NBER WORKING PAPER SERIES

PRODUCING HEALTH:
MEASURING VALUE ADDED OF NURSING HOMES

Liran Einav
Amy Finkelstein
Neale Mahoney

Working Paper 30228
http://www.nber.org/papers/w30228

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2022, revised November 2024

We are grateful to Sean Gao, Sam Wang, Alicia Weng, Xuyang Xia, Connie Xu, Yuci Zhou, and especially James Okun and Carine You, for superb research assistance, and to the Laura and John Arnold Foundation, and the National Institute on Aging (Finkelstein R01-AG032449) for financial support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Producing Health: Measuring Value Added of Nursing Homes
Liran Einav, Amy Finkelstein, and Neale Mahoney
NBER Working Paper No. 30228
August 2022, revised November 2024
JEL No. I1,I11,L15,L8,L88

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 almost a week sooner, or one quarter 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.

Liran Einav
Stanford University
Department of Economics
579 Jane Stanford Way
Stanford, CA 94305-6072
and NBER
leinav@stanford.edu

Neale Mahoney
Department of Economics
Stanford University
579 Jane Stanford Way
Stanford, CA 94305
and NBER
nmahoney@stanford.edu

Amy Finkelstein
Department of Economics, E52-442
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
afink@mit.edu

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 measures of patient health and estimate nursing home value added in improving these health measures for about 14,000 nursing homes. Our estimation approach is designed to account 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 its size and because of 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 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 to aid in the recovery from a hospitalization or medical condition.

Specifically, we focus on the approximately 85% of Medicare nursing home patients with coverage from Traditional Medicare (henceforth "Medicare"), as opposed to coverage from a private Medicare Advantage plan. 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).

Estimating a nursing home's value added requires overcoming two econometric challenges, which we do by following approaches developed by the prior literature. First, there is potential 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 excluded instruments. We specify a model in which distance shifts nursing home choice but does not affect value added except via the choice of nursing home. Conditional on the validity of our instrument and our functional form assumptions, this approach allows us to control for unobserved determinants of demand that could otherwise bias our value added estimates.

A second challenge in our setting is that patients may leave nursing homes 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 they 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). Our SNF-level parameterization of downstream discharge accommodates any mechanism that creates variation across SNFs in discharge propensities for patients with similar health. This could include differences in clinical practice styles or differences in the extent of strategic discharge (Hackmann et al. July 2024) among other factors.

Our estimate of nursing home value added can be interpreted as the average improvement in a patient's health between the assessment at admission and the one taken 30 days later. As noted, an attraction of our setting is that we can analyze value-added for any of the many different health measures recorded in nursing home assessments, either individually or in combination. For our baseline analysis, we combine all the consistently and comprehensively recorded health measures in a one-dimensional "health index" that is a weighted average of the individual measures, where the weights reflect the measure's importance in predicting whether the patient is discharged home to the community within a week of their 30-day

assessment. We consider our baseline measure appealing given that, as mentioned above, the purpose of Medicare-covered nursing home care is to help the patient recover to the point where they can be discharged home. However, it is straightforward to use our approach to estimate value added for other aggregations of the individual health measures. To illustrate this, we also use our framework to estimate and present SNFs' value added in improving several alternative health indices.

We estimate our value added 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.[1] Our results indicate substantial heterogeneity in value added. While the (patient-weighted) average nursing home increases the health index – i.e., the probability of community discharge within a week – by 4.9 percentage points between admission and 30 days, there is a 5.6 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. This means that the 5.6 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 almost an entire week sooner, a one-quarter 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 more than $2,500 per patient relative to the 10th percentile nursing home. Interestingly, our estimates of nursing homes' value added are only weakly correlated with their CMS five-star rating (correlation of 0.23).

We find considerable variation in (patient-weighted) nursing home value added both across and within geographic markets. Value added is on average 2.8 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 4.9 percentage points, almost as large as the corresponding nationwide difference (of 5.6). 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. While the exact estimates naturally vary, our general finding of considerable variation in value added both across and within markets is robust to measuring value added using different health indices (such as the extent of the patients' limitations to activities of daily living). The substantial within- and across-market

[1]The value-added measures for each nursing home are available to researchers on our individual web pages.

heterogeneity also persists when we separately estimate value added using our baseline health measure for different groups of patients based on their age or their socioeconomic status. We also find that SNF value added for different patient demographic groups are highly (positively) correlated within SNF, which suggests that our baseline specification which assumes a one-dimensional metric for SNF value added across patients may not be a bad approximation.

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.[2] 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 lines of work. It complements the literature on value added for hospitals (e.g., Geweke et al. 2003; Chandra et al. 2016; Doyle et al. 2019; Hull February 5, 2018) and nursing homes (e.g., Olenski and Sacher April 2024) which, like most of the prior health literature, has focused on mortality as 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. November 2023; Hackmann et al. July 2024) 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 econometric 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

[2]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).

and data. In Section 3 we present our econometric model of value added. Section 4 discusses implementation and estimation, and Section 5 presents the main results. In Section 6 we present results from alternative specifications and briefly compare our estimates to other measures of nursing home quality (including the widely-used CMS five-star ratings). Section 7 concludes. All appendix material referenced in the paper is available in Einav et al. (2024).

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" for our baseline health index.

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. While over 60% of nursing home patients are covered by Medicare at the time of admission,[3] only 14% of nursing home patients have Medicare as their primary payer at a given point in time. By contrast, 60% of nursing home patients have Medicaid as their primary payer at a given point in time, because Medicaid patients have much longer stays than Medicare patients. 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 rest 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

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

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 the 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 (e.g., 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 2024). Certified SNFs are also required to assess their 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 2024a). Extremely poorly-performing SNFs – fewer 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) (CMS 2024b). 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. November 2023; Hackmann et al. July 2024). 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 on whom we

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.[4] The rich, longitudinal information that the assessments provide on patient physical health, mental health, and cognition will form 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 that provides the patient's 5-digit zip code of residence, an indicator for dual eligibility for Medicaid (a proxy for low income), separate indicators for 27 different chronic conditions at the start of the year of their SNF admission, an indicator for discharge to the SNF from an acute care hospital, and diagnoses from the prior hospital admission if the patient was discharged directly from a hospital.[5] We also use facility-level data from the OSCAR/CASPER system, created during the annual re-certification process, to measure the SNF's number of beds, occupancy rate, for-profit status, and whether the SNF is based within a hospital (LTCFocus 2024). Lastly, we used a variety of geographic crosswalks to match zipcodes with various other geographic identifiers (HUD 2024; NBER 2024; Dartmouth Atlas 2024).

2.3 Sample Construction and Summary Statistics

We start with the full sample of all patient 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 22.5 million patient-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 whose stay is covered by Traditional Medicare. This leaves 10.7 million patient-stays at 16,306 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 A.I provides more detail

[4]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.

[5]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 in the Medicare data 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 (CMS 2024c, 2024d).

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.[6] Second, we require that the patient is admitted to the SNF when discharged from 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 I 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. The 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).

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 health between the initial assessment and the assessment conducted at 30 days. We chose the 30-day assessment as the endpoint because, relative to the 14-,

[6]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; although in principle facilities could game whether or not a return is anticipated, in practice, of the over 3.7 million patients in our sample that are discharged to the community with no anticipated return, only about 56,500 (1.5%) of these patients actually return within 30 days of discharge. 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).

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 that there was 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 II 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 SNF 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 fewer 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 SNF has more than a thousand.[7]

For computational tractability, we restrict the choice set of SNFs that each patient considers, rather than allow each patient to choose across all SNFs in the US. We begin with the standard market definition for hospital choice, namely the 306 Hospital Referral Regions (HRRs).[8] Each HRR is composed of one or more Hospital Service Areas (HSAs). We then partition the 118 largest HRRs into smaller choice sets, which are geographically connected HSAs within each HRR (see Appendix B for more details). These, along with the 188 smallest HRRs, yield a total of 655 choice sets, which we refer to as "sub-HRRs." About 19% of SNF stays occur outside of the patient's sub-HRR. The average sub-HRR has about 20 SNFs and 7,300 stays, and 5% of them have more than 40 SNFs.

[7]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 A.I).

[8]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 A.II 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 (0-4) 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 III lists a selected group of the 109 measures and provides summary statistics for them at admission. It shows the mean and standard deviation across patient stays, and the standard deviation across SNFs (weighted by patient stays).[9]

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 they require extensive assistance but are not totally dependent on the staff. There is considerable heterogeneity across patients in these measures (column (2)). Although not as heterogeneous, there is also a non-trivial variation in the averages of these measures across SNFs (column (3)). This across-SNF heterogeneity highlights the importance of grappling with potential selection on initial health.

More broadly, the large number of measures underscores the need for combining these measures into an interpretable health index. As we discuss in more detail below, our baseline approach will be to construct an index guided by the SNF's objectives for these patients, which is to help them recover to the point where they can be discharged home. We will also examine a number of alternative indices, which have different conceptual motivations and exclude or change the relative weights on various health measures.

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

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 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

$$h_{i2} = \alpha_j + \theta_h h_{i1} + x_i'\theta_x + \varepsilon_i, \tag{1}$$

where h_{i1} controls for the effect of baseline health, x_i is a vector of observable patient demographics, and ε_i is a mean-zero i.i.d. error term. 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 and patient characteristics.[10] 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 is straightforward to estimate the model separately for different groups of patients, and in Section 6 we will estimate the value-added model separately for patients of different age groups and by whether the patient is dually eligible for Medicaid (a proxy for being low-income). Second, it assumes that the serial correlation in health (captured by θ_h) does not vary across SNFs. This assumption is needed for the cross-SNF differences in α_j to be interpretable; if we allowed θ_h to vary by SNF, then the same value of α_j would lead to different health improvements as a function of SNF-specific θ_h, so we would not be able to characterize the health improvement across SNFs with α_j alone.

[10] In practice, we will estimate θ_h to be very close to 1 (0.965). 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 econometric approach.

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 savvy patients are more likely to choose higher-quality SNFs and are also 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 strive to 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).[11] 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.

[11] 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 day 30.

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 i.i.d. error term, drawn from $N(0, \sigma_\nu^2)$. 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.[12] Recent work has shown evidence that SNFs engage in strategic admission and discharge of less profitable patients (Hackmann et al. July 2024). Nursing homes may also weigh differently non-strategic factors, such as clinical guidelines and patients preferences, in their discharge decisions. For our purpose of producing unbiased estimates of SNF-specific value added (α_j), the SNF-specific discharge threshold (λ_j) is sufficient to purge our SNF-specific value added estimates from any SNF-level sample selection bias that would result from differences in (strategic or non-strategic) downstream discharge behavior across SNFs (conditional on our functional form and parametric assumptions).

We model discharge upstream in a more statistical way, assuming that upstream discharges are the result of a probit stochastic process, which is a (presumably declining)

[12]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 C, this parsimonious model performs well compared to richer specifications.

function of h_{i2}. That is, we assume

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

where $\Phi(\cdot)$ is the standard normal cumulative density function and k indexes markets. We allow γ_k^0 and γ_k^1 to be market-specific in order to account for any potential differences in hospital access across markets.[13] We note that, from a modeling perspective, although we describe the upstream model as a more passive decision than the downstream model, 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

$$u_{ij} = \delta_j(h_{i1}, x_i) - \tau m_{ij} + \eta_{ij}, \tag{4}$$

where $\delta_j(h_{i1}, x_i)$ is the average utility from SNF j for all patients with baseline health h_{i1} and demographics x_i, m_{ij} is the log distance between patient i and SNF j, and η_{ij} is an error term drawn i.i.d. 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 choice, such as available capacity.[14]

[13] In principle, it might be desirable to allow the parameters of the discharge model to also vary across SNFs within a market. In practice, however, because upstream discharge is naturally harder to predict as it is predominantly driven by unexpected health events, and, as we describe below, our health index is not designed to predict upstream discharge, our market-level estimates of the γ's are already quite variable, making us wary of estimating γ's at a more granular level. If, in the extreme, upstream discharge is completely uncorrelated with our health index, the upstream discharge model would have no power. However, in that case we would not expect any differential selection across SNFs that is driven by upstream discharges.

[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

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

$$\mathbb{E}\left[h_{i2} \mid c_{ij}, h_{i1}, x_i, \eta_{i1}, \ldots, \eta_{i|J_i|}\right] = \alpha_j + \theta_h h_{i1} + x_i'\theta_x + \left[\sum_{\ell \in J_i} \phi_\ell(\eta_{i\ell} - \mu_\eta)\right] + \varphi(\eta_{ij} - \mu_\eta), \quad (5)$$

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

$$\mathbb{E}\left[h_{i2} \mid c_{ij}, h_{i1}, x_i, m_{i1}, \ldots, m_{i|J_i|}\right] = \alpha_j + \theta_h h_{i1} + x_i'\theta_x + \left[\sum_{\ell \in J_i} \phi_\ell \beta_{i\ell}\right] + \varphi\beta_{ij}, \quad (6)$$

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

$$\beta_{i\ell}(j) = \begin{cases} -\log \hat{p}_{i\ell} & \ell = j \\ \frac{\hat{p}_{i\ell}}{1-\hat{p}_{i\ell}} \log(\hat{p}_{i\ell}) & \text{otherwise} \end{cases}, \quad (7)$$

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

$$\hat{p}_{ij} = \frac{\exp\left(\delta_j(x_i) - \tau m_{ij}\right)}{\sum_{\ell \in J_i} \exp\left(\delta_\ell(x_i) - \tau m_{i\ell}\right)}. \quad (8)$$

The estimating equation (6) integrates over an error term ϵ that is distributed $N(0, \sigma_\epsilon^2)$. Appendix D provides the full derivation.

The additional parameters in equation (6) have an economic interpretation. The $\phi_\ell\beta_{i\ell}$ terms control for any unobservable correlation between the choice 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 $\phi_\ell\beta_{i\ell}$'s. This term would capture any Roy-type

allocation to nursing homes (Gandhi 2020). In our context, any impact of capacity constraints will be reflected in 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 sub-HRR. We normalize the average utility from the outside option (which includes the disutility from travel distance) to zero; that is, $\delta_0 = 0$ and $m_{i0} = 0$ in equation (4) for all i.

selection, in which a patient has an idiosyncratically higher preference for a SNF at which they are idiosyncratically more likely to improve.

The key assumptions underlying our econometric approach are the validity of the distance instrument and the functional form assumptions embedded in the choice and value added models. In terms of the functional form assumptions, our control function approach assumes that the unobserved determinants of choice enter the value added equation in a linear, additive separable manner. This follows the literature i.e., Dubin and McFadden 1984; Abdulkadiroğlu et al. 2020 and can be thought of as a first order approximation of a specification that allows for a richer relationship between choice and value added. In terms of the validity of the distance instrument, in Section 4.2, we provide evidence in support of the exogeneity of the distance instrument with respect to SNF value added. Specifically, in the spirit of Altonji et al. (2005) and Oster (2019), we show the coefficient on distance is stable when we add a richer set of controls to the SNF choice model. This is reassuring of the identifying assumption, but it is also useful to think through the potential bias that would result from an invalid instrument. If our instruments were invalid because distance to the SNF directly improved patient health (e.g., due to familiarity with the neighborhood or more frequent visits from family), our value added estimates would be upward biased with ambiguous effects on the distribution of value added across SNFs. If our instrument were invalid because high value added SNFs happened to be located closer to patients who are unobservably more likely to improve, our value added estimates for high value-added SNFs would be overstated (and vice versa for low value-added SNFs), artificially stretching the distribution of value added. The same logic, mutatis mutundis, applies to the scenario where high value added SNFs happened to be located closer to patients who are unobservably less likely to improve.

3.3 Intuition for Identification

To gain some intuition for the identification properties of the model, we consider first a simpler version of the model in which there is a health production function (equation (1)) and only downstream discharge (equation (2)), and where we have already conditioned on observable patient characteristics x_i and there is conditional random assignment to SNFs (so that we do not yet have to worry about selection in). The objects of interest are 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. In our case, the selection problem is

that period-2 health is not observed for patients who are discharged before period 2. In this simplified setting, we would estimate $h_{i2} = \alpha_j + \theta_h h_{i1} + \varepsilon_i$, taking into account the fact that h_{i2} is observed if and only if $\varepsilon_i < \lambda_j + \nu_i - \alpha_j - \theta_h h_{i1}$ (see equation (2)). The identification challenge arises because SNFs may vary not only in their value added α_j, but also in their discharge threshold λ_j. Thus, if we were to run a simple regression of h_{i2} on SNF fixed effects, the fixed effects would depend on both α_j and λ_j, and we will not be able to separately identify the SNF's value added.

To address this, we take advantage of the (observed) heterogeneity in patient health at admission, h_{i1}. The key to identification in our setting arises from the fact that we observe many patients with different initial health within a SNF, and each SNF is characterized by a single discharge threshold λ_j.[16] To see this, note that – provided sufficiently large support of h_{i1} – 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.[17] Since the probability of downstream discharge is a sufficient statistic for the bias from selection,[18] 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 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.

This core identification argument is preserved when we introduce patient demographics x_i and the possibility of upstream discharges. Introducing patient demographics is straightforward; we can apply the same argument conditional on a set of patients that have the same value of x_i to recover value added. We modelled upstream discharge with a probit equation with parameters γ_k^0 and γ_k^1 that do not vary within markets (see equation (3)). Given this assumption, it is sufficient to observe a single SNF in the market with some patients with a low enough value of h_1 (or a high enough value of λ_j) such these patients are never discharged downstream before period 2. This allows us to recover the market level

[16]The fact that we observe discharge decisions for many patients within each SNF is also a key difference from 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.

[17]Given the model, the probability of a patient with initial health h_{i1} to be discharged prior to period 2 out of SNF j is monotonically (and continuously) increasing in h_{i1}, and as long as the support of h_{i1} spans sick enough patients who do not get discharged before period 2 as well as healthy enough patients who always do, one can invert the discharge probability at SNF j' to obtain the appropriate value for h'_{i1}.

[18]That is, if $Pr(\varepsilon_i > \lambda_j + \nu_i - \alpha_j - \theta_h h_{i1}) = Pr(\varepsilon_i > \lambda_{j'} + \nu_i - \alpha_{j'} - \theta_h h'_{i1})$ then $E(\varepsilon_i | \varepsilon_i < \lambda_j + \nu_i - \alpha_j - \theta_h h_{i1}) = E(\varepsilon_i | \varepsilon_i < \lambda_{j'} + \nu_i - \alpha_{j'} - \theta_h h'_{i1})$.

upstream discharge parameters for each market. With the upstream discharge model in hand, we can map from the observed distribution of h_{i2} in each SNF to the "pre-upstream discharge" distribution of h_{i2}. We can then apply the identification argument above (for the setting without upstream discharge) using this adjusted pre-upstream distribution of h_{i2} to identify SNF value added.

Finally, we account for potential selection into the SNF in a more standard way, with a control function for selection into the SNF that uses 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 from versus 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.

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). Unlike test scores in education, which (debatably) provide a low-dimensional summary of student achievement, there is no "off the shelf" summary measure of health to use for our analysis.

Our econometric framework can be applied to any aggregate (or individual) health measure. For our baseline analysis, we construct a health index that is a weighted average of the individual health measures, where the weights are determined by the importance of these measures for predicting whether a patient will soon be discharged home. We consider this a natural and appealing measure because it aligns with the SNF's objective which is, as we have noted, to help Medicare patients recover to the point where they can be discharged home. But of course one may well be interested in other measures of health, and we therefore

also compute value-added using several alternative health measures and compare the results to those from our baseline approach.

Baseline health index

The weights on the individual health measures that we use to construct our baseline health index are taken from a prediction model – trained on our large sample of patients – that maps from the 109 health measures recorded during the 30-day assessment to an indicator for whether the patient was discharged to the community within 7 days of their assessment.[19] We then apply these weights to the health measures recorded at the initial and 30-day assessments to construct the health index for these time periods. A higher value added SNF – i.e., one that increases its patients health index by more between the initial and 30-day assessments – can be thought of as a SNF that more quickly raises patient health to the level conducive to community discharge.

Operationally, we use a regression tree for our prediction model, since it is able to capture potentially important interactions among the health measures.[20] 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.[21]

To give some sense of what drives the health index, the last column of Table III 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

[19]The choice to use the 30-day assessment is motivated by two considerations. First, it 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. In practice, the choice of 30-day assessment is not essential; we find that using the initial assessment creates a highly-correlated health index. 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 were too short. About 18% of the patients who have a 30-day assessment are discharged downstream within the next week (see Figure 1).

[20]As described in Appendix A, about 500,000 patient stays (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 E for more details.

[21]Appendix Figure A1 plots the observed 7-day discharge rate against the period 2 health index, both overall and for patients with different values of period 1 health. Panel A shows that period 2 health is unsurprisingly extremely strongly correlated with discharge rates; when the data is binned into 100 groups, the correlation is virtually perfect. The subsequent figures show this correlation is just as strong when we split the data by terciles of period 1 health, demonstrating the stability of the model for different patient groups.

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).[22]

Table IV Panels A and B summarize the resulting health index and how it correlates with subsequent outcomes.[23] Panel A shows the index for the patients on which we trained the regression tree.[24] 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. Patients who are actually discharged to the community within the next 7 days have a higher average health index (0.201) than patients who remain at the SNF (0.127). Patients who are discharged elsewhere (predominantly to the hospital) within 7 days have a lower average health index (0.113). These patients also have a slightly lower health index than those who remain in the SNF (0.127), although the difference is small. The small difference is consistent with our health index not being trained to predict discharge to the hospital and the fact that the sudden health shocks that precipitate hospital discharge are inherently hard to predict.

Panel B shows a similar qualitative pattern when we apply the health index construction to the initial health assessment for all patients, a sample different 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). Those who remain in the SNF or are discharged elsewhere are sicker (both groups have an average health index at admission of 0.104); the identical initial health index for those who remain or are discharged elsewhere is again consistent with low predictive power of the health index for upstream discharge.

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.[25]

[22] In Section 6.1 we explore the sensitivity of our key results to using alternative ways to construct the health index.

[23] We defer a discussion of Panel C to the next section.

[24] Appendix Table A.IV 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 E.

[25] A potential conceptual concern with our baseline health index is that it might reflect some degree of

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 an alternative measure, known in the clinical literature as the long-form ADL (Carpenter et al. 2006). It is the sum of the scores of seven measures of limitations to activities of daily living (ADLs): hygiene, walking, dressing, locomotion, transfer, toilet, bed-mobility, and eating (see Appendix Table A.II).[26] For our purposes, we negate the long-form ADL so that a higher score corresponds to better health.

We also consider a modification to our baseline health index to deal with a natural concern that the nursing home staff in the SNFs may have an incentive to strategically misreport some or all of the health measures in the MDS assessments. A similar concern 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) or even fraudulently change students' answers (Jacob and Levitt 2003). In our setting, the existence or direction of any bias to our value-added estimates is unclear. SNFs face opposing incentives to report better scores (to increase their quality ratings) and worse scores (to generate larger payments). Moreover, our value-added estimates are based on within-patient changes in health assessment, so that any persistent bias between the initial and 30-day assessments will be differenced out in the construction of value added. Nonetheless, to assess any potential bias from strategic misreporting, we construct a modified version of our baseline health index that excludes health measures where nursing homes may have an incentive to shade their scores.

To do so, we note that not all 109 health measures in our baseline health index are used in the CMS quality star ratings or the CMS case-mix reimbursement criteria. We therefore restrict attention to the 45 of the original 109 health measures that are not used in either the

reverse causality. If discharge is planned several days in advance, and the SNF uses the days before the planned discharge to train patients to be more self-sufficient, this higher self-sufficiency would be heavily weighted in the health index, and the baseline health index might be measuring short-term preparation for discharge rather than underlying health. To reduce the influence of short-term preparation, we construct an alternative index that excludes soon-to-be discharged patients. Specifically, rather than train the model on discharge within 7 days of the 30-day assessment, we restrict the sample to patients who receive a 30-day assessment and remain at the SNF for at least 7 additional days. We then construct an alternative health index that is based on predicting downstream discharge on days 8-14 after the 30-day assessment versus remaining at the SNF for at least 15 days. Reassuringly, the alternative health index has a similar goodness of fit, with an AUC of 0.70 (versus 0.71 for the baseline health index), a similar distribution to the baseline index (see Appendix Table A.VIII), and is highly correlated with the baseline index both within and across markets (see Appendix Figure A10).

[26]Since each measure takes a value from 0-4, the aggregate measure ranges from 0 to 28, with higher values indicating that the patient needs more assistance.

star ratings or the reimbursement criteria; we refer to these as "non-incentivized measures" (see Appendix Table A.II for which measures are "non-incentivized").[27] We train a regression tree that uses these 45 health measures (instead of the 109) to predict whether the patient will be discharged to the community within 7 days of their 30-day assessment and use the resulting prediction model to construct an alternative health index. The new regression tree has an AUC of 0.70, which is very close to 0.71 AUC for our baseline health index.

4.2 Distance as an Instrument for Nursing Home Choice

We use patients' 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. November 2023), hospital (e.g., McClellan et al. 1994; Geweke et al. 2003; Cutler 2007; Card et al. May 2023), 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 add increasingly rich controls for patient demographics and health measures.[28] 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 over 80% as distance increases from 1 to 5 miles. It also shows that the relationship between distance and SNF choice is well approximated with a logarithmic 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

[27] We also drop measures categorized as "supporting measures" (the two measures for which inclusion in the reimbursement criteria is ambiguous) or those introduced in January 2015 or July 2016 (near the end of our sample period).

[28] 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.

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 $\hat{p}_{i\ell}$ that are inputs into the control function terms $\beta_{i\ell}$ in equation (6). For computational tractability, we limit the shifters of δ_j in the choice model to the baseline health index h_{i1} and the demographics x_i to be binned age a_i and Medicaid dual-eligibility (a proxy for being low-income).[29]

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)).[30] In equation (6), we control for binned age and Medicaid dual-eligibility, similar to the demand model.[31]

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: seven "national" parameters $(\theta, \varphi, \sigma_\epsilon, \sigma_\nu, \theta_{x1}, \theta_{x2}, \theta_{x3})$, and all other parameters, which vary at either the market- or SNF-level. This allows us to have a nested estimation procedure. Conditional on the "national" parameters, we estimate the model market by market, which is relatively standard and fast. We then search numerically over the seven "national" parameters to maximize the likelihood. In Appendix F, we describe the likelihood function and our estimation approach in more detail.

To obtain standard errors for the parameters estimated from equation (6), we perform a two-step score bootstrap procedure similar to the approach in Abdulkadiroğlu et al. (2020). Appendix F.4 provides more details. This approach allows estimation error in the control

[29] That is, we define x_i to be three mutually exclusive age bin indicators ($a_{1i} = 1\{65 \leq a_i \leq 74\}$, $a_{2i} = 1\{75 \leq a_i \leq 84\}$, and $a_{3i} = 1\{85 \leq a_i\}$) and an indicator for Medicaid dual-eligibility g_i. Thus $\delta_j(x_i) = \delta_{0j} + \delta_{1j}h_{i1} + \delta_{2j}g_i + \delta_{3j}a_{2i} + \delta_{4j}a_{3i}$. In Appendix Figure A3 we show that adding more shifters does not affect the estimated coefficient on log-distance.

[30] As shown in Table I, there is a small share (approximately 7%) 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 of the health production function.

[31] As in the demand model, we divide age a_i into three mutually exclusive bins a_{1i}, a_{2i}, and a_{3i}. The vector x_i in equation (6) then contains three demeaned indicators $(\tilde{g}_i, \tilde{a}_{2i}, \tilde{a}_{3i})$; $\theta_x := (\theta_{x1}, \theta_{x2}, \theta_{x3})$ is defined conformably.

function terms to propagate through to the value-added and downstream discharge parameter estimates.

Finally, the variation in estimated value added across SNFs can be biased upwards by noise in the individual SNF value added estimates. To adjust for this potential bias, we perform an empirical Bayes (EB) shrinkage procedure (Morris 1983), shrinking our SNF- and market-level parameter estimates toward their global means using their standard errors.[32]

5 Main Results

5.1 Parameter Estimates and Model Fit

Panel A of Figure 4 reports estimates of the "national" parameters that we restrict to be the same across SNFs in all markets, while Table V reports summary statistics for the γ_k^0's and γ_k^1's, which are allowed to vary across markets (indexed by k). The θ_h coefficient on initial health is close to 1, implying that α_j can be interpreted as the average health improvement at the given SNF. The σ_ϵ coefficient for the standard deviation of the health shocks is 0.075, which is slightly smaller than the 0.095 cross-sectional standard deviation of initial health (Table IV). The γ_k^1 market-level slope coefficients in the upstream discharge model are negative in most markets; this is consistent with lower upstream discharge probability for patients with higher initial health, which seems natural. However, as we discussed earlier, the fact that the health index is not designed to predict upstream discharge, as well as that upstream discharges are naturally harder to predict as they are predominantly driven by unexpected health events, may explain the substantial dispersion in the slope parameter across markets.[33]

Panel C of Table IV shows some evidence on model fit, displaying model-predicted discharge locations and reporting the mean and standard deviation of health conditional on those discharge locations.[34] The shares discharged to different locations are very similar

[32]More details are provided in Appendix G. For the baseline specification, we estimate the standard errors using a bootstrap methodology that allows estimation error in the control function terms to propagate through to the value added estimates. Since the bootstrap approach is computationally intensive, and the standard errors are virtually indistinguishable to those from an approach that calculates standard errors using the Hessian from the second stage likelihood function, we use the Hessian-based approach for the non-baseline specifications.

[33]The range of estimated slope coefficients is misleading in terms of their implications for heterogeneity in upstream discharge patterns across markets. The large standard deviation of γ_k^1 is driven by the long tails of the distribution. Moreover, the intercept and slope coefficients are negatively correlated: markets with a large intercept have a more negative slope, such that the predicted upstream probabilities are similar within the fairly concentrated range of h_1.

[34]Appendix H provides more details on how these predicted discharge destinations are computed.

in the actual data (Panel B) and when predicted by the model (Panel C). The initial health distributions of health conditional on these discharge locations is also similar in the actual data and when model-generated. This suggests that our model is capturing selection upstream and downstream reasonably well.

5.2 Estimates of Nursing Home Value Added

Figure 5 shows our estimates of the distribution of value-added across SNFs. 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. Panel A plots a kernel density of the baseline value added estimates and a second kernel density of the estimates shrunk towards the mean using the empirical Bayes methodology. The empirical Bayes adjustment attenuates the dispersion in value added by about 40% (reducing the standard deviation in value-added across SNFs from 0.038 to 0.023). We therefore use the empirical-Bayes adjusted estimates of our SNF-specific parameters for all subsequent analysis.

Panel B shows the distribution of value added for the SNFs in our baseline sample, along with summary statistics for this distribution. The average (median) SNF value added is 0.049 (0.050). Combined with our estimate of nearly 1 for the coefficient on health at admission ($\theta_h = 0.965$), this implies that the average SNF increases the weekly probability of discharge to the community by about 4.9 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 difference in value added between the 10th- and 90th-percentile SNFs is 0.056; the difference in value added between the 25th- and 75th-percentile SNFs is 0.029. 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). That is, 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.[35]

[35]CMS stipulates that the initial assessments be conducted by day 8 and the day-30 assessments be conducted during days 27-33, leaving SNFs some flexibility as to when exactly to conduct patient health assessments. If patients' health improved at a constant daily rate, then SNFs that had a longer lag between assessments would incorrectly appear to be higher value added. Reassuringly, Appendix Figure A11 shows that there is little systematic variation in the average lag in assessment dates across SNFs (inter-quartile range of 20.5 to 21.3; 10-90th percentile range of 19.8 to 22.1) and that the average lag is not correlated with

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 5.6 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 $2,600 in Medicare savings per SNF admission.

Panels B and C 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 B shows that the discharge threshold is uncorrelated with the value added. This does not imply that patients have on average equal length of stay across low versus high value SNFs. Indeed, because patient health improves more quickly at high value added SNFs, and discharge thresholds do not systematically vary, patients on average have shorter stays at higher value added facilities.

Panel C shows that the correlation between the value added and the selection coefficient ϕ_j is negative and large (correlation of -0.42). 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 reflect the fact that patients with the lowest expected health improvements are disproportionately likely to travel longer distances to high-value-added SNFs. By contrast, the estimate of the "Roy selection" coefficient φ is negligible (Panel A of Figure 4), implying that there is little correlation between patients' unobserved expected health improvements and patients' idiosyncratic preferences for SNFs.

5.3 Sources of Heterogeneity in Nursing Home Value Added

Heterogeneity Across and Within Markets

We analyze heterogeneity in value added across and within markets. To do this, we use the standard market definition used in the Dartmouth Atlas literature on geographic variation in health care productivity, namely the 306 Hospital Referral Regions (HRRs) defined in Section 2. Figure 6 plots average SNF value added by market (Panel A) and the difference

our value added estimates (correlation of 0.00).

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, the Midwest, and the Pacific Northwest. The average difference in value added between the 90th- and 10th-percentile markets is about 2.8 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 HRR, the difference in value added between the 90th- and 10th-percentile nursing homes is 4.9 percentage points (third row of Panel C), which is almost as large as the 5.6 percentage point 90-10 difference nationwide (top row of Panel C). This finding regarding substantial within-market dispersion in value added is robust to using another standard, but smaller market definition, namely Hospital Service Areas (HSAs).[36] Unlike market-level average valued added, the within-market dispersion in value added does not exhibit any consistent regional pattern.[37]

To further examine the within-market dispersion in value added, Figure 7 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 (i.e. top) decile of value added, roughly 25% of patients are in SNFs that perform below the national median. Likewise, in markets in the second decile of value added, roughly 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 4.9 days sooner (relative to a median length of stay of 22 days), saving Medicare approximately $2,300 per patient stay. We view this within-market dispersion as potentially promising. Relative to reallocation across markets, it seems more feasible to imagine policies – such as

[36]HSAs are designed to represent local health-care markets for community-based inpatient care (https://www.ncbi.nlm.nih.gov/books/NBK585005/). There are over 3,000 HSAs, an order of magnitude more than the number of HRRs. The within-market difference in value added between the 90th- and 10th-percentile nursing homes when markets are defined as HSAs remains quite large (3.9 percentage points).

[37]As shown in Appendix Figure A4, there is a fairly weak positive correlation (0.06) between average value added at the market level and the within-market variation.

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

Correlates of Nursing Home Value Added

A natural question is whether the variation in SNF value added is associated with underlying characteristics of SNFs or their patients. Table VI examines these correlations both unconditionally (column (1)) and after conditioning out market fixed-effects (column (2)).[38]

The relationship between SNF value added and SNF-level characteristics is weak, both overall as well as 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 77% of facilities, have slightly lower value added (average of 0.048) than non-profit facilities (average of 0.055). However, the negative correlation between for-profit status and value added is not very strong (−0.134). Within markets, SNF value added is weakly negatively correlated with chain ownership (−0.064) and with whether the SNF has an Alzheimer's special care unit (−0.011). Chain-owned SNFs account for 70% of all SNFs, and approximately 24% of SNFs have an Alzheimer's unit. The SNF's share of Medicare patients shows the strongest (positive) correlation (0.248).

SNF value added shows a stronger correlation with patient characteristics, both unconditionally and within markets (Panel B). There is some sorting to SNFs based on baseline patient health, with healthier patients (those with a 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, a lower share of patients who are dually-eligible for Medicaid at the time of admission, a higher share of patients that are female or married at admission, and a higher average patient age.

6 Additional Analyses

We report results for SNF-value added for several alternative health measures and several alternative econometric models. Our key finding of substantial within-market heterogeneity persists across these alternatives. Table VIII summarizes the correlation between our baseline measure and alternative measures; once again, we report both the unconditional correlation between our baseline value added and each of the alternatives (column (1)), and the partial

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

correlation conditional on market fixed effects (column (2)).[39]

6.1 Estimating value added for alternative health measures

We re-estimate our value-added model for the two alternative health measures described in Section 4.1: limitations to activities of daily living (ADLs) and a health index constructed using the subset of the baseline measures for which SNF strategic misreporting incentives should be limited ("excluding incentivized measures"), since these measures are not used for either CMS quality measures or reimbursement rates. Appendix Table A.IX provides the parameter estimates for each of these separate models.

The first three panels of Table VII show that our key finding that the within-market variation in SNF quality is substantial relative to the national variation or the cross-market variation persists under both measures. Panel A reproduces our baseline results (from Panel C of Figure 6), while Panels B and C show the estimated variation for the alternative health measures. In addition, Panel A of Table VIII shows that our baseline SNF value-added measure is fairly highly correlated with value added measured using ADLs (correlation of 0.56), and even more highly correlated (correlation of 0.85) with the "excluding incentivized measures" version.

6.2 Allowing SNF value added to vary across patients

Our baseline model estimates a single value-added measure for each SNF. In practice, of course, SNF value-added may vary across patients. One possibility is that SNFs specialize, so that they are relatively more effective in treating some types of patients and less effective in treating others. Another possibility is that SNF quality is a general characteristic, with higher quality SNFs exhibiting higher value added for all types of patients.

To investigate this, we estimate value-added separately for patients of different age groups (65-74, 75-84, and 85 or older) and separately by whether the patients are dually eligible for Medicaid (which is a proxy for being low-income). Specifically, we re-estimate the SNF demand model in equation (4) and the health production function in equation (6) separately for each of the five groups. Appendix Table A.IX provides the parameter estimates for each demographic sub-group respectively.

The results paint a fairly clear picture in favor of the latter hypothesis: good SNFs tend to be good across the board, for all patient groups. Appendix Figures A7, A8, and A9

[39] Appendix Figure A6 produces scatter plots that correspond to column (2) of each row of Table VIII.

summarize these results. Each panel in these figures pairs up two groups of patients, and then presents bin scatters of the value added estimates for one group of patients against the corresponding estimates (of the same SNFs) for another group of patients. The pairwise correlations for each pair are always positive and quite high (predominantly between 0.5 and 0.7), suggesting that summarizing SNFs' quality with a single value added (as we do in our baseline analysis) captures the lion's share of the quality variation across SNFs. In addition, Panel D of Table VII shows that our finding of substantial within-market heterogeneity in value-added across SNFs (relative to the national variation or the cross-market variation) persists even when value-added is estimated separately for each demographic sub-group.

6.3 Relationship to Alternative Nursing Home Ratings

We compare our baseline SNF value-added estimates to the "five-star" SNF 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 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 Panel B of Table VIII reports a fairly low positive correlation of 0.23 between our value added and the overall five-star rating. The low correlation between our measure of value added and the overall government rating 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; Yaraghi et al. 2016; Rau 2018). The other rows in Panel B show the correlation between our value added and each of the star-rating's sub-components.[40] The correlation between our measure and the quality index component, which uses the same health assessment data we do, is very low (0.02), reflecting a number of underlying differences in the construction of these measures.[41]

To get a sense of how different markets might fare under our value-added measure relative

[40] These sub-components are only weakly correlated with each other, with correlations ranging from a low of -0.02 between the inspection and staffing sub-components to a high of 0.16 between the inspection and quality components.

[41] 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.

to the current star ratings, Figure 8 plots the average star rating in each market against the average value-added estimate in each market. There is considerable 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 0.005 and 0.026, respectively, relative to the average of 0.049). 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 0.061.

6.4 Relationship to 90-Day Outcomes

We also 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 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, the comparison of these measures to our value-added measure can still serve as an informative sniff test (ahem) of our estimates, and allow 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 original issue of 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, patient demographics (three age bins and Medicaid dual-eligibility), and the control functions for SNF choice, as in equation (6), which gives our estimates a slightly different interpretation.

Panel C of Table VIII 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 from 0.08 to 0.42 across outcomes 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 patients 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). 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 otherwise show up in a higher death rate or a higher readmission rate.

6.5 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 and patient demographics: $h_{i2} = \alpha_j + x_i'\theta_x + \epsilon_{i2}$. A slightly more sophisticated measure, which we call "average health improvement," is based on a regression of period 2 (i.e. 30-day assessment) health on initial health and patient demographics: $h_{i2} = \alpha_j + \theta_h h_{i1} + x_i'\theta_x + \epsilon_{i2}$. The first two rows in Panel D of Table VIII report the correlation between our preferred measure of value added and these two simpler measures. Our preferred value-added measure has 0.63 correlation with average health and 0.72 correlation with the health improvement; these correlations are slightly lower (0.58 and 0.69, respectively) within markets.

The final two rows in Panel D of Table VIII 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. Controlling for selection into the SNF by itself has a small impact on the correlation with the preferred measure. Controlling for selection out actually marginally reduces the correlation with the preferred measure, despite

the fairly strong adverse selection documented in Panel C of Figure 4. These results indicate that it is important to account for both selection in and selection out to credibly recover SNF value added.

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 to the 90th percentile of the nursing home value-added distribution is equivalent to being able to discharge a patient to the community almost one week faster (or about one-quarter of 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 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, as well as 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 a number of ways to extend our analysis of nursing home value added. While we examined some SNF-level correlates of SNF quality, there is likely scope to examine the correlation between our baseline and alternative value added measures and a richer set of measures of SNF behavior. Indeed, with the right data and variation, it may be possible to estimate the causal effect of various aspects of SNF behavior on relevant value added metrics. We conducted some basic analysis of heterogeneity in value added within SNFs across patients of different ages or by a proxy for income status. Extending this analysis to better understand how nursing homes contribute to heterogeneity in patient outcomes would be another promising area for inquiry.

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.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools". *Journal of Political Economy* 113 (1): 151–84.

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/.

Bring, Johan. 1995. "Variable importance by partitioning R2". *Quality & Quantity* 29 (2): 173–189.

Card, David, Alessandra Fenizia, and David Silver. May 2023. "The Health Impacts of Hospital Delivery Practices". *American Economic Journal: Economic Policy* 15, number 2 (): 42–81. https://doi.org/10.1257/pol.20210034. https://www.aeaweb.org/articles?id=10.1257/pol.20210034.

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.

Centers for Medicare & Medicaid Services. 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.

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

———. 2024a. "Compare Nursing Home Quality". https://www.medicare.gov/what-medicare-covers/what-part-a-covers/compare-nursing-home-quality.

———. 2024b. *Long Term Care Minimum Data Set (MDS) 3.0.* https://resdac.org/cms-data/files/mds-30.

———. 2024c. *Master Beneficiary Summary File.* https://resdac.org/cms-data/files/mbsf-base.

———. 2024d. *Medicare Provider Analysis and Review.* https://resdac.org/cms-data/files/medpar.

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.

Dartmouth Atlas. 2024. *Geographic Boundary Files.* https://data.dartmouthatlas.org/supplemental/#boundaries.

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 Neale Mahoney. 2024. "Producing Health: Measuring Value Added of Nursing Homes". NBER Working Paper No. 30228.

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. November 2023. "Owner Incentives and Performance in Healthcare: Private Equity Investment in Nursing Homes". *The Review of Financial Studies* 37, number 4 (): 1029–1077. ISSN: 0893-9454. https://doi.org/10.1093/rfs/hhad082. https://doi.org/10.1093/rfs/hhad082.

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 R. Ziebarth. July 2024. "Patient versus Provider Incentives in Long-Term Care". *American Economic Journal: Applied Economics* 16, number 3 (): 178–218. https://doi.org/10.1257/app.20210264. https://www.aeaweb.org/articles?id=10.1257/app.20210264.

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. February 5, 2018. *Estimating Hospital Quality with Quasi-Experimental Data,* 3118358, Rochester, NY, February 5, 2018. Visited on November 2, 2024. https://doi.org/10.2139/ssrn.3118358. https://papers.ssrn.com/abstract=3118358.

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.

Long-Term Care: Facts on Care in the US. 2024. "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.

National Bureau of Economic Research. 2024. *SSA to Federal Information Processing Series (FIPS) Core-Based Statistical Area (CBSA) and Metropolitan and Micropolitan Statistical Area (MSA) County Crosswalk.* https://www.nber.org/research/data/ssa-federal-information-processing-series-fips-core-based-statistical-area-cbsa-and-metropolitan-and.

Olenski, Andrew, and Szymon Sacher. April 2024. "Estimating Nursing Home Quality with Selection". *The Review of Economics and Statistics* (): 1–31. ISSN: 0034-6535. https://doi.org/10.1162/rest_a_01449. https://doi.org/10.1162/rest%5C_a%5C_01449.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence". *Journal of Business & Economic Statistics* 37 (2): 187–204.

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 Department of Housing and Urban Development. 2024. *Zipcode Population Weighted Centroids.* https://catalog.data.gov/dataset/zip-code-population-weighted-centroids.

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.

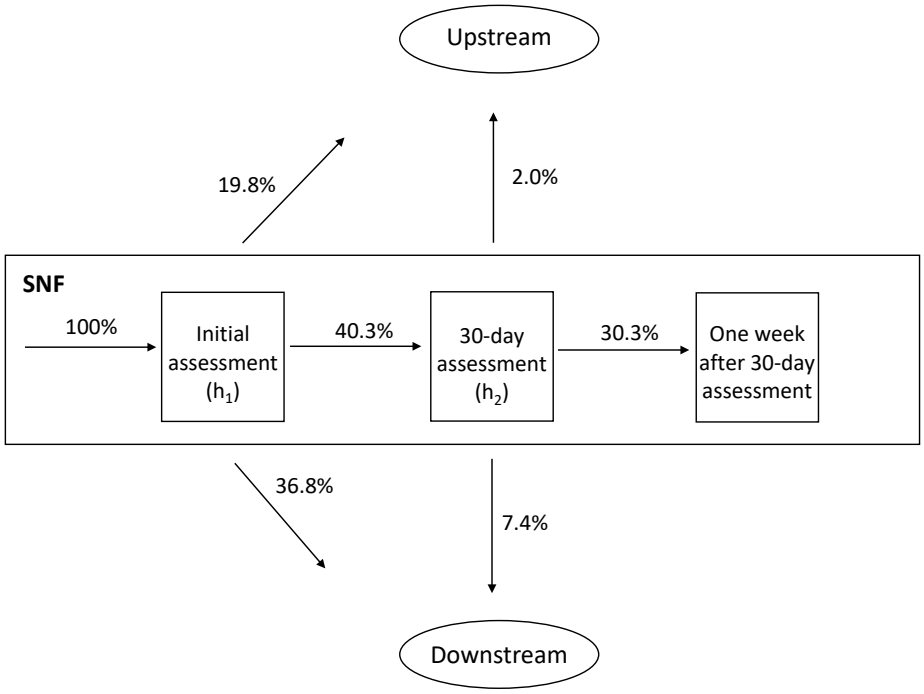
Wennberg, John E, Megan M Cooper, JD Birkmeyer, et al. 1998. "The Dartmouth Atlas of Health Care 1998".

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).

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

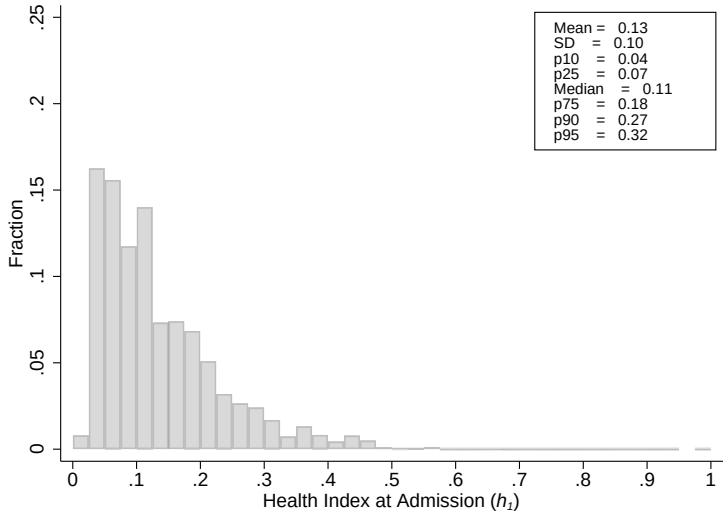
Figure 1: Movement and Measurement of Patients



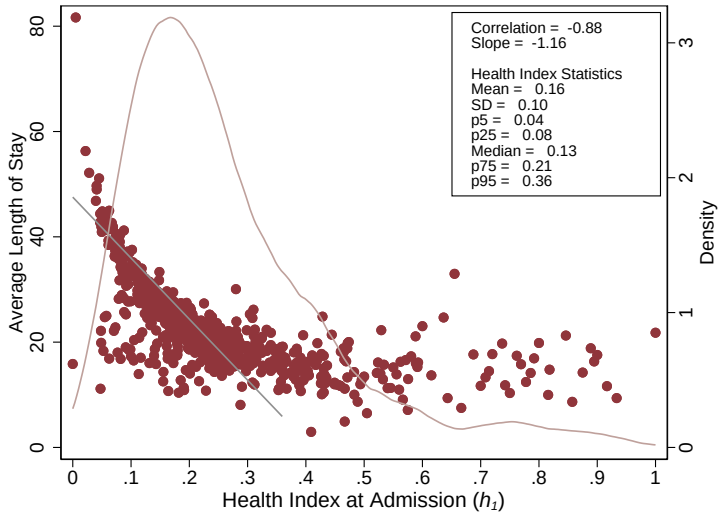
Notes: Figure shows the movement of SNF patient stays and the timing of health assessments. "Downstream" refers to patients who are discharged to the community. "Upstream" refers to patients who die in the nursing home, are sent to hospice, or are sent to an acute care hospital. Between the initial assessment 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 locations (not shown). Statistics on discharge after the 30-day assessment are for discharges within 7 days of the 30-day assessment. In the one-week period following the 30-day assessment, 0.6% of patient stays are discharged to destinations 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



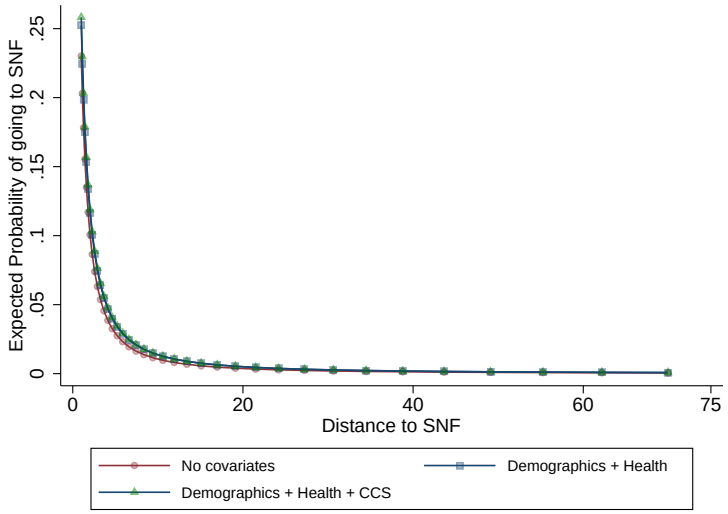
(B) Health Index at Admission and Length of Stay



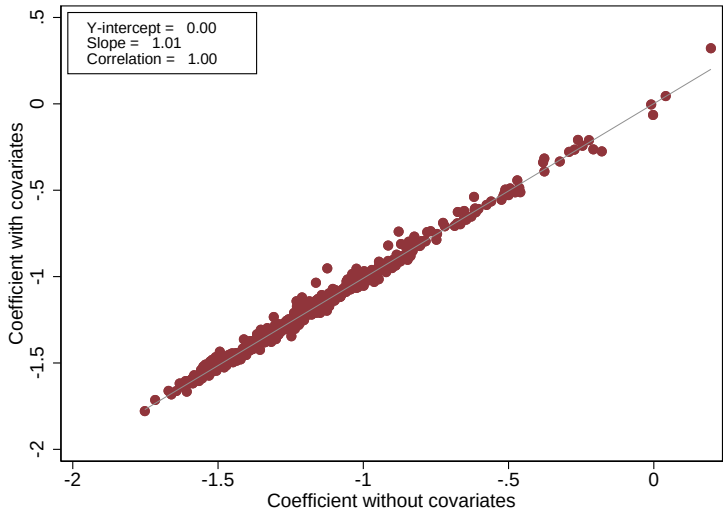
Notes: Panel (A) shows the distribution of our health index at admission (h_1) in the baseline sample ($N = 6,246,686$). Panel (B) shows the relationship between length of stay and health index at admission (h_1) conditional on discharge to the community during our study period ($N = 3,847,120$); it 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



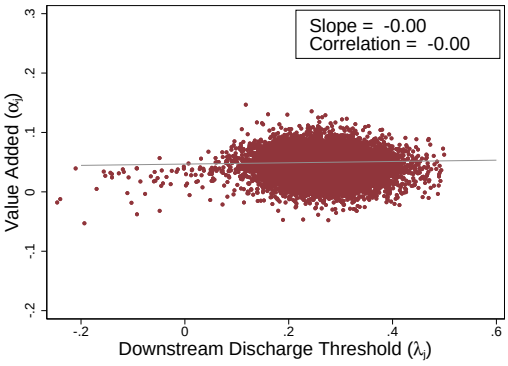
Notes: Panel A plots the non-parametric relationship between the probability of going to a SNF and distance to that SNF in the greater Chicago market from specifications that sequentially add observables interacted with choice-specific indicators ($N = 212,937$ patient stays; 294 SNFs). Panel B plots the coefficient on log distance from estimates of demand model shown in equation (4) estimated on the baseline sample ($N = 6,246,686$). Specifically, each point denotes a choice set, with the 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 and health shifters. The shifters for demographics and health are the first five 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).

Figure 4: Value Added

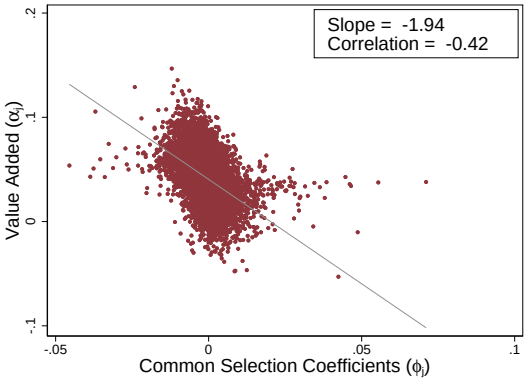
(A) National Parameters

Parameter	Estimate
θ_h	0.965
φ	0
σ_ϵ	0.075
σ_v	0.145
θ_{x1}	-0.005
θ_{x2}	-0.01
θ_{x3}	-0.02

(B) Value Added vs. Discharge Threshold



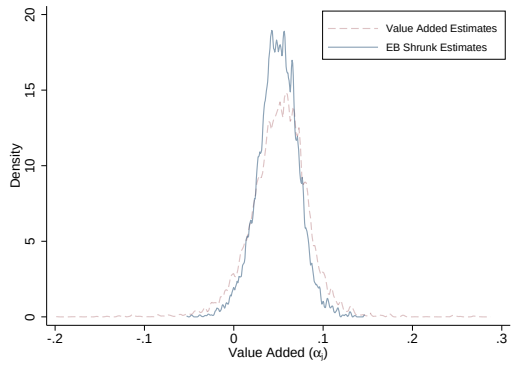
(C) Value Added vs. Selection Coefficient



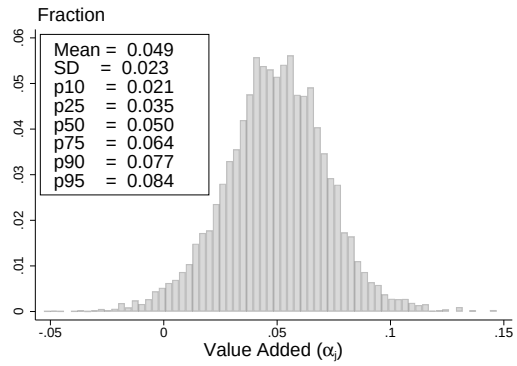
Notes: Figure reports the parameter estimates from jointly estimating equations (2), (3), and (6). Panel A reports estimates of the "national parameters" from our model. Parameter θ_{x1} is the coefficient for Medicaid dual-eligibility, θ_{x2} is the coefficient for ages 75–84, and θ_{x3} is the coefficient for ages 85+. Panel B plots the relationship between value added (α_j) and the SNF-specific downstream discharge thresholds (λ_j). Panel C 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 patient stays. The analysis weights each SNF by the number of patient stays we observe at that facility.

Figure 5: Distribution of Value Added across SNFs

(A) Kernel Densities with and without empirical Bayes (EB) Shrinkage



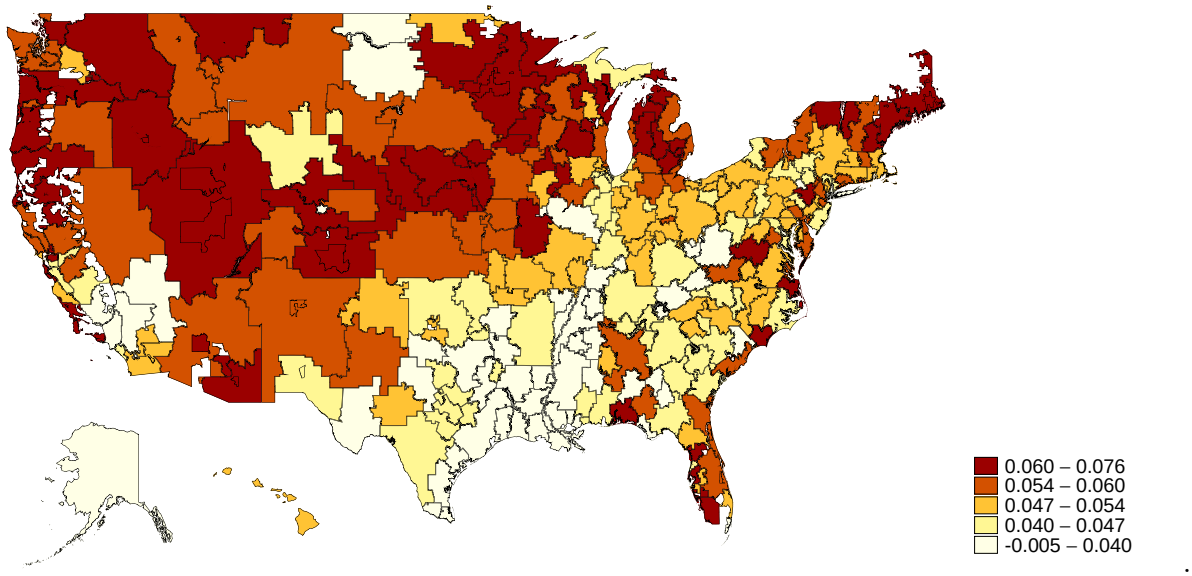
(B) Distribution of EB-Adjusted Value Added



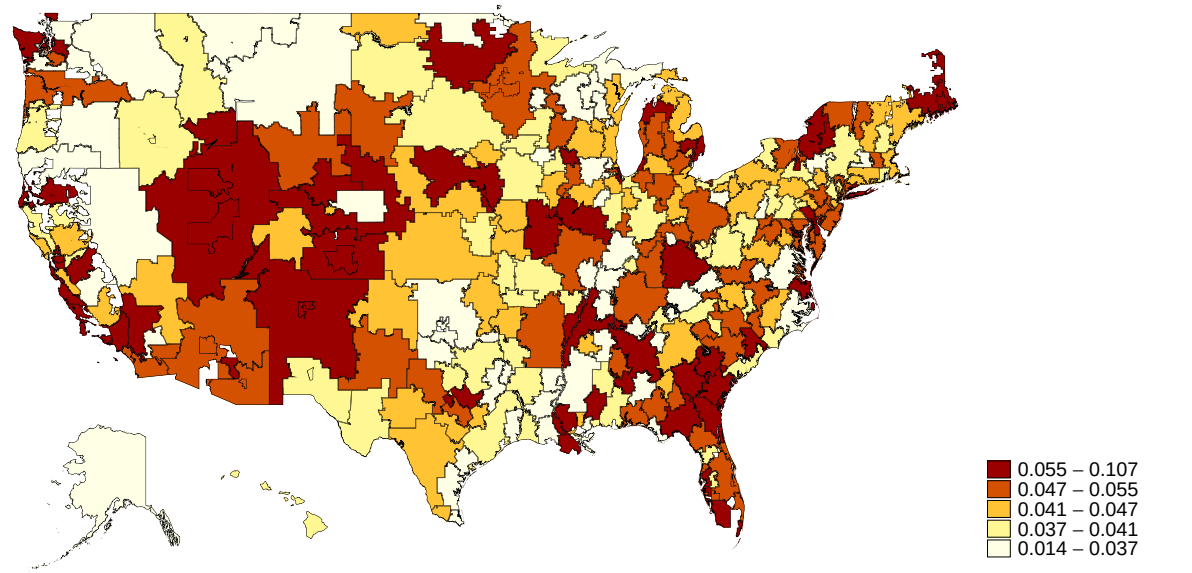
Notes: Panel A shows a kernel density of value added before and after performing empirical Bayes shrinkage to account for estimation error in the value-added estimates. The correlation between the value added estimates and their shrunken counterparts is 0.81. The standard errors of value added are the critical inputs to the empirical Bayes procedure. The average standard error is 0.015. Panel B shows the distribution of (empirical-Bayes adjusted) value added (α_j) across SNFs. The analyses weight each SNF by the number of patient stays we observe at that facility.

Figure 6: Heterogeneity Across and Within Markets

(A) Average Value Added by Market



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

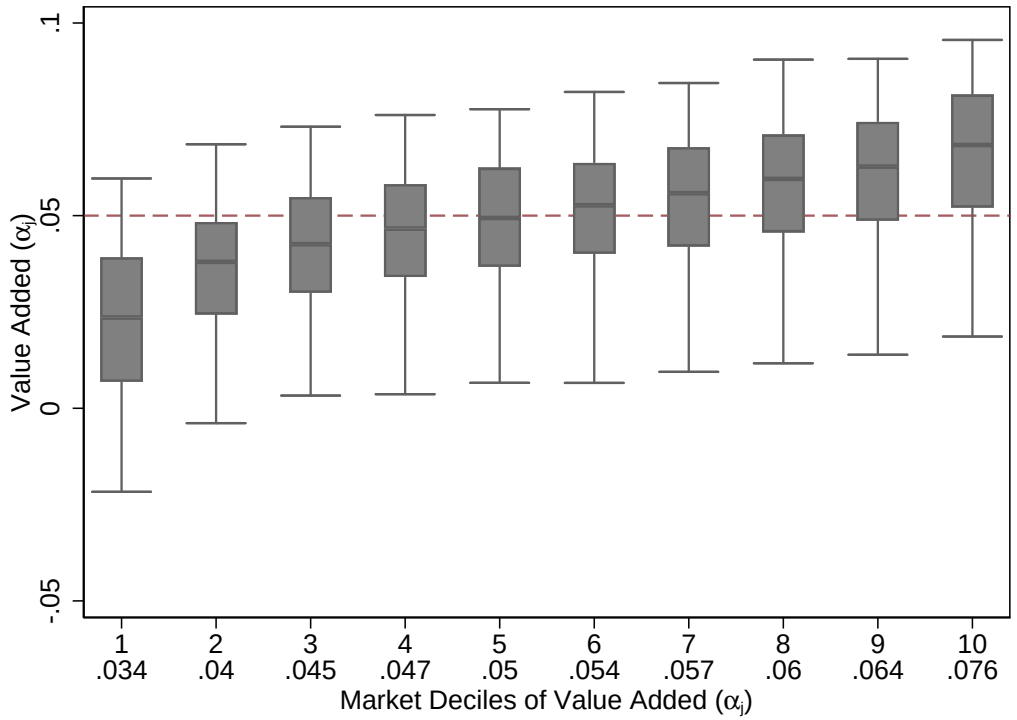


(C) Distributions of Value-Added

	Mean	SD	10th	25th	50th	75th	90th
Across SNFs	0.049	0.023	0.021	0.035	0.050	0.064	0.077
Across markets	0.049	0.012	0.035	0.044	0.050	0.057	0.063
Within markets	0.049	0.019	0.025	0.036	0.050	0.062	0.074

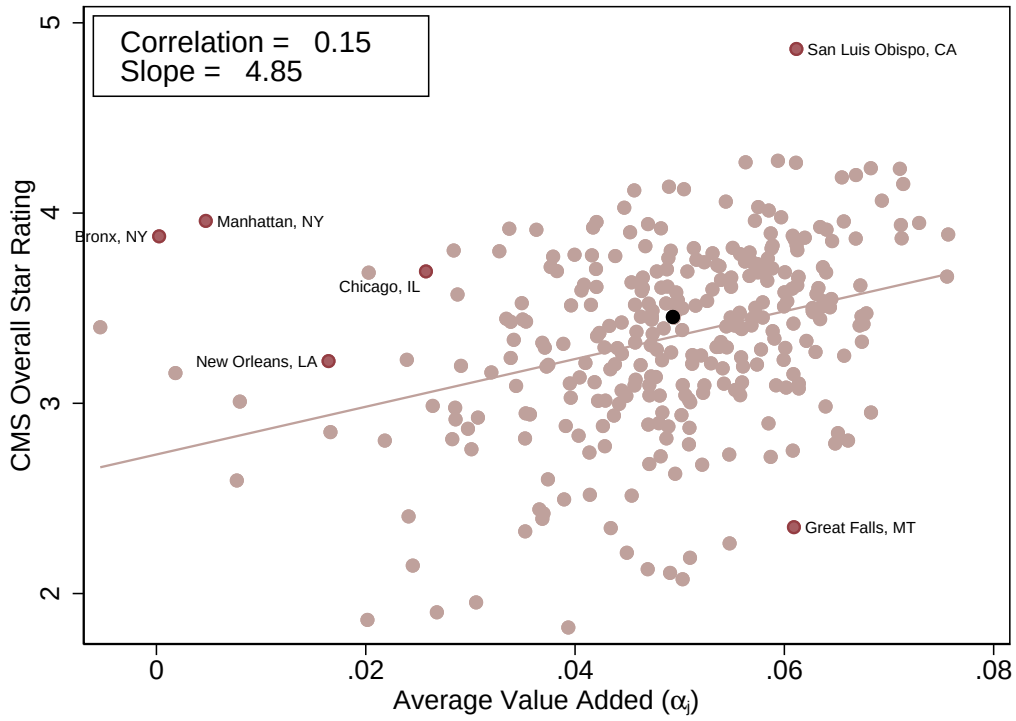
Notes: Panel (A) plots the average value added in each market. Panel (B) plots the difference between the 90th percentile and 10th percentile of the value-added distribution within each market. Summary statistics are constructed after weighting each SNF by the number of patient stays. The color gradients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

Figure 7: 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, where markets are defined by the 306 HRRs. 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 patient stays. The dashed horizontal line is the median across all SNFs in our sample, weighted by patient stays.

Figure 8: 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 patient stays. The points represent 306 markets covering the 13,646 SNFs for which we observe CMS star ratings. The black dot shows the patient-stay-weighted average of the overall rating and value added. Certain outlier markets and other markets of interest are labeled. The line is based on a linear regression with observations are weighted by the number of patient stays in each market.

TABLE I: Patient Characteristics

	Mean (1)		Std. dev. across SNFs (2)
A. Patient characteristics at admission (N = 6,246,686)			
Age	81.0	(8.3)	1.9
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 (days) (prior to SNF admission)	7.6	(7.7)	2.0
B. SNF length of stay			
Length of stay (days)	44.7	(100.9)	20.8
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 patient stays in that SNF.

TABLE II: 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 (958.5)	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 patient stays in the SNF.

TABLE III: Selected Health Index Measures

Health Measure		Mean	Std. dev.	Std. dev. across SNFs	Importance
		(1)	(2)	(3)	(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 the BIMS score, where higher values are 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 patient stays 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 of the health index model 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 IV: 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	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
C. Model-predicted decisions (h as of initial assessment, N = 6,215,699):			
All patients	1.000	0.134	0.095
Discharged to community before day-30 assessment	0.365	0.185	0.112
Still at the SNF for day-30 assessment	0.425	0.106	0.069
Discharged elsewhere before day-30 assessment	0.209	0.103	0.068

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 patient stays with non-missing 30-day assessment data. Panel B shows the health index at admission for our full sample of 6,246,686 patient stays. Panel C shows the health index at admission for model-predicted discharge groups for patients in SNFs that have estimates for all parameters.

TABLE V: Upstream discharge parameters

	Mean	SD	10th	25th	50th	75th	90th
γ^0	-0.37	0.39	-0.64	-0.54	-0.43	-0.31	-0.14
γ^1	-4.95	37.39	-2.92	-1.12	-0.15	0.47	1.07

Notes: Table reports summary statistics of the market-level upstream discharge parameters (γ_k^0 and γ_k^1) in our model for 655 markets. The patient-stay-weighted correlation between γ_k^0 and γ_k^1 is -0.96. The patient-stay-weighted correlation between winsorized γ_k^0 (dropping the bottom 5%) and γ_k^1 is -0.85, and between γ_k^0 and winsorized γ_k^1 is -0.85.

TABLE VI: Correlations of Value-Added Estimates with SNF and Patient Characteristics

	Correlation	Partial correlation controlling for market FEs
	(1)	(2)
A. SNF characteristics		
SNF occupancy	-0.015	0.000
Number of beds	-0.180	-0.063
For profit	-0.134	-0.140
Hospital-based	0.015	0.045
Chain ownership	0.013	-0.064
Medicare share	0.248	0.296
Has Alzheimer's special care unit	0.038	-0.011
B. Patient characteristics		
Health index at admission	0.480	0.386
Black patient share	-0.274	-0.233
Medicaid (aka "dually eligible") at admission	-0.463	-0.452
Female patient share	0.242	0.318
Married patient share	0.333	0.265
Mean age	0.172	0.211

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 patient stays in the SNF and calculated based on value-added estimates for 13,938 SNFs covering 306 markets. "Chain ownership" refers to whether a facility is owned or leased by a multi-facility (chain) organization.

TABLE VII: Distribution of value-added for alternative measures and specifications

	Mean	SD	10th	25th	50th	75th	90th
A. Baseline							
Across SNFs	0.049	0.023	0.021	0.035	0.050	0.064	0.077
Across HRRs	0.049	0.012	0.035	0.044	0.050	0.057	0.063
Within HRRs	0.049	0.019	0.025	0.036	0.050	0.062	0.074
B. ADLs							
Across SNFs	3.988	0.642	3.251	3.548	3.938	4.341	4.778
Across HRRs	3.988	0.244	3.712	3.831	3.965	4.146	4.285
Within HRRs	3.988	0.583	3.285	3.600	3.958	4.323	4.732
C. Excluding incentivized measures							
Across SNFs	0.045	0.020	0.022	0.033	0.044	0.057	0.070
Across HRRs	0.045	0.011	0.033	0.041	0.046	0.052	0.055
Within HRRs	0.045	0.017	0.024	0.034	0.044	0.056	0.066
D. By patient characteristics							
i. Ages 65-74							
Across SNFs	0.064	0.033	0.026	0.043	0.061	0.084	0.106
Across HRRs	0.064	0.014	0.049	0.057	0.066	0.071	0.081
Within HRRs	0.064	0.029	0.029	0.044	0.062	0.083	0.105
ii. Ages 75-84							
Across SNFs	0.055	0.033	0.020	0.035	0.053	0.075	0.094
Across HRRs	0.055	0.015	0.040	0.048	0.058	0.065	0.071
Within HRRs	0.055	0.027	0.024	0.037	0.054	0.073	0.091
iii. Ages 85+							
Across SNFs	0.047	0.025	0.019	0.032	0.047	0.063	0.078
Across HRRs	0.047	0.012	0.030	0.042	0.051	0.055	0.060
Within HRRs	0.047	0.021	0.022	0.033	0.047	0.060	0.074
iv. Not dual-eligible							
Across SNFs	0.062	0.032	0.024	0.042	0.061	0.082	0.101
Across HRRs	0.062	0.015	0.044	0.056	0.065	0.072	0.078
Within HRRs	0.062	0.028	0.028	0.043	0.062	0.080	0.098
v. Dual-eligible							
Across SNFs	0.025	0.036	-0.018	0.002	0.024	0.047	0.071
Across HRRs	0.025	0.021	-0.001	0.010	0.021	0.039	0.052
Within HRRs	0.025	0.029	-0.010	0.006	0.024	0.043	0.062

Notes: Table reports summary statistics for value added estimates for various models across and within markets.

TABLE VIII: Correlations of Value-Added Estimates with Alternative Measures

	Correlation (1)	Partial correlation controlling for market FEs (2)
A. Alternative health indices		
ADLs	0.56	0.53
Excluding incentivized health measures	0.85	0.81
B. CMS star ratings		
Overall	0.23	0.25
Quality	0.02	0.07
Inspection	0.14	0.19
Staffing	0.35	0.28
C. 90-day outcomes (*signed so that higher is better*)		
Share of days in facilities 90 days	0.39	0.41
Home at 90 days	0.40	0.42
Spending per day 90 days	0.35	0.32
SNF readmission 90 days	0.22	0.22
Hospital readmission 90 days	0.15	0.08
Alive at 90 days	0.21	0.29
D. Other specifications		
Average health	0.63	0.58
Average health improvement	0.72	0.69
Average health improvement + select in	0.73	0.72
Average health improvement + select out	0.69	0.61

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 correlations with alternative health indices defined using the same health assessment data as in our baseline estimates. See Section 6.1 for details. Panel B shows the correlation with the overall five-star rating and its three sub-components. Panel C 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. See Section 6.4 for details. Panel D shows correlations with alternative econometric specifications of value added using the health index. Correlations are weighted by the number of patient stays in the SNF and calculated based on value-added estimates for 13,938 SNFs covering 306 markets, except for Panel B which is based on the 13,646 SNFs (covering 306 markets) for which we observe CMS five-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 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 and 8). Most dropped stays are missing responses to all of the questions on the health assessment.

Of the 6.2 million patient stays covered by our final sample (see Appendix Table A.I), 2.5 million are 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 E).

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 A.II. The vast majority of the 109 measures are missing for fewer 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 A.II, 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 (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 seven during which the resident received walking training for at least 15 minutes. 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 Choice set ("sub-HRR") construction

To define the choice set of SNFs for each patient, we begin with the standard market definition for hospital choice, namely the 306 Hospital Referral Regions (HRRs), each of which is composed of one or more Hospital Service Areas (HSAs). There are 3,436 HSAs in the United States. Of those, 276 (8%) HSAs have no SNF in our data. Of the remaining 3,160 HSAs, 315 (9% of all HSAs) have some SNFs but no SNF with more than 50 episodes, leaving 2,845 HSAs in our baseline sample restriction.

To make our choice sets more computationally tractable, we split the largest HRRs into smaller markets comprised of geographically connected HSA groups as follows: We would like each sub-HRR to have more than 80% of its episodes in the top 20 largest SNFs. Out of the 306 HRRs, 188 already satisfy this criterion, and 118 are split into sub-HRRs. To ensure that the sub-HRRs to be geographically connected, we use k-means clustering on the longitude and latitude of HSA centroids. For HRRs with less than 80% of their episodes in the top 20 largest SNFs, we split HRRs by the following recursive method: Split each HRR into two groups by k-means clustering. Then, for each sub-HRR, if it has less than 80% of its episodes in the top 20 largest SNFs, split that sub-HRR further into two groups. Stop until every sub-HRR either has more than 80% of its episodes in the top 20 largest SNFs, or has only one HSA in it.

In summary, of the 306 original HRRs, 188 of them are not split. The 118 HRRs that are split are comprised of 467 sub-HRRs. This yields a total of 655 sub-HRRs.

C 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 had 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 A.III 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 estimate 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 so, recall that we measure health at admission (typically around day 5), at 14 days (for people still in the SNF), and at 30 days (see Section 2). 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 A.III. 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 A2 shows the predicted probability of being discharged to the community at $t = 1$, $t = 1.33$, and $t = 2$. We see these functions lie essentially 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.

D 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.[42]

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 patient i's idiosyncratic preferences for SNF j. SNF value added, α_j, is given by

$$\mathbb{E}\left[h_{i2} \mid h_{i1}, x_i, \eta_{i1}, \ldots, \eta_{i|J_i|}, c_i\right] = \alpha_j + \theta_h h_{i1} + x_i'\theta_x + g_j(\eta_{i1}, \ldots, \eta_{i|J_i|}),$$

where $g_j(\eta_{i1}, \ldots, \eta_{i|J_i|})$ allows period 2 health to vary flexibly with unobserved preferences (i.e., to allow for arbitrary selection patterns), c_i is a categorical variable that indicates the SNF chosen by patient i, and J_i is patient i's choice set. 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 $\eta_{i1}, \ldots, \eta_{ij}$ according to

$$\mathbb{E}\left[h_{i2} \mid h_{i1}, x_i, \eta_{i1}, \ldots, \eta_{i|J_i|}, c_i\right] = \alpha_j + \theta_h h_{i1} + x_i'\theta_x + \underbrace{\sum_{\ell \in J_i} \phi_\ell(\eta_{i\ell} - \mu_\eta)}_{\text{Preferences not specific to } j} + \underbrace{\varphi(\eta_{ij} - \mu_\eta)}_{\text{Selection specific to } j},$$

where μ_η is the mean of the logit shocks (Euler's constant) and acts as a normalization. The term $\sum_{\ell \in J_i} \phi_\ell(\eta_{i\ell} - \mu_\eta)$ captures preferences that have effects on the outcome that are not

[42] 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 $\delta_k(x_i)$.

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 $\eta_{i1}, \ldots, \eta_{ij}$ yields

$$\mathbb{E}\left[h_{i2} \mid h_{i1}, x_i, \eta_{i1}, \ldots, \eta_{i|J_i|}, c_i\right] = \alpha_j + \theta_h h_{i1} + x_i'\theta_x + \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell}(j) + \varphi \beta_{ij}(j),$$

where $\beta_{i\ell}(j) = E[\eta_{i\ell} - \mu_\eta \mid c_i = j]$ are the control functions. Letting j indicate the chosen SNF, we will show below that the control functions are

$$\beta_{i\ell}(j) = \begin{cases} -\log \hat{p}_{i\ell} & l = j \\ \frac{\hat{p}_{i\ell}}{1-\hat{p}_{i\ell}} \log \hat{p}_{i\ell} & \text{otherwise} \end{cases}, \tag{9}$$

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

We will now derive the expressions for the control functions in (9). Again, letting j denote the chosen SNF, we know from rearranging the utility function that

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

From Small and Rosen (1981), we know

$$\mathbb{E}[u_{ij} \mid c_i = j] = \log \left[\sum_{\ell \in J_i} \exp(\delta_\ell - \tau m_{i\ell})\right] + \mu_\eta. \tag{10}$$

Substitution yields

$$\begin{aligned} \mathbb{E}[\eta_{ij} - \mu_\eta \mid c_i = j] &= \log \left[\sum_{\ell \in J_i} \exp(\delta_\ell - \tau m_{i\ell})\right] + \mu_\eta - \delta_j + \tau m_{ij} - \mu_\eta \\ &= \log \left[\sum_{\ell \in J_i} \exp(\delta_\ell - \tau m_{i\ell})\right] - \log\left[\exp(\delta_j - \tau m_{ij})\right] \\ &= -\log \left[\frac{\exp(\delta_j - \tau m_{ij})}{\sum_{\ell \in J_i} \exp(\delta_\ell - \tau m_{i\ell})}\right] \\ &= -\log \hat{p}_{ij}, \end{aligned}$$

which is the first expression in (9).

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

$$\mathbb{E}[\eta_{i\ell} - \mu_\eta \mid c_i = j] = \mathbb{E}[u_{i\ell} \mid c_i = j] - \delta_\ell + \tau m_{i\ell} - \mu_\eta. \tag{11}$$

By properties of conditional expectation, we know

$$\mathbb{E}[u_{i\ell}] = \mathbb{P}(c_i = \ell)\mathbb{E}[u_{i\ell} \mid c_i = \ell] + \mathbb{P}(c_i \neq \ell)\mathbb{E}[u_{i\ell} \mid c_i \neq \ell]. \tag{12}$$

Using the result (10) from Small and Rosen (1981), as before, and substituting the above expression into the expected utility equation yields

$$\delta_\ell - \tau m_{i\ell} + \mu_\eta = \mathbb{P}(c_i = \ell)\left[\log\left(\sum_{\ell' \in J_i} \exp(\delta_{\ell'} - \tau m_{i\ell'})\right) + \mu_\eta\right] + \mathbb{P}(c_i \neq \ell)\mathbb{E}[u_{i\ell} \mid c_i \neq \ell],$$

which implies

$$\begin{aligned}\mathbb{E}[u_{i\ell} \mid c_i \neq \ell] = \frac{1}{1 - \mathbb{P}(c_i = \ell)}\Bigg\{&\delta_\ell - \tau m_{i\ell} + \mu_\eta \\ &- \mathbb{P}(c_i = \ell)\left[\log\left(\sum_{\ell' \in J_i} \exp(\delta_{\ell'} - \tau m_{i\ell'})\right) + \mu_\eta\right]\Bigg\}.\end{aligned}$$

Substitution of the above into (11) then yields

$$\begin{aligned}E[\eta_{i\ell} - \mu_\eta \mid c_i = j] &= \frac{1}{1 - \mathbb{P}(c_i = \ell)}\Bigg\{\delta_\ell - \tau m_{i\ell} + \mu_\eta \\ &\quad - \mathbb{P}(c_i = \ell)\left[\log\left(\sum_{\ell' \in J_i} \exp(\delta_{\ell'} - \tau m_{i\ell'})\right) + \mu_\eta\right]\Bigg\} - \delta_\ell + \tau m_{i\ell} - \mu_\eta \\ &= \frac{1}{1 - \mathbb{P}(c_i = \ell)}\Bigg\{\delta_\ell - \tau m_{i\ell} + \mu_\eta - (1 - \mathbb{P}(c_i = \ell))\left[\delta_\ell - \tau m_{i\ell} + \mu_\eta\right] \\ &\quad - \mathbb{P}(c_i = \ell)\left[\log\left(\sum_{\ell' \in J_i} \exp(\delta_{\ell'} - \tau m_{i\ell'})\right) + \mu_\eta\right]\Bigg\} \\ &= \frac{\mathbb{P}(c_i = \ell)}{1 - \mathbb{P}(c_i = \ell)}\left[\delta_\ell - \tau m_{i\ell'} - \log\left(\sum_{\ell' \in J_i} \exp\left(\delta_{\ell'} - \tau m_{i\ell'}\right)\right)\right] \\ &= \frac{\mathbb{P}(c_i = \ell)}{1 - \mathbb{P}(c_i = \ell)}\left[\log\left[\exp\left(\delta_\ell - \tau m_{i\ell}\right)\right] - \log\left(\sum_{\ell' \in J_i} \exp\left(\delta_{\ell'} - \tau m_{i\ell'}\right)\right)\right] \\ &= \frac{\mathbb{P}(c_i = \ell)}{1 - \mathbb{P}(c_i = \ell)} \log\left(\frac{\exp\left(\delta_\ell - \tau m_{i\ell}\right)}{\sum_{\ell' \in J_i} \exp\left(\delta_{\ell'} - \tau m_{i\ell'}\right)}\right) \\ &= \frac{\hat{p}_{i\ell}}{1 - \hat{p}_{i\ell}} \log \hat{p}_{i\ell}. \end{aligned} \tag{13}$$

E 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 A.V). 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 within 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 A.VI provides more details on the procedure.

Appendix Table A.VII 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 A.V, 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.

F Estimation Details

F.1 Deriving the Likelihood

The health process and the 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

$$L = \prod_i \underbrace{\left[\mathbb{P}(d_i^D = 1)\right]^{d_i^D}}_{\text{downstream}} \underbrace{\left[\mathbb{P}(d_i^U = 1)\right]^{d_i^U}}_{\text{upstream}} \underbrace{\left[\mathbb{P}(d_i^D = 0 \cap d_i^U = 0 \mid h_{i2}) f(h_{i2})\right]^{(1-d_i^U)(1-d_i^D)}}_{\text{still in SNF}}, \quad (14)$$

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

We will derive an explicit formula for L by considering each partition separately. 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

$$\begin{aligned}
\mathbb{P}(d_i^D = 1) &= \mathbb{P}(h_{i2} \geq \lambda_j + \nu_i) \\
&= \mathbb{P}\left(\alpha_j + \theta_h h_{i1} + x_i'\theta_x + \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} + \varphi\beta_{ij} + \epsilon_i \geq \lambda_j + \nu_i\right) \\
&= \mathbb{P}\left(\epsilon_i - \nu_i \geq \lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} - \varphi\beta_{ij}\right) \\
&= 1 - \mathbb{P}\left(\epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} - \varphi\beta_{ij}\right) \\
&= 1 - \Phi\left(\frac{\lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} - \varphi\beta_{ij}}{\sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}}\right) \\
&= \Phi\left(\frac{\alpha_j + \theta_h h_{i1} + x_i'\theta_x + \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} + \varphi\beta_{ij} - \lambda_j}{\sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}}\right).
\end{aligned}$$

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 $\epsilon_i - \nu_i \sim N(0, \sigma_\epsilon^2 + \sigma_\nu^2)$.

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

$$\begin{aligned}
\mathbb{P}(d_i^U = 1) &= \mathbb{P}(d_i^D = 0)\mathbb{P}(d_i^U = 1 \mid d_i^D = 0) \\
&= \Phi\left(\frac{\lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} - \varphi\beta_{ij}}{\sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}}\right) \mathbb{P}(d_i^U = 1 \mid d_i^D = 0).
\end{aligned}$$

Now consider $\mathbb{P}(d_i^U = 1 \mid d_i^D = 0)$. Since we model $\mathbb{P}(d_i^U = 1 \mid d_i^D = 0, h_{i2}) = \Phi(\gamma_k^0 + \gamma_k^1 h_{i2})$, we have that

$$\mathbb{P}(d_i^U = 1 \mid d_i^D = 0) = \int \tilde{\phi}(\epsilon_i)\Phi\left(\gamma_k^0 + \gamma_k^1\left(\alpha_j + \theta_h h_{i1} + x_i'\theta_x + \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} + \varphi\beta_{ij} + \epsilon_i\right)\right) d\epsilon_i, \tag{15}$$

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

$$\begin{aligned}
\mathbb{P}(d_i^U = 1) = \Phi&\left(\frac{\lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} - \varphi\beta_{ij}}{\sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}}\right) \\
&\cdot \int \tilde{\phi}(\epsilon_i)\Phi\left(\gamma_k^0 + \gamma_k^1\left(\alpha_j + \theta_h h_{i1} + x_i'\theta_x + \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} + \varphi\beta_{ij} + \epsilon_i\right)\right) d\epsilon_i
\end{aligned}$$

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

0. We know

$$\mathbb{P}(d_i^D = 0 \cap d_i^U = 0 \mid h_{i2}) f(h_{i2}) = \mathbb{P}(d_i^D = 0 \mid h_{i2}) \mathbb{P}(d_i^U = 0 \mid d_i^D = 0, h_{i2}) f(h_{i2}). \tag{16}$$

It is trivial that

$$f(h_{i2}) = \phi\left(\frac{h_{i2} - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} - \varphi \beta_{ij}}{\sigma_\epsilon}\right), \tag{17}$$

where $\phi(\cdot)$ is the standard normal probability density function. Similarly, by assumption we know that $\mathbb{P}(d_i^U = 0 \mid d_i^D = 0, h_{i2}) = 1 - \Phi(\gamma_k^0 + \gamma_k^1 h_{i2})$. Now consider $\mathbb{P}(d_i^D = 0 \mid h_{i2})$. We know that

$$\begin{aligned}
\mathbb{P}(d_i^D = 0 \mid h_{i2}) &= \mathbb{P}(h_{i2} < \lambda_j + \nu_i \mid h_{i2}) \\
&= \mathbb{P}\left(\epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} - \varphi \beta_{ij} \;\middle|\; h_{i2}\right) \\
&= \mathbb{P}\left(\epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} - \varphi \beta_{ij} \;\middle|\; \epsilon_i\right).
\end{aligned}$$

In order to derive an expression for the above quantity, 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,

$$\begin{pmatrix} \epsilon_i \\ \epsilon_i - \nu_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 & \sigma_\epsilon^2 \\ \sigma_\epsilon^2 & \sigma_\epsilon^2 + \sigma_\nu^2 \end{pmatrix}\right). \tag{18}$$

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

$$\epsilon_i - \nu_i \mid \epsilon_i \sim N\left(\epsilon_i, \sigma_\nu^2\right). \tag{19}$$

Hence, we have

$$\begin{aligned}
\mathbb{P}(d_i^D = 0 \mid h_{i2}) &= \Phi\left(\frac{\lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} - \varphi \beta_{ij} - \epsilon_i}{\sigma_\nu}\right) \\
&= \Phi\left(\frac{\lambda_j - h_{i2}}{\sigma_\nu}\right).
\end{aligned} \tag{20}$$

Together this yields

$$\begin{aligned}
\mathbb{P}(d_i^D = 0 \cap d_i^U = 0 \mid h_{i2}) f(h_{i2}) = \Phi\left(\frac{\lambda_j - h_{i2}}{\sigma_\nu}\right) &\left(1 - \Phi(\gamma_k^0 + \gamma_k^1 h_{i2})\right) \\
&\cdot \phi\left(\frac{h_{i2} - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} - \varphi \beta_{ij}}{\sigma_\epsilon}\right).
\end{aligned}$$

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

$$L = \prod_i \underbrace{\left[\Phi\left(\frac{\alpha_j + \theta_h h_{i1} + x_i'\theta_x + \sum_{\ell \in J_i}\phi_\ell \beta_{i\ell} + \varphi\beta_{ij} - \lambda_j}{\sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}}\right)\right]^{d_i^D}}_{\text{downstream}}$$
$$\cdot \underbrace{\left[\Phi\left(\frac{\lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i}\phi_\ell\beta_{i\ell} - \varphi\beta_{ij}}{\sqrt{\sigma_\epsilon^2 + \sigma_\nu^2}}\right)\int \tilde{\phi}(\epsilon_i)\Phi\left(\gamma_k^0 + \gamma_k^1\left(\alpha_j + \theta_h h_{i1} + x_i'\theta_x + \sum_{\ell \in J_i}\phi_\ell\beta_{i\ell} + \varphi\beta_{ij} + \epsilon_i\right)\right) d\epsilon_i\right]^{d_i^U}}_{\text{upstream}}$$
$$\cdot \underbrace{\left[\Phi\left(\frac{\lambda_j - h_{i2}}{\sigma_\nu}\right)\left(1 - \Phi(\gamma_k^0 + \gamma_k^1 h_{i2})\right)\phi\left(\frac{h_{i2} - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i}\phi_\ell\beta_{i\ell} - \varphi\beta_{ij}}{\sigma_\epsilon}\right)\right]^{(1-d_i^U)(1-d_i^D)}}_{\text{still in SNF}}. \tag{21}$$

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

F.2 Estimation

For computational reasons, we estimate the model in two steps by partitioning parameters into two groups: "national" parameters, $(\theta_h, \varphi, \sigma_\epsilon, \sigma_\nu, \theta_{x1}, \theta_{x2}, \theta_{x3})$, and market-level or SNF-level parameters, $(\alpha_j, \lambda_j, \phi_j, \gamma_k^0, \gamma_k^1)$. 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. This approach is computationally attractive since it is easily parallelizable.

The estimation process is as follows.

1. Specify grids for the "national" parameters $(\theta_h, \varphi, \sigma_\epsilon, \sigma_\nu, \theta_{x1}, \theta_{x2}, \theta_{x3})$.

For all combinations of points on these grids,

2. Maximize the likelihood with respect to $(\alpha_j, \lambda_j, \phi_j, \gamma_k^0, \gamma_k^1)$ within each market holding $(\theta_h, \varphi, \sigma_\epsilon, \sigma_\nu, \theta_{x1}, \theta_{x2}, \theta_{x3})$ 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

- $\theta_h \in \{0.96, 0.965, 0.97\}$
- $\sigma_\epsilon \in \{0.07, 0.075, 0.08\}$
- $\sigma_\nu \in \{0.14, 0.145, 0.15\}$
- $\varphi \in \{-0.005, 0, 0.005\}$

- $\theta_{x1} \in \{-0.01, -0.005, 0\}$
- $\theta_{x2} \in \{-0.015, -0.01, -0.005\}$
- $\theta_{x3} \in \{-0.025, -0.02, -0.015\}$.

Another problem to tackle is that of estimating the integral in (equation (15)), which enters the likelihood for patients who are discharged upstream. Due to runtime concerns, we estimate this integral using rectangular quadrature. We explored the sensitivity of our results to this method by focusing on a large market and estimating the error 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 also tried limiting the boundary to within $4\sigma_\epsilon$, $5\sigma_\epsilon$, and $6\sigma_\epsilon$ of the mean. Our results were essentially invariant to the choice of boundary, so in practice, we limit the domain of integration to be within $4\sigma_\epsilon$ of the mean. The error percentage was less than 0.01% across all of the tests with equally spaced nodes.

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

In the likelihood derived in Appendix F.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:

$$\begin{aligned}
\tilde{\phi}(x) &= \frac{d}{dx}\left[\mathbb{P}\left(\epsilon_i \leq x \mid d_i^D = 0\right)\right] \\
&= \frac{d}{dx}\left[\mathbb{P}\left(\epsilon_i \leq x \mid h_{i2} < \lambda_j + \nu_i\right)\right] \\
&= \frac{d}{dx}\left[\mathbb{P}\left(\epsilon_i \leq x \;\middle|\; \epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} - \varphi\beta_{ij}\right)\right].
\end{aligned} \tag{22}$$

To understand the functional form of this expression, we must first revisit the joint distribution of ϵ_i and $\epsilon_i - \nu_i$. Recall from Appendix F.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

$$\begin{pmatrix} \epsilon_i \\ \epsilon_i - \nu_i \end{pmatrix} \;\middle|\; \epsilon_i - \nu_i < \lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell \in J_i} \phi_\ell \beta_{i\ell} - \varphi\beta_{ij}$$

follows a bivariate truncated normal distribution, so (22) is the marginal density of a bivariate truncated normal cumulative density function. Hence, we have that

$$\tilde{\phi}(x) = \frac{1}{C}\phi\left(\frac{x}{\sigma_\epsilon}\right)\Phi\left(\frac{\lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell\in J_i}\phi_\ell\beta_{i\ell} - \varphi\beta_{ij} - x}{\sigma_\nu}\right), \tag{23}$$

where $\phi(\cdot)$ is the standard normal probability density function, and C is the total probability in the truncated distribution of $(\epsilon_i, \epsilon_i - \nu_i)$ and is given by

$$C = \int_{-\infty}^{\infty}\phi\left(\frac{x}{\sigma_\epsilon}\right)\Phi\left(\frac{\lambda_j - \alpha_j - \theta_h h_{i1} - x_i'\theta_x - \sum_{\ell\in J_i}\phi_\ell\beta_{i\ell} - \varphi\beta_{ij} - x}{\sigma_\nu}\right)dx.$$

F.4 Standard Errors

To obtain standard errors for the parameters estimated from equation (6), we perform a two-step score bootstrap procedure similar to the approach in Abdulkadiroğlu et al. (2020). This procedure adjusts inference for the extra uncertainty introduced by the choice model estimates without recalculating the first-step estimates or requiring that we analytically derive the influence of the first-step estimates on the second-step estimates.

Recall that our first-step choice model in equation (4) has average utility of the form

$$\delta_j(x_i) = \delta_{0j} + \delta_{1j}h_{i1} + \delta_{2j}g_i + \delta_{3j}a_{2i} + \delta_{4j}a_{3i}, \tag{24}$$

where g_i is an indicator for whether patient i is dually-eligible for Medicaid, a_i is patient i's age, $a_{2i} = 1\{75 \leq a_i \leq 84\}$, and $a_{3i} = 1\{85 \leq a_i\}$. Let Δ denote the vector of choice model parameters for a given market (which includes all $\{\delta_{\ell j}\}_{\ell=0}^5$ and τ) and $\hat{\Delta}$ denote the parameters estimated via maximum likelihood.

Our second-step model is specified by equations (2), (3), and (6). Let

$$\Gamma = (\alpha_1, \ldots, \alpha_J, \lambda_1, \ldots, \lambda_J, \phi_0, \ldots, \phi_J, \gamma_k^0, \gamma_k^1)'$$

be the parameters that are estimated in the inner loop of the second-step optimization (omitting the national parameters $(\theta_h, \varphi, \sigma_\epsilon, \sigma_\nu, \theta_{x1}, \theta_{x2}, \theta_{x3})$ optimized in the grid search) and $\hat{\Gamma}$ be the parameters estimated via maximum likelihood.

Fix the national parameters at their optimized values. For a given market of n patients, we generate market-specific bootstrap distributions for Δ and Γ via the following steps for each bootstrap trial $b \in \{1, \ldots B\}$:

1. Generate a bootstrap distribution for Δ by taking repeated Newton-Raphson steps from the full-sample estimates and randomly reweighting each observation's score contribution. The first-step bootstrap estimate of Δ in trial b is

$$\hat{\Delta}^b = \hat{\Delta} - H_1^{-1}(\hat{\Delta}) \times \sum_{i=1}^{n}\zeta_i^b s_1(i, \hat{\Delta}), \tag{25}$$

   where $H_1(\hat{\Delta})$ is the first-step Hessian and $s_1(i, \hat{\Delta})$ is the first-step score contribution

of patient i at the optimal parameter vector $\hat{\Delta}$, while $\{\zeta_i^b\}_{i=1}^n \sim N(0,1)$ is a set of i.i.d. standard normal weights.

2. Using an additional set of Newton-Raphson steps, generate a second-step bootstrap estimate of Γ as

$$\hat{\Gamma}^b = \hat{\Gamma} - H_2^{-1}(\hat{\Gamma};\hat{\Delta}) \times \sum_{i=1}^{n} \left[\zeta_i^b s_2(i,\hat{\Gamma};\hat{\Delta}) - s_2(i,\hat{\Gamma};\hat{\Delta^b})\right], \tag{26}$$

where $H_2(\hat{\Gamma};\hat{\Delta})$ and $s_2(i,\hat{\Gamma};\hat{\Delta})$ are the second-stage Hessian and score contribution of individual i, respectively, at the two-step optima $(\hat{\Gamma},\hat{\Delta})$. The term $s_2(i,\hat{\Gamma};\hat{\Delta}^b)$ plugs the bootstrap estimate $\hat{\Delta}^b$ from (25) into the second-step score contribution function $s_2(i,\hat{\Gamma};\cdot)$ to account for the additional variability in the second-step score due to the first-step estimate $\hat{\Delta}$.

We then construct standard errors for Γ using the distribution of $\hat{\Gamma^b}$ across bootstrap trials. This approach yields market-by-market standard errors for all parameters aside from the national parameters.

G Empirical Bayes shrinkage

Suppose true value added for SNF j is α_j. Our estimates of value added are

$$\hat{\alpha}_j = \alpha_j + \eta_j,$$

where α_j is the true underlying value added and η_j is measurement noise. The goal of empirical Bayes (EB) shrinkage is to adjust value added estimates to reduce sampling variance at the cost of increased bias, yielding a minimum mean squared error (MSE) prediction of α_j (Morris 1983). Aside from reducing MSE, EB shrinkage also eliminates attenuation bias that would arise in models using α_j as a regressor (Jacob and Lefgren 2008).

Assume first that the parameters of the underlying distribution of known. Specifically, we assume the following distribution of estimated value added $\hat{\alpha}_j$:

$$\hat{\alpha}_j \mid \alpha_j, \pi_j^2 \sim N(\alpha_j, \pi_j^2), \quad \alpha_j \sim N(\mu, \sigma^2),$$

where π_j^2, μ, and σ^2 are known. Note that μ and σ^2 are common across SNFs. The posterior distribution of α_j is then

$$\alpha_j \mid \hat{\alpha}_j, \pi_j^2, \mu, \sigma^2 \sim N(\alpha_j^{EB}, \pi_j^2(1-b_j)),$$

where

$$\alpha_j^{EB} = (1-b_j)\hat{\alpha}_j + b_j\mu$$
$$b_j = \frac{\pi_j^2}{\pi_j^2 + \sigma^2}.$$

Written in this form, we see that the EB estimator α_j^{EB} "shrinks" the original estimates $\hat{\alpha}_j$ towards the prior mean μ.

The primary issue that arises from this framework is that π_j^2, μ, and σ^2 are unknown and must instead be estimated. First, we can estimate σ^2 by squaring the standard errors from the two-step score bootstrap to produce $\hat{\sigma}^2$.

Next, we estimate μ and σ along the lines of of Morris (1983). Quantities $\hat{\mu}$ and $\hat{\sigma}^2$ are simultaneously determined, so we initialize uniform weights $w_j = 1$ for all j and then estimate via iteration:

1. Compute $\hat{\mu}$ then $\hat{\sigma}^2$ using the expressions

$$\hat{\mu} := \frac{\sum_j w_j \hat{\alpha}_j}{\sum_j w_j}$$
$$\hat{\sigma}^2 = \max\left\{0, \frac{\sum_j w_j \left\{\left(\frac{J}{J-1}\right)(\hat{\alpha}_j - \hat{\mu})^2 - \hat{\pi}_j^2\right\}}{\sum_j w_j}\right\}.$$

2. If on the second or greater iteration and $\hat{\sigma}^2$ has converged, exit. Otherwise, fix new w_j using the expression

$$w_j = \frac{1}{\hat{\pi}_j^2 + \hat{\sigma}^2}$$

and return to step 1.

This algorithm yields the feasible posterior mean

$$\begin{aligned} \alpha_j^{EB(f)} &= (1 - \hat{b}_j)\hat{\alpha}_j + \hat{b}_j \hat{\mu} \\ \hat{b_j} &= \left(\frac{J-2}{J}\right)\left(\frac{\hat{\pi}_j^2}{\hat{\pi}_j^2 + \hat{\sigma}^2}\right). \end{aligned} \tag{27}$$

The same procedure can be used to shrink SNF-specific parameters other than value added.

H Model-Predicted Discharge Destinations

One potential test of the fit of our model is to pass the original value of h_{i1} through the model to produce a set of discharge decisions, then compare the share of patients discharged downstream, discharged upstream, and still in the SNF to the actual shares observed in the data.

Specifically, we simulate the discharge decisions as follows. First, taking the initial distribution of h_{i1} for all patients, simulate the distribution of 30-day health as

$$\hat{h}_{i2} = \hat{\alpha}_j + \hat{\theta}_h h_{i1} + x_i' \hat{\theta}_x + \sum_{\ell \in J_i} \hat{\phi}_\ell \hat{\beta}_{i\ell} + \hat{\varphi} \hat{\beta}_{ij}.$$

Next, draw ϵ_i and ν_i for each patient as

$$\begin{pmatrix} \epsilon_i \\ \nu_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_\epsilon^2 & 0 \\ 0 & \hat{\sigma}_\nu^2 \end{pmatrix} \right).$$

Then simulate the downstream discharge decision according to the rule

$$\hat{d}_i^D = 1 \iff \hat{h}_{i2} \geq \hat{\lambda}_j + \nu_i.$$

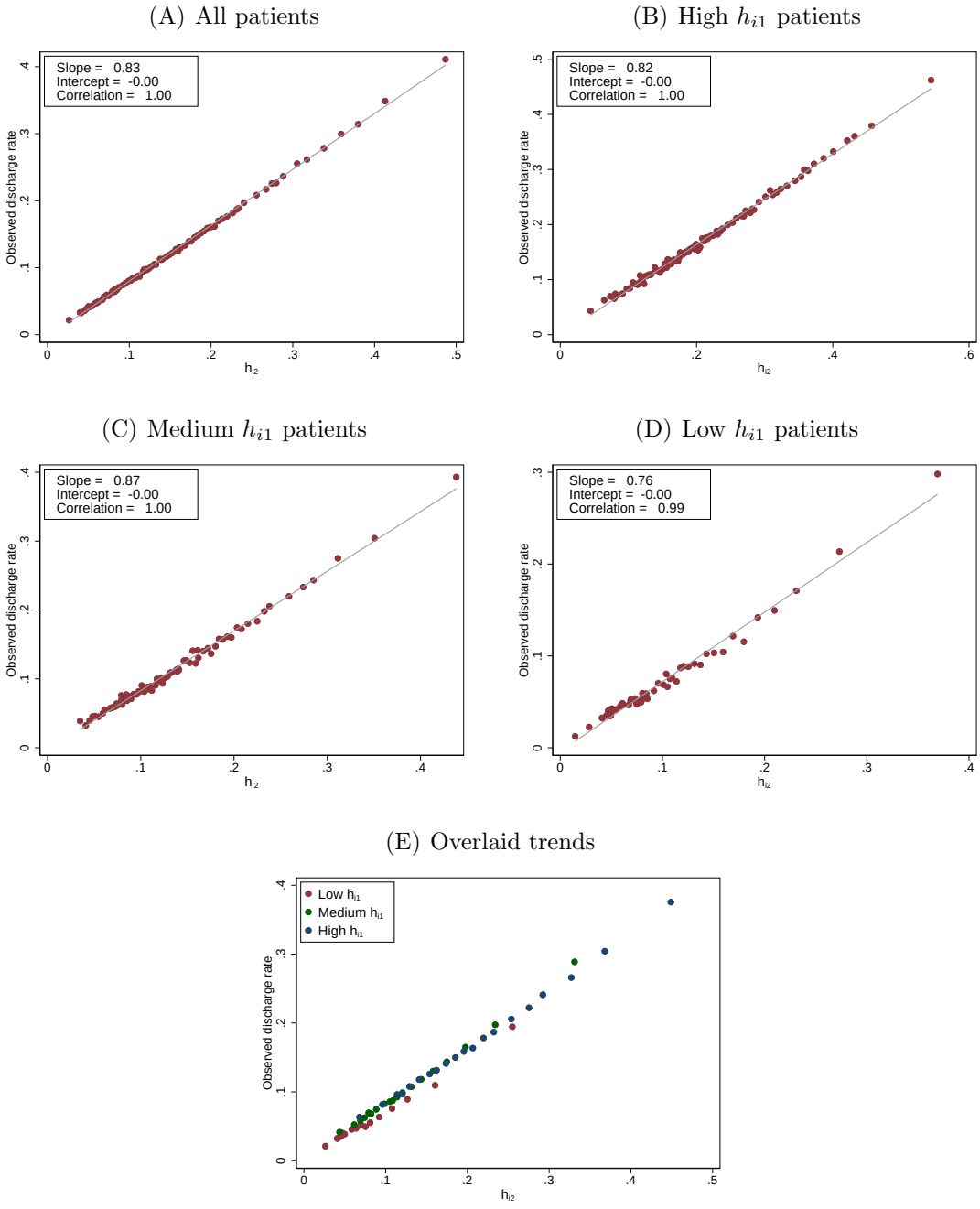
The upstream discharge decision (conditional on $\hat{d}_i^D = 0$) is then simulated as

$$\mathbb{P}(\hat{d}_i^U = 1 \mid \hat{d}_i^D = 0) = \Phi(\hat{\gamma}_k^0 + \hat{\gamma}_k^1 \hat{h}_{i2}).$$

If $\hat{d}_i^D = \hat{d}_i^U = 0$, the patient remains in the SNF.

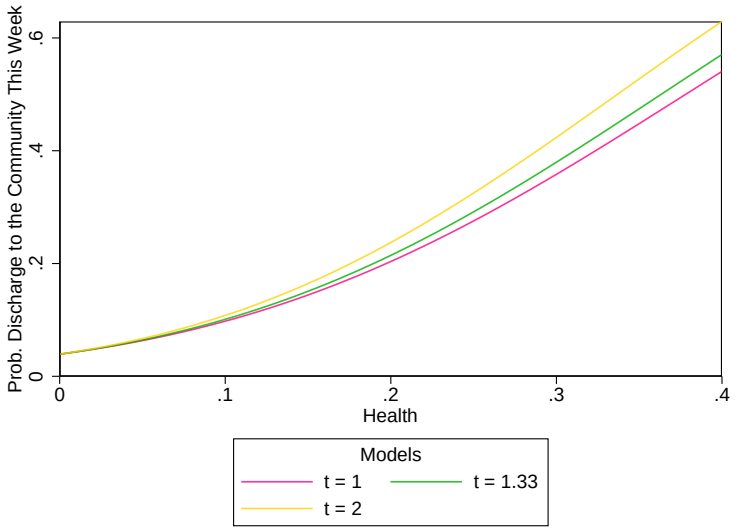
Panel C of Table IV shows summary statistics for health for various model-predicted discharge decision groups. The results in Panel C (model-predicted) and Panel B (external fit) are very similar. The sample in Panel C is slightly smaller than the sample in Panel B because a small fraction of SNFs in our sample do not have a full set of parameter estimates.

Appendix Figure A1: Discharge within 7 days of 30-day assessment vs. h_{i2}



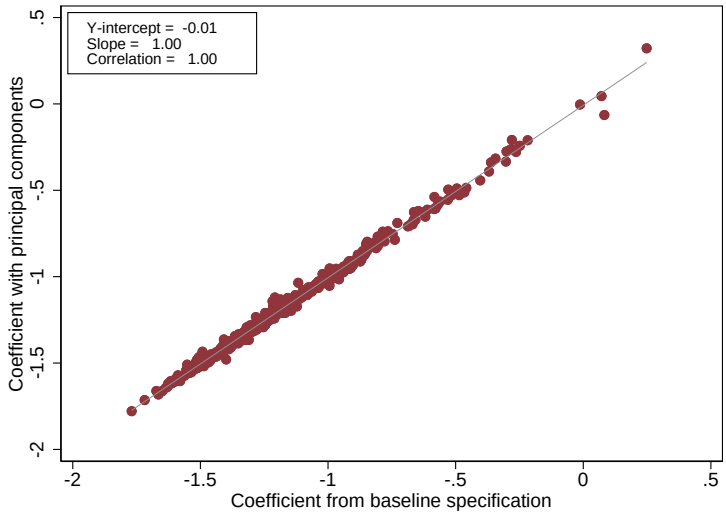
Notes: Figure shows binscatters of the share of patient-stays discharged within 7 days of the 30-day assessment against week-4 health h_{i2}. For each binscatter in (a)-(d), we divide the distribution of h_{i2} into 100 equally-sized bins, compute the share of patient-stays within each bin discharged to the community within 7 days, and scatter this share against the mean h_{i2} of each bin. For each panel, the line shown is the fitted regression line with the slope and intercept reported in the corner; the correlation is the unweighted linear correlation for the 100 points shown in the plot. Panel (a) shows the relationship for all 2,028,014 patient-stays with non-missing 30-day assessment data. Panels (b), (c), and (d) divide the distribution of these patient-stays into terciles of h_{i1} and repeat the same exercise; Panel (c) has 682,749 patient-stays, Panel (c) has 670,673 patient-stays, and panel (d) has 674,592 patient-stays. Panel (e) overlays Panels (b)–(d) but divides the distribution of h_{i2} in each tercile into 20 equally-sized bins (instead of 100).

Appendix Figure A2: Weekly Discharge Rule



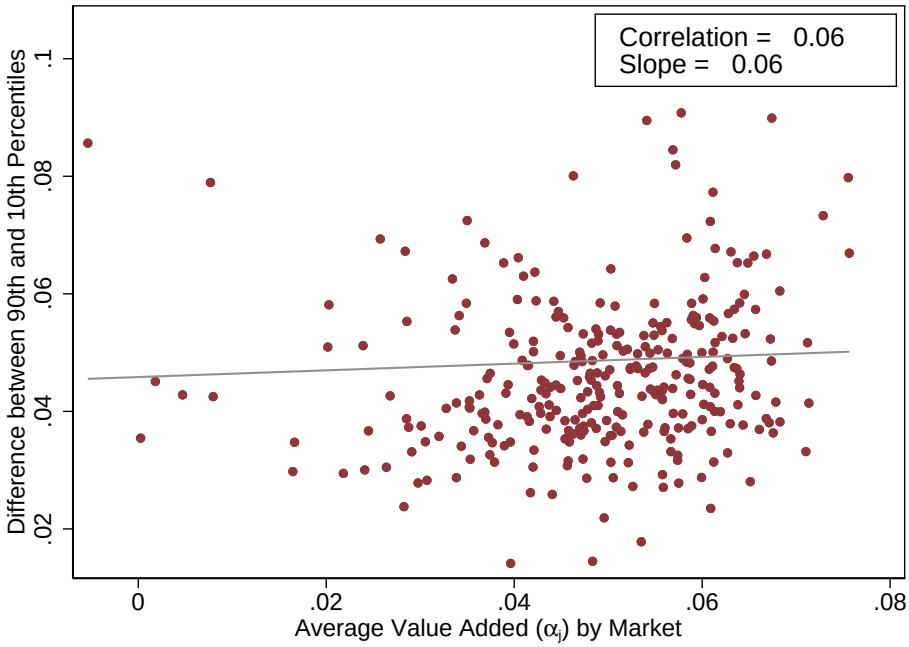
Notes: Figure plots the results from estimating an assessment-level probit discharge model that allows for the discharge decision to depend (linearly) on time in the SNF. We estimate a probit of the probability of discharge in the week following the assessment at $t = 1$ (the 5-day assessment), $t = 1.33$ (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 A.III.

Appendix Figure A3: 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, age bins, and Medicaid-dual-eligibility. The y-axis shows the same coefficient for each market from an alternative specification that uses the first five principal components of the following observables: 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). Coefficients are shown for all 655 choice sets.

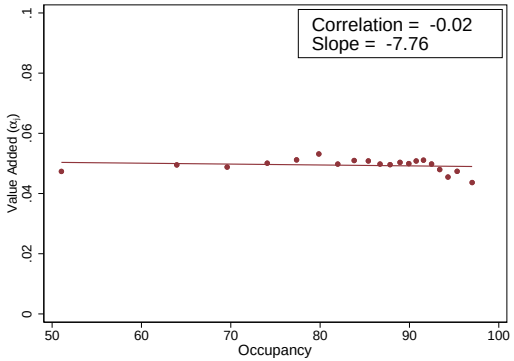
Appendix Figure A4: Correlating Difference between 90th and 10th Percentile Value-Added Within Market with Average Value-Added in Market



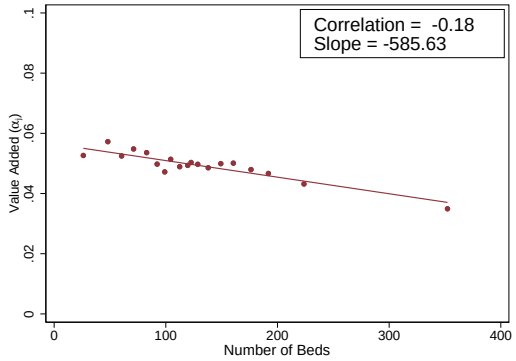
Notes: Figure shows a scatter plot for each of the 306 markets of the difference in value added between the 90th- and 10th-percentile SNF in the market weighted by patient-stays and the average value added of SNFs in the market weighted by patient-stays. The correlation is weighted by the number of patient-stays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

Appendix Figure A5: Correlates of SNF Value Added

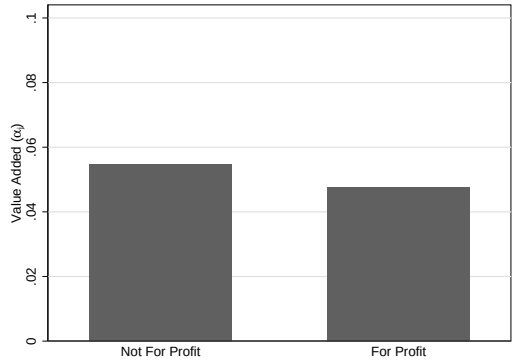
(A) SNF occupancy



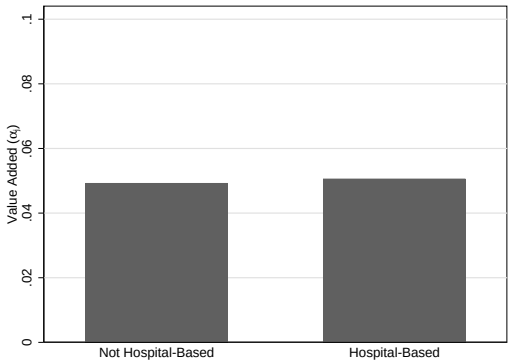
(B) Number of beds



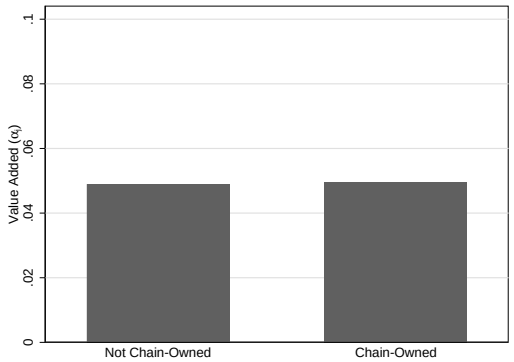
(C) For-profit status



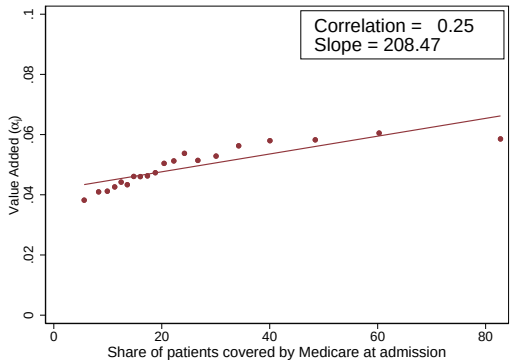
(D) Hospital-based



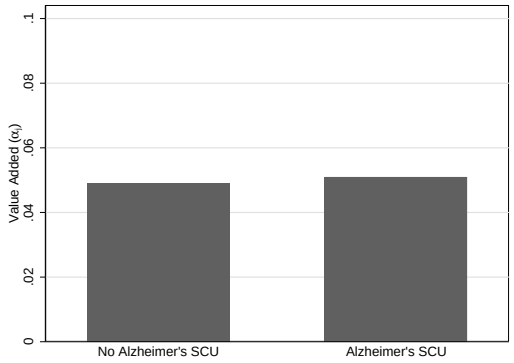
(E) Chain ownership



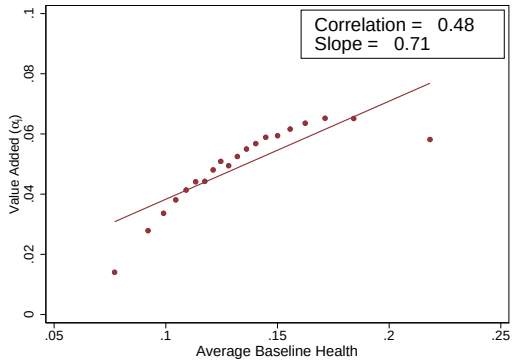
(F) Medicare share



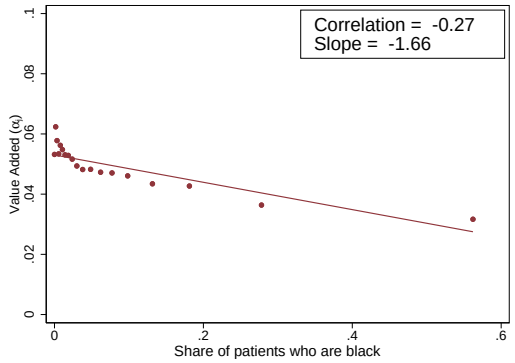
(G) Has Alzheimer's special care unit



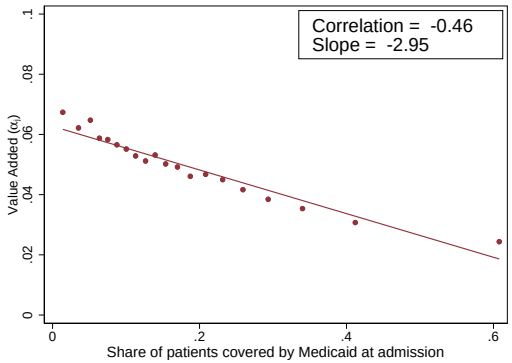
(H) Health index at admission



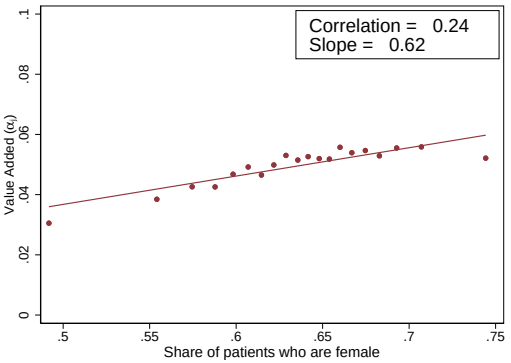
(I) Black patient share



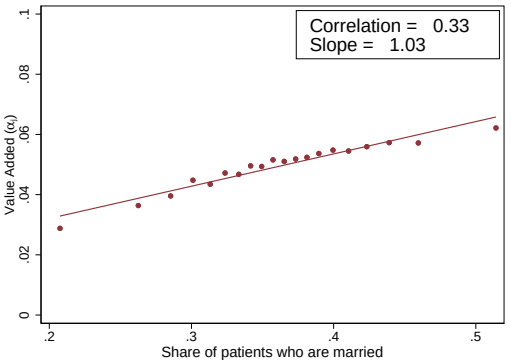
(J) Dually-eligible for Medicaid at admission share



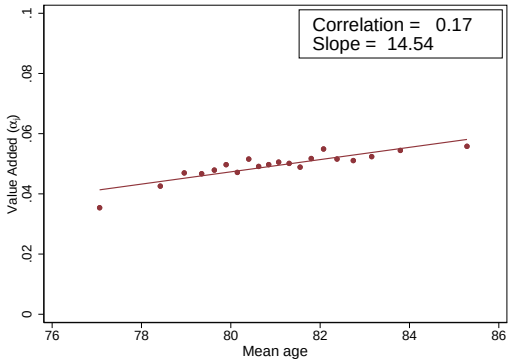
(K) Female patient share



(L) Married patient share



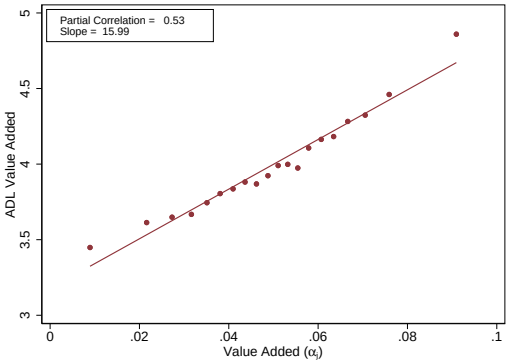
(M) Mean age



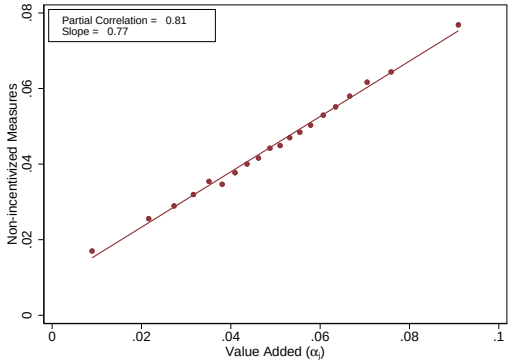
Notes: Figures show binscatters of value added on SNF characteristics using 20 bins for continuous SNF characteristics (discrete characteristics are shown as bar plots). The Medicare share, SNF occupancy, number of beds, for-profit status indicator, hospital-based indicator, chain ownership indicator, and Alzheimer's special care unit indicator come from OSCAR data. The black patient share, female patient share, married patient share, mean age, and dual-eligibility for Medicaid at admission share are calculated on 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

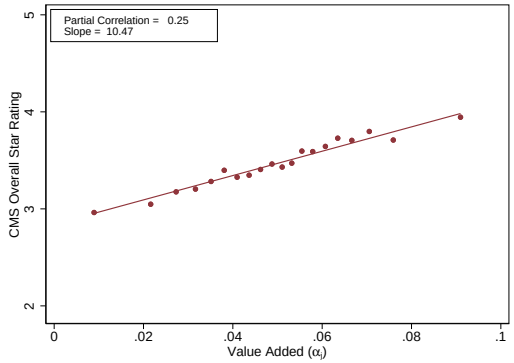
(A) ADLs



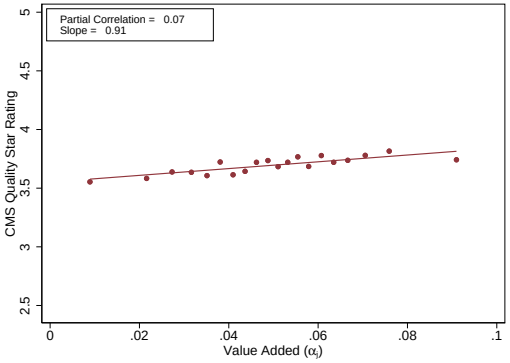
(B) Excluding incentivized health measures



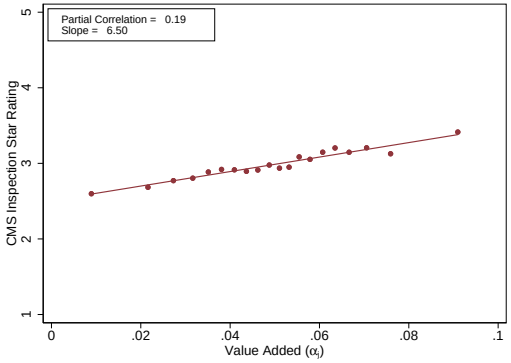
(C) Overall Ratings



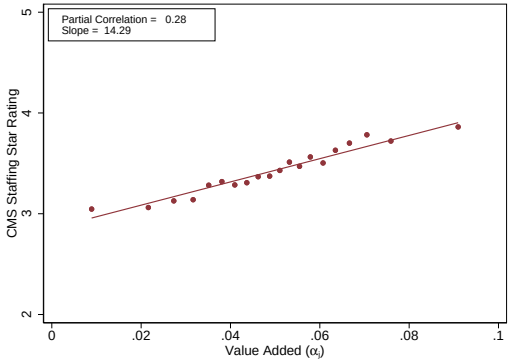
(D) Quality Ratings



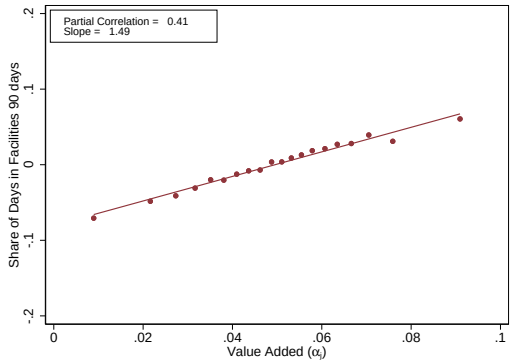
(E) Inspection Ratings



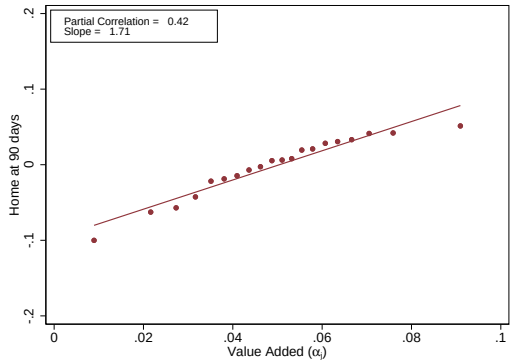
(F) Staffing Ratings



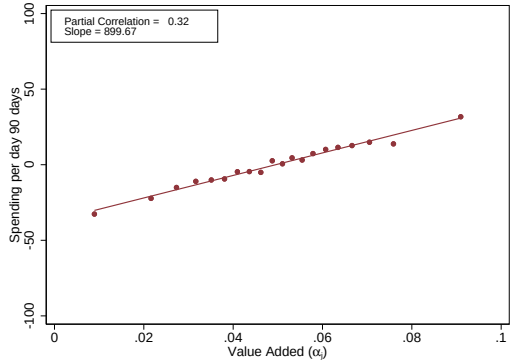
(G) Share of Days in Facilities at 90 Days



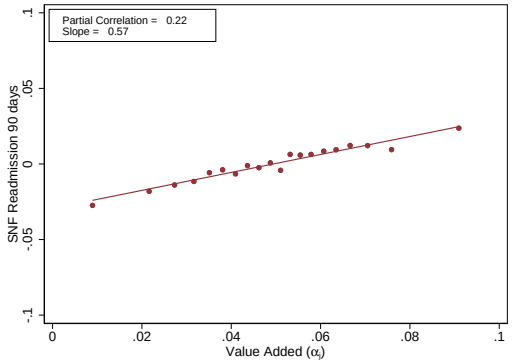
(H) Home at 90 Days



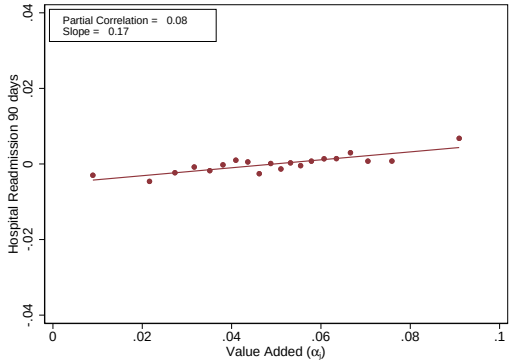
(I) Spending per Day at 90 Days



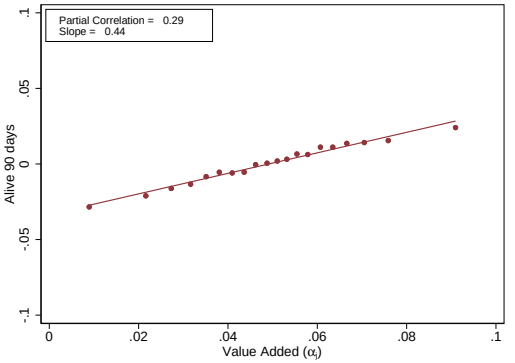
(J) SNF Readmission within 90 Days



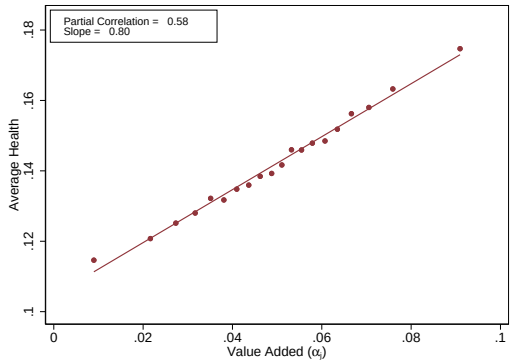
(K) Hospital Readmission within 90 Days



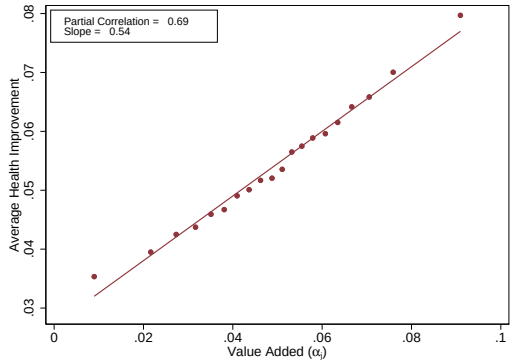
(L) Alive at 90 Days



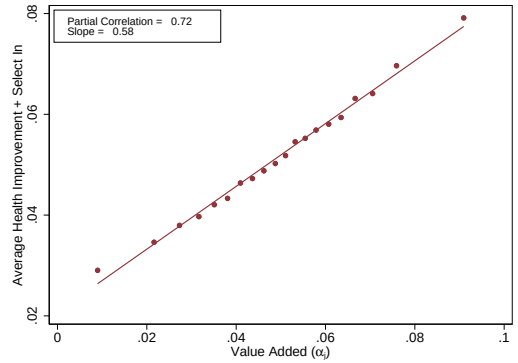
(M) Average Health



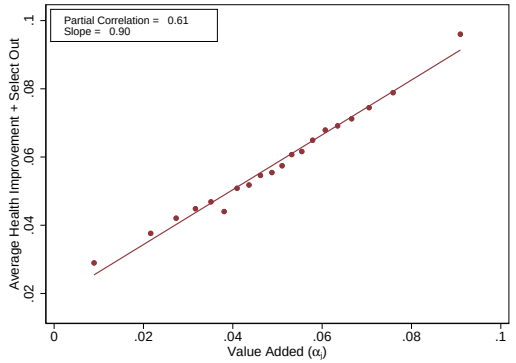
(N) Average Health Improvement



(O) Average Health Improvement + Select In



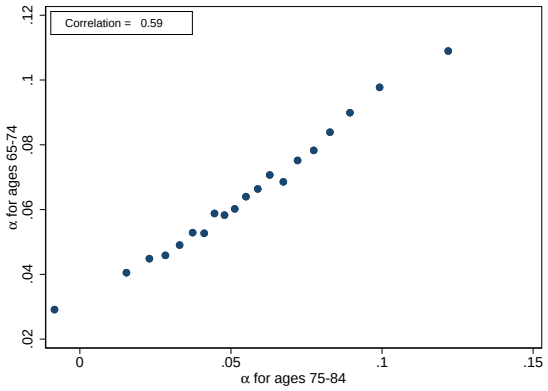
(P) Average Health Improvement + Select Out



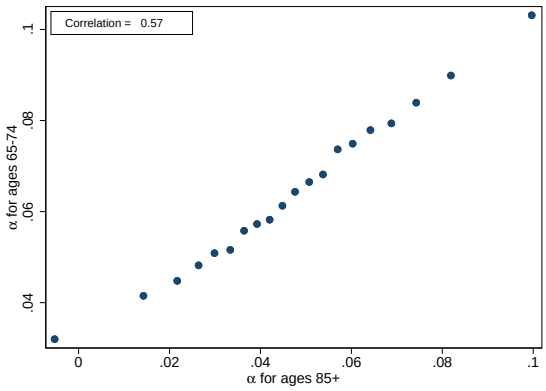
Notes: Figures show binscatters of alternative measures of SNF quality on SNF value added (α_j) controlling for market fixed effects. The alternative measures are (A) sum of activities of daily living (ADLs), (B) the health index excluding incentivized health measures, (C) CMS Overall Star Ratings, (D)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

Appendix Figure A7: Correlating value added estimates by age group

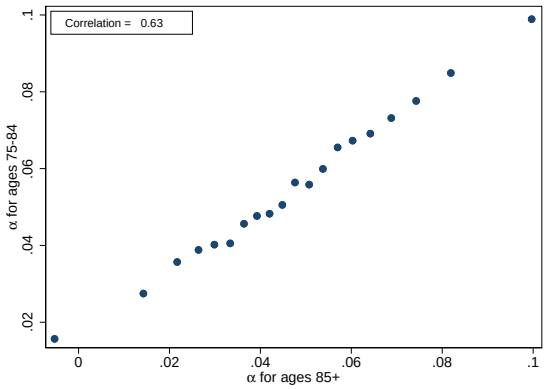
(A) Ages 65–74 vs. ages 75–84



(B) Ages 65–74 vs. Ages 85+

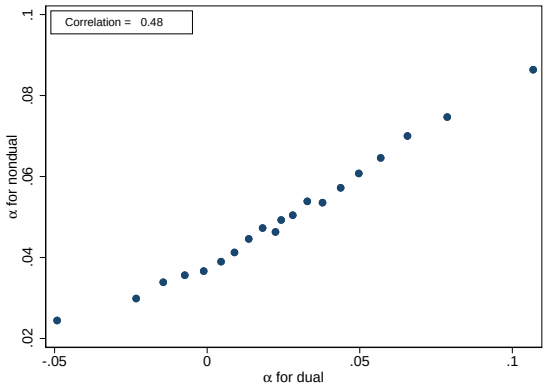


(C) Ages 75–84 vs. Ages 85+



Notes: Figure shows patient-stay-weighted SNF-level binscatters and correlations between value added estimates from the specifications estimated on subgroups determined by patient age. The sample consists of the 13,368 SNFs with value added estimates for all subgroups.

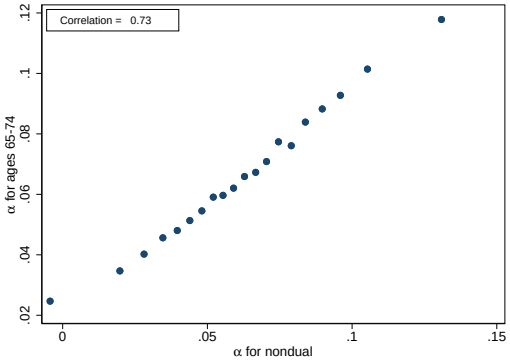
Appendix Figure A8: Correlating value added estimates by Medicaid status



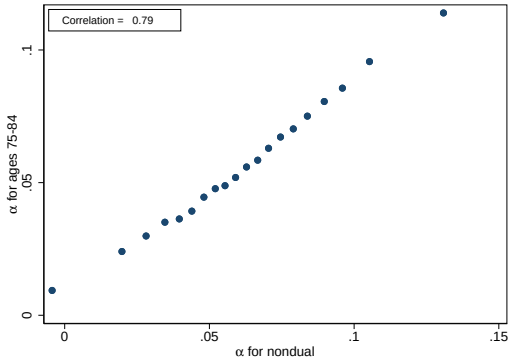
Notes: Figure shows patient-stay-weighted SNF-level binscatters and correlations between value added estimates from the specifications estimated on subgroups restricted by Medicaid dual-eligibility status. The sample consists of the 13,368 SNFs with value added estimates for all subgroups.

Appendix Figure A9: Correlating estimates by age group against estimates by Medicaid status

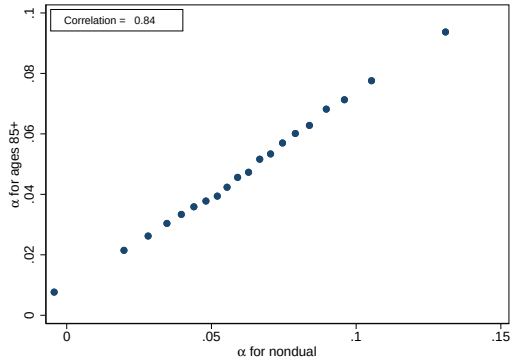
(A) Ages 65–74 vs. nondual



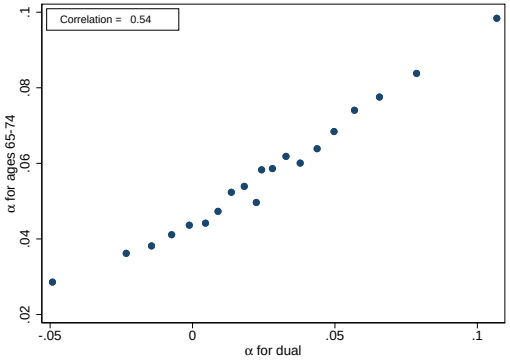
(B) Ages 75–84 vs. nondual



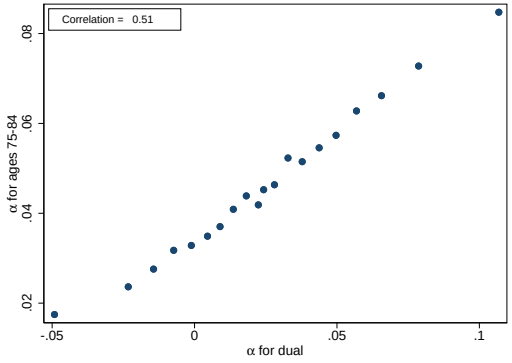
(C) Ages 85+ vs. nondual



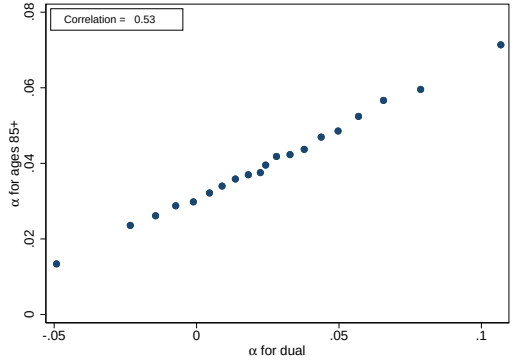
(D) Ages 65–74 vs. dual



(E) Ages 75–84 vs. dual



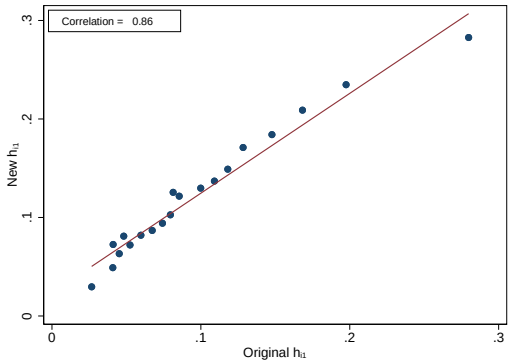
(F) Ages 85+ vs. dual



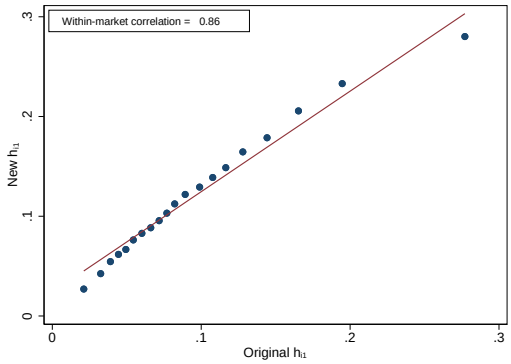
Notes: Figure shows patient-stay-weighted SNF-level binscatters and correlations between value added estimates from the specifications estimated on subgroups restricted by patient age (vertical axes) and Medicaid dual-eligibility (horizontal axes). The sample consists of the 13,368 SNFs with value added estimates for all subgroups.

Appendix Figure A10: Health index with alternative discharge definition

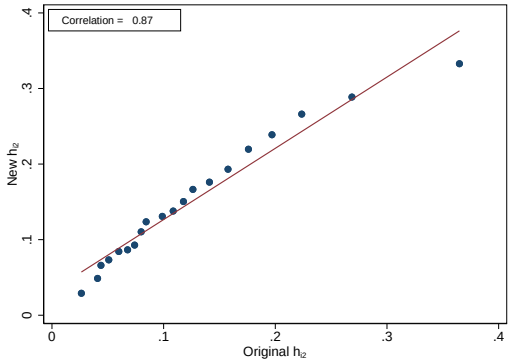
(A) Overall h_{i1}



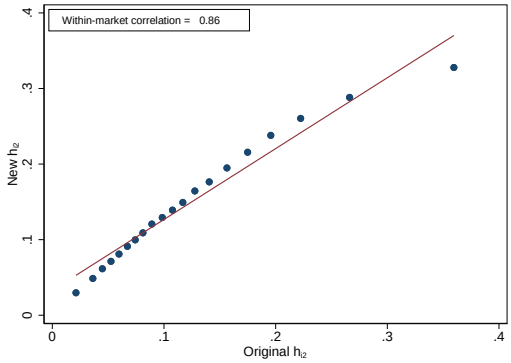
(B) Within-market h_{i1}



(C) Overall h_{i2}



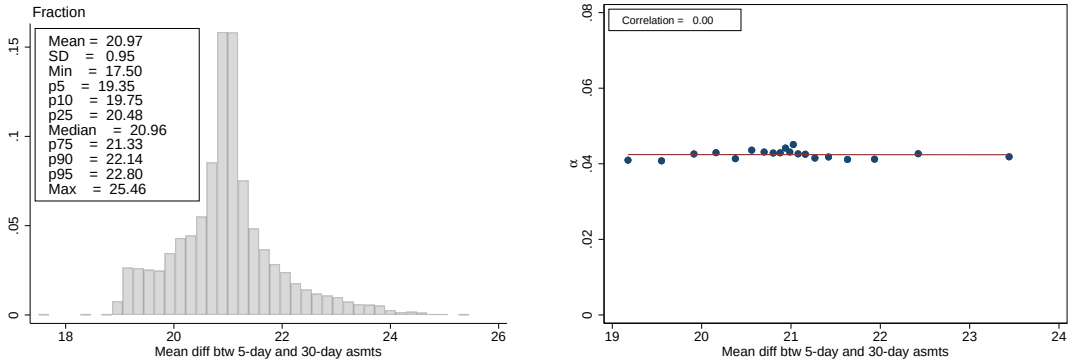
(D) Within-market h_{i2}



Notes: Figure shows binscatters of the new health index plotted against the baseline health index. Panels (a) and (b) show fitted lines for $6,246,151$ observations of h_{i1} both overall and within market (HRR fixed effects), while panels (c) and (d) show fitted lines for $2,516,942$ observations of h_{i2} both overall and within market.

Appendix Figure A11: Differences in the timing of health assessments

(A) Distribution of mean timing difference (B) Correlating mean timing difference with α_j



Notes: Figure shows (a) the distribution of the SNF-level mean difference (in days) between the 5-day and 30-day assessments and (b) a scatterplot of value added α_j against this mean difference.

Appendix Table A.I: Sample Restrictions

	Stays	SNFs
Full sample	22,520,577	16,355
Medicare and >= 65	10,721,447	16,306
First stays and discharged directly from ACH	7,213,728	15,730
CCS codes	6,598,040	15,680
Has 5-day health assessment	6,326,937	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" consists of all Medicare and Medicaid patient-stays reported in MDS 3.0 between October 1, 2011 and September 30, 2016. The next row ("Medicare and $\geq$ 65") restricts the sample to patient-stays whose admission is covered by Traditional Medicare and are at least 65 years old. The next row ("First stays and discharged directly from ACH") restricts to the first stay of each episode, as well as patients who enter the SNF directly from an acute-care hospital. The "CCS codes" row restricts to patient-stays with non-missing Clinical Classification Software (CCS) codes from 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 over the 5-year period.

Appendix Table A.II: 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 percentage 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.

* Non-incentivized measure.

[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 A.III: 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 C. The first two specifications 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 estimated on the 2,028,014 patient-stays 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 A.IV: 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 patient-stays 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 E.

Appendix Table A.V: 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 patient-stays conditional on being in the SNF at 30 days ($N = 2,517,006$) for both patient-stays who are missing and not missing the 30-day assessment.

Appendix Table A.VI: Imputation Procedure

	(1) Case 1 — observe $\bar{h}_{1.33}$ (N = 464,404)	(2) Case 2 — do not observe $\bar{h}_{1.33}$ (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 E. 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 discretized health. Time $t = 1$ corresponds to the 5-day assessment, $t = 1.33$ corresponds to the 14-day assessment, and $t = 2$ corresponds to the 30-day assessment. Variable d is a categorical variable that indicates the discharge destination of a patient-stay within 7 days of the 30-day assessment: still in the SNF (0), discharged downstream (1), or discharged upstream/elsewhere (2).

Appendix Table A.VII: 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	243,208	149,955	71,241	11,221	53,748	5,945
N Imputed		533,451	241,570	149,949	71,213	11,039	53,739	5,941
Share Imputed		0.9965	0.9933	1.0000	0.9996	0.9838	0.9998	0.9993
Discretized h_2								
Mean	5.66	6.81	8.15	5.94	4.78	9.55	5.57	4.79
SD	3.32	3.78	3.68	3.52	2.85	3.80	3.53	3.05
Number of candidate comparison observations chosen from								
Mean		8.56	2.86	13.35	2.00	6.50	30.55	2.70
SD		22.15	5.10	26.69	2.02	14.16	44.29	3.43

Notes: Table shows summary statistics by imputation cases for patient-stays who are still in the SNF at the time of the 30-day health assessment. Column (1) shows the non-imputed sample, while column (2) shows the sample for which 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.

Appendix Table A.VIII: Alternative health index with new discharge definition

	Mean	SD	10th	25th	50th	75th	90th
A. Baseline model							
h_{i1}	0.10	0.06	0.04	0.05	0.08	0.12	0.18
h_{i2}	0.12	0.09	0.04	0.06	0.10	0.16	0.24
B. New model							
h_{i1}	0.12	0.07	0.05	0.06	0.11	0.16	0.23
h_{i2}	0.15	0.09	0.05	0.08	0.13	0.21	0.28

Notes: Table reports summary statistics for h_{i1} and h_{i2} constructed using two different regression trees. The baseline regression tree predicts the probability of discharge within 7 days after the 30-day assessment, while the new tree predicts the probability of discharge within 8–14 days after the 30-day assessment. The sample consists of the $1,752,505$ patient-stays that remain in the SNF until at least 7 days after the 30-day assessment.

Appendix Table A.IX: National parameters estimated for alternative measures and specifications

Parameter	Baseline	ADLs	Excluding incentivized	By age			By Medicaid	
				Ages 65-74	Ages 75-84	Ages 85+	Dual	Nondual
θ_h	0.965	4	0.99	0.90	0.95	0.92	0.86	0.94
φ	0	1.1	0	0	0	0	0	0
σ_ϵ	0.075	10	0.07	0.08	0.08	0.07	0.10	0.08
σ_v	0.145	0	0.13	0.16	0.17	0.17	0.14	0.14
θ_{x1}	-0.005	0	-0.01	-0.01	-0.01	-0.01	--	--
θ_{x2}	-0.01	-0.5	-0.01	--	--	--	-0.01	-0.01
θ_{x3}	-0.02	-1	-0.02	--	--	--	-0.01	-0.02

Notes: Table reports the national parameter estimates for various alternative models. Parameter θ_{x1} is the coefficient for Medicaid dual-eligibility, θ_{x2} is the coefficient for ages 75–84, and θ_{x3} is the coefficient for ages 85+.