 to be negative, so that patients with better health are less likely to be discharged upstream. We allow γ_0 and γ_1 to be market specific in order to account for any potential differences in hospital access across markets.

[13]This model is obviously a simplification, abstracting for example from the possibility that the discharge decision also directly depends on the length of time spent in the SNF or the improvement in health since admission. However, as shown in Appendix B, this parsimonious model performs well compared to richer specifications.

We should note that although the way we describe the downstream model (as an active decision) and the upstream model (as a more passive decision) is different (and arguably more realistic), the two situations are very similar from a statistical and estimation perspective, giving rise to a "probit-like" model downstream and a probit model upstream.

Selection Into the Nursing Home

The possibility that patients do not sort into SNFs randomly, even conditional on observables, is a standard concern in the value-added literature. Following the literature (Dubin and McFadden 1984; Abdulkadiroğlu et al. 2020), we use an instrumental variable – the distance between the patient's residence and her potential SNF choices – to construct a control function that accounts for this selection. Specifically, we first specify a discrete choice model of SNF choice which depends on distance to the SNF. We then show how this choice model yields closed-form selection correction terms in the value-added equation. We provide evidence in support of the exclusion restriction in Section 4.2.

Let the utility of patient i from SNF j be

$$u_{ij} = \delta_j(x_i) - \tau m_{ij} + \eta_{ij}, \tag{4}$$

where $\delta_j(x_i)$ is the average utility from SNF j for all patients with (observable) characteristics x_i, m_{ij} is (the log of) the distance between patient i and SNF j, and η_{ij} is a an error term drawn iid from a Type I Extreme Value (logit) distribution. Note that we impose no restriction on δ_j; it may be correlated with SNF value-added α_j and with any other characteristics of the SNF that affect patient demand, such as available capacity.[14]

Following Dubin and McFadden (1984), we account for selection on unobservables with a linear control function in the (de-meaned) logit errors. Conditional on the choice of SNF c_{ij}, period 1 health h_{i1}, and the demand shocks η_{ij}, expected period 2 health can be written as:

$$\mathbb{E}[h_{i2}|c_{ij}, h_{i1}, \eta_{i1}, ..., \eta_{iJ_i}] = \alpha_j + \theta h_{i1} + \sum\nolimits_{k \in J_i} \phi_k(\eta_{ik} - \mu_\eta) + \varphi(\eta_{ij} - \mu_\eta), \tag{5}$$

where J_i is patient i's choice set,[15] and μ_η is the mean of the logit errors (Euler's constant).

[14]Recent work has drawn attention to the challenges in estimating demand under latent choice constraints (Agarwal and Somaini 2022) and to the importance of nursing home capacity constraints in affecting patient allocation to nursing homes (Gandhi 2020). In our context, capacity constraints will be reflected in the SNF fixed effects, the δ_j's, in the choice model (equation (4)). Since we do not attempt to interpret these coefficients as reflecting demand, this concern is not directly relevant for our analysis.

[15]The choice set J_i always includes the outside option ($j = 0$), which in our context is a SNF that is outside patient i's market (HRR). We assume that all out-of-market SNFs are equally far from all the

Integrating out the η's yields the estimating equation:

$$\mathbb{E}[h_{i2}|c_{ij}, h_{i1}, m_{i1}, ..., m_{iJ_i}] = \alpha_j + \theta h_{i1} + \sum\nolimits_{k \in J_i} \phi_k \beta_{ik} + \varphi \beta_{ij}, \tag{6}$$

where the β's are functions of the logit choice probabilities $\hat{p}_{ik}$:

$$\beta_{ik}(j) = \begin{cases} -\log \hat{p}_{ik} & k = j \\ \frac{\hat{p}_{ik}}{1-\hat{p}_{ik}} \log(\hat{p}_{ik}) & otherwise \end{cases}, \tag{7}$$

and the logit choice probabilities are predicted values from the choice model described in equation (4):

$$\hat{p}_{ij} = \frac{\exp\left(\delta_j(x_i) - \tau m_{ij}\right)}{\sum_{k \in J_i} \exp\left(\delta_k(x_i) - \tau m_{ik}\right)}. \tag{8}$$

The estimating equation (6) integrates over an error term ϵ that is distributed $N(0, \sigma_\epsilon)$. Appendix C provides the full derivation.

The additional parameters in equation (6) have an economic interpretation. The $\phi_k \beta_{ik}$ terms control for any unobservable correlation between the demand shocks for SNF k and the value added at SNF k. This term would correct, for example, for the bias that would result from a SNF that is known to generate larger improvements in health receiving patients who are unobservably less likely to improve. The $\varphi \beta_{ij}$ term soaks up any correlation between the patient's demand shock at the chosen SNF j and patient's health improvement at the chosen SNF, beyond that captured by the $\phi_k \beta_{ik}$'s. This term would capture any Roy-type selection in which a patient has an idiosyncratically higher preference for a SNF at which they are idiosyncratically more likely to improve.

3.3 Intuition for Identification

To gain some intuition for the identification properties of the model, we consider a simpler version of the model in which there is a health production function (equation (1)) and only downstream discharge (equation (2)). The object of interest is the SNF value-added parameters, the α_j's in equation (1).

This version of the model closely resembles the standard Heckman (1979) selection model for estimating the impact of wages on hours worked. There, the selection problem is that wages are not observed for individuals who work zero hours. In our case, the problem is

patients in the market, and normalize the average utility (which includes the disutility from travel distance) to zero; that is, $\delta_0 = 0$ and $m_{i0} = 0$ in equation (4) for all i.

that period 2 health is not observed for patients who are discharged before period 2. Our key identification challenge arises because SNFs may vary not only in their value added (the α_j's in equation (1)), but also in their discharge thresholds (the λ_j's in equation (2)). Thus, if we were to run a simple probit of discharge on initial health and SNF fixed effects, the SNF fixed effects would capture the difference $\alpha_j - \lambda_j$.

To address this, we take advantage of the (observed) heterogeneity in patient health at admission. The key to identification in our setting arises from the fact that we observe many patients with different initial health h_{i1} within a SNF, and each SNF is characterized by a single discharge threshold λ_j, thus allowing us to identify α_j and λ_j separately.[16]

To see this, note that for any two SNFs j and j', there exists a pair of initial health levels h_{i1} at SNF j and h'_{i1} at SNF j', such that the probability of downstream discharge is equalized. Since the probability of downstream discharge is a sufficient statistic for the bias from selection, it follows that for these two cases the selection correction terms in the value-added equation are identical. This, in turn, implies that the differential improvements for patients h_{i1} and h'_{i1} at their respective SNFs allow us to recover the SNFs' differential value added. We can then apply this argument for each pair of SNFs in order to identify the full set of α_j's up to a level and scale normalization. The normalization is not needed if there exists a SNF k and a health level h, such that discharge probability for that health level at that SNF is zero.

Our control function for selection into the SNF is more standard, using distance as an instrument to identify the selection correction terms. Intuitively, people who live farther away from a given SNF j will only choose it if it provides them with a large idiosyncratic utility shock (η_{ij} in equation (4)) relative to people who live nearby. By comparing health outcomes for people who live further vs. closer to the SNF, we can identify any correlation between unobserved determinants of choice and improvements in health.

4 Implementation

In this section, we discuss some key implementation decisions we make to bring the model to the data. Specifically, we describe the construction of our baseline health index, provide support for our choice of distance as an instrument for SNF demand, and discuss our specification and estimation in more detail.

[16]The fact that we observe discharge decisions for many patients within each SNF is also a key difference with the original Heckman (1979) setup, in which each worker makes only one labor force participation decision; in that setting, an instrument that shifts that decision without affecting wages is typically needed.

4.1 Health Index

An attraction of our setting is the rich, consistent, and comprehensively measured data from patient health assessments – including measures of physical health (e.g, shortness of breath), physical independence (e.g., degree of assistance needed with dressing), mental health (e.g., depression), and cognitive ability (e.g., delirium). Naturally, these measures speak to different dimensions of health, with differing (and unknown) utility weights. Likewise, SNF value-added may also differ across these dimensions.

Unlike test scores in education, which (debatably) provide a low-dimensional summary of student achievement, there is no established measure of health that we can take "off the shelf" for our analysis. Our baseline approach, which follows the spirit of Morris et al. (2018), is to construct a univariate health index that combines the health assessment data in a way that is guided by SNFs' purpose with regard to Medicare patients: shepherding them to the point where they can safely be discharged back to the community. Specifically, we estimate the probability that a patient is discharged to the community within 7 days of their 30-day assessment, and use these predicted values as our health index.[17] From this perspective, a higher quality SNF is one that can nurture a patient more quickly to the level of health that is conducive to community discharge. While we think this approach has its merits, our econometric framework can be applied to any single health measure or combination of measures.

Operationally, we construct the heath index with a regression tree that uses the 30-day health assessment data to predict whether the patient is discharged to the community within 7 days of their 30-day assessment.[18] We use a regression tree to capture potentially important interactions among the health measures. We estimate the tree using five-fold cross validation to tune the complexity parameter. The resulting tree has 635 terminal nodes ("leaves") and an Area Under the Curve (AUC) of 0.71.

Loosely speaking, our health index can be thought of as a weighted average of the

[17]The choice to use the 30-day assessment rather than the initial assessment is not essential, but is motivated by two factors. First, this makes the construction of the health index orthogonal to our primary analysis, which relies on what happens to patients between admission and the 30-day assessment, but not after. Second, over the first two weeks in the SNF there is a non-trivial set of patients who get transferred to other SNFs for, we suspect, non-health reasons. The choice of a 7-day outcome horizon is admittedly arbitrary, attempting to trade off greater noise that can be generated from additional health events over longer periods against lower prediction quality of "rare" outcomes, which would be the case if the horizon is too short. About 18% of the patients who have a 30-day assessment are discharged downstream within the next week (see Figure 1).

[18]As described in Appendix A, about 600,000 patients (out of a total of 2,517,006 who are present for the 30-day assessment) are missing values for their 30-day assessment. For these patients, we use a hot-deck procedure to impute missing values; see Appendix D for more details.

underlying 109 health measures, with the weights reflecting the importance of these measures in increasing the likelihood that the patient can be discharged back to the community. To give some sense of what drives the health index, the last column of Table 3 shows the "importance" of each individual measure, defined as the sum of the incremental R^2 for that variable across all the leaves of the tree, scaled so that the "most important" variable has an importance value of 100. The most important variable for predicting discharge to the community is pain intensity; other important variables include mood score, needing help with toileting, and needing help with locomotion (i.e., moving between locations).

Table 4 summarizes the resulting health index and how it correlates with subsequent outcomes. Panel A shows the index for the patients on which we trained the regression tree.[19] The average patient in this sample has a health index of 0.135; that is, they have a 13.5% chance to be discharged to the community within 7 days. However, those who actually get discharged to the community within the next 7 days have a higher average health index of 0.201. Likewise, those who get discharged elsewhere (predominantly to the hospital) within 7 days have a lower average health index (0.113) relative to those who remain in the SNF (0.127).

Panel B shows a similar qualitative pattern when we apply the health index construction to the initial health assessment for all patients, a different sample from the one used to construct the index. Relative to the average health index at admission (0.134), patients who are discharged to the community before the 30-day assessment are healthier (average health index at admission of 0.186), while those who remain in the SNF or are discharged elsewhere are sicker (average health index at admission of 0.104). The full distribution of the health index at admission is shown in Figure 2, Panel A. The standard deviation is 0.10; only 5% of patients have a probability of discharge at admission of 32% or more, which is consistent with them being admitted for care.

Panel B of Figure 2 shows a scatter plot of the average SNF length of stay (among those ultimately discharged to the community during our study period) against health index at admission for each of the 635 unique values of the health index. The relationship between health at admission and length of stay is monotone, with a slope slightly steeper than -1. Thus, one can alternatively think about a SNF with a 1 percentage point higher value added as one that is able to get a patient to the same level of health one day faster.

[19] Appendix Table A4 shows the fit for all patients present at 30 days, which includes the patients whose 30-day health assessment was missing, and was imputed using the hot-decking procedure described in Appendix D.

4.2 Distance as an Instrument for Nursing Home Choice

We use the patient's distance to each nursing home as the excluded instruments in our control function for SNF choice. This is a common instrument for choice of health-care provider, including choice of nursing home (e.g., Grabowski et al. 2013; Gupta et al. 2021), hospital (e.g., McClellan et al. 1994; Geweke et al. 2003; Cutler 2007; Card et al. 2019), and other medical providers (e.g., Einav et al. 2016). Our identifying assumption is that distance affects health improvements only through the choice of SNF.

To investigate the plausibility of this assumption, we examine the stability of the relationship between distance and SNF choice as we saturate the specification with patient demographics and health measures.[20] Panel A of Figure 3 plots the probability of going to a SNF as a non-parametric function of distance for the sample of patients and SNFs in greater Chicago. The figure shows that distance is strongly predictive of SNF choice, with the expected probability dropping by almost 50% as distance increases from 1 to 5 miles. It also shows that the relationship between distance and SNF choice is well approximated with a log relationship, which motivates our functional form choice in the demand model (equation (4)). Finally, it shows that this relationship is virtually identical across specifications with different sets of observables interacted with choice-specific indicators (described in the figure notes).

Panel B extends the exercise in Panel A to all of the markets in our data. We estimate the model of SNF choice as a function of log distance (see equation (4)) in each market both without and with observable shifters of δ_j, and plot the market-specific coefficients on log distance from the specification without observable shifters (x-axis) against the specification with the full set of shifters (y-axis). The coefficients are almost identical across specifications, suggesting that the stability of the SNF demand-distance relationship we saw for the Chicago market in Panel A generally holds for all markets.

4.3 Specification and Estimation

Estimation proceeds in two steps. We first estimate the SNF demand model in equation (4) market-by-market to construct the choice probabilities $\hat{p}_{ik}$ that are inputs into the control function terms β_{ik} in equation (6). For computational tractability, we estimate the demand

[20] This coefficient stability test can be thought of as a type of "balance test" for our identifying variation. In quasi-experimental settings, it is typical to look at whether observables are uncorrelated with treatment assignment (the excluded variable). In our context, the excluded variables are the vector of distances to each SNF (m_{ij}) in the choice set, so a simple balance test is infeasible. However, a similar test can be constructed by examining how the coefficient on distance τ in the choice model varies with the inclusion of covariates; if it does not, this is consistent with our exclusion restriction.

model using only the health index at admission as a (linear) shifter of δ_j.[21]

With the control function estimates in hand, we then jointly estimate the remaining components of the model by maximum likelihood. Specifically, we jointly estimate the health process (equation (6)), the downstream discharge model (equation (2)), and the upstream discharge model (equation (3)).[22] One numerical challenge is that with thousands of SNFs, we have thousands of parameters (recall that α_j, λ_j, and ϕ_j are all SNF-specific). To overcome this challenge, we partition the model parameters to two groups: four "national" parameters $(\theta, \varphi, \sigma_\epsilon, \sigma_\nu)$, and all other parameters, which are all market or SNF specific. This allows us to have a nested estimation procedure. Conditional on these "national" parameters, we estimate the model market by market, which is relatively standard and fast. We then search numerically over the four "national" parameters that maximize the likelihood. In Appendix E we describe the likelihood function and our estimation approach in more detail.

5 Results

5.1 Estimates of Nursing Home Value Added

Panel A of Figure 4 shows the distribution of value added for the SNFs in our baseline sample, along with summary statistics for this distribution. Panel B reports estimates of the "national" parameters that we restrict to be the same across SNFs in all markets, as well as summary statistics for the γ_0's and γ_1's, which are allowed to vary across markets. In this figure, and all subsequent results, each SNF is weighted by the number of patient-stays so that the statistics are representative of the nursing home sector.

The average (median) SNF value added is 0.045 (0.047). Combined with our estimate of almost exactly 1 for the coefficient on health at admission ($\theta = 0.99$), this implies that the average SNF increases the weekly probability of discharge to the community by about 4.5 percentage points between the initial and 30-day health assessments. The fact that value added is on average positive (health improves on average while at a nursing home) is consistent with the role of nursing homes in the Medicare system to aid patient rehabilitation and recovery (rather than to provide long-term custodial care).

The figure also indicates substantial heterogeneity across SNFs in value added. The

[21] That is, we assume that $\delta_j(x_i) = \delta_{0j} + \delta_{1j} h_{i1}$. In Appendix Figure A2 we show that adding more shifters does not affect the estimated coefficient on log-distance.

[22] As shown in Table 1, there is a small share (approximately 3%) of patients who are discharged to another SNF or to "other" locations prior to the 30-day assessment. Because it is unclear how to classify such discharges, we do not use these patients for estimation.

difference in value added between the 10th and 90th percentile SNF is 0.063; the difference in value added between a 25th and 75th percentile SNF is 0.031. The inter-quartile range in value added across SNFs is over one quarter of the inter-quartile range of health in admission across patients (of 0.11; see Figure 2). Alternatively, going from a SNF at the 25th percentile to a SNF at the 75th percentile would generate the same health improvement as moving a patient from the 10th to the 25th percentile in health at admission.[23]

The heterogeneity in value added across SNFs can also be interpreted in terms of variation in length of stay and the associated Medicare spending at the SNF. Recall from Figure 2 that a one percentage point increase in the health index at admission is associated with about one less day on average in a SNF. Given this relationship, moving from a 10th percentile value-added SNF to a 90th percentile one can also be thought of as getting a patient to the same level of health almost one week (6.3 days) faster, relative to a median length of stay of 22 days. With median Medicare spending per SNF day of about $470 (MedPAC 2018), this corresponds to about $3,000 in Medicare savings per SNF admission.

Panels C and D of Figure 4 show the relationship between SNF value added and the two other SNF-specific parameters estimated in the model: the discharge threshold λ_j and the selection coefficients ϕ_j. Panel C shows that the discharge threshold is weakly positively correlated with the value added (correlation of 0.18). A positive correlation indicates that higher value-added SNFs tend to set a higher health threshold for discharging their patients back to the community; this creates an equalizing force for length of stay across SNFs.

Panel D shows that the correlation between the value added and the selection coefficient ϕ_j is negative and large (correlation of -0.46). This is consistent with a standard adverse selection mechanism, in which the most challenging patients – those with the lowest unobserved expected health improvements – have greater demand for the highest value-added SNFs. Since distance is our excluded instrument, the estimates more specifically imply that patients with the lowest expected health improvements are disproportionately likely to travel longer distances to high-value-added SNFs. By contrast, the "Roy selection" coefficient φ is small

[23]It is natural to be concerned that the estimated variation in value added across SNFs is biased upward due to estimation error. However, in practice, we find that this is not a concern. Specifically, in Appendix Figure A3, we plot a kernel density of the baseline value added and a second kernel density of estimates where we use an empirical Bayes methodology to shrink our estimates towards the mean to account for this noise. The adjustment has essentially no effect on the dispersion in value added. This is not particularly surprising since we restricted our sample to SNFs with at least 50 patient stays. Given the similarity of the estimates, we use the non-adjusted estimates for the subsequent analysis. Additionally, to test for variation in value added over time within SNFs, we separately estimate our model on the first and second half of our sample after restricting to 5.5 million patients at 10,828 SNFs with sufficient sample size in both periods. We find that the correlation between value added estimated on these two samples is 0.72, suggesting value added is fairly stable over time.

and economically insignificant (Panel B). This implies that there is little correlation between patients' unobserved expected health improvements and patients' idiosyncratic preferences for SNFs.

5.2 Sources of Heterogeneity in Nursing Home Value Added

Figure 5 plots average SNF value added by market (Panel A) and the difference in value added between the 90th and 10th percentile SNFs within each market (Panel B), as well as associated summary statistics (Panel C). Nursing home value added is lower on average across a band of markets that stretches from Texas northeast through the Deep South and up through Appalachia. It is highest in the New England and the Mountain states. The average difference in value added between the 90th and 10th percentile markets is about 3.2 percentage points, or about half of the 90th to 10th percentile difference across SNFs.

While the substantial cross-market variation will be familiar to many health economists, the considerable variation in nursing home value added *within* markets may be more surprising. In the average market, the difference in value added between the 90th and 10th percentile nursing homes is 5.6 percentage points (bottom row of Panel C), which is almost as large as the 6.3 percentage point 90-10 difference nationwide (top row of Panel C).[24]

To further examine the within-market dispersion in value added, Figure 6 plots the average within-market distribution of value added for markets at different deciles in the national distribution of value added. It shows substantial overlap in the distribution of SNF quality across markets of different average quality. For example, in markets in the 10th decile of value added, roughly 25% of patients are in SNFs that perform below the national median. Likewise, in markets in the third decile of value added, 25% of patients are in SNFs that perform better than the national median.

This considerable within-market dispersion suggests potentially large gains from within-market reallocation of patients. For example, if patients at the 10th percentile SNF within a market could be moved to the 90th percentile SNF within the same market, the gains would be on average equivalent to getting a patient home at the same health level 5.6 days sooner (relative to a median length of stay of 22 days), saving Medicare approximately $2,600 per patient. We view this within-market dispersion as potentially promising. Relative to reallocation across markets, it seems more feasible to imagine policies – such as payment

[24]Unlike market-level average valued added, the within-market dispersion in value added does not exhibit any consistent regional pattern. As shown in Appendix Figure A4, there is a fairly weak negative correlation (-0.16) between average value added at the market level and the within-market variation.

incentives or information to consumers – that could reallocate patients across SNFs within their geographic market.

A natural question is whether the variation is SNF value added is associated with underlying characteristics of the SNF or its patients. Table 5 examines these correlations both unconditionally (column (1)) and after conditioning out market fixed-effects (column (2)).[25] The relationship between SNF value added and SNF-level characteristics is weak overall and within markets (Panel A). SNF value added is virtually uncorrelated with occupancy rate and size (as measured by the number of beds). For-profit SNFs, which account for 71% of facilities, have slightly lower value added than non-profit facilities; value-added is on average 0.044 for for-profit SNFs versus 0.048 for non-profits. However, the correlation between for-profit status and value added is virtually zero (-0.055).

SNF value added shows a stronger correlation with SNF patient characteristics unconditionally, but the correlations become fairly weak once we condition on market fixed effects (column (2)). There is some sorting to SNFs based on baseline patient health, with healthier patients (higher health index at admission) being more likely to go to higher value-added SNFs; higher value-added SNFs also receive a lower share of Black patients, and a lower share of patients who are dually-eligible for Medicaid at the time of admission.

6 Relationship to Alternative Nursing Home Ratings

We compare our value-added estimates to alternative measures of SNF quality, including the government "five-star" ratings and alternative measures we can construct with our data. Table 6 summarizes our findings. Once again we report both the unconditional correlation between our value added and each of the alternatives (column (1)), and the partial correlation conditional on market fixed effects (column (2)).[26]

6.1 Comparison to Government "Five-Star" Rating

Panel A compares our value-added estimates to the "five-star" ratings produced by the government and regularly used by patients, providers, and insurers. Specifically, we compare our quality measure to the overall five-star rating and its three sub-components: a quality index which relies on the same health-assessment data we use in this paper, an inspection score based on deficiencies (e.g., abuse and neglect by staff and medication errors) identified

[25] Appendix Figure A5 shows scatter plots that correspond to column (1) of each row of Table 5.

[26] Appendix Figure A6 produces scatter plots that correspond to column (1) of each row of Table 6.

during the three most recent annual state health inspections and the last three years of complaints, and a SNF-reported staffing measure (CMS 2008b).

The first row of Table 6 reports a fairly low positive correlation of 0.12 between our value added and the overall five-star rating. The other rows in Panel A show the correlation between our value added and the star-rating's sub-components; the low correlation between our measure and the quality index component may reflect a number of underlying differences in the construction of these measures.[27]

The low correlation between our measure of value added and the government ratings is consistent with widespread concerns about the reliability of the government ratings, highlighted in a scathing New York Times critique (Thomas 2014). Among other issues, there is evidence that SNFs game the staffing measure by inflating self-reported estimates and up-staffing around the (known) inspection dates (Thomas 2014; Boccuti et al. 2015; Han et al. 2016; Rau 2018).

To get a sense of how different markets might fair under our value-added measure relative to the current star ratings, Figure 7 plots the average star ratings in each market against the average value-added estimate in each market. There is a lot of dispersion around the regression line, suggesting that a number of markets would benefit or suffer substantially under the alternative measures. For example, Manhattan (New York) and Chicago (Illinois) both score well with the star ratings (averaging 4.0 and 3.7 stars respectively, relative to the nationwide average of 3.5), but perform poorly with our value-added measure (with average scores of -0.0002 and -0.0007 respectively, relative to the average of 0.045). On the other hand, Great Falls (Montana) averages a fairly low star rating of only 2.3 but an above-average value-added estimate of 0.06.

6.2 Comparison to 90-day Outcomes

We examine the correlation between our value-added measure and quality measures based on outcomes that we observe at 90 days after SNF admission, such as hospital readmission or death. These measures are available for all patients, and therefore do not raise concerns about bias from "selection out" (of the SNF); however, they are relatively indirect measures of what the SNF is supposed to be doing – namely, getting a patient to the point where they can safely be discharged to the community. Another limitation is that – because they are measured at 90 days – they are likely affected by events that happen to the patient outside

[27] Unlike our measure of health improvement, the government's quality rating is based on a cross-section of patient health. The government quality rating is also based only on a partially overlapping set of measures, and it includes all patients rather than just the short-stay Medicare patients that we focus on.

of the SNF (recall that the median length of stay is 22 days), and therefore provide only a noisy measure of SNF quality. These caveats notwithstanding, their comparison to our value-added measure can still serve as an informative sniff test (ahem) of our estimates, and allows us to investigate potential concerns about multi-tasking.

To estimate SNF quality for these outcomes, we make two modifications to the baseline value-added specification described in Section 3. First, as mentioned above, we do not face the same issue of sample selection bias from patient discharge and therefore do not need to correct for selection out. Second, unlike the health index that we observe at admission and at 30 days, we do not observe these outcomes at admission and therefore cannot estimate a value-added model that uses the value of the outcome at admission as a control. Instead, we control for the health index at admission as well as the control functions for SNF choice, as in Equation (6), which gives our estimates a slightly different interpretation.

Panel B of Table 6 shows these estimates for six different 90-day outcomes. In reporting the results, we "flip" the sign of "bad" outcomes, so that a positive correlation has the same qualitative interpretation across the six 90-day outcomes, regardless of whether they are "good" (e.g., home in 90 days) or "bad" (e.g., death or readmission). All six 90-day outcomes are positively correlated with our value-added estimates both overall and within markets. These correlations range across outcomes from 0.05 to 0.26 conditional on market fixed effects. Our measure of value added was designed to capture the extent to which SNFs are able to improve patient health in a way that allows them to be discharged to their homes. Consistent with this intent, our value-added measure has the largest correlation with the health measures that also reflect whether the patient is able to transition to their homes (indicator for whether a patient is home at 90 days and share of days in facilities), and a weaker correlation for the other measures. The fact that SNFs that score better on our value-added measure also tend to score better on all of these other measures suggests that higher SNF value added does not come at the expense of addressing other unmeasured health issues that would show up in a higher death rate or higher readmission rate.

6.3 Comparison to Alternative Health Indices

As noted at the outset, our approach can be used to estimate value added for any alternative health indices that can be constructed from the health assessment data. To illustrate this, we construct two alternative indices and compare the correlation between our preferred measure of value added with the measures of value added that use these alternative indices (Panel C of Table 6).

The first alternative measure we consider, known in the clinical literature as the long-form ADL (Carpenter et al. 2006), is the sum of the scores of 7 of the ADL measures reported in Appendix Table A2.[28] Our preferred value added measure has a moderately strong correlation with value added measured using ADLs (correlation of 0.45), which is consistent with the importance of ADLs in our health index (see Table 3).

The second measure we consider is the quality measure proposed by Morris et al. (2018), which is constructed using the ADL measures.[29] We find a fairly weak 0.21 correlation between our preferred value added measure and value added estimated using the Morris et al. (2018) quality measure.

6.4 Comparison to More Basic Econometric Models

Finally, we consider the quantitative importance of different components of our econometric model for our value-added estimates. We start with a model in which value added is simply the average value of the 30-day health index at the SNF. We recover this measure, which we call "Average health," from a regression of period 2 health on SNF fixed effects: $h_{i2} = \alpha_j + \epsilon_{i2}$. A slightly more sophisticated measure, which we call "Average health improvement," is based on a regression of period 2 health on initial health: $h_{i2} = \alpha_j + \theta h_{i1} + \epsilon_{i2}$.[30] The first two rows in Panel D of Table 6 report the correlation between our preferred measure of value added and these two simpler measures. Our preferred value-added measure has 0.42 correlation with average health and 0.60 correlation with the health improvement; these correlations are slightly lower (0.34 and 0.55 respectively) within markets.

The final two rows in Panel D of Table 6 explore the quantitative importance of the corrections we employ in our construction of SNF value added. In the third row, we report the relationship between our preferred value-added measure and a version of the average improvement model that accounts for the selection into the SNF (as we do in our baseline model), but does not account for the selection out (via discharges from the SNF before the 30 day assessment). The fourth row reports on the reverse exercise, accounting for selection out but without correcting for selection into the SNF. Starting with the measure of average

[28] These ADLs are hygiene, dressing, locomotion, transfer, toilet, bed-mobility, and eating. Since each measure takes value from 0-4, the aggregate measure ranges from 0 to 28, with higher values indicating that the patient needs more assistance.

[29] To construct their measure, Morris et al. (2018) divide each ADL measure into terciles, coded from 0 to 2 where 0 represents the tercile with the worst outcomes and 2 represents the tercile with the best outcomes. The index is constructed by summing these values together.

[30] Additionally controlling for patient demographics (above median age, female, married, Hispanic, Black, white, dual eligible) in this model does not affect the estimates of α_j.

health improvement, correcting for selection into the SNF increases the correlation with our preferred measure from 0.60 to 0.68, while correcting for selection out increases the correlation to 0.73. The importance of correcting for selection into the SNF is consistent with the fairly strong adverse selection documented in Panel D of Figure 4. These results suggest that failure to account for either selection in or selection out can have a meaningful impact on value-added estimates.

7 Conclusions

In this paper we developed an econometric model that allowed us to estimate and compare the health production process at almost 14,000 nursing homes. We found substantial heterogeneity in value added across nursing homes; moving from the 10th percentile of the nursing home value-added distribution to the 90th percentile is equivalent to being able to discharge a patient to the community almost one-week faster (or almost one-third the median length of stay in our sample).

Strikingly, we found that the dispersion in nursing home value added *within* markets is almost as large as the nationwide dispersion. In other words, there are at least some high-quality nursing homes in low-quality markets and at least some low-quality nursing homes in high-quality markets. This points to the potential for substantial gains from policies that encourage reallocation of patients to higher quality nursing homes within their market.

Our findings therefore suggest that an important area for further work is to analyze the economic and policy forces that can achieve such market reallocation. Using our value-added measure as the dependent variable, it would be useful to study how it is affected by mergers or other changes in competition, or by the types of policies that are used to try to improve nursing home quality, such as changes in Medicare reimbursement, the imposition of minimum staffing ratios, or information campaigns (Wunderlich and Kohler 2001; Konetzka 2020).

There are also a number of natural ways to extend our analysis of nursing home value added. Most obviously, while we focus on differences across nursing homes in improving health to the point that the patient can be safely discharged to the community, our approach can be applied to any of the many different underlying measures of physical health, mental health, or cognitive capacity that may be of particular interest. In addition, while we have estimated a single value-added measure for each nursing home, future work could explore whether there are different dimensions of value added – e.g., across patients needing different types of rehabilitation or patients of different demographics – and whether these dimensions are positively or negatively correlated within a nursing home.

References

Abdulkadiroğlu, Atila, Parag A Pathak, Jonathan Schellenberg, and Christopher R Walters. 2020. "Do parents value school effectiveness?" *American Economic Review* 110 (5): 1502–39.

Agarwal, Nikhil, and Paulo Somaini. 2022. "Demand Analysis under Latent Choice Constraints". *NBER Working Paper,* number 29993.

Boccuti, Cristina, Giselle Casillas, and Tricia Neuman. 2015. "Reading the Stars: Nursing Home Quality Star Ratings, Nationally and by State", https://www.kff.org/medicare/issue-brief/reading-the-stars-nursing-home-quality-star-ratings-nationally-and-by-state/.

Bowblis, John R, and Christopher S Brunt. 2014. "Medicare Skilled Nursing Facility Reimbursement and Upcoding". *Health Economics* 23 (7): 821–840.

Bring, Johan. 1995. "Variable importance by partitioningR 2". *Quality & Quantity* 29 (2): 173–189. Visited on May 20, 2022.

Card, David, Alessandra Fenizia, and David Silver. 2019. "The Health Impacts of Hospital Delivery Practices". *NBER Working Paper,* number 25986.

Carpenter, G. Iain, Charlotte L. Hastie, John N. Morris, Brant E. Fries, and Joel Ankri. 2006. "Measuring change in activities of daily living in nursing home residents with moderate to severe cognitive impairment". *BMC Geriatrics* 6 (7).

Castle, Nicholas G., and Jamie C. Ferguson. 2010. "What Is Nursing Home Quality and How Is It Measured?" *The Gerontologist* 50 (4): 426–442.

Centers for Medicare & Medicaid Services. 2008a. *Talking Points – Improving the CMS Nursing Home Compare Website — Guidance Portal.* https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/SurveyCertificationGenInfo/downloads/improvingnhcompare.pdf.

———. 2008b. "The Five-Star Nursing Home Rating System – Questions and Answers". https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/SurveyCertificationGenInfo/downloads/SCLetter09-18.pdf.

———. 2016. *Long-Term Care Facility Resident Assessment Instrument 3.0 User's Manual.* https://downloads.cms.gov/files/mds-30-rai-manual-v114-october-2016.pdf.

———. 2017. "Medicare-Required SNF PPS Assessments". https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/SNF-PPS/SNFAssessHTML022817f.html.

———. 2018. "Survey and Enforcement Process for Skilled Nursing Facilities and Nursing Facilities". Chapter 7 in *State Operations Manual.* https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/som107c07pdf.pdf.

———. 2019a. "Medicare Coverage of Skilled Nursing Facility Care". https://www.medicare.gov/Pubs/pdf/10153-Medicare-Skilled-Nursing-Facility-Care.pdf.

———. 2019b. "National Health Expenditures by Type of Service and Source of Funds: Calendar Years 1960 to 2019". https://www.cms.gov/files/zip/national-health-expenditures-type-service-and-source-funds-cy-1960-2019.zip.

Centers for Medicare & Medicaid Services. 2022. "Special Focus Facility ("SFF") Program". https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/CertificationandComplianc/Downloads/SFFList.pdf.

———. No date. "Compare Nursing Home Quality". https://www.medicare.gov/what-medicare-covers/what-part-a-covers/compare-nursing-home-quality.

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.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014a. "Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates". *American Economic Review* 104 (9): 2593–2632.

———. 2014b. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood". *American Economic Review* 104 (9): 2633–79.

42 C.F.R. §483. 2016.

Congressional Budget Office. 2008. *Geographic Variation in Health Care Spending.* Washington, DC: Congress of the United States. https://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/89xx/doc8972/02-15-geoghealth.pdf.

Cornell, Portia Y, David C Grabowski, Edward C Norton, and Momotazur Rahman. 2019. "Do report cards predict future quality? The case of skilled nursing facilities". *Journal of Health Economics* 66:208–221.

Cutler, David M. 2007. "The lifetime costs and benefits of medical technology". *Journal of Health Economics* 26 (6): 1081–1100.

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 (5): 841–852.

Dubin, Jeffrey A, and Daniel L McFadden. 1984. "An econometric analysis of residential electric appliance holdings and consumption". *Econometrica* 52 (2): 345–362.

Einav, Liran, Amy Finkelstein, and Heidi Williams. 2016. "Paying on the Margin for Medical Care: Evidence from Breast Cancer Treatments". *American Economic Journal: Economic Policy* 8 (1): 52–79.

Gandhi, Ashvin. 2020. "Picking your patients: Selective admissions in the nursing home industry". *Working Paper,* visited on January 8, 2021. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3613950.

Gandhi, Ashvin, Huizi Yu, and David C Grabowski. 2021. "High Nursing Staff Turnover In Nursing Homes Offers Important Quality Information: Study examines high turnover of nursing staff at US nursing homes." *Health Affairs* 40 (3): 384–391.

Gawande, Atul. 2009. "The Cost Conundrum: What a Texas Town Can Teach Us about Health Care". *The New Yorker,* https://www.newyorker.com/magazine/2009/06/01/the-cost-conundrum.

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

Goldstein, Matthew, and Robert Gebeloff. 2019. "Dozens of Nursing Homes With HUD-Backed Mortgages Have 'Serious Deficiencies'". *The New York Times,* https://www.nytimes.com/2019/06/24/business/nursing-homes-hud.html.

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 (1): 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 (4): 754–764.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health". *Journal of Political Economy* 80 (2): 223–255.

Gupta, Atul, Sabrina T Howell, Constantine Yannelis, and Abhinav Gupta. 2021. "Does private equity investment in healthcare benefit patients? Evidence from nursing homes". *NBER Working Paper.*

Hackmann, Martin B. 2019. "Incentivizing better quality of care: The role of Medicaid and competition in the nursing home industry". *American Economic Review* 109 (5): 1684–1716.

Hackmann, Martin B, R Vincent Pohl, and Nicolas Ziebarth. 2021. "Patient vs. provider incentives in long term care". *Working Paper.*

Han, Xu, Niam Yaraghi, and Ram Gopal. 2016. "Five-star ratings for sub-par service: Evidence of inflation in nursing home ratings". *Governance Studies at Brookings,* 1–22.

Harrington, Charlene, Helen Carrillo, Rachel Garfield, MaryBeth Musumeci, and Ellen Squires. 2018. "Nursing Facilities, Staffing, Residents and Facility Deficiencies, 2009 Through 2016", https://www.kff.org/medicaid/report/nursing-facilities-staffing-residents-and-facility-deficiencies-2009-through-2016/.

Heckman, James J. 1979. "Sample selection bias as a specification error". *Econometrica* 47 (1): 153–161.

Hoffman, James V., Lori Czop Assaf, and Scott G. Paris. 2001. "High-Stakes Testing in Reading: Today in Texas, Tomorrow?" *The Reading Teacher* 54 (5): 482–492.

Hull, Peter. 2020. "Estimating Hospital Quality with Quasi-Experimental Data". *Working Paper,* visited on July 15, 2022. https://www.google.com/url?q=https%3A%2F%2Fwww.dropbox.com%2Fs%2Fhb54rrz3vte8gij%2FRAM_012020.pdf%3Fraw%3D1&sa=D&sntz=1&usg=AOvVaw0Lbo2EeXd1bMjdm9zRILa9.

Institute of Medicine. 2013. *Variation in Health Care Spending: Target Decision Making, Not Geography.* Redacted by Joseph P. Newhouse, Alan M. Garber, Robin P. Graham, Margaret A. McCoy, Michelle Mancher, and Ashna Kibria. Washington, DC: The National Academies Press.

Jacob, Brian A., and Steven D. Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating". *The Quarterly Journal of Economics* 118 (3): 843–877.

Jacobs, Andrew, and Matt Richtel. 2019. "Nursing Homes Are a Breeding Ground for a Fatal Fungus". *The New York Times,* https://www.nytimes.com/2019/09/11/health/nursing-homes-fungus.html.

Koedel, Cory, Kata Mihaly, and Jonah E Rockoff. 2015. "Value-added modeling: A review". *Economics of Education Review* 47:180–195.

Konetzka, R. Tamara. 2020. "The Challenges of Improving Nursing Home Quality". *JAMA Network Open* 3 (1): e1920231.

Lazear, Edward P. 2006. "Speeding, Terrorism, and Teaching to the Test*". *The Quarterly Journal of Economics* 121 (3): 1029–1061.

Levinson, Daniel. 2015. *The Medicare Payment System for Skilled Nursing Facilities Needs to be Reevaluated.* https://oig.hhs.gov/oei/reports/oei-02-13-00610.pdf.

Long-Term Care: Facts on Care in the US. No date. "About Us". http://ltcfocus.org/about.

McClellan, Mark, Barbara J. McNeil, and Joseph P. Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality?: Analysis Using Instrumental Variables". *JAMA* 272 (11): 859–866.

Medicare Payment Advisory Commission. 2018. "Skilled nursing facility services". Chapter 8 in *Report to the Congress: Medicare Payment Policy.* Medicare Payment Advisory Commission.

———. 2020. "Post-acute care". Chapter 8 in *Health Care Spending and the Medicare Program.* Medicare Payment Advisory Commission.

———. 2021. "Skilled Nursing Facility Services Payment System". https://www.medpac.gov/wp-content/uploads/2021/11/medpac_payment_basics_21_snf_final_sec.pdf.

Morris, Carl N. 1983. "Parametric Empirical Bayes Inference: Theory and Applications". *Journal of the American Statistical Association* 78 (381): 47–55.

Morris, John N, Katherine Berg, Eva Topinkova, Leonard C Gray, and Erez Schachter. 2018. "Developing quality indicators for in-patient post-acute care". *BMC Geriatrics* 18 (1): 161.

Morris, John N, Catherine Hawes, Brant E Fries, Charles D Phillips, Vincent Mor, Sidney Katz, Katharine Murphy, Margaret L Drugovich, and Alan S Friedlob. 1990. "Designing the national resident assessment instrument for nursing homes". *The Gerontologist* 30 (3): 293–307.

Olenski, Andrew, and Szymon Sacher. 2022. *Estimating Nursing Home Quality with Selection.* SSRN Scholarly Paper 4054786. Rochester, NY: Social Science Research Network. Visited on May 4, 2022. https://doi.org/10.2139/ssrn.4054786. https://papers.ssrn.com/abstract=4054786.

Rau, Jordan. 2018. "'It's Almost Like a Ghost Town.' Most Nursing Homes Overstated Staffing for Years". *The New York Times,* https://www.nytimes.com/2018/07/07/health/nursing-homes-staffing-medicare.html.

Rau, Jordan, and Elizabeth Lucas. 2018. "Medicare Slashes Star Ratings for Staffing at 1 in 11 Nursing Homes". *The New York Times,* https://www.nytimes.com/2018/07/27/health/medicare-nursing-homes.html.

Ryskina, Kira L., R. Tamara Konetzka, and Rachel M. Werner. 2018. "Association Between 5-Star Nursing Home Report Card Ratings and Potentially Preventable Hospitalizations". *INQUIRY: The Journal of Health Care Organization, Provision, and Financing* 55.

Silver-Greenberg, Jessica, and Robert Gebeloff. 2021. "Maggots, Rape and Yet Five Stars: How U.S. Ratings of Nursing Homes Mislead the Public". *The New York Times,* https://www.nytimes.com/2021/03/13/business/nursing-homes-ratings-medicare-covid.html.

Skinner, Jonathan. 2011. "Causes and Consequences of Regional Variations in Health Care". In *Handbook of Health Economics,* 2:45–93. Elsevier.

Small, Kenneth, and Harvey Rosen. 1981. "Applied welfare economics with discrete choice models". *Econometrica* 49 (1): 105–130.

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

Todd, Petra E, and Kenneth I Wolpin. 2003. "On the specification and estimation of the production function for cognitive achievement". *The Economic Journal* 113 (485): F3–F33.

12 U.S.C. §1715w. 2014.

United States Government Accountability Office. 2016. *Nursing Homes: Consumers Could Benefit from Improvements to the Nursing Home Compare Website and Five-Star Quality Rating System.* https://www.gao.gov/products/gao-17-61.

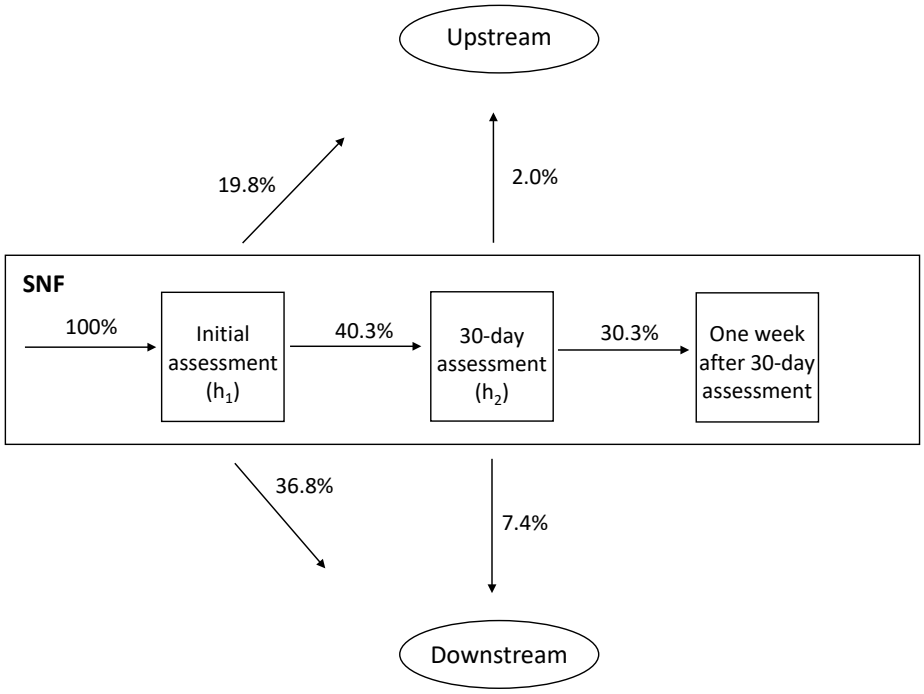
———. 2019. *Nursing Homes: Improved Oversight Needed to Better Protect Residents from Abuse.* https://www.gao.gov/products/gao-19-433.

Wennberg, John E, Megan M Cooper, JD Birkmeyer, et al. 1998. "The Dartmouth atlas of health care 1998".

Wunderlich, Gooloo S, and Peter O Kohler. 2001. *Improving the quality of long-term care.* National Academies Press (US).

Wunderlich, Gooloo S., Frank Sloan, and Carolyne K. Davis. 1996. *Staffing and Quality of Care in Nursing Homes.* National Academies Press (US).

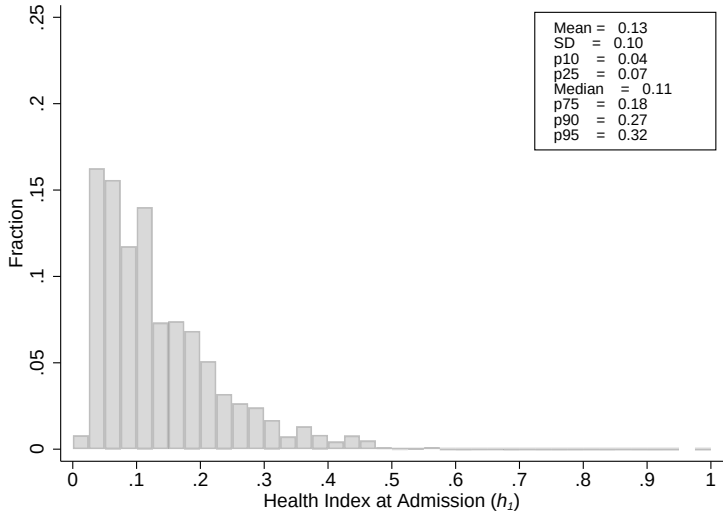
Figure 1: Movement and Measurement of Patients



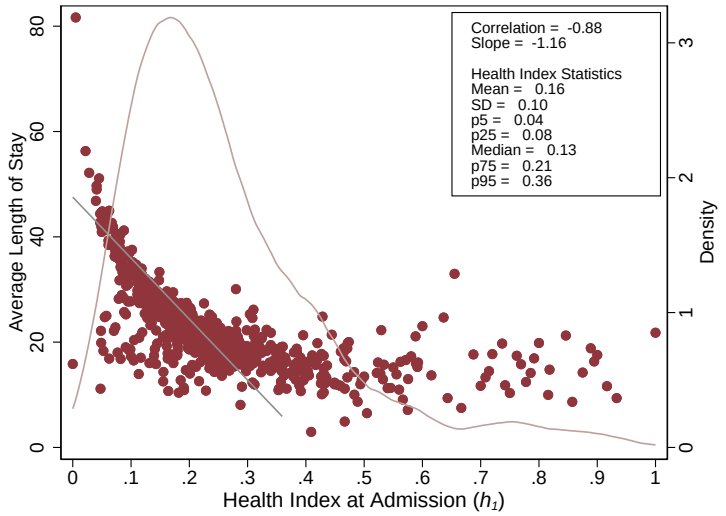
Notes: Figure shows movement of SNF patients and the timing of health assessments. Downstream refers to patients who are discharged to the community. Upstream refers to patients who die in the nursing home, are sent to hospice, or are sent to an acute care hospital. Between the initial assessment and the 30-day assessment 3.1 percent of patients are discharged to a destination not accounted for by our upstream and downstream definitions; these include discharges to other SNFs, facilities for intellectual and developmental disabilities, inpatient rehabilitation facilities, and other/unknown (not shown). Statistics on discharge after the 30-day assessment are for discharges within 7 days of the 30-day assessment. Between the 30-day assessment and one week after the assessment, 0.6% of patients are discharged to a destination not accounted for by our upstream and downstream destinations (not shown). Statistics are calculated using our baseline sample (N=6,246,686).

Figure 2: Health Index

(A) Distribution of Health Index at Admission



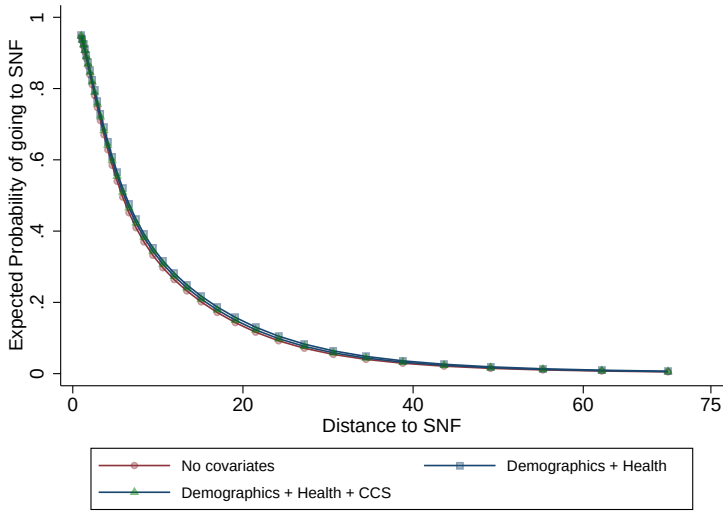
(B) Health Index at Admission and Length of Stay



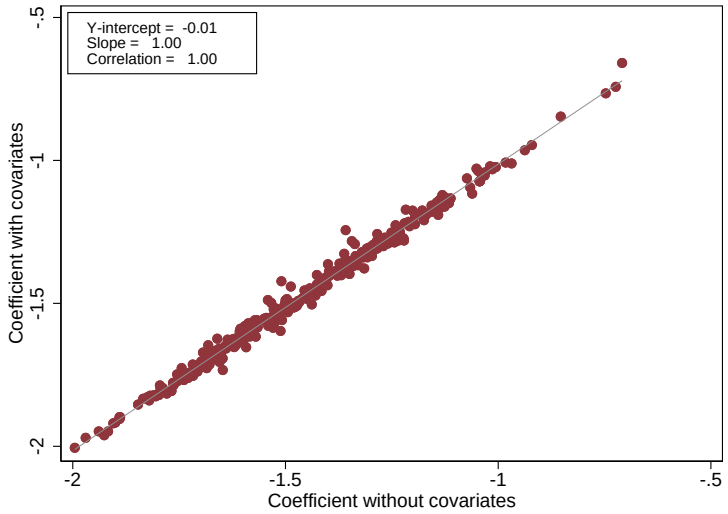
Notes: Panel (A) shows the distribution of our health index at admission (h_1) in the baseline sample (N=6,246,686). Panel (B) shows the relationship between length of stay and health index at admission (h_1) conditional on discharge to the community during our study period (N=3,847,120); it shows the relationship for each of the 635 unique values of h_1 that come from our regression tree. The line in Panel (B) is the linear fit between average length of stay and baseline health for patients with a baseline health measure below the 95th percentile (0.36) conditional on discharge to the community.

Figure 3: Distance Instrument for Selection In

(A) Chicago HRR



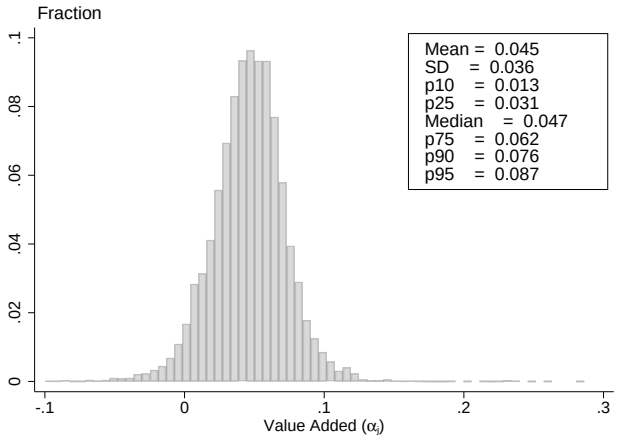
(B) All Markets



Notes: Panel (A) plots the non-parametric relationship between the probability of going to a SNF and distance to that SNF in the greater Chicago market from specifications that sequentially add observables interacted with choice-specific indicators (N = 212,937 patients; 294 SNFs). Panel (B) plots the coefficient on log distance from market-by-market estimates of demand model shown in equation (4) estimated on the baseline sample (N=6,246,686). Specifically, each point denotes a market, with horizontal axis displaying the coefficient on distance from the specification without observable shifters of δ_j and the vertical axis displaying the coefficient on distance from the specification with demographics, health, and CCS shifters. The shifters for demographics and health are the first ten principal components of a set of demographics (above median age, female, married, Hispanic, Black, white, dual eligible) and health measures at admission (walking, indwelling, falls, shortness of breath, depressed, delirium, vomiting, fever, dehydration, weight loss, and long-form ADL). The CCS shifters are indicators for the 20 most common Clinical Classification codes.

Figure 4: Value Added

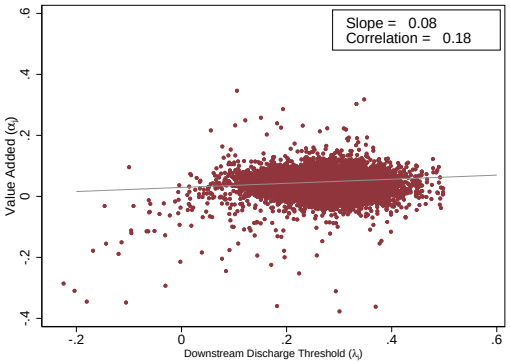
(A) Distribution of Value Added



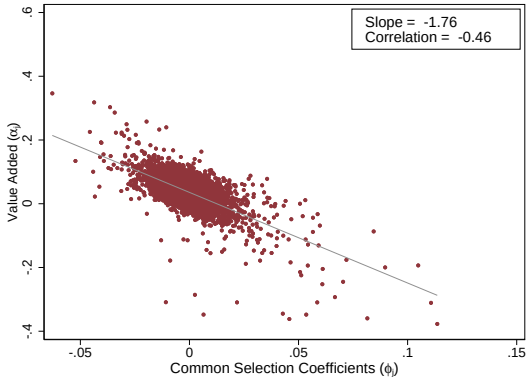
(B) Additional Parameters

Parameter	Estimate
A. "National" parameters:	
θ	0.99
φ	0
σ_ϵ	0.07
σ_v	0.15
B. Upstream discharge parameters:	
Average γ_0	-0.38
Std. Dev. of γ_0	0.26
Average γ_1	-1.2
Std. Dev. of γ_1	8.2

(C) Value Added vs. Discharge Threshold



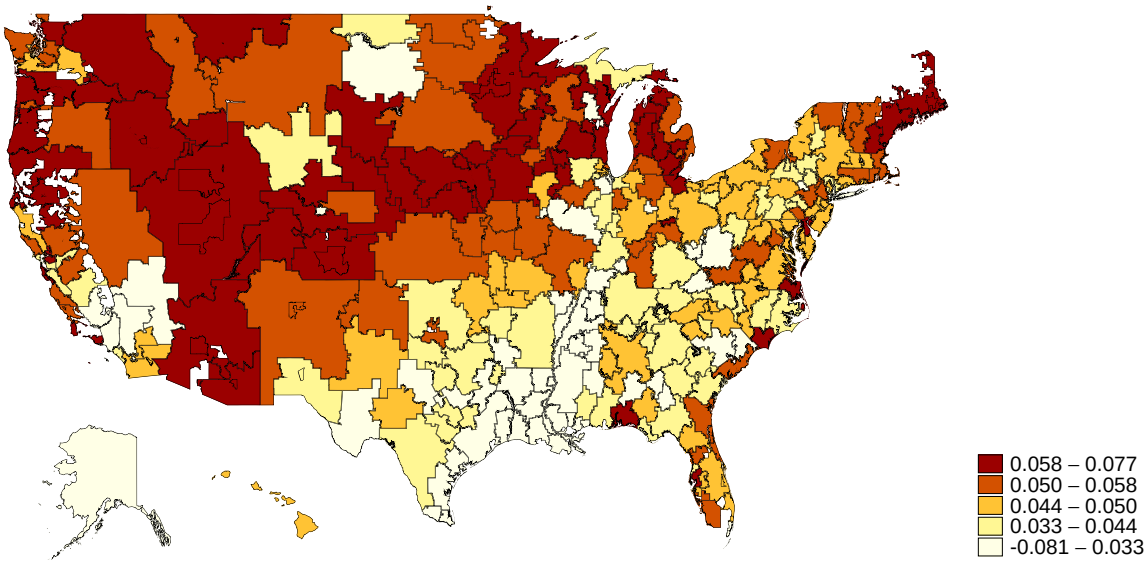
(D) Value Added vs. Selection Coefficient



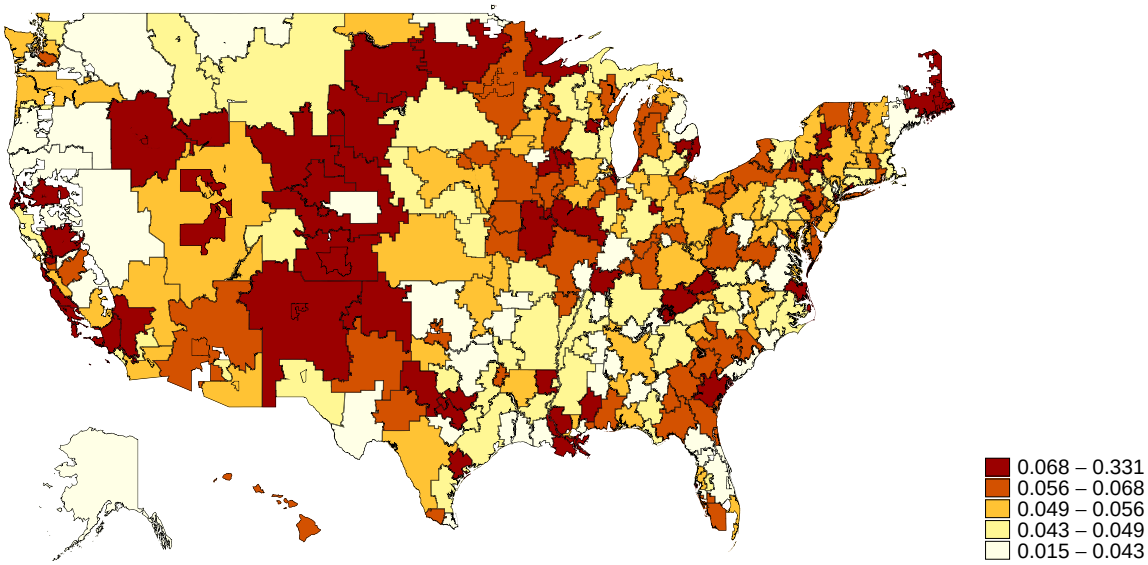
Notes: Figure reports the parameter estimates from jointly estimating equations (2), (3), and (6). Panel (A) shows the distribution of value added (α_j) across SNFs. Panel (B) reports estimates of the "national parameters" from our model and summary statistics of the market-level upstream discharge parameters (γ_0 and γ_1). Panel (C) plots the relationship between value added (α_j) and the SNF-specific downstream discharge thresholds (λ_j). Panel (D) plots the relationship between value added (α_j) and the SNF-specific coefficients on the selection control function (ϕ_j). The sample is comprised of 13,938 SNFs and 6,246,686 patients. The analysis weights each SNF by the number of patients we observe at that facility.

Figure 5: Heterogeneity Across and Within Markets

(A) Average Value Added by Market



(B) 90-10 Difference in Value Added By Market

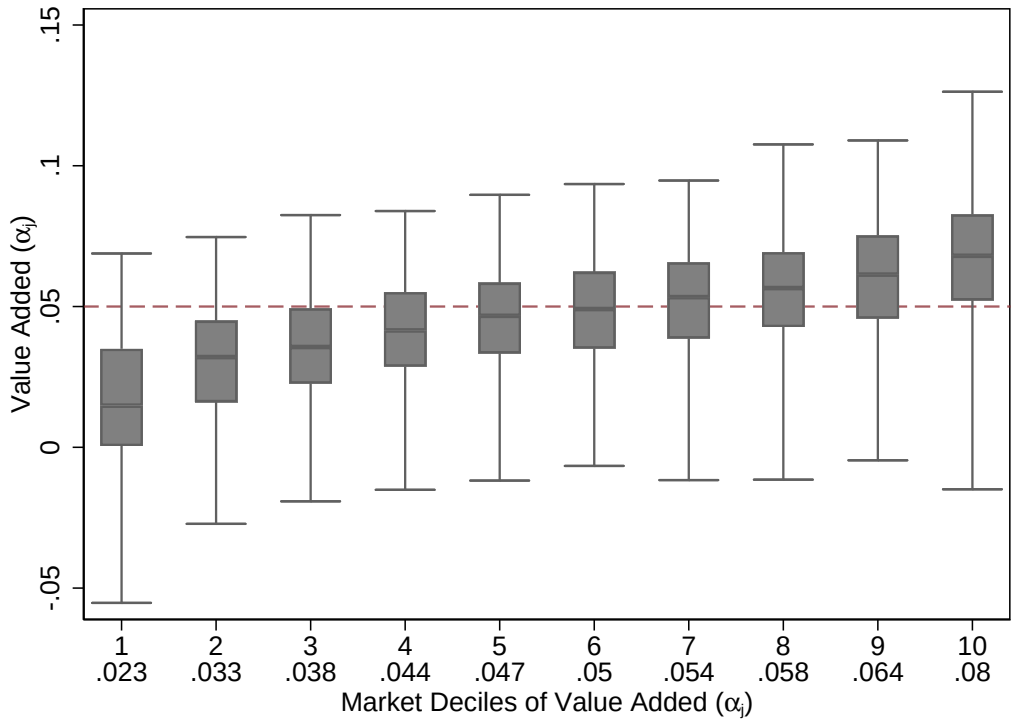


(C) Distributions of Value-Added

	Mean	SD	10th	25th	50th	75th	90th
Across SNFs	0.045	0.036	0.013	0.031	0.047	0.062	0.076
Across Markets	0.045	0.015	0.029	0.039	0.048	0.054	0.061
Within Markets	0.045	0.027	0.017	0.032	0.047	0.06	0.073

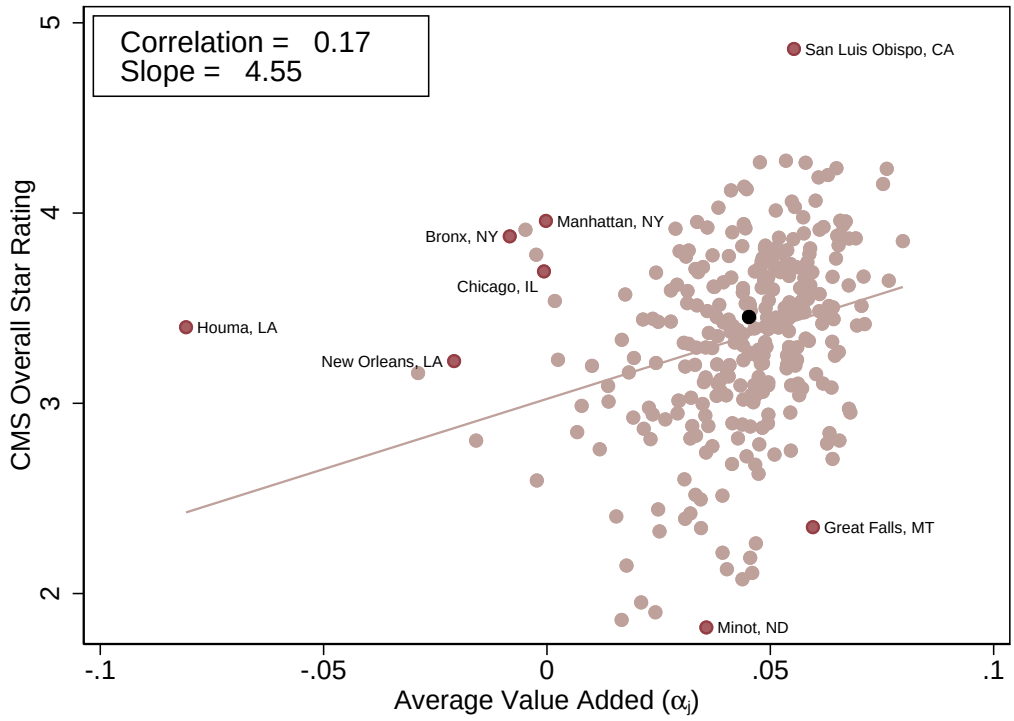
Notes: Panel (A) plots the average value added in each market. Panel (B) plots the difference between the 90th percentile and 10th percentile of the value added distribution within each market. Panel (C) shows summary statistics of value added across SNFs, across markets, and within markets; Both maps and summary statistics are constructed after weighting each SNF by the number of patients. The color gradients follow quartiles of the corresponding market-level distribution. Data is shown for 306 markets covering 13,938 SNFs. Blank regions in the maps have no SNFs with patients in our sample.

Figure 6: Dispersion in Value Added Within Markets



Notes: Figure shows box-and-whiskers plots of SNF value added by deciles of the market-level distribution of value added. The lower (upper) "whisker" shows the 2.5th (97.5th) percentile. The bottom of each box is the 25th percentile, the middle is the median, and the top is the 75th percentile. The distribution depicted by each box-and-whisker plot is weighted by patients. The dashed horizontal line is the median across all SNFs in our sample, weighted by patients.

Figure 7: Correlating CMS Star Ratings and Value Added at the Market Level



Notes: Figure shows a scatter of the average CMS overall star rating against average value added at the market level. Market-level averages are weighted by the number of patients in each SNF. The points represent 306 markets covering the 13,646 SNFs for which we observe CMS star ratings. The black dot shows the patient-weighted average of the overall rating and value added. Certain outlier markets and other markets of interest are labeled. The line is a linear fit weighted by the number of patients in each market.

Table 1: Patient Characteristics

	Mean (1)		Std. dev. across SNFs (2)
A. Patient characteristics at admission (N = 6,246,686)			
Age	81.0	(8.33)	1.92
Female	0.63		0.06
Married	0.36		0.07
White	0.86		0.18
Black	0.08		0.14
Asian	0.02		0.06
Hispanic	0.03		0.09
Other/Unknown	0.01		0.04
Medicaid (aka "dually eligible") at admission	0.18		0.14
Length of hospital stay (prior to SNF admission)	7.58	(7.67)	1.96
B. SNF length of stay			
Length of stay	44.7	(100.93)	20.81
Share discharged before 30-day assessment	0.59		0.12
Share discharged before 60-day assessment	0.85		0.07
Share discharged before 90-day assessment	0.92		0.04
C. Discharge destination for patients discharged prior to day-30 assessment			
Community (downstream)	0.62		0.14
Acute care hospital (upstream)	0.28		0.11
Deceased or hospice (upstream)	0.03		0.03
Another nursing home	0.05		0.04
Other	0.02		0.02

Notes: Table shows summary statistics of patient characteristics in our baseline sample (N = 6,246,686 patients at 13,996 SNFs). Column (1) reports the mean across patients, with patient-level standard deviations in parentheses. Column (2) reports the standard deviation of the SNF-level average across SNFs, with each SNF weighted by the number of patients.

Table 2: SNF Characteristics

	No. of obs.	Mean (weighted mean)[a]	Std. dev.	5th pct'ile	Median	95th pct'ile
Number of beds	13,646	111.8 (128.3)	60.5	40	104	210
Occupancy %	13,646	82.2 (83.7)	12.6	56.8	85.9	95.9
For profit status	13,646	0.71 (0.70)				
Hospital-based	13,646	0.04 (0.06)				
Admissions	13,996	446.3 (911.2)	478.1	65	290	1,343

Notes: Table shows summary statistics of SNF characteristics in our baseline sample. All data except for the number of admissions in the bottom row come from Online Survey Certification And Reporting (OSCAR), which are collected during annual inspections of SNFs and are missing for 350 SNFs.
[a] Weighted means are weighted by the number of patients in the SNF.

Table 3: Selected Health Index Measures

Health Measure		Mean (1)	Std. dev. (2)	Std. dev. across SNFs (3)	Importance (4)
Physical					
Has fallen	(0-1)	0.07	0.25	0.03	4.54
Vomiting	(0-1)	0.02	0.15	0.01	0.00
Turning balance	(0-3)	2.11	0.79	0.29	14.84
Upper extremity range	(0-2)	0.20	0.52	0.15	4.38
Surgical wounds	(0-1)	0.32	0.46	0.12	15.51
Mental					
Mood score (PHQ-9)	(0-27)	2.59	3.50	1.59	82.18
Antipsychotics	(0-1)	0.09	0.29	0.05	2.81
Antidepressants	(0-1)	0.34	0.48	0.07	19.35
Psychosis	(0-1)	0.98	0.14	0.03	0.00
Brief Interview for Mental Status (BIMS)	(0-15)	12.12	3.80	1.06	18.98
Restraints					
Bed rail restraints	(0-2)	0.03	0.24	0.18	0.00
Trunk rail restraints	(0-2)	0.00	0.02	0.00	0.00
Limb rail restraints	(0-2)	0.00	0.04	0.01	0.00
Trunk chair restraints	(0-2)	0.00	0.06	0.01	0.00
Limb chair restraints	(0-2)	0.00	0.03	0.00	0.00
Activities of Daily Living (ADL)					
Walking	(0-4)	2.96	1.06	0.36	31.36
Hygiene	(0-4)	2.51	0.90	0.40	1.40
Dressing	(0-4)	2.75	0.69	0.22	9.65
Locomotion	(0-4)	2.75	0.97	0.34	53.00
Toilet	(0-4)	2.79	0.69	0.23	43.32
Pain					
Pain management	(0-1)	0.71	0.45	0.12	2.77
Pain presence	(0-1)	0.60	0.49	0.13	7.41
Pain swallowing	(0-1)	0.03	0.16	0.03	0.00
Mouth/face pain when chewing	(0-1)	0.01	0.11	0.02	0.00
Pain intensity	(0-10)	3.11	3.36	1.19	100.00
Treatement and Equipment					
Walker	(0-1)	0.65	0.48	0.20	3.87
Wheelchair	(0-1)	0.86	0.35	0.15	10.97
Limb prosthesis	(0-1)	0.00	0.05	0.00	0.00
Insulin	(0-7)	1.13	2.46	0.35	6.70
Oxygen therapy	(0-1)	0.26	0.44	0.09	24.78
Interaction					
Speech clarity	(0-2)	0.09	0.40	0.09	0.00
Makes self understood	(0-3)	0.25	0.64	0.19	3.18
Ability to understand others	(0-3)	0.29	0.64	0.21	6.45
Verbal behavior symptoms	(0-3)	0.04	0.23	0.03	0.08
Other behavior symptoms	(0-3)	0.03	0.24	0.04	0.84

Notes: Table shows summary statistics at admission for a selected subset of the 109 health measures that are used in our health index. The health measures are binary or categorical variables. The range for each measure is shown in parentheses; higher values indicate worse health, except for BIMS score, where higher is better. Appendix A provides more detail on the measurement of some variables. Standard deviation across SNFs in column (3) is weighted by the number of patients in the SNF. The importance measure in column (4) is defined as the incremental R^2 of that variable; that is, the difference in R^2 when regressing our outcome – discharge to the community – on all health measures, with and without that variable. The incremental R^2 is then scaled such that the "most important" variable has an importance of 100 (see Bring (1995) for a discussion of variable importance).

Table 4: Health Index

	Share of baseline sample	Mean h	Std. dev. of h
A. Internal fit (h as of day-30 assessment, N = 2,028,014):			
All patients assessed at day 30	0.325	0.135	0.093
Discharged to community w/in 7 days	0.035	0.201	0.115
Still at the SNF 7 days after	0.275	0.127	0.087
Discharged elsewhere w/in 7 days	0.014	0.113	0.083
B. External fit (h as of initial assessment, N = 6,246,686):			
All patients assessed at day 30	1.000	0.134	0.095
Discharged to community before day-30 assessment	0.368	0.186	0.106
Still at the SNF for day-30 assessment	0.403	0.104	0.069
Discharged elsewhere before day-30 assessment	0.229	0.104	0.082

Notes: Table shows summary statistics for our health index. Panel A shows the health index at 30 days for the sample of 2,028,014 patients with non-missing 30-day assessment data; Panel B shows the health index at admission for our full sample of 6,246,686 patients.

Table 5: Correlations of Value Added Estimates with SNF and Patient Characteristics

	Correlation	Partial correlation controlling for market FEs
	(1)	(2)
A. SNF characteristics		
SNF occupancy	0.039	0.054
Number of beds	-0.079	0.017
For profit	-0.056	-0.036
B. Patient characteristics		
Black patient share	-0.186	-0.110
Medicaid (aka "dually eligible") at admission	-0.299	-0.221
Health index at admission	0.220	0.084

Notes: Table shows SNF-level correlations between our value-added estimates and characteristics of both SNFs (Panel A) and their patients (Panel B). Correlations are weighted by the number of patients in the SNF and calculated based on value-added estimates for 13,938 SNFs covering 306 markets.

Table 6: Correlations of Value Added Estimates with Alternative Measures

	Correlation	Partial correlation controlling for market FEs
	(1)	(2)
A. CMS star ratings		
Overall	0.12	0.11
Quality	0.03	0.07
Inspection	0.06	0.07
Staffing	0.20	0.11
B. 90-day outcomes (*signed so that higher is better*)		
Share of days in facilities 90 days	0.24	0.21
Home at 90 days	0.27	0.26
Spending per day 90 days	0.22	0.15
SNF readmission 90 days	0.17	0.14
Hospital readmission 90 days	0.11	0.05
Alive at 90 days	0.18	0.22
C. Alternative health indices		
ADL measure	0.45	0.40
Morris et al. (2018) measure	0.21	0.15
D. Other specifications		
Average health	0.42	0.34
Average health improvement	0.60	0.55
Average health improvement + select in	0.68	0.66
Average health improvement + select out	0.73	0.69

Notes: Table shows correlations between our baseline value added estimates and alternative measures of SNF quality, with and without controlling for market fixed effects. Panel A shows the correlation with the overall five-star rating and its three sub-components. Panel B shows correlations with outcomes based on the 90-day period following SNF admissions. For instance, "home at 90 days" is an indicator for whether the patient is at home 90 days post admission and "SNF readmission 90 days" is an indicator for whether the patient was readmitted to the SNF within 90 days of the initial admission. Panel C shows correlations with alternative health indices defined using the same health assessment data as in our baseline estimates. See Section 6.2 for details. Panel D shows correlations with alternative econometric specifications of value added using the health index. See Section 6.3 for details. Correlations are weighted by the number of patients in the SNF and calculated based on value-added estimates for 13,938 SNFs covering 306 markets, except for Panel A which is based on the 13,646 SNFs (covering 306 markets) for which we observe star ratings.

Producing Health: Measuring Value Added of Nursing Homes

Online Appendix

A Health Measures Used in Health Index

We observe 132 separate health measures in both the 5-day and the 30-day assessment in every year in our data. Following the clinical literature (Morris et al. 2018), we remove six redundant measures that relate to activities of daily living (ADL) and urinary tract health. We also combine closely related health measures into composites. Specifically, we sum together those indicators for shortness of breath and those for pain management. This leaves us with 122 separate health measures.

We restrict our sample to stays that have a 5-day health assessment, which we define as having non-missing responses for at least half of the 122 health measures and occurring in the first 8 days of the stay (CMS requires the 5-day assessment to be conducted between days 1-8). This reduces our sample from approximately 6.6 million to 6.2 million stays (see Appendix Table A1). Most of dropped stays are missing responses to all of the questions on the health assessment.

Of the remaining 6.2 million stays, 2.5 million were long enough to cover the 30-day assessment period. Among these, approximately 500,000 are missing more than half (and typically all) of the health measures at 30 days. When this is the case, we treat the entire assessment as missing and use a hot-deck procedure to impute values for these missing assessments (see Appendix D).

We drop 13 of the 122 measures because they are missing for over 80% of the assessments at both 5-days and 30-days. The remaining 109 measures are summarized in Appendix Table A2. The vast majority of the 109 measures are missing for less than 5% of the 5-day assessments. Only three measures (pain frequency, pain effect on sleep, and pain effect on activity) are missing for between 40-50% of the 5-day assessments, and 50-60% of the 30-day assessments. When a measure is missing, we fill it in with a distinct "99" value, so that all measures in each assessment have numeric values.

As shown in Appendix Table A2, most of the health measures are either binary variables that indicate whether or not a patient receives some treatment or experiences some condition or categorical variables that reflect the intensity of the patient's condition. For example, the measures of activities of daily living (ADLs) – such as walking, dressing, or hygiene – take values from 0 to 4, with 0 meaning independent with no help or staff oversight at any time, 1 meaning supervision (oversight, encouragement or cueing), 2 meaning limited assistance (resident highly involved in activity and staff provide guided maneuvering of limbs or other non-weight-bearing assistance), 3 meaning extensive assistance (resident involved in activity, staff provide weight-bearing support), and 4 meaning total dependence (full staff performance every time during entire 7-day period.) There are also some cardinal measures that reflect the number of days in the last week a condition or treatment happened, and range from 0 to 7. For example, the walking training measure reflects the number of days in the last 7

during which resident received walking training for at least 15 minutes a day. Finally, there is a measure that counts the number of venous and arterial ulcers present on a patient; the maximum number in our sample is 9.

B Characterizing Discharge Decisions

We model the SNF's decision of whether to discharge someone to the community before day 30 as a function of their current health (see equation (2)). This abstracts from the possibility that this discharge decision could depend not only on their current health, but also on their health improvement, or the length of time they have been at the SNF.

Here we provide empirical support for our modeling decision by estimating alternative discharge models. Specifically, we examine whether or not a patient is discharged to the community in the week following the 30-day assessment. For this discharge outcome, we estimate both our baseline model – in which this decision is only a function of current health – and a variety of richer models. The results suggest that our baseline model is a good approximation of the discharge decision.

Appendix Table A3 shows the results. For purposes of exploring the appropriate functional form for the discharge model, we pool data across SNFs, and ignore any SNF-level heterogeneity. The first column shows results for our baseline discharge model. The second column shows results of estimating an augmented version of our baseline discharge model, in which we allow the discharge decision to depend not only on the patient's current health level (as in our baseline model), but also on their health improvement since admission. Controlling for health improvement has almost no effect on the impact of the health level on the discharge probability, and the coefficient on health improvement is an order of magnitude smaller than that on the health level.

Additionally, we estimated a probit model of discharge to the community that allows for discharge to depend (linearly) on the length of time in the SNF. To do this, recall (see Section 2) that we measure health at admission (typically around day 5), and (for people still in the SNF) at 14 days, and at 30 days. We can therefore think of our health measures as occurring roughly at week 1 ($t = 1$), week 2 ($t = 1.33$), and week 4 ($t = 2$) in the SNF.

The results are reported in the third column of Appendix Table A3. They show similar effects of health each week on the probability of discharge, indicating that discharge probabilities as a function of health do not vary with length of time in the SNF. To see this more clearly, Appendix Figure A1 shows the predicted probability of being discharged to the community at $t = 1$, $t = 1.33$, and $t = 2$. We see these functions all lie basically on top of each other, supporting the idea that discharge to the community one week after a given assessment is predominantly a function of health at the beginning of the week, and not length of time in the SNF.

C Controlling for Selection In

In Section 3, we briefly described the discrete choice framework used to construct the control function for selection into SNFs. In this section, we provide more detail on the derivation of

this control function.[31]

Recall that utility is given by $u_{ij} = \delta_j - \tau m_{ij} + \eta_{ij}$ where δ_j is a SNF fixed effect, m_{ij} is the (log) distance patient i must travel to SNF j, and η_{ij} is the patient i's idiosyncratic preferences for SNF j. SNF value added, α_j, is given by

$$E[h_{i2}|h_{i1}, \eta_{i1}, ..., \eta_{ij}, c_i] = \alpha_j + \theta h_{i1} + g_j(\eta_{i1}, ..., \eta_{ij}), \tag{9}$$

where $g_j(\eta_{i1}, ..., \eta_{ij})$ allows period 2 health to vary flexibly with unobserved preferences (i.e., to allow for arbitrary selection patterns) and c_i is a categorical variable that indicates the SNF chosen by patient i. We assume that distance does not affect health conditional on these unobserved preferences so m_{ij} is excluded from the outcome equation above.

Following Dubin and McFadden (1984), we parameterize $g_j(\cdot)$ as a linear function of the unobserved logit shocks $\eta_{i1}, ..., \eta_{ij}$ according to

$$E[h_{i2}|h_{i1}, \eta_{i1}, ..., \eta_{ij}, c_i] = \alpha_j + \theta h_{i1} + \underbrace{\sum\nolimits_{l \in J_i} \psi_l(\eta_{il} - \mu_\eta)}_{\text{Preferences not specific to } j} + \underbrace{\varphi(\eta_{ij} - \mu_\eta)}_{\text{Selection specific to } j}, \tag{10}$$

where μ_η is the mean of the logit shocks (Euler's constant) and acts as a normalization. The term $\sum_{l \in J_i} \psi_l(\eta_{il} - \mu_\eta)$ captures preferences that have effects on the outcome that are not specific to SNF j while $\varphi(\eta_{ij} - \mu_\eta)$ captures selection that is specific to SNF j. Roy-type selection is indicated by $\varphi > 0$.

Integrating over the unobserved logit shocks $\eta_{i1}, ..., \eta_{ij}$ yields

$$E[h_{i2}|h_{i1}, \eta_{i1}, ..., \eta_{ij}, c_i] = \alpha_j + \theta h_{i1} + \sum\nolimits_{l \in J_i} \psi_l \beta_{il}(j) + \varphi \beta_{ij}(j), \tag{11}$$

where $\beta_{ik}(j) = E[\eta_{ik} - \mu_\eta | c_i = j]$ are the control functions. Letting j indicate the chosen alternative, the control functions are

$$\beta_{ik}(j) = \begin{cases} -\log \hat{p}_{ik} & k = j \\ \frac{\hat{p}_{ik}}{1-\hat{p}_{ik}} \log \hat{p}_{ik} & otherwise \end{cases}, \tag{12}$$

where $\hat{p}_{ik}$ is the predicted probability that patient i chooses SNF k based on their distance to SNFs. Since $\hat{p}_{ik} < 1$, $\log \hat{p}_{ik} < 0$, meaning that the control function takes a positive value when $k = j$ and a negative value otherwise.

We will now derive the resulting expressions for these control functions. Again, letting j denote the chosen alternative, we know from rearranging the utility function that

$$E[\eta_{ij} - \mu_\eta | c_i = j] = E[u_{ij} | c_i = j] - \delta_j + \tau m_{ij} - \mu_\eta. \tag{13}$$

From Small and Rosen (1981), we know

$$E[u_{ij} | c_i = j] = \log \left[\sum\nolimits_{l \in J_i} \exp(\delta_l - \tau m_{il}) \right] + \mu_\eta. \tag{14}$$

[31] To simplify the notation, we use δ_j and δ_k below instead of explicitly denoting the conditioning on observable patient characteristics used in the actual estimation, $\delta_j(x_i)$ and $\delta_k(x_i)$.

Substitution yields,

$$\begin{aligned} E[\eta_{ij} - \mu_\eta | c_i = j] &= \log\left[\sum\nolimits_{l \in J_i} \exp(\delta_l - \tau m_{il})\right] + \mu_\eta - \delta_j + \tau m_{ij} - \mu_\eta \\ &= \log\left[\sum\nolimits_{l \in J_i} \exp(\delta_l - \tau m_{il})\right] - \log\left[\exp(\delta_j - \tau m_{ij})\right] \\ &= -\log\left[\frac{\exp(\delta_j - \tau m_{ij})}{\sum_{l \in J_i} \exp(\delta_l - \tau m_{il})}\right] = -\log \hat{p}_{ij}. \end{aligned} \tag{15}$$

For a non-chosen alternative $k \neq j$ we have

$$E[\eta_{ik} - \mu_\eta | c_i = j] = E[u_{ik} | c_i = j] - \delta_k + \tau m_{ik} - \mu_\eta. \tag{16}$$

By proprieties of conditional expectation, we know

$$E[u_{ik}] = \Pr(c_i = k) E[u_{ik} | c_i = k] + \Pr(c_i \neq k) E[u_{ik} | c_i \neq k]. \tag{17}$$

Using the result from Small and Rosen (1981), as before, and substituting the above expression yields

$$\delta_k - \tau m_{ik} + \mu_\eta = \Pr(c_i = k)\left[\sum\nolimits_{l \in J_i} \exp(\delta_l - \tau m_{il}) + \mu_\eta\right] + \Pr(c_i \neq k) E[u_{ik} | c_i \neq k], \tag{18}$$

which implies

$$E[u_{ik} | c_i \neq k] \;=\; \frac{1}{1 - \Pr(c_i = k)}\left[\delta_k - \tau m_{ik} + \mu_\eta - \Pr(c_i = k)\left[\sum\nolimits_{l \in J_i} \exp(\delta_l - \tau m_{il}) + \mu_\eta\right]\right].$$

Substitution then yields

$$\begin{aligned} E[\eta_{ik} - \mu_\eta | c_i = j] =& \frac{1}{1 - \Pr(c_i = k)}\left[\delta_k - \tau m_{ik} + \mu_\eta - \Pr(c_i = k)\left[\sum\nolimits_{l \in J_i} \exp(\delta_l - \tau m_{il}) + \mu_\eta\right]\right] \\ & - \delta_k + \tau m_{ik} - \mu_\eta \\ =& \frac{1}{1 - \Pr(c_i = k)}\Big[\delta_k - \tau m_{ik} + \mu_\eta - (1 - \Pr(c_i = k))\left[\delta_k - \tau m_{ik} + \mu_\eta\right] \\ & - \Pr(c_i = k)\left[\sum\nolimits_{l \in J_i} \exp(\delta_l - \tau m_{il}) + \mu_\eta\right] - \delta_k + \tau m_{ik} - \mu_\eta\Big] \\ =& \frac{\Pr(c_i = k)}{1 - \Pr(c_i = k)}\left[\delta_k - \tau m_{ik} - \left[\log\left(\sum\nolimits_{l \in J_i} \exp\left(\delta_l - \tau m_{il}\right)\right)\right]\right] \\ =& \frac{\Pr(c_i = k)}{1 - \Pr(c_i = k)}\left[\log\left(\exp\left(\delta_k - \tau m_{ik}\right)\right) - \left[\log\left(\sum\nolimits_{l \in J_i} \exp\left(\delta_l - \tau m_{il}\right)\right)\right]\right] \\ =& \frac{\Pr(c_i = k)}{1 - \Pr(c_i = k)}\left[\log\left(\frac{\exp\left(\delta_k - \tau m_{ik}\right)}{\sum_{l \in J_i} \exp\left(\delta_l - \tau m_{il}\right)}\right)\right] \\ =& \frac{\hat{p}_{ik}}{1 - \hat{p}_{ik}} \log \hat{p}_{ik}. \end{aligned} \tag{19}$$

D Hot-Deck Imputation Procedure

Approximately 20% of patients, who are still in the SNF 30 days after admission, do not receive a 30-day health assessment. Moreover, assessments are not missing at random: patients with missing 30-day assessments are more likely to be discharged within 7 days of when their assessment would have been (Appendix Table A5). We use a "hot-deck" procedure to impute the 30-day health index for patients with missing information.

The hot-deck procedure takes patients with missing 30-day assessments and finds a set of "donor" patients who share observable information with the missing patients but are not missing their 30-day assessments. The missing health index is imputed by random selection from these donors. We require donor patients to be from the same SNF as the recipient patient, and from this pool select the donors by matching to the recipient based on the health index at admission, the health index at 14 days (if available), and discharge direction (upstream, downstream, or stay) in the week following what should have been the 30-day assessment. To facilitate matching, we discretize the health index into 16 bins of width 0.025, ranging from 0 to 0.4. We observe initial health assessment data for all patients (h_1); 14-day assessment for 68% of patients ($h_{1.33}$), and discharge destination with 7 days of the 30-day assessment (if any) for all patients (d_2). We have two different imputation procedures depending on whether or not we observe 14-day health (case 1 or case 2). Appendix Table A6 provides more details on the procedure.

Appendix Table A7 shows summary statistics for patients who are still in a SNF at the time of the 30-day assessment for the two cases. We successfully impute health at 30 days for 99.7% of patients with missing information. Consistent with the differential discharge patterns documented in Appendix Table A5, patients with missing health at 30 days are on average healthier than those with non-missing health. Imputed health varies in the expected way based on the patients location as 7 days after their (imputed) 30 day assessment, with the highest values for those discharged downstream, lowest values for those discharged upstream, and middle values for those who remain in the SNF.

E Estimation Details

E.1 Deriving the Likelihood

The health process and upstream and downstream discharge equations are specified in Section 3. In our setting, patients can be partitioned into three groups:

1. Discharged downstream before period 2 ($d_i^D = 1$)
2. Discharged upstream before period 2 ($d_i^U = 1$)
3. Still in the SNF by period 2 ($d_i^D = 0 \cap d_i^U = 0$).

Given this partition, the likelihood for a market is given by

$$L = \Pi_i \underbrace{\left[\Pr(d_i^D = 1)\right]^{d_i^D}}_{\text{downstream}} \underbrace{\left[\Pr(d_i^U = 1)\right]^{d_i^U}}_{\text{upstream}} \underbrace{\left[\Pr(d_i^D = 0 \cap d_i^U = 0 | h_{i2}) f(h_{i2})\right]^{(1-d_i^U)(1-d_i^D)}}_{\text{still in SNF}}. \quad (20)$$

where $f(h_{i2})$ be the probability density function of h_{i2}.

We will derive an explicit formula for L by considering each partition separately. Additionally, recall our definitions of the health production (equation (1)) and the downstream and upstream discharge equations (equations (2) and (3)) provided in Section 3. First, consider patients with $d_i^D = 1$. We know

$$\begin{aligned}
\Pr(d_i^D = 1) =& \Pr(h_{i2} \geq \lambda_j + \nu_i) \\
=& \Pr(\alpha_j + \theta h_{i1} + \sum\nolimits_{k \in J_i} \phi_k \beta_{ik} + \varphi \beta_{ij} + \epsilon_i \geq \lambda_j + \nu_i) \\
=& \Pr(\epsilon_i - \nu_i \geq \lambda_j - \alpha_j - \theta h_{i1} - \sum\nolimits_{k \in J_i} \phi_k \beta_{ik} - \varphi \beta_{ij}) \\
=& 1 - \Pr(\epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta h_{i1} - \sum\nolimits_{k \in J_i} \phi_k \beta_{ik} - \varphi \beta_{ij})) \\
=& 1 - \Phi\left(\frac{\lambda_j - \alpha_j - \theta h_{i1} - \sum_{k \in J_i} \phi_k \beta_{ik} - \varphi \beta_{ij}}{\sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}} \right) \\
=& \Phi\left(\frac{\alpha_j + \theta h_{i1} + \sum_{k \in J_i} \phi_k \beta_{ik} + \varphi \beta_{ij} - \lambda_j}{\sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}} \right).
\end{aligned} \tag{21}$$

where $\Phi(\cdot)$ is the standard normal cumulative density function. This is the case because ϵ_i and ν_i are independent normal random variables, meaning that $\epsilon_i - \nu_i \sim N\left(0, \sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}\right)$.

Next, consider patients with $d_i^U = 1$. Notice that we can rewrite the portion of the likelihood for these patients as follows

$$\begin{aligned}
\Pr(d_i^U = 1) =& \Pr(d_i^U = 1 | d_i^D = 0) \Pr(d_i^D = 0) \\
=& \Pr(d_i^U = 1 | d_i^D = 0) \Phi\left(\frac{\lambda_j - \alpha_j - \theta h_{i1} - \sum_{k \in J_i} \phi_k \beta_{ik} - \varphi \beta_{ij}}{\sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}} \right).
\end{aligned} \tag{22}$$

Now consider $\Pr(d_i^U = 1 | d_i^D = 0)$. Since we model $\Pr(d_i^U = 1 | d_i^D = 0, h_{i2}) = \Phi(\gamma_0 + \gamma_1 h_{i2})$ we have that

$$\Pr(d_i^U = 1 | d_i^D = 0) = \int \tilde{\phi}(\epsilon_i) \Phi(\gamma_0 + \gamma_1(\alpha_j + \theta h_{i1} + \sum\nolimits_{k \in J_i} \phi_k \beta_{ik} + \varphi \beta_{ij} + \epsilon_i)) d\epsilon_i \tag{23}$$

where $\tilde{\phi}(\cdot)$ is the probability density function of ϵ_i conditional on $d_i^D = 0$ (see Section E.3 for a derivation of $\tilde{\phi}(\cdot)$). Together, this yields

$$\begin{aligned}
\Pr(d_i^U = 1) = & \int \tilde{\phi}(\epsilon_i) \Phi(\gamma_0 + \gamma_1(\alpha_j + \theta h_{i1} + \sum\nolimits_{k \in J_i} \phi_k \beta_{ik} + \varphi \beta_{ij} + \epsilon_i)) d\epsilon_i \\
& \cdot \Phi\left(\frac{\lambda_j - \alpha_j - \theta h_{i1} - \sum_{k \in J_i} \phi_k \beta_{ik} - \varphi \beta_{ij}}{\sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}} \right).
\end{aligned} \tag{24}$$

Lastly, consider patients who remain in the SNF in period 2, meaning that $d_i^D = 0 \cap d_i^U = 0$.

We know

$$\Pr(d_i^D = 0 \cap d_i^U = 0|h_{i2}) f(h_{i2}) = \Pr(d_i^D = 0|h_{i2}) \Pr(d_i^U = 0|d_i^D = 0, h_{i2}) f(h_{i2}). \tag{25}$$

It is trivial that $f(h_{i2}) = \phi\left(\frac{h_{i2} - \alpha_j - \theta h_{i1} - \sum_{k \in J_i} \phi_k \beta_{ik} - \varphi \beta_{ij}}{\sigma_\epsilon}\right)$ where $\phi(\cdot)$ is the standard normal probability density function. Similarly, by assumption we know that $\Pr(d_i^U = 0|d_i^D = 0, h_{i2}) = 1 - \Phi(\gamma_0 + \gamma_1 h_{i2})$. Now consider $\Pr(d_i^D = 0|h_{i2})$. We know that

$$\begin{aligned}
\Pr(d_i^D = 0|h_{i2}) &= \Pr(h_{i2} < \lambda_j + \nu_i|h_{i2}) \\
&= \Pr(\epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta h_{i1} - \sum\nolimits_{k \in J_i} \phi_k \beta_{ik} - \varphi \beta_{ij}|h_{i2}) \\
&= \Pr(\epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta h_{i1} - \sum\nolimits_{k \in J_i} \phi_k \beta_{ik} - \varphi \beta_{ij}|\epsilon_i).
\end{aligned} \tag{26}$$

In order to derive an expression for $\Pr(\epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta h_{i1} - \sum_{k \in J_i} \phi_k \beta_{ik} - \varphi \beta_{ij}|\epsilon_i)$ we must first understand the joint distribution of ϵ_i and $\epsilon_i - \nu_i$. Since these two random variables can be expressed as a linear combination of two independent normal random variables (namely, ϵ_i and ν_i), we know they are jointly normally distributed. Specifically,

$$\begin{pmatrix} \epsilon_i \\ \epsilon_i - \nu_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 & \sigma_\epsilon^2 \\ \sigma_\epsilon^2 & \sigma_\epsilon^2 + \sigma_\nu^2 \end{pmatrix}\right). \tag{27}$$

The conditional distribution of a bivariate normal is also normal. Specifically,

$$\epsilon_i - \nu_i|\epsilon_i \sim N(\epsilon_i, \sigma_\nu). \tag{28}$$

Hence, we have

$$\begin{aligned}
\Pr(d_i^D = 0|h_{i2}) =& \Phi\left(\frac{\lambda_j - \alpha_j - \theta h_{i1} - \sum_{k \in J_i} \phi_k \beta_{ik} - \varphi \beta_{ij} - \epsilon_i}{\sigma_\nu}\right) \\
=& \Phi\left(\frac{\lambda_j - \alpha_j - \theta h_{i1} - \sum_{k \in J_i} \phi_k \beta_{ik} - \varphi \beta_{ij} - h_{i2} + \alpha_j + \theta h_{i1} + \sum_{k \in J_i} \phi_k \beta_{ik} + \varphi \beta_{ij}}{\sigma_\nu}\right) \\
=& \Phi\left(\frac{\lambda_j - h_{i2}}{\sigma_\nu}\right).
\end{aligned} \tag{29}$$

Together this yields

$$\begin{aligned}
&\Pr(d_i^D = 0 \cap d_i^U = 0|h_{i2}) f(h_{i2}) = \\
&\Phi\left(\frac{\lambda_j - h_{i2}}{\sigma_\nu}\right)(1 - \Phi(\gamma_0 + \gamma_1 h_{i2}))\phi\left(\frac{h_{i2} - \alpha_j - \theta h_{i1} - \sum_{k \in J_i} \phi_k \beta_{ik} - \varphi \beta_{ij}}{\sigma_\epsilon}\right).
\end{aligned} \tag{30}$$

Combining the three cases we considered above, the likelihood for any market is given by

$$L = \Pi_i \underbrace{\left[\Phi\left(\frac{\alpha_j + \alpha_h h_1 - \lambda_j}{\sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}}\right)\right]^{d_i^D}}_{\text{downstream}}$$

$$\cdot \underbrace{\left[\int \tilde{\phi}(\epsilon_i)\Phi(\gamma_0 + \gamma_1(\alpha_j + \theta h_{i1} + \sum\nolimits_{k\in J_i} \phi_k \beta_{ik} + \varphi\beta_{ij} + \epsilon_i))d\epsilon_i \Phi\left(\frac{\lambda_j - \alpha_j - \theta h_{i1} - \sum_{k\in J_i}\phi_k\beta_{ik} - \varphi\beta_{ij}}{\sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}}\right)\right]^{d_i^U}}_{\text{upstream}}$$

$$\cdot \underbrace{\left[\Phi\left(\frac{\lambda_j - h_{i2}}{\sigma_\nu}\right)(1 - \Phi(\gamma_0 + \gamma_1 h_{i2}))\,\phi\left(\frac{h_{i2} - \alpha_j - \theta h_{i1} - \sum_{k\in J_i}\phi_k\beta_{ik} - \varphi\beta_{ij}}{\sigma_\epsilon}\right)\right]^{(1-d_i^U)(1-d_i^D)}}_{\text{still in SNF}}.$$

For estimation purposes, we minimize the negative log-likelihood.

E.2 Estimation

For computational reasons, we estimate the model in two steps by partitioning parameters into two groups: "national" parameters, $(\theta, \varphi, \sigma_\epsilon, \sigma_\nu)$, and market-level (or SNF-level) parameters, $(\alpha_j, \lambda_j, \phi_j, \gamma_0, \gamma_1)$. This procedure takes advantage of the fact that the market-level parameters only show up in the likelihood for the specific market, meaning that conditional on the national parameters (which show up in the likelihood for all markets), we can maximize the likelihood market-by-market which is computationally attractive since this is easily parallelizable.

The estimation process is as follows.

1. Specify grids for the "national" parameters $(\theta, \varphi, \sigma_\epsilon, \sigma_\nu)$

For all combinations of points on these grids,

2. Maximize the likelihood with respect to $(\alpha_j, \lambda_j, \phi_j, \gamma_0, \gamma_1)$ within each market holding $(\theta, \varphi, \sigma_\epsilon, \sigma_\nu)$ fixed

After searching over all grid points,

3. Maximize the sum of the market-level likelihoods over the grid of "national" parameters.

Steps 1 through 3 are repeated as the grids specified in step 1 are fine tuned. Step 3 is accomplished by using a gradient based nonlinear solver within markets. The final grids for each parameter are

- $\theta \in \{0.945, 0.9562, 0.9675, 0.9788, 0.99, 1.1\}$
- $\sigma_\epsilon \in \{0.05, 0.0563, 0.0625, 0.0688, 0.0750\}$
- $\sigma_\nu \in \{0.1, 0.15, 0.2, 0.25, 0.3\}$
- $\varphi \in \{-0.003, -0.0015, 0, 0.0015, 0.003\}$

Another problem to tackle is estimating the following integral which enters the likelihood for patients who are discharged upstream, $\int \tilde{\phi}(\epsilon_i)\Phi(\gamma_0+\gamma_1(\alpha_j+\theta h_{i1}+\sum_{k\in J_i}\phi_k\beta_{ik}+\varphi\beta_{ij}+\epsilon_i))d\epsilon_i$. Due to runtime concerns, we estimate this integral using rectangular quadrature. We explore sensitivity to this method of estimating the integral by focusing on a large market and estimating the error that is introduced when estimating the integral via rectangular quadrature relative to Monte Carlo integration. We tested a variety of methods for choosing the location of nodes, the number of nodes, and the boundary of the integral (since the domain of integration is unbounded). Equally spaced nodes performed best and the percent error was not sensitive to the number of nodes. We tested limiting the boundary to within $4\sigma_\epsilon$, $5\sigma_\epsilon$, and $6\sigma_\epsilon$ of the mean. Results were essentially invariant to the choice of boundary this far out so in practice we limit the domain of integration to be within $4\sigma_\epsilon$ of the mean. The percent error was less than 0.01% across all of the tests with equally spaced nodes.

E.3 Deriving the Marginal Density of ϵ_i Conditional on $d_i^D = 0$

In the likelihood derived in Appendix E.1, we saw that the probability density function of ϵ_i conditional on $d_i^D = 0$ shows up in the likelihood for patients discharged upstream. This is a result of modeling the upstream decision conditional on $d_i^D = 0$. For estimation purposes, we need an explicit formula for $\tilde{\phi}(\epsilon_i)$ since we estimate the integral by numerical quadrature. Specifically, we are interested in the following probability density function

$$\begin{aligned}\tilde{\phi}(x) =& \frac{d}{dx}\left[\Pr\left(\epsilon_i \leq x | d_i^D = 0\right)\right] \\ =& \frac{d}{dx}\left[\Pr\left(\epsilon_i \leq x | h_{i2} < \lambda_j + \nu_i\right)\right] \\ =& \frac{d}{dx}\left[\Pr\left(\epsilon_i \leq x | \epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta h_{i1} - \sum\nolimits_{k\in J_i}\phi_k\beta_{ik} - \varphi\beta_{ij}\right)\right].\end{aligned}$$

To understand the functional form of $\frac{d}{dx}\left[\Pr\left(\epsilon_i \leq x | \epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta h_{i1} - \sum_{k\in J_i}\phi_k\beta_{ik} - \varphi\beta_{ij}\right)\right]$ we must first revisit the joint distribution of ϵ_i and $\epsilon_i - \nu_i$. Recall from Section 1 that

$$\begin{pmatrix} \epsilon_i \\ \epsilon_i - \nu_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 & \sigma_\epsilon^2 \\ \sigma_\epsilon^2 & \sigma_\epsilon^2 + \sigma_\nu^2 \end{pmatrix}\right).$$

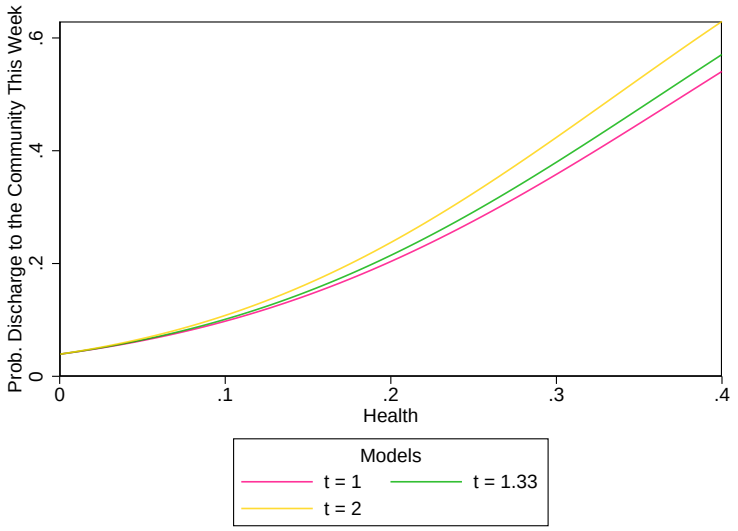
Thus, the joint distribution of $\epsilon_i, \epsilon_i - \nu_i | \epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta h_{i1} - \sum_{k\in J_i}\phi_k\beta_{ik} - \varphi\beta_{ij}$ follows a bivariate truncated normal. Hence, $\frac{d}{dx}\left[\Pr\left(\epsilon_i \leq x | \epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta h_{i1} - \sum_{k\in J_i}\phi_k\beta_{ik} - \varphi\beta_{ij}\right)\right]$ is the marginal density of a bivariate truncated normal cumuilative density function. Thus, we have that

$$\tilde{\phi}(x) = \frac{1}{C}\phi\left(\frac{x}{\sigma_\epsilon}\right)\Phi\left(\frac{\lambda_j - \alpha_j - \theta h_{i1} - \sum_{k\in J_i}\phi_k\beta_{ik} - \varphi\beta_{ij} - x}{\sigma_\nu}\right),$$

where C is the total probability in the truncated distribution of $(\epsilon_i, \epsilon_i - \nu_i)$ and is given by

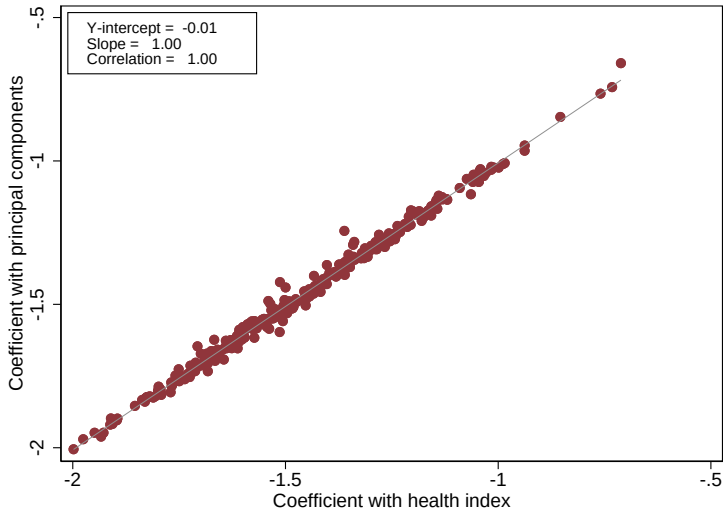
$$\mathrm{C} = \int_{-\infty}^{\infty}\int_{-\infty}^{\lambda_j - \alpha_j - \theta h_{i1} - \sum_{k\in J_i}\phi_k\beta_{ik} - \varphi\beta_{ij}} \exp\left\{-\frac{\sigma_\epsilon^2+\sigma_\nu^2}{2\sigma_\nu^2}\left[\left(\frac{x}{\sigma_\epsilon}\right)^2 - 2\rho\left(\frac{x}{\sigma_\epsilon}\right)\left(\frac{y}{\sqrt{\sigma_\epsilon^2+\sigma_\nu^2}}\right) - \left(\frac{y}{\sqrt{\sigma_\epsilon^2+\sigma_\nu^2}}\right)^2\right]\right\} dydx.$$

Appendix Figure A1: Weekly Discharge Rule



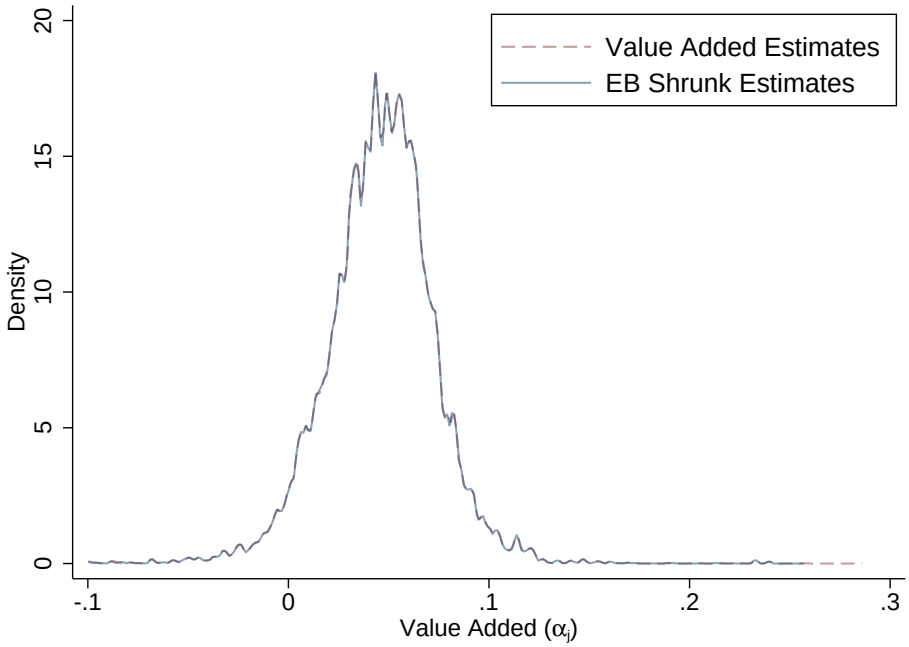
Notes: Figure plots the results from estimating an assessment-level probit discharge model that allows for the discharge decision to depend (linearly) on time in the SNF. Specifically, we estimate a probit of the probability of discharge in the week following the assessment at $t = 1$ (the 5-day assessment), $t = 1.33$ (the 14-day assessment) and $t = 2$ (the 30 day assessment). Specifically, we estimate $d_{it}^D = 1 \iff h_{id}^* = \bar{\rho} h_{it} + \Delta\rho t h_{it} + \xi_i > \lambda$ where $d_{it}^D = 1$ if patient i is discharged downstream at time t, h_{id}^* is patient i's latent health at discharge and h_{it} is health at time t. We estimate this model on 13,231,304 patient-assessment observations. This is the same model shown in Column (3) of Appendix Table A3.

Appendix Figure A2: Correlating Coefficient on Log-Distance



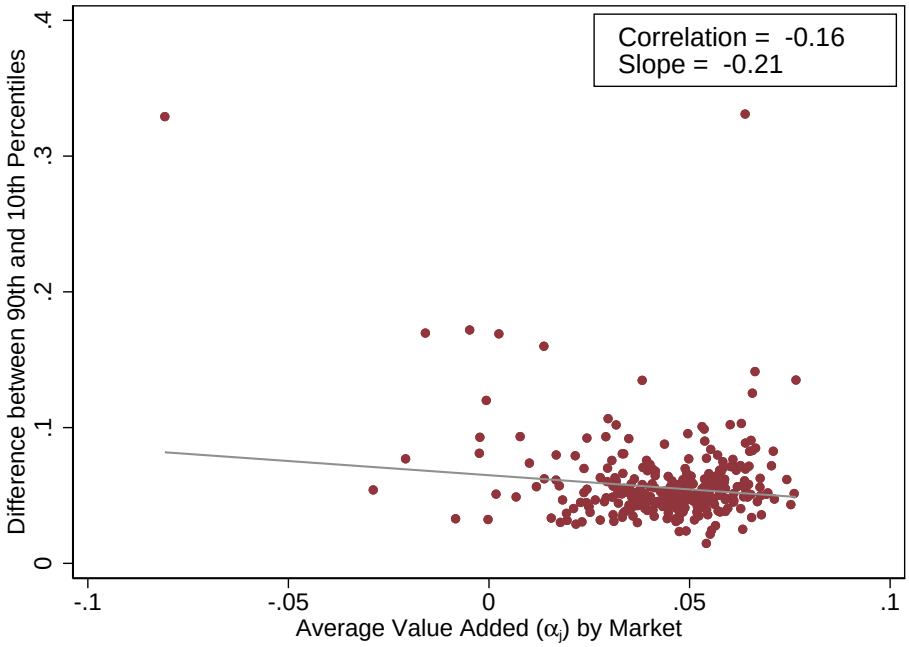
Notes: Figure correlates the coefficient on log-distance from the demand model in equation (4), which we estimate (market-by-market) using two specifications with different sets of observable shifters of δ_j. The x-axis shows the coefficient for each market from our main specification which only uses the health index at admission. The y-axis shows the same coefficient for each market from an alternative specification that also uses (in addition to the health index at admission) the first ten principal components of the following observables: demographics (above median age, female, married, Hispanic, Black, white, dual eligible), health measures at admission (walking, indwelling, falls, shortness of breath, depressed, delirium, vomiting, fever, dehydration, weight loss, and long-form ADL), and the top 20 CCS codes. Coefficients are shown for all 306 markets.

Appendix Figure A3: Empirical Bayes Shrinkage



Notes: Figure shows a kernel density of value added as well as the analogous density after performing empirical Bayes shrinkage to account for estimation error in the value-added estimates. The correlation between the value added estimates and their shrunken counterparts (weighted by stays) is 0.96. The standard errors (SEs) of value added are the critical inputs to the empirical Bayes procedure and are used to estimate the underlying variance of the distribution of value added within markets. The average SE is 0.0002 with a standard deviation of 0.003 indicating that our value added estimates are precisely estimated. We perform shrinkage market-by-market, shrinking estimated value added to market specific means. To perform shrinkage we follow the procedure described in Morris (1983). This procedure results in nearly the same distribution of value added with a mean of 0.045 and a standard deviation of 0.033.

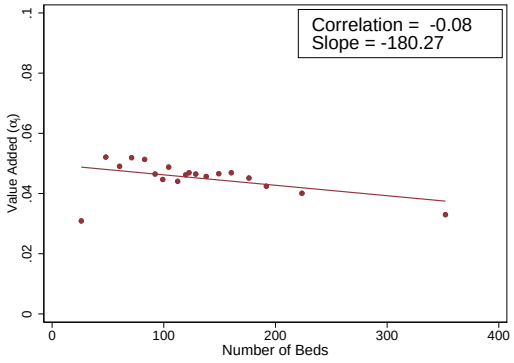
Appendix Figure A4: Correlating Difference between 90th and 10th Percentile VA Within Market with Average VA in Market



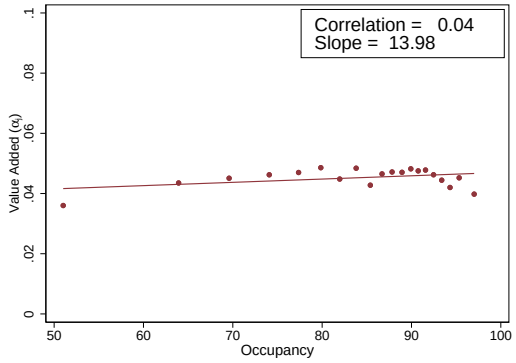
Notes: Figure shows a scatter plot for each of the 306 markets of the difference in value added between the 90th and 10th percentile SNF in the market weighted by patients and the average value of SNFs in the market weighted by patients. The correlation is weighted by the number of patients in the market. The line is a linear fit weighted by the number of patients in each market.

Appendix Figure A5: Correlates of SNF Value Added

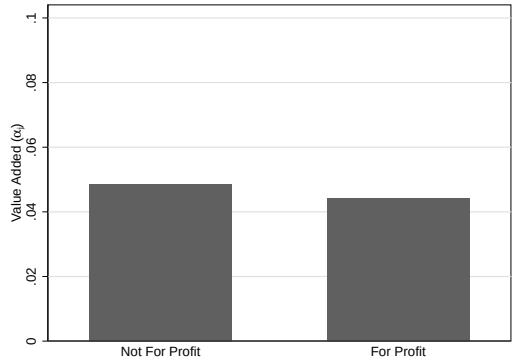
(A) Number of Beds



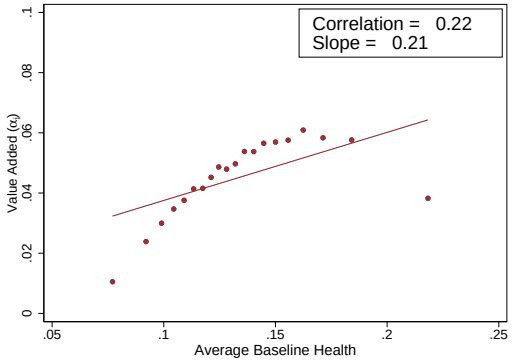
(B) SNF Occupancy



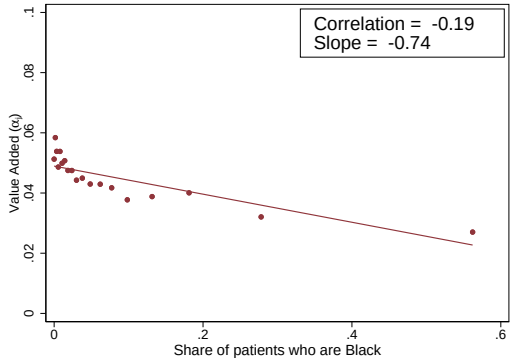
(C) For Profit Status



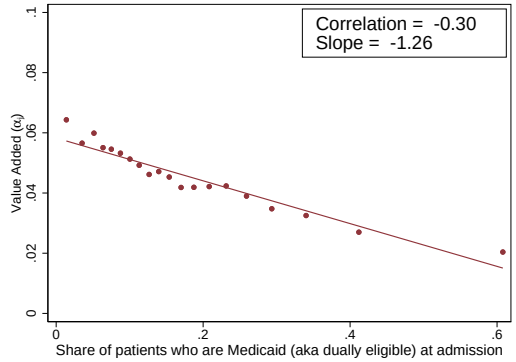
(D) Health Index at Admission



(E) Black Patient Share



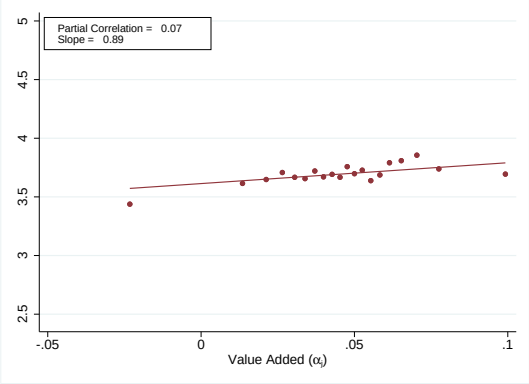
(F) Dually-eligible for Medicaid at Admission Share



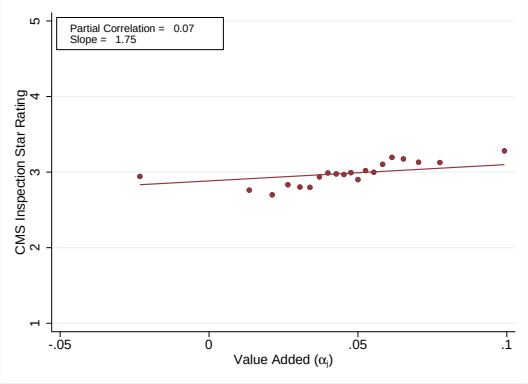
Notes: Figures show binscatters of value added on SNF characteristics using 20 bins for continuous SNF characteristics (discrete characteristics are shown as box plots). Number of beds, SNF occupancy, and for-profit status come from OSCAR data. The health index at admission, black patient share, and dually-eligible for medicaid at admission share are calculated using all 6,246,686 patient-stays in our full sample of Medicare patient-stays.

Appendix Figure A6: Correlations of Value Added Estimates with Alternative Measures

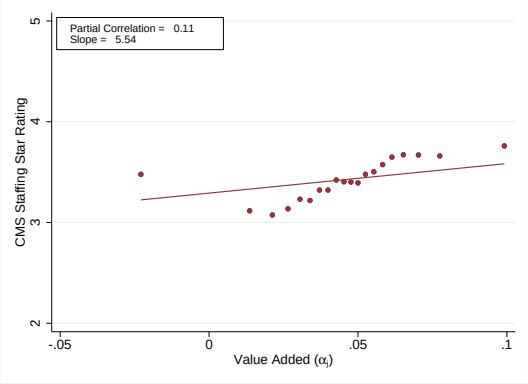
(A) Quality Ratings



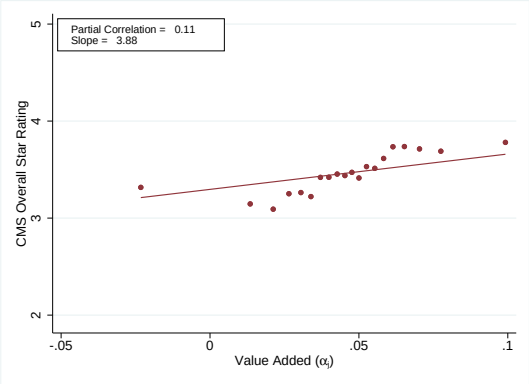
(B) Inspection Ratings



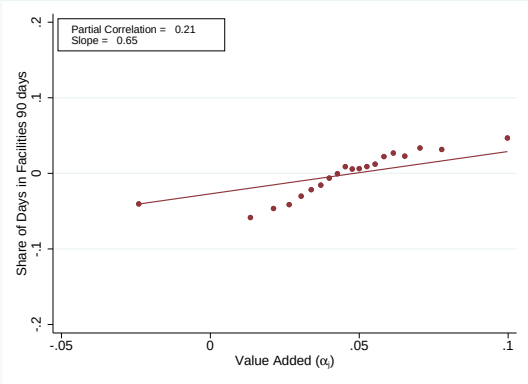
(C) Staffing Ratings



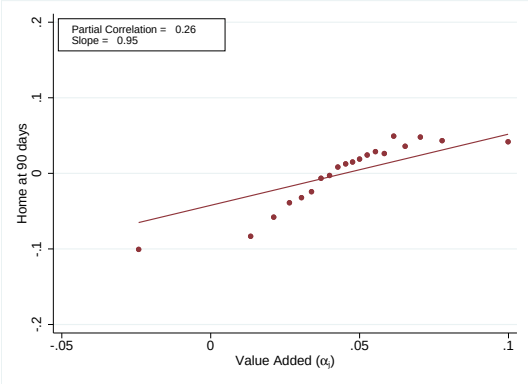
(D) Overall Ratings



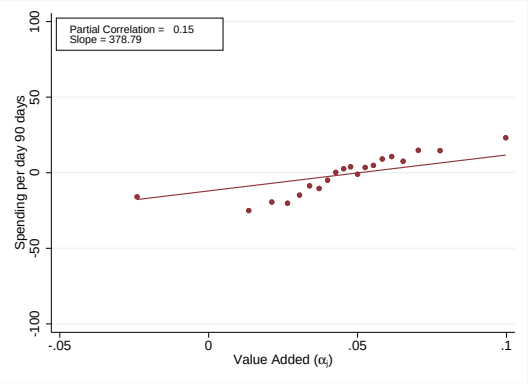
(E) Share of Days in Facilities 90 days



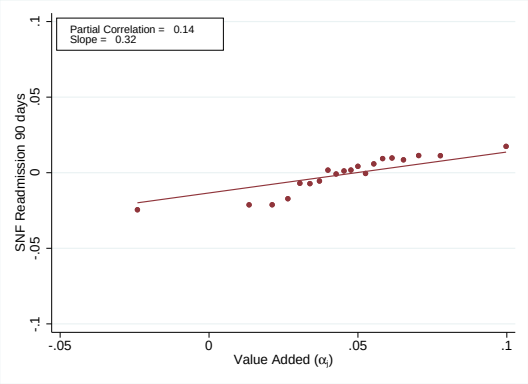
(F) Home at 90 days



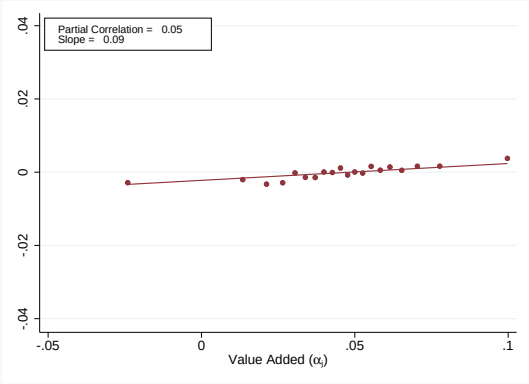
(G) Spending per day 90 days



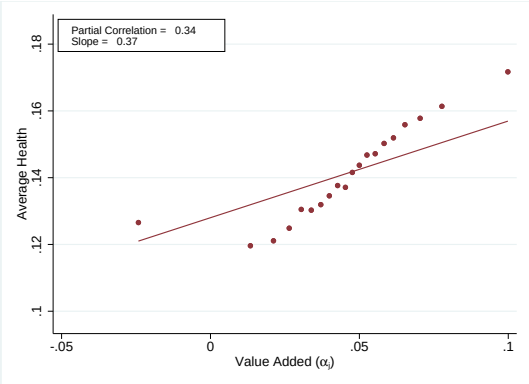
(H) SNF Readmission 90 days



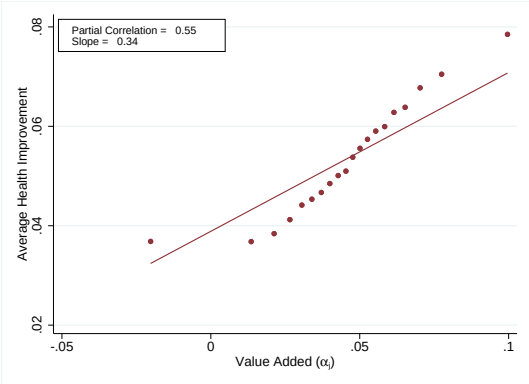
(I) Hospital Readmission 90 days



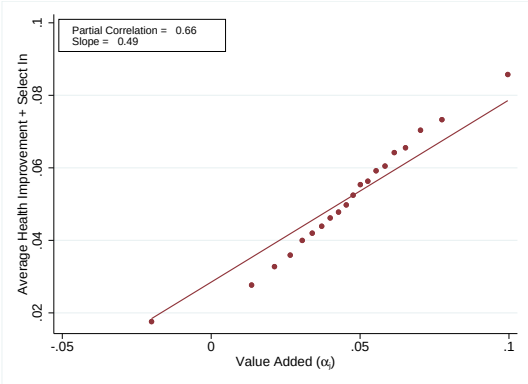
(J) Average Health



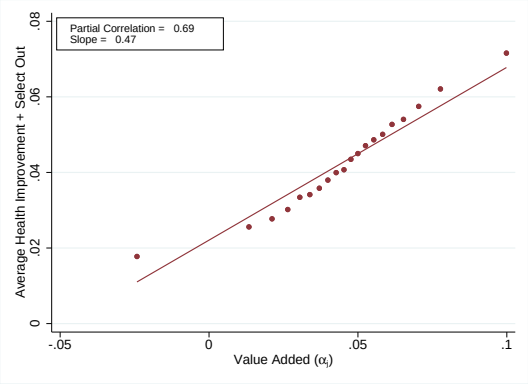
(K) Average Health Improvement



(L) Average Health Improvement + Select In



(M) Average Health Improvement + Select Out



Notes: Figures show binscatters of alternative measures of SNF quality on SNF value added (α_j) controlling for market fixed effects. The alternative measures are (A) CMS Quality Star Ratings, (B) CMS Inspection Star Ratings, (C) CMS Staffing Star Ratings, (D) CMS Overall Star Ratings, (E) share of days spent in facilities within 90 days of SNF admission, (F) the probability of being home 90 days after SNF admission, (G) spending per day within 90 days of SNF admission, (H) the probability of readmission to a SNF within 90 days of SNF admission, (I) the probability of hospital readmission within 90 days of SNF admission, (J) average health value added, (K) average health improvement value added, (L) average health improvement + select in value added, and (M) average health improvement + select out value added. Binscatters and correlations are weighted by the number of patients in the SNF. The y-axis of each binscatter starts from the mean of the variable minus 1.5 times the standard deviation to the mean plus 1.5 times the standard deviation.

Appendix Table A1: Sample Restrictions

	Stays	SNFs
Full sample	24,078,049	16,355
Medicare and >= 65	10,726,854	16,307
First stays	7,213,728	15,730
CCS codes	6,598,040	15,680
Has 5-day health assessment	6,327,400	15,623
SNF has at least 50 episodes (analysis sample)	6,246,686	13,996

Notes: Table shows the sequential sample restrictions applied to go from the full sample (top row) to our analysis sample (bottom row); at each step it shows the remaining number of patient-stays and distinct SNFs. The "Full sample" is all Medicare and Medicaid patient stays in in MDS 3.0 between October 01, 2011 to September 30, 2016. The next row ("Medicare and $\geq$ 65") restricts the sample to patients who, on admission, have Traditional Medicare and are at least 65 years old. The next row ("First stays") restricts to the first stay of each episode. The "CCS codes" row restricts to patients with a non-missing Clinical Classification Software (CCS) code from an acute hospital stay prior to SNF admission. The "Has 5-day assessment" row restricts to patient stays with a 5-day assessment (as defined in Appendix A). The final row excludes SNFs with fewer than 50 episodes.

Appendix Table A2: List of Health Measures Used

Measure	ResDAC Variable Name(s)	Range	% Missing at 5-day	% Missing at 30-day
Physical				
Has fallen	J1800_FALL_LAST_ASMT_CD	(0-1)	0.11	0.03
Vomiting	J1550B_VMTG_CD	(0-1)	0.06	0.01
Fever	J1550A_FVR_CD	(0-1)	0.06	0.01
Dehydration	J1550C_DHYDRT_CD	(0-1)	0.06	0.01
Life prognosis < 6 months	J1400_LIFE_PRGNS_CD	(0-1)	0.27	0.09
Weight loss	K0300_WT_LOSS_CD	(0-2)	1.51	0.26
Urinary continence	H0300_URNRY_CNTNC_CD	(0-4)	0.19	0.04
Bowel continence	H0400_BWL_CNTNC_CD	(0-4)	0.31	0.03
Hearing	B0200_HEARG_CD	(0-3)	0.87	0.21
Vision	B1000_VSN_CD	(0-4)	1.50	0.46
Seated to standing balance	G0300A_BAL_SEAT_STNDG_CD	(0-3)	0.81	0.30
Walking balance	G0300B_BAL_WLKG_CD	(0-3)	0.99	0.44
Turning balance	G0300C_BAL_TRNG_ARND_CD	(0-3)	1.82	1.16
Moving on/off toilet balance	G0300D_BAL_TOILT_CD	(0-3)	0.93	0.38
Surface-to-surface balance	G0300E_BAL_SRFC_TRNSFR_CD	(0-3)	0.80	0.29
Upper extremity range	G0400A_UPR_XTRMTY_MTN_CD	(0-2)	0.74	0.17
Lower extremity range	G0400B_LWR_XTRMTY_MTN_CD	(0-2)	0.74	0.17
Has unhealed pressure ulcer	M0210_STG_1_HGHR_ULCR_CD	(0-1)	0.14	0.04
# of venous/arterial ulcers	M1030_ARTRL_ULCR_NUM	(0-9)	0.16	0.08
Foot infection	M1040A_FT_INFCTN_CD	(0-1)	0.07	0.02
Diabetic foot ulcer	M1040B_DBTC_FT_ULCR_CD	(0-1)	0.07	0.02
Open foot lesions	M1040C_OTHR_LSN_FT_CD	(0-1)	0.07	0.02
Other open lesions	M1040D_OPEN_LSN_CD	(0-1)	0.07	0.02
Surgical wounds	M1040E_SRGCL_WND_CD	(0-1)	0.06	0.02
Burns	M1040F_BRN_CD	(0-1)	0.06	0.01
Skin and ulcer treatments	M1200Z_NO_SKIN_TRMNT_CD	(0-1)	0.04	0.01
Tracheostomy care	O0100E2_TRCHOSTMY_POST_CD	(0-1)	0.22	0.54
Physical behavior symptoms	E0200A_PHYS_BHVRL_CD	(0-3)	1.21	0.21
Internal bleeding	J1550D_INTRNL_BLEDG_CD	(0-1)	0.06	0.01
Loss of liquids/solids in mouth	K0100A_LOSS_MOUTH_EATG_CD	(0-1)	0.36	0.11
Holding food in mouth	K0100B_HLD_FOOD_MOUTH_CD	(0-1)	0.36	0.11
Coughing/choking	K0100C_CHOK_DRNG_MEAL_CD	(0-1)	0.36	0.11
No swallowing issue	K0100Z_NO_SWLWG_CD	(0-1)	0.36	0.11
Passive range of motion program	O0500A_PSV_ROM_NUM	(0-7)	0.19	0.09
Active range of motion program	O0500B_ACTV_ROM_NUM	(0-7)	0.19	0.09
Bed mobility training	O0500D_BED_MBLTY_TRNG_NUM	(0-7)	0.19	0.09
Transfer training	O0500E_TRNSFR_TRNG_NUM	(0-7)	0.19	0.09
Walking training	O0500F_WLKG_TRNG_NUM	(0-7)	0.19	0.09
Dressing training	O0500G_DRSG_TRNG_NUM	(0-7)	0.19	0.09
Eating training	O0500H_EATG_TRNG_NUM	(0-7)	0.19	0.09
Amputation training	O0500I_AMPUTTN_TRNG_NUM	(0-7)	0.20	0.09
Short-breathness[a]	J1100A_SOB_EXRTN_CD, J1100B_SOB_SITG_CD, J1100C_SOB_LYG_CD	(0-3)	0.23	0.08
Activities of Daily Living (ADLs)				
Walking	G0110C1_WLK_ROOM_SELF_CD	(0-4)	0.24	0.11
Hygiene	G0110J1_PRSNL_HYGNE_SELF_CD	(0-4)	0.16	0.03
Dressing	G0110G1_DRESS_SELF_CD	(0-4)	0.16	0.03
Locomotion	G0110E1_LOCOMTN_ON_SELF_CD	(0-4)	0.20	0.04
Transfer	G0110B1_TRNSFR_SELF_CD	(0-4)	0.08	0.00
Toilet	G0110I1_TOILTG_SELF_CD	(0-4)	0.09	0.01
Bed-mobility	G0110A1_BED_MBLTY_SELF_CD	(0-4)	0.08	0.01
Eating	G0110H1_EATG_SELF_CD	(0-4)	0.09	0.01
Interaction				
Speech clarity	B0600_SPCH_CLRTY_CD	(0-2)	0.69	0.13
Makes self understood	B0700_SELF_UNDRSTOD_CD	(0-3)	0.79	0.12
Ability to understand others	B0800_UNDRST_OTHR_CD	(0-3)	0.83	0.14
Verbal behavior symptoms	E0200B_VRBL_BHVRL_CD	(0-3)	1.21	0.21
Other behavior symptoms	E0200C_OTHR_BHVRL_CD	(0-3)	1.21	0.21
Communication training	O0500J_COMMUN_TRNG_NUM	(0-7)	0.20	0.09

Measure	ResDAC Variable Name(s)	Range	% Missing at 5-day	% Missing at 30-day
Mental				
Mood score (PHQ-9)	D0300_MOOD_SCRE_NUM	(0-27)	3.00	0.66
Antipsychotics	N0400A_ANTIPSYCHTC_CD	(0-1)	0.00	0.00
Anti-anxiety meds	N0400B_ANTINXTY_CD	(0-1)	0.00	0.00
Antidepressants	N0400C_ANTIDPRSNT_CD	(0-1)	0.00	0.00
Hypnotics	N0400D_HPNTC_CD	(0-1)	0.00	0.00
Delirium - inattention	C1310B_INTNTN_IND	(0-2)	1.99	0.42
Delirium - disorganized thinking	C1310C_DSRGNZD_THKNG_IND	(0-2)	2.09	0.49
Delirium - altered consciousness	C1310D_LVL_OF_CNSCSNS_IND	(0-2)	1.95	0.40
Psychosis	E0100Z_NO_PSYCHOSIS_CD	(0-1)	0.90	0.18
Wandering presence	E0900_WNDR_CD	(0-3)	1.20	0.21
Brief Interview for Mental Status (BIMS)	C0500_BIMS_SCRE_NUM	(0-15)	13.53	8.74
Restraints				
Bed rail restraints	P0100A_BED_RAIL_CD	(0-2)	0.05	0.03
Trunk rail restraints	P0100B_TRNK_RSTRNT_BED_CD	(0-2)	0.05	0.03
Limb rail restraints	P0100C_LMB_RSTRNT_BED_CD	(0-2)	0.05	0.03
Other bed restraints	P0100D_OTHR_RSTRNT_BED_CD	(0-2)	0.06	0.04
Trunk chair restraints	P0100E_TRNK_RSTRNT_CHR_CD	(0-2)	0.05	0.03
Limb chair restraints	P0100F_LMB_RSTRNT_CHR_CD	(0-2)	0.05	0.03
Chair prevent rise restraint	P0100G_CHR_PRVNT_RISE_CD	(0-2)	0.05	0.03
Other chair restraints	P0100H_OTHR_RSTRNT_CHR_CD	(0-2)	0.07	0.05
Pain				
Pain presence	J0300_PAIN_CD	(0-1)	12.27	8.10
Pain swallowing	K0100D_CMPLNT_SWLWG_CD	(0-1)	0.35	0.10
Mouth/face pain when chewing	L0200F_MOUTH_PAIN_CD	(0-1)	0.59	0.99
Pain intensity	J0600A_PAIN_INTNSTY_NUM	(0-10)	25.91	21.03
Pain frequency	J0400_PAIN_FREQ_CD	(1-4)	48.51	55.40
Pain effect on sleep	J0500A_PAIN_EFCT_SLEEP_CD	(0-1)	48.10	55.02
Pain effect on activities	J0500B_PAIN_EFCT_ACTVTY_CD	(0-1)	48.11	55.03
Pain management[b]	J0100A_SCHLD_PAIN_MDCTN_CD, J0100B_PRN_PAIN_MDCTN_CD, J0100C_OTHR_PAIN_INTRVTN_CD	(0-1)	0.00	0.00
Treatment and Equipment				
Anticoagulants	N0400E_ANTICOAGLNT_CD	(0-1)	0.00	0.00
Antibiotics	N0400F_ANTBTC_CD	(0-1)	0.00	0.00
Diuertics	N0400G_DRTC_CD	(0-1)	0.00	0.00
Indwelling catheter	H0100A_INDWLG_CTHTR_CD	(0-1)	0.03	0.01
Hospice care	O0100K2_HOSPC_POST_CD	(0-1)	0.23	0.54
Hearing aid	B0300_HEARG_AID_CD	(0-1)	0.77	0.15
Corrective lens	B1200_CRCTV_LENS_CD	(0-1)	1.02	0.21
Cane/crutch	G0600A_CANE_CD	(0-1)	0.09	0.03
Walker	G0600B_WLKR_CD	(0-1)	0.12	0.05
Wheelchair	G0600C_WHLCHR_CD	(0-1)	0.09	0.03
Limb prosthesis	G0600D_LIMB_PRSTHTC_CD	(0-1)	0.08	0.02
Ventilator or respirator	O0100F2_VNTLTR_POST_CD	(0-1)	0.22	0.54
Quarantine for disease	O0100M2_ISLTN_POST_CD	(0-1)	0.24	0.60
Rejection of care	E0800_RJCT_EVALTN_CD	(0-3)	1.21	0.22
In bowel toileting program	H0500_BWL_TOILTG_PGM_CD	(0-1)	0.19	0.08
Broken/loose denture	L0200A_BRKN_DNTR_CD	(0-1)	0.60	1.01
Injections	N0300_INJCT_MDCTN_DAY_NUM	(0-7)	0.06	0.02
Insulin	N0350A_INSLN_INJCT_DAY_NUM	(0-7)	0.07	0.02
Chemo	O0100A2_CHMTHRPY_POST_CD	(0-1)	0.23	0.58
Radiation	O0100B2_RDTN_POST_CD	(0-1)	0.23	0.54
Oxygen therapy	O0100C2_OXGN_POST_CD	(0-1)	0.20	0.54
Suctioning	O0100D2_SCTNG_POST_CD	(0-1)	0.22	0.54
IV medications	O0100H2_IV_MDCTN_POST_CD	(0-1)	0.22	0.54
Transfusions	O0100I2_TRNSFSN_POST_CD	(0-1)	0.22	0.54
Dialysis	O0100J2_DLYS_POST_CD	(0-1)	0.22	0.54
Splint or brace assistance	O0500C_BRC_ASTNC_NUM	(0-7)	0.19	0.09

Notes: Table shows the list of the 109 health measures used, the variable names from the MDS dataset, their range, and the percenatge missing in the 5-day assessment and 30-day assessment samples. A detailed description of the variables can be found at https://resdac.org/cms-data/files/mds-30/data-documentation.

[a] Short breathness is a sum of the associated three variables.

[b] Pain management is 1 if any of the three associated variables is 1.

Appendix Table A3: Alternative Discharge Models

	(1) Baseline specification	(2) Allowing for health improvement	(3) Allowing for impact of time
Intercept	1.72	1.73	1.76
	(0.002)	(0.002)	(0.001)
h_2	4.02	4.12	
	(0.011)	(0.016)	
h_t			4.47
			(0.006)
h_t*t			0.19
			(0.002)
Δh		-0.17	
		(0.021)	
N	2,028,014	2,028,014	13,231,304

Notes: Table shows the estimated parameters from the three alternative discharge models described in Appendix B. The first two specification are probit models of the probability of discharge in the week following the 30-day assessment. Column (1) shows our baseline model, in which this discharge probability is a function only of the 30-day health assessment (h_2). Column (2) augments this baseline model to allow the discharge probability to also depend on a health improvement term (Δh) that reflects the change in the health index between the initial assessment and the 30-day assessment. Both these models are estimate on the 2,028,014 patients who are assessed at 30 days. Column (3) augments the discharge model further to allow for the discharge decision to depend (linearly) on time in the SNF; it therefore specifies an assessment-level probit of the probability of discharge in the week following the assessment at $t = 1$ (the 5-day assessment), $t = 1.33$ (the 14-day assessment) and $t = 2$ (the 30-day assessment). Specifically, we estimate $d_{it}^D = 1 \iff h_{id}^* = \bar{\rho} h_{it} + \Delta\rho t h_{it} + \xi_i > \lambda$ where $d_{it}^D = 1$ if patient i is discharged downstream at time t, h_{id}^* is patient i's latent health at discharge and h_{it} is health at time t. We estimate this model on 13,231,304 patient-assessment observations.

Appendix Table A4: Health Index Fit for All Patients in a SNF at 30 days

	Share of baseline sample	Mean h	Std. dev. Of h
Fit for those who have a 30-day assessment (h as of day-30 assessment, N = 2,517,006):			
All patients assessed at day 30	0.403	0.138	0.093
Discharged to community w/in 7 days	0.074	0.195	0.104
Still at the SNF 7 days after	0.303	0.127	0.086
Discharged elsewhere w/in 7 days	0.026	0.110	0.078

Notes: Table shows summary statistics for our health index at 30-days for all patients assessed at 30-days. This includes patients for whom we hot-deck impute their health index at 30 days due to missing health measures. Details of this imputation procedure are presented in Appendix D.

Appendix Table A5: Discharge Destinations by 30-Day Assessment Availability

	Has 30-day assessment	Missing 30-day assessment
Share of patients in SNF at 30 days	0.802	0.198
Still in SNF 7 days later	0.847	0.365
Share discharged within 7 days		
Community (downstream)	0.110	0.482
ACH (upstream)	0.029	0.091
Hospice (upstream)	0.000	0.002
Death (upstream)	0.004	0.026
Elsewhere	0.010	0.035

Notes: Table shows the discharge destinations of patients conditional on being in the SNF at 30-days (N = 2,517,006) for both patients who are missing and not missing the 30-day assessment.

Appendix Table A6: Imputation Procedure

	(1) Case 1 -- observe $h_{1.33}$ (N = 464,404)	(2) Case 2 -- do not observe $h_{1.33}$ (N = 70,914)
Step 1	Match on $\bar{h}_1$, $\bar{h}_{1.33}$, d	Match on $\bar{h}_1$, d
Step 2	Match on $\bar{h}_1$, d, adjacent $\bar{h}_{1.33}$	Match on d and closest $\bar{h}_1$
Step 3	Match on $\bar{h}_{1.33}$, d_2, adjacent $\bar{h}_1$	Match on d, but allow d = 0 if d=2, closest $\bar{h}_1$
Step 4	Match on d, closest $\bar{h}_1$, $\bar{h}_{1.33}$	
Step 5	Match on d but allow d = 0 if d =2, closest $\bar{h}_1$, $\bar{h}_{1.33}$	

Notes: Table describes our hot-deck imputation procedure, which is described in Appendix D. For all imputations, we observe health at admission (h_1) and discharge destination within 7 days of the 30-day assessment (d). Column (1) shows the procedure when we also observe health at 14 days ($h_{1.33}$). Column (2) shows the procedure when we do not observe health at 14 days. In the table, $\bar{h}_t$ corresponds to the discretized health. $t = 1$ corresponds to the 5-day assessment, $t = 1.33$ corresponds to the 14-day assessment, and $t = 2$ corresponds to the 30-day assessment. d is a categorical variable that indicates the discharge destination of a patient within 7 days of the 30-day assessment: still in the SNF (0), discharged downstream (1), or discharged upstream/elsewhere (2).

Appendix Table A7: Summary Statistics by Imputation Cases

	(1)	(2)	(3)			(4)		
Sample	Not missing h_2	Missing h_2	Case 1			Case 2		
Subsample			downstream	still in SNF	upstream	downstream	still in SNF	upstream
Statistic								
N	1,983,464	535,318	71,241	149,955	243,208	11,221	53,748	5,945
N Imputed		533,451	71,213	149,949	241,570	11,039	53,739	5,941
Share Imputed		0.9965	0.9996	1.0000	0.9933	0.9838	0.9998	0.9993
Discretized h_2								
Mean	5.66	6.81	4.78	5.94	8.15	9.55	5.57	4.79
SD	3.32	3.78	2.85	3.52	3.68	3.80	3.53	3.05
Number of candidate comparison observations chosen from								
Mean		8.56	2.00	13.35	2.86	6.50	30.55	2.70
SD		22.15	2.02	26.69	5.10	14.16	44.29	3.43

Notes: Table shows summary statistics for patients who are still in the SNF at the time of the 30-day health assessment, by imputation cases. Column (1) shows the non-imputed sample, while Column (2) shows the sample for whom h_2, health at 30 days, is imputed. Columns (3) and (4) break up the summary statistics in column (2) for two different cases: Case 1, not missing health at 14 days ($h_{1.33}$), and Case 2, missing health at 14 days; within each of these cases we further break down summary statistics by whether and where they were discharged within the next 7 days of the imputed 30-day health assessment: downstream discharge, still in the SNF, and upstream discharge.