NBER WORKING PAPER SERIES

DO REPORT CARDS PREDICT FUTURE QUALITY?
THE CASE OF SKILLED NURSING FACILITIES

Portia Y. Cornell
David C. Grabowski
Edward C. Norton
Momotazur Rahman

Working Paper 25940
http://www.nber.org/papers/w25940

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2019

Do Report Cards Predict Future Quality? The Case of Skilled Nursing Facilities
Portia Y. Cornell, David C. Grabowski, Edward C. Norton, and Momotazur Rahman
NBER Working Paper No. 25940
June 2019
JEL No. I11,I18

ABSTRACT

Report cards on provider performance are intended to improve consumer decision-making and address information gaps in the market for quality. However, inadequate risk adjustment of report-card measures often biases comparisons across providers. We test whether going to a skilled nursing facility (SNF) with a higher star rating affects outcomes for a patient. We exploit variation over time in the distance from a patient's residential ZIP code to SNFs with different ratings to estimate the causal effect of admission to a higher-rated SNF on health care outcomes, including mortality. We found that patients who go to higher-rated SNFs experience lower mortality, fewer days in the nursing home, and fewer hospital readmissions.

Portia Y. Cornell
Brown University
121 S. Main St.
Providence, RI 022902
portia_cornell@brown.edu

David C. Grabowski
Harvard University
Department of Health Care Policy
Harvard Medical School
180 Longwood Avenue
Boston, MA 02115
grabowski@med.harvard.edu

Edward C. Norton
Department of Health Management and Policy
Department of Economics
University of Michigan
School of Public Health
1415 Washington Heights, M3108 SPHII
Ann Arbor, MI 48109-2029
and NBER
ecnorton@umich.edu

Momotazur Rahman
Center for Gerontology and Health Care Research
121 South Main Street, S-6
Providence, RI-02912
momotazur_rahman@brown.edu

Although quality is an essential determinant of the demand for goods and services, consumers often have incomplete information about quality prior to purchase. As such, in many markets consumers rely on published firm ratings to help choose a particular provider of goods or services. The quality of health care services is especially difficult for consumers to evaluate. Therefore, it can be efficient for the government to overcome this market failure by collecting information about quality and reporting it publicly. In the case of health care, the government publishes report cards online for hospitals, dialysis facilities, nursing homes, and home health agencies. Proponents of public report cards believe that these ratings overcome the asymmetric information problem, allow consumers to make informed decisions about quality of care, force health care providers to compete on quality, and create incentives for providers to improve quality of care.

Critics of public report cards, however, are concerned about several practical problems. Report cards publish ratings based on past measures, which may not predict future outcomes well. In particular, past measures may not predict future quality of care if those measures do not directly measure the aspects of quality that consumers care about, if they are based on small samples, or if they are not adequately risk adjusted. Health care providers may have an incentive to improve quality, but they also have an incentive to cream skim and endogenously select patients who are expected to have better outcomes. Risk adjustment is never perfect.

Given the theoretical arguments both in favor of and against public report cards, what matters is how they work in practice (Werner et al., 2012). Do public report card ratings predict future quality of care? In the case of a Medicare patient seeking a high-

2

quality nursing home following a hospital stay, if a patient chooses a skilled nursing facility (SNF) with a higher star rating, will she less likely to have adverse outcomes? Or, does a higher star rating simply reflect a more favorable mix of patients in the past or good performance on measures that are not valued by that patient? This paper addresses those empirical research questions.

As way of background, a large literature documents the prevalence of low-quality SNF care (Grabowski and Norton 2012). A parallel literature has observed wide geographic variation in Medicare spending on post-acute services such as SNFs (Newhouse and Garber 2013). In linking these two literatures, the general sentiment is that a great deal of low-value SNF care is delivered (Medicare Payment Advisory Commission 2016). To discourage low-value care, a number of market-based approaches have been introduced to encourage a market for quality. These approaches include the use of report cards (Konetzka, Grabowski et al. 2015), pay-for-performance (Grabowski, Stevenson et al. 2016; Norton, 1992), and alternative payment models like bundled payment (Sood, Huckfeldt et al. 2011) and accountable care organizations (McWilliams, Gilstrap et al. 2017). If these market-based approaches are going to be successful, the measures of SNF quality of care need to be reliable and valid.

Researchers and stakeholders have expressed concerns over the different SNF quality measures used for public reporting and payment (Mor 2006). The primary overall quality measure for SNFs is the Nursing Home Compare star rating, which is collected by the Centers for Medicare and Medicaid Services (CMS) and posted publicly on a website. Nursing homes (including SNFs) are rated on a variety of measures. This information is summarized in one overall rating, from one star (worst) to five stars (best). Given that

3

Nursing Home Compare's overall star rating is intended to provide consumers accurate information about quality of care, it is important to know if a patient's outcomes will improve if she chooses a higher-rated nursing home.

The greatest empirical challenge to evaluating the star ratings is overcoming the endogeneity of choice of SNF by patients who are not randomly assigned to nursing homes of varying star ratings. Prior studies have measured the correlation between star ratings and outcomes (e.g., Kimball et al., 2018; Ogunneye et al., 2015; Unroe et al., 2012, Ryskina at al., 2018), but they have not addressed the issue of patient selection across high- and low-star facilities. As such, any differences observed in these earlier studies may be an artifact of the different types of patients treated at high- and low-star facilities. Although these studies have often included a broad set of controls, unmeasured factors are likely correlated with both the selection of a high-quality SNF and patient outcomes, which leads to biased estimates.

The contribution of this paper is to address the potential bias introduced by patient case mix by using instrumental variables. This approach exploits variation over time in the distance from the patient's home address to the closest nursing home of each star rating. We ask, if a patient who needs post-acute care goes to a SNF with a higher star rating, then what effect does that choice have on their mortality, hospital readmission, and length of stay? We condition our main analyses on the patient's neighborhood. Using ZIP code fixed effects, the model compares patients from the same ZIP codes who are discharged at different times but make different choices about which SNF to enter because star ratings, and the relative distances to SNFs of different star ratings, change over time. Unlike the earlier literature, our approach allows us to estimate the causal

4

effect on outcomes of going to a higher-rated facility. The local average treatment effect for compliers is the effect of choosing a higher-rated SNF because of the changes in proximity between a person's home and SNFs in each star category.

To preview our results, using data from nearly 1.3 million new SNF Medicare patients in 2012-2013, we find that being admitted to a SNF with one additional star leads to significantly lower mortality, fewer days in the nursing home, fewer hospital readmissions, and more days at home or with home health care during the first six months post SNF admission. These results are robust to different modeling choices, such as the inclusion of SNF or hospital fixed effects; specification of star rating as linear or ordinal; and different ways of measuring outcomes.

I. Background

A. Skilled nursing facilities

Consumers in the nursing home market consist of both chronic (long-stay) and post-acute SNF (short-stay) residents. Each year, nearly 4 million elderly persons are admitted to SNFs for post-acute care, which is covered by Medicare. On any given day, about 1 million people reside in these facilities and receive long-term care, which is financed by Medicaid, private insurance, and by out-of-pocket payments. On a typical day, about 58% of nursing home patients were financed by Medicaid, 16% were financed by Medicare and 26% were financed privately through long-term care insurance or pay out-of-pocket.

SNFs offer skilled nursing care and rehabilitation services, such as physical and occupational therapy and speech-language pathology services, to Medicare beneficiaries

following an acute-care hospital stay. The supply of SNFs in the United States has remained relatively constant over the past few years, composed of roughly 15,000 facilities that are two-thirds for-profit. Almost all SNFs are also Medicaid-certified nursing homes that care for chronically ill or disabled residents for long-term stays.

In 2015, Medicare spent $29.8 billion on 2.4 million covered SNF admissions for 1.7 million fee-for-service beneficiaries (MedPAC 2017). Between 2007 and 2009, spending on post-acute care accounted for only 5% of the level of spending on Medicare parts A and B, but 73% of the variation in spending, adjusted for input prices and case mix (Institute of Medicine 2013, Table 2-10). The approximate median Medicare payment per SNF stay was just under $18,361, with an average length of stay of 26.4 days (MedPAC 2017). Medicare's prospective payment system (PPS) for SNF services was implemented based on the start of the facility fiscal year on or after July 1, 1998. Under the Medicare SNF PPS, facilities are paid a predetermined daily rate, up to 100 days, but only after a qualifying hospital stay of at least three days. The per diem prospective payment rate for SNFs covers routine, ancillary, and capital costs related to the services provided under part A of the Medicare program. Adjustments to the SNF Medicare payment rates are made according to a resident's case-mix and geographic factors associated with wage variation (Medicare Payment Advisory Commission 2009). Importantly, SNF payment is not adjusted based on their star rating.

B. Nursing Home Compare

In October 1998, CMS introduced a web-based nursing home report card initiative called Nursing Home Compare (www.medicare.gov/NHCompare). Nursing Home Compare was designed with the goal of harnessing "market forces to encourage

6

poorly performing homes to improve quality or face the loss of revenue" (page 3) (U.S. General Accounting Office 2002). In addition to information on facility characteristics (e.g., size, ownership status) and location, Nursing Home Compare reports data on various dimensions of quality. The initial report cards introduced in 1998 included only reports of survey deficiencies, but CMS has expanded the quality information available on the website. Information on professional and nurse aide staffing were introduced in June 2000, and the Nursing Home Quality Initiative (NHQI) in 2002 added Minimum Data Set (MDS)-based quality indicators to the website. These quality indicators encompass both short- and long-stay measures of patient outcomes.

Beginning in December 2008, the Nursing Home Compare website now reports four new composite quality measures: an overall 5-star rating along with specific 5-star ratings for inspections (deficiencies), staffing, and the MDS-based quality measures (e.g. restraint rate, anti-psychotic use rate, etc.). In 2016, CMS introduced three claims-based short-stay measures to the website and the 5-star rating: successful discharge to the community, an emergency department visit, and readmission to the hospital. Our analysis, which spans 2012–2013, does not include these updated measures.

Star ratings are composed of three sub-domains: health inspections, staffing, and quality measures (QM) (CMS 2017). The health inspection domain, which has the most influence on the overall rating, is calculated from within-state rankings to control for between-state differences but it is not risk-adjusted. The staffing domain is case-mix adjusted using Resource Utilization Group (RUG)-III scores in the quarter closest to the date of the most recent staffing survey. In the quality measures domain, ratings are risk-adjusted using a logistic regression with covariates from claims and MDS assessments.

7

The models include age, sex, length of stay, comorbidity indexes, previous

hospitalizations, and diagnoses. They do not include socioeconomic or race variables.

Several factors make standard risk adjustment techniques challenging. First, a key

concern is that these measures do not adequately account for selection of patients across

facilities due to the weak explanatory power of the variables used in the risk adjustment

modeling. There may be unobserved measures excluded from the risk adjustment that

bias cross-facility comparisons. Second, the available measures may not accurately

reflect quality that is important to patients if the inputs to final star ratings are not are

clinically relevant. Finally, some of the reported measures may be susceptible to up-

coding and gaming by SNFs due to their use in public reporting or payment (Bowblis and

Brunt 2013, Ryskina at al., 2018).

Several papers have questioned whether the Nursing Home Compare measures

are reliable and valid (Mor 2006). A number of studies have focused on improving the

risk adjustment variables or methodology employed on Nursing Home Compare. More

extensive risk adjustment of the measures was shown to change the rank ordering of the

facilities (Mukamel, Glance et al. 2008). Other research has argued that the Nursing

Home Compare measures can be improved by multilevel modeling (Arling, Lewis et al.

2007) or multivariate risk adjustment (Li, Cai et al. 2009).

These studies all show ways in which the existing measures on Nursing Home

Compare might be made more consistent relative to other measures. An issue is that these

studies lack a comparison measure that is purged of selection. As Arling and colleagues

(2007) acknowledged, "we have no 'gold standard' for validating our risk adjustment

methodology" (p. 1193). This lack of a gold standard raises two issues. The first is that,

with inadequate risk adjustment, any measure of quality is contaminated because it partly reflects selection. That is, high quality ratings based on patient outcomes may reflect a healthier mix of patients in ways that cannot be adjusted through statistics. The other related issue is that, if patients endogenously select into certain SNFs, it is hard to measure the causal effect of quality rating on patient outcomes. Patient selection is challenging to overcome.

Our goal is not to examine whether we can improve the existing risk adjustment, but rather to evaluate the validity of the existing measures. Our approach follows a new literature that employs instrumental variables to validate quality measures. Doyle and colleagues (2017) studied hospital quality measures by exploiting ambulance company preferences as an instrument for patient assignment. They found that assignment to a higher-scoring hospital resulted in better patient outcomes. In a study of SNF re-hospitalization, Rahman and colleagues (2016) instrumented for selection to a nursing home using empty beds in a patient's local market. The authors found that assignment to a nursing home with a historically low re-hospitalization rate led to fewer readmissions. We will use a similar approach to answer our research question with a causal estimate of the effect of going to a facility with a higher star rating on patients' outcomes.

II. Data and sample

A. Data sources

This study relies on several sources for individual-level characteristics including the Medicare enrollment denominator file, Medicare claims and the Minimum Data Set (MDS) for nursing home resident assessments, and SNF provider data, including Online

9

Survey Certification and Reporting (OSCAR) and quality ratings from Nursing Home Compare.

The Medicare Standard Analytic File includes all claims related to inpatient, skilled nursing facility care, home health, and hospice services for Medicare fee-for-service enrollees. All Part A claims (inpatient, SNF) include dates of service and up to 25 diagnoses. The Medicare enrollment file identifies individuals enrolled in Medicare within a given year and includes demographic data, survival status, residential ZIP code, and program eligibility information for Parts A, B and D, Medicare Advantage (managed care), and Medicaid.

The MDS contains clinical assessments of all residents in Medicare- or Medicaid-certified nursing homes (regardless of the payment source). They are given upon admission to the facility and then periodically, at least quarterly, thereafter. The MDS includes summary measures of cognitive and physical functioning, continence, pain, mood state, diagnoses, health conditions, mortality risk, special treatments, and medication use.

The Online Survey Certification and Reporting (OSCAR) System is a compilation of all the data elements collected by surveyors during the inspection survey conducted at nursing facilities for the purpose of certification for participation in the Medicare and Medicaid programs. The database includes organizational characteristics such as the number of beds, ownership, and chain membership, staffing availability and aggregate patient characteristics. SNF star rating data is obtained from CMS website (CMS 2017). We also used American Hospital Association (AHA) data for year 2013 for several hospital characteristics.

B. Study sample

Applying the Residential History File methodology (Intrator, et al. 2011), which concatenates MDS assessment and Medicare claims into individual beneficiary trajectories, we identified all Medicare fee-for-service (FFS) beneficiaries who were discharged directly from an acute general hospital to a SNF for post-acute care between January of 2012 and June of 2013. We started with hospital and SNF claims data that identifies 1,576,010 SNF admissions with valid hospital and SNF identification, discharged from 4,706 hospitals with no SNF claims during the previous 12 months. We excluded 28,740 (1.82%) who were not discharged from a general acute care hospital. We then merged this data to MDS. We excluded 182,176 (11.77%) observations because either we could not find matched assessment in MDS based on individual identification and the admission date, or MDS indicated that the individual was in nursing home during the previous one-year period. We excluded patients with any SNF residence history in the one year prior to admission because they would be frailer than post-acute care patients from the community. Finally, we dropped 86,638 (6.34%) of the individuals who did not reside in the 48 contiguous states or did not have a valid residential ZIP code. Our final sample consisted of 1,278,456 Medicare FFS beneficiaries discharged from 4,332 acute care hospitals to 15,166 SNFs.

C. Outcomes measures

In order to assess the validity of star ratings, we measure their effect on outcomes that are relevant to patient welfare and system costs. In this paper, we consider the time spent in six mutually exclusive settings: at home without home health, at home with home health, in skilled nursing care, in hospital inpatient care, hospice care, and

11

deceased. These outcomes are key because hospitalization from SNF is not only viewed as a signal of potential inefficiencies and cost-shifting by nursing homes, but also can cause stress and disorientation for patients. Mortality is commonly used as a marker for hospital quality of care and has implications for hospital reimbursement, although it is not part of the Nursing Home Compare rating.

Using the Residential History File (RHF) methodology (Intrator et al. 2011), we follow patients for 180 days following their admission to a SNF by concatenating Medicare enrollment, Medicare claims, and MDS. We assign the patient's location on each day to one of the six settings: at home without home health, at home with home health care, in the SNF, in hospital, in hospice, and deceased. Table 1 shows mean statistics for the outcomes variables. Our main outcomes of interest are the following dichotomous variables: any hospitalization within 30 and 180 days; death within 30 and 180 days; and becoming a long-stay nursing resident, defined as staying in a SNF or nursing home for longer than 100 days. At the end of 180 days, 54% of patients had hospitalized at least once; 21% were deceased; and 16% had spent more than 100 days in the nursing home.

We also construct six integer outcome variables, each quantifying the number of days in the specified setting in the 180 days following discharge from the hospital, also summarized in Table 1. The mean number of days in a nursing home is 48. Figure 1 plots the share of patients in each setting on each of the 180 days following SNF admission. On day 0 everyone was in a SNF, as determined by the inclusion criteria for the analysis. On day 60, 35% of them were at home without any Medicare paid support, 24% of them

were at home with home health care support, 12% were dead, 3% were in hospital, 2% were in hospice, and the remaining 23% were in a nursing home.

D. Nursing home quality measure

Our main explanatory variable is the five-star rating from the admitting SNF's Nursing Home Compare Report Card, coded as an integer between 1 and 5 where 1 indicates the lowest possible quality rating. For our preferred analysis, we enter star rating as a continuous variable, assuming that an increase in star rating of the admitting SNF by one star has linear effect on outcomes. In a sensitivity analysis, we treat each star rating as a binary indicator (four dummies for star ratings 2-5 with rating 1 as the benchmark).

A facility may see a change in its overall rating whenever new data are available in any of the three domains—inspections, staffing, and quality measures (QM). The inspection and staffing domains are updated approximately yearly after an inspection; in addition, the health-inspection score can also change with a new complaint, a revisit, resolution of disputed deficiencies, or exclusion of old complaints after a set period of time. Quality measures are updated quarterly for measures based on the MDS and every six months for claims measures. Thus, a given SNF could get an updated rating approximately every quarter. Because inspections are distributed throughout the year, consumers may face changes each month in the SNF ratings of the SNFs near to them.

Figure 2 plots the proportion of SNFs with each star rating during the 18 months when individuals in our sample were admitted to SNFs (January of 2012 and June of 2013). During this time period, the share of SNFs with one star declined from 15% to 12% and the share of SNFs with five-star increased from 16% to 22%. These changes are

13

due to updates of different components of star-ratings and not due change in cut-off points or re-basing. Such change in cut-off points occurred in October 2010 (transitioning from MDS 2.0 to MDS 3.0), in February 2015 when CMS changed weights for staffing and QM components and later in March-July 2016. Therefore, our instrumental variable has much variation over time within geographic area and these variations are due to actual changes in measured quality of care, not arbitrary changes in thresholds.

Figure 2 also plots the share of patients in our sample that were admitted in SNFs with different rating. The share of post-acute patients going to one- and two-star SNFs declined more steeply than the decline in proportion of facilities in these categories; whereas the share of patients going to four- and five-star SNFs increased at a higher rate than the share of facilities. Because nearly all facilities admit a mix of short- and long-stay patients, the trends could be explained by 4- and 5-star SNFS may be increasing their market share of more profitable post-acute patients while 1- and 2-star facilities are increasing the proportion of their beds occupied by long-stay residents.

E. Control variables

We include a series of patient-level control variables. Demographic characteristics include age, sex, and race from the enrollment file and language spoken and marital status from the MDS. We also include dual-Medicaid eligibility on the month of SNF admission from the Medicare enrollment file. We distinguish between partial- and full-dual eligibility where partial-dual eligible beneficiaries have some cost-sharing that are not paid by Medicaid.

From the index hospitalization claims prior to SNF admission, we include measures of health status. These measures included the Elixhauser (Elixhauser et al.,

14

1998) and Deyo (Deyo et al., 1992) co-morbidity indexes, hospital length of stay, the number of cardiac care unit (CCU) days, and the number of intensive care unit (ICU) days during the hospitalization. Other clinical characteristics were obtained from the MDS and include indicators for common diagnoses (e.g., diabetes, serious mental illness etc.), and the Morris late loss activities of daily living (ADL) scale (Morris et al., 1999), and the cognitive performance scale (CPS) (Morris et al., 1994). We summarize the control variables for the full sample (see Table 2).

III. Empirical Strategy

A. Empirical model

Our main regression of interest is the relationship between star rating and patient outcomes, controlling for patient factors and residential ZIP code effects. This relationship can be described by equation (1)

$$(1) \qquad Y_{izhnt} = \beta_0 + \beta_1 Q_{int} + X_i \gamma + \theta_t + \zeta_z + \epsilon_{izhnt}$$

where Y_{izhnt} refers to the outcome of person i residing in ZIP code z discharged from hospital h to nursing home n in month t. Q_{int} is the star rating (on a scale of 1–5) of the admitting nursing facility on month t; later we will explain how we deal statistically with the endogeneity of this variable. The vector X_i includes the patient's demographics and clinical characteristics; θ_t are month-of-admission fixed effects; ζ_z are patient's residential ZIP code fixed effects; and ϵ is the residual.

This model does not include any SNF-level variables other than the SNF's star rating. This is partly because the star rating is a composite measure that summarizes

15

different quality aspects of a SNF. Additionally, our objective is to inform stakeholders

(patients, hospitals, accountable care organizations and Medicare advantage plans) about

whether this one-dimensional measure can be used to choose a SNF in an informed

manner. Of note, related prior research also has not accounted for other provider

characteristics when estimating similar models (Doyle et al., 2017; Rahman et al., 2016).

B. Inference problem: patient selection

A key issue when estimating the effect of an increase in star rating on days in a

post-acute setting is that patients are not randomly selected into nursing homes of varying

quality, reflecting both demand- and supply-side differential selection. On the consumer

side, patients who select a SNF based on quality information may be more savvy

consumers of health care in other ways, such as the selection of doctors and hospitals.

Patients who are more strongly motivated to recover at home may select a SNF where

they think they have the best chance of a successful discharge. On the other hand,

readmissions-reduction and hospital value-based purchasing, which penalize high

readmission rates, give hospitals incentives to select particular patients for care in more

highly-rated SNFs (Norton et al., 2018). A hospital's discharge planner may

preferentially send patients that are at high risk of readmission, or who will have high

post-discharge expenses, toward high-star facilities to reduce costly re-hospitalizations.

These examples of selection bias would exert opposing biases, leading a naïve model to

either over- or under-state the influence of star rating on the outcome. Traditional

methods to control for this bias, such hospital or SNF fixed effects to estimate the effect

of within-facility changes in star rating, do not address patient sorting over time on

unobserved characteristics in response to quality ratings.

One way to provide suggestive evidence of this kind of selection is to compare descriptive characteristics of patients admitted to a SNF above, equal to, and below the median for their ZIP code (see last three columns of Table 2). Even within ZIP codes, there are important differences between patients who go to SNFs with above-average versus below-average star ratings, including greater than 20% differences in prevalence of full-dual eligibility, Alzheimer's disease, schizophrenia, and bipolar disease. These differences suggest potential bias from other unobserved differences. Therefore, our main statistical concern is how to control for the endogeneity bias in star rating to estimate causal effects. In the following section, we propose an instrumental variable approach to address patient selection. The IV approach estimates the causal effect of star rating for the patient whose choice of SNF of a particular star rating is influenced by distance from home. The IV models also include ZIP code, SNF, and hospital fixed effects.

C. Proximity to SNF in each star category as the instrumental variable

To address selection on unobserved health status, we leverage variation in the quality rating of nursing homes near a patient's home ZIP code during the month of SNF admission. Our IV is the distance from patient's residential ZIP code to the nearest SNF with a particular star rating in the month of SNF admission. The intuition is straightforward (Gowrisankaran and Town, 1999). While choosing a SNF, a patient faces a choice of SNFs distinguished by distance and by star-ratings. Because star-ratings change month-to-month, patients from same ZIP code face different choices depending on the month of admission. For example, even when the choice of SNFs remains the

17

same, the proximity to a 5-star SNF can vary month-to-month. Because proximity to a

SNF is one of the main determinant of SNF admission (Hirth et al. 2014; Rahman,

Norton, Grabowski 2016), we use the variation in proximities of SNFs with different star-

rating as IVs.

To operationalize patient preferences based on distance, we created five

instrumental variables based on the log-distance from the patients' home to the nearest

SNF in each quality category (following a similar approach by Gowrisankaran and Town,

1999). We denote this as D_{zt}^k, which is the natural logarithm of the distance from the

patients' ZIP code z to the nearest k-star SNF (where k=1,2,3,4,5) from in month t. For

each patient in the dataset, we obtain geo-coordinates for the centroid of the ZIP code of

home residence reported in the Medicare denominator file, and for each SNF, we know

the coordinates of the exact address. For each patient we calculate the distance (using the

great-circle formula) from the centroid of the patient's home ZIP code to nearest facility

of quality level k. Thus, a D_{zt}^k can have really large value if the nearest k-star SNF is far

away from patient's ZIP code. We use the natural logarithm of distance because

preliminary analysis showed that the relationship was linear on a log scale. Because we

have the coordinates of the exact address of each nursing home, the distance from SNF to

centroid is greater than zero even when the patient and facility are in the same ZIP code.

Thus each ZIP code in our dataset has a 5-dimensional vector of instrumental variables,

where the first value is the log distance to the nearest 1-star facility in that month, the

second is the log distance to the nearest 2-star facility, etc.

There are two ways that the instruments can change value over time. First, the

star rating of existing SNFs can change each month. Between 1 January 2012 and 30

June 2013, out of the 15,166 SNFs used by our patient cohort, 12,080 of them changed star rating at least once. Second, there can be entry or exit of a nearby SNF. However, there were only 153 SNF entries into the market and 79 exits. Thus, the variation in IVs over time is mostly driven by changes in star ratings of existing SNFs. It is important to note that a change in star-rating of one SNF can change the value of more than one IV. For example, an increase in a rating of a nearby SNF from 3 stars to 4 stars can change the values of both distance to the nearest 3-star SNF (D_{zt}^3) and the distance to the nearest 4-star SNF (D_{zt}^4). Alternatively, this SNF may no longer be one of the nearest SNFs because there of another 4-star SNF that is nearer to that ZIP code; in this case only the distance to the nearest 3-star SNF (D_{zt}^3) changes.

Our instruments exploit both cross-sectional and longitudinal variation in star ratings, not just within-SNF changes in star ratings. For example, an alternative IV could be defined as the star rating of the nearest SNF from patient's ZIP code in the month of admission (in that case, there will be just one IV). In such case, after controlling for ZIP code fixed effects, such an IV would rely only on the change in star rating of the nearest SNF from a ZIP code. In contrast, our instruments incorporate variation in star ratings over the entire distribution of stars, and as a result have much greater strength in predicting the first-stage model of choice of number of stars.

Another important aspect of this study design is that it controls for residential ZIP code fixed effects. Prior studies using differential distances (Grabowski et al. 2013, Hirth et al. 2014) assume that patients generally do not choose their home residence with regard to the quality of nearby post-acute-care facilities — assignment to quality is therefore effectively random. However, we suspect that any difference in average

19

characteristics of patients between ZIP codes (for example, demand for quality) is partially reflected in the star rating of the nearest SNF. Therefore, we control for such differences using residential ZIP code fixed effects.

The first-stage equation predicts the star rating of the SNF chosen by the patient as a function of the log-distance to the closest SNF of each star rating and other individual, time, and residential ZIP code fixed effects.

$$(2) \qquad Q_{int} = \alpha_0 + \sum_{k=1}^{5} \alpha^k D_{zt}^k + X_i \gamma + \theta_t + \zeta_z + \epsilon_{izhnt}$$

Here D_{zt}^k is the log-distance of the nearest k-star SNF from ZIP code z in month t. We expect the coefficients on distance, representing the relationship between distance and average quality, to be negative for high-star SNFs, positive for low-star SNFs, and monotonically declining from low- to high-star. As the distance to a one-star or two-star SNF increases, a patient is less likely to choose a low-star SNF and more likely to choose a high-star SNF. As the distance to a four-star or five-star SNF increases, the predicted star rating of the chosen SNF will decrease. To estimate the causal effect of star rating on the outcome, the second stage regresses the outcome on predicted star rating $\widehat{Q_{int}}$ from equation (2).

$$(3) \qquad Y_{int} = \alpha_0 + \widehat{Q_{int}} + X_i \gamma + \theta_t + \zeta_z + \epsilon_{izhnt}$$

We estimate this system of equations as two-stage-least-squares using the `xtivreg2` command in Stata with standard errors clustered by ZIP code.

D. Robustness checks

We also estimate these models substituting residential ZIP code fixed effects with hospital fixed effects, including both residential ZIP code and hospital fixed effects, and including both ZIP code and SNF fixed effects. A SNF-fixed-effects model measures

20

differences in outcomes between two patients who go to the same provider at different times, conditioning on time-invariant differences in quality between providers. Hospital fixed effects control for differences in hospital quality associated with SNF star rating. For the two-way (hospital and ZIP code, and SNF and ZIP code) fixed effect models, we use the `ivreg2hdfe` command (Bahar 2014) because `xtivreg2` command is designed for one set of fixed effects. Test statistics are based on robust standard errors clustered by residential zip code (with SNF FE models, clustered by SNF).

In most of our analyses, we consider star rating as a continuous variable, which assumes that an increase in star rating has a linear relationship with outcomes. As a sensitivity analysis, to allow for a nonlinear relationship, we examine the effect of star rating using separate binary variables for each star rating with one-star SNF as the benchmark category. We also estimate our model including diagnosis related groups (DRG) fixed effects.

Several studies (Rahman, et al. 2013, Rahman, et al. 2016) found that admission to hospital-based SNFs yields better health outcomes and that the better outcome is likely due to better coordination between the hospital and the SNF. It is possible that effects of such coordination are captured in the quality ratings. So we estimated separate models for patients treated in hospitals with and without hospital-based SNFs.

We also perform a robustness check to address concern over the competing risk of death. We estimate the binary outcome models on the subset of patients who were alive on day 30 and 180, respectively, for the 30-day rehospitalization and for the 180-day rehospitalization and long-stay nursing variables. We also estimate the continuous-days

21

models adjusted for exposure time. To create these variables, we calculate the mortality-adjusted days as $Y_a = Y / \text{days\_alive} * 180$.

IV. Results

A. Descriptive results to assess the validity of the IVs

We perform several analyses to assess the validity of our IVs in addition to performing standard statistical tests. The main objective of these analyses is to assess the within and between variation in SNF choice and patient characteristics across residential ZIP codes. To demonstrate the variation over time within ZIP codes, we calculate the ZIP code level median distance (logged) to the nearest 5-star rated SNF and compare patients who experiences different values of the IV relative to their zip code level median. The idea is that, for each ZIP code, we have 18 values (one per month) of an IV varying by month and can calculate the median of these 18 values. Then we compare patients within a ZIP code who were admitted in months when the IV is smaller, equal to and greater than the ZIP code level median.

The first step is to quantify the within-ZIP code variation of the instruments. Because we have five different instruments that are similar in nature, we just focus on the natural log of distance of the nearest 5-star SNF. About 20 percent of the patients experienced longer distance to the nearest 5-star SNF relative to the median value this distance in their ZIP code. These individuals experienced, on average, 2.44 times higher value of the fifth IV relative to the individuals experiencing the median value. On the other hand, 15 percent experienced a shorter distance to the nearest 5-star SNF relative to the median value this distance in their ZIP code.

22

The first descriptive analysis provides visual evidence that our IVs strongly predict the endogenous variable, quality rating. Figure 3 compares the share of patients admitted to a 5-star SNF during months when the distance to the nearest 5-star rated SNF is above the ZIP code level median to those admitted during months when the distance was below the ZIP code level median. In addition to confirming that demand declines with distance (the two lines slope downwards), it also shows that within-ZIP-code variation over time is important. When the distance is below the median (upper line) then the share of patients is larger. There is a significant difference between the two lines for a wide range of distances. This implies that the first-stage relationship of the IV analysis is likely to hold even with ZIP code fixed effects included.

The second descriptive analysis assesses the validity of our exclusion-restriction assumption. A common technique to assess the assumption of ignorable treatment assignment is to stratify the sample by high and low values of the IV and evaluate covariate balance. Characteristics of patients experiencing different values of the fifth IV relative to the ZIP code level median are presented in Table 3. While Table 2 shows that patients discharged to SNFs with different star-rating from same ZIP code look fairly different, Table 3 shows that patients experiencing different level of the fifth IV look almost identical.

To facilitate an intuitive visual display, we aggregate the patient covariates listed in Table 2 into one measure, the predicted likelihood of entering a 5-star SNF, which is similar to a propensity score. To create the measure, we estimate a multinomial logit model with star rating of chosen SNF as the outcome and predicted the likelihood of entering a 5-star SNF for each patient in the sample. This prediction score can be

23

interpreted as the demand for a 5-star-rated SNF. Figure 4 plots the predicted likelihood of admission to a 5-star SNF for two groups of patients: those who were admitted to a SNF during months when the distance to the nearest 5-star rated SNF is above the ZIP code median, and those below the ZIP code median. If the exclusion restriction assumption is correct, then the groups stratified by distance within ZIP code will be balanced on their covariates and the predicted admission will be the same for the two groups. As expected, these lines are overlapping, implying that patients within a ZIP code experiencing lower and higher value of IV are observably the same. Thus, conditional upon ZIP code fixed effects, patients with different values of the IV are similar. Of note, both of the lines in Figures 3 and 4 are slightly downward sloping, implying that patients in ZIP codes with different values of the IVs are somewhat different. This justifies the inclusion of ZIP code fixed effects.

Besides comparing observable characteristics between patients with different values of the IVs, we also use two falsification checks as additional evidence that the instrument satisfies the exclusion-restriction assumptions. First, we assess whether changes in star rating of SNFs in the vicinity attracts more patients to receive post-acute care in SNFs instead of going to other settings such as home, home health care, or a hospital-based rehabilitation facility. A change in the balance of patients discharged to a SNF could suggest that patient selection associated with the IVs might be driving our results. We analyze all hospital discharges during our study period to assess whether the likelihood of a discharge to a SNF is associated with the IVs conditional upon residential ZIP code fixed effects. We find no such statistical association (see Appendix Table A5).

24

Second, using a sample of patients who had no previous SNF stay and were discharged from an inpatient hospital stay directly to home in 2012-13, we estimate the reduced-form effect of the IVs on the 90-day mortality and 30-day hospital readmissions, conditional on ZIP code fixed effects (see Appendix Table A4). An association of post-discharge outcomes with variation in proximity to SNFs with different star rating would implicate some violation of the instrument's exclusion restriction, perhaps related to hospital quality or post-discharge practices. The association of the IVs with these outcomes was both statistically insignificant and negligibly small (see Appendix Table A4). The null finding in this non-SNF sample is consistent with our assumptions about the instrument's exogeneity.

B. Regression results

Before describing the main results, we discuss the results from the first-stage models to show how strong our instrumental variables are (see Table 4). The outcome in the first-stage model is star rating, an integer from 1 to 5. We run five versions of the model, differing in whether we included fixed effects or not for ZIP code, SNF and hospital. The first column has no fixed effects. The second and third columns have either hospital fixed effects or patient ZIP code fixed effects. The fourth column has both hospital and patient ZIP code fixed effects. The fifth column has patient ZIP code and SNF fixed effects. The results are broadly similar across all five columns, although not surprisingly, when adding fixed effects the *R*-squared increases and the *F*-statistics for the instruments decrease. Our preferred model is the third column, with patient ZIP code fixed effects, because this appears to provide the most control for unobserved patient-

level characteristics; adding hospital or SNF fixed effects has little additional predictive power, while using up many degrees of freedom.

The five instrumental variables individually and collectively strongly predict the endogenous variable of star rating. Each instrument is the natural logarithm of the distance from the patient's ZIP code centroid to the nearest SNF of a particular star rating (1 through 5, for five instruments). The t-statistics in the model with patient ZIP code fixed effects range from -46 to 35. Therefore, each instrument is highly statistically significant. Collectively for the five instruments, the F-statistic is greater than 365 in all models with and without fixed effects (Table 4), far exceeding the minimum recommended number (Staiger and Stock, 1997). The pattern of coefficient magnitudes is also revealing and plausible. The coefficients on the distance to a one-star SNF is positive, and the magnitudes decline monotonically with the coefficient on the distance to a five-star SNF being negative.

Table 5 shows the first-stage regression results where admission to SNFs with different star ratings are modeled as separate outcomes. The four columns of this table represent regressions for admission to SNF with $k = 2, 3, 4$ and 5 star rating, with 1-star SNF as the benchmark. These models include ZIP code fixed effects. The likelihood of admission to a SNF with star rating k is negatively associated with the distance to the nearest SNF with star rating k and positively associated with the distances to the SNFs with star rating other than k. For example, an increase in the distance of the nearest 2 star-SNF by one mile decreases the likelihood of admission to a 2 star SNF by 1.5 percentage points. On the other hand, an increase in the distance of the nearest 3 star-SNF by one

26

mile increases the likelihood of admission to a 2 star SNF by 0.5 percentage points (see Appendix Table A1).

The main results show that patients discharged from a hospital who go to a higher rated SNF have significantly lower rates of death, rehospitalization, and long-term nursing stay (see Table 6). Each coefficient in Table 6 is from a different regression model. After discussing the results from our preferred specification — 2SLS with patient ZIP code fixed effects — we describe how the results are similar or different across different model specifications.

The outcomes are any acute hospitalizations in 30 and 180 days, death within 30 and 100 days, and becoming a long-stay (>100 days) nursing-home resident. The results from the 2SLS model with ZIP code fixed effects and patient characteristics show that going to a SNF with one additional star leads to statistically significant decreases in the probability of 30-day and 180-day hospitalizations of 0.005 and 0.004, although our estimates are not statistically different from zero. The probability of death within 30 and 180 days decreases by 0.005 and 0.01. The probability of a nursing-home stay over 100 days decreases by 0.008. These results for mortality and long-term nursing-home stay are statistically significant and clinically striking. To put the magnitudes in context, compared to the sample means reported in Table 1, these effects represent 6% and 5% decreases from the average 30-day and 180-day levels of mortality of 0.07 and 0.21, respectively; and 5% decrease from average incidence of long-term nursing stay, which was 0.16. A change of two stars would have twice the predicted effect.

Table 6 also supports the assumption that our instruments are exogenous to patient characteristics, conditional on patient residential ZIP code. Coefficients change

27

notably with the addition of ZIP code fixed effects, but adding patient characteristics in the third set of models, estimates are similar, suggesting that the IV estimates are not conditional on patient characteristics. (Reduced-form estimates in Appendix Table A2 are consistent with this result.) After adjusting for ZIP code fixed effects, the OLS and IV results are similar. The Hausman test statistics also do not reject the null hypothesis of exogenous regressors, suggesting that with ZIP fixed effects OLS gives similar estimates to our IV estimator. The final set of models adds SNF fixed effects. Addition of SNF fixed effects to the OLS models does substantially change the effect finding, resulting effects of star-rating to be statistically insignificant. However, in the 2SLS models, estimates do not change with the addition of SNF fixed effects. Similarly, the Hausman test statistics reject the null hypothesis of exogenous regressor. We proceed with the robustness checks based on the (more parsimonious) instrumental variable model with patient characteristics and ZIP code fixed effects.

Table 7 shows the main 2SLS regression results (zip code fixed effect model) with star rating as categorical variables, where 1-star SNF is the reference category. An admission to a 5-star SNF instead of a 1-star SNF, reduced 30-day mortality by 2 percentage points, 180-day mortality by 4.5 percentage points, and long-term nursing stay by 4 percentage points. These results are generally consistent with a fourfold increases of the effects with ZIP code fixed effects in Table 6, suggesting that specifying a linear effect of star rating is a reasonable approximation. The addition of SNF fixed effects (Appendix Table A3) gives similar results.

Table 8 shows an alternative specification with the dependent variables defined as the number of days in one of six mutually exclusive and exhaustive states (death, hospice,

28

inpatient, nursing home, home health, and home). Because the total number of days is 180, the coefficients in each column corresponding to the same set of models (either OLS or 2SLS) sum approximately to zero. The magnitudes of the coefficients are directly interpretable as days. For example, the causal effect of an increase in star rating by one star is to spend 1.44 fewer days deceased in the first six months post discharge. Given that on average patients spend 25 days in death, a 1.44 day change is a 5.8% change. The percentage changes from the mean for the other outcomes are 3.3% increase in inpatient hospital days, 2.4% decrease in SNF days, 4.0% increase in days at home with home health, and 2.5% increase in days at home. We also examined the mortality-adjusted days. These results, presented in Appendix Table A7, are consistent in magnitude and direction with our main results.

It is instructive to see how the results are a function of time since discharge. We re-ran all the models 180 times, by day, and graphed the results (see Figure 5). The *x*-axis for each of the six graphs are time in days, and the *y*-axis is the increase in probability of being in that state given a one-star increase in rating. The 2SLS coefficients are roughly equal to the integral of the effects shown in the graphs. The effect of an increase in star rating increases the chance of being at home, and that effect gets gradually stronger. Higher rated SNFs also increase the use of home health, lower the use of SNFs, and lower the probability of death. These results are not short run; they get larger in magnitude and stronger in statistical significance over time.

Table 9 presents specification checks. The first two rows show binary outcomes with hospital fixed effects; these are similar to our main results. We report 2SLS results

29

with separately for patients in hospitals with and without hospital based SNF. The direction of the effects are the same for both samples. However, the size of the effect varies. Finally, inclusion of diagnoses related group (DRG) fixed effects does not change the results substantially.

V. Discussion

We find that discharge to a higher star SNF led to significantly lower mortality, fewer days in the nursing home, fewer hospital readmissions, and more days at home or with home health care during the first six months post SNF admission. SNF star ratings matter for patient outcomes. This is intuitive, given that star ratings are based in part on quality of care measures, and consistent with previous estimates of the association of star rating with hospitalization and mortality, some of which we summarize in Appendix Table A7. But prior studies have not been able to control for the clear endogeneity of choice of SNF, leaving uncertain whether simple correlations are due to patient selection or actual causal effects. Our results show a strong causal effect in the expected direction. Our instrument identifies variation in star rating both from within-provider change in star rating and from patients who go to a different facility because of changes in relative distances of each quality rating. The IV results are robust to the inclusion of SNF fixed effects, suggesting that within-SNF changes in star rating do reflect differences in patient outcomes.

Some limitations in the interpretation of these results are important to note. We estimate a local average treatment effect among patients whose choice of SNF quality is

30

influenced by variation in distance to the SNF in each quality category. Thus our results may be driven by patients (or their discharge planners) with stronger preferences for quality.

Star ratings are not based on patient mortality, and yet we find that higher ratings are nonetheless powerful predictors of survival. In 2016, CMS introduced new quality measures into the star rating, including short-stay, unplanned hospital readmission and successful discharge to the community. We do not find strong effects of star rating on hospital admissions, but these new measures may improve the performance of star rating to predict hospitalization.

These findings give rise to two important questions. First, given that higher star ratings lead to lower rates of death and hospital readmission,, are patients and their advocates using this measure effectively when selecting SNFs? And if not, what policy measures could increase the use of the star ratings?

A large literature has suggested that the use of the star ratings on the Nursing Home Compare website is relatively low (e.g., Konetzka and Perraillon, 2016; Shugarman and Brown, 2007). Consumers lack awareness of the ratings but also have some mistrust of the rankings. Patients also choose SNFs based on more than just quality of care. As Shugarman and Brown (2007) suggest, "patients and their families are likely to rely upon lists of facilities (when available), more obvious physical and sensory characteristics of the facilities, and word of mouth, and be more concerned with the location of the facility than with the technical aspects of the clinical quality of care provided" (p.23). Thus, many SNF patients do not select the highest quality SNF in their choice set (Medicare Payment Advisory Commission 2018).

31

Given our results, which suggest that star ratings predict important outcomes, and the fact that few patients use these ratings to select a nursing home, what would be the consequence if Medicare were to adopt new policies to more actively encourage use of star ratings? Such policies might include providing assistance to patients who are referred to SNFs for post-acute care in identifying those higher-star providers. Fee-for-service beneficiaries have a "basic freedom of choice" to select any SNF participating in Medicare, although this choice may be constrained by out-of-pocket costs for patients who stay in a SNF beyond 90-day Medicare benefit (Rahman et al. 2014). Our study does not observe those costs and whether they correlate with star rating. However, Medicare could expand the authority of hospital discharge planners to recommend higher quality SNFs by mandating that they provide star ratings to patients at the time of discharge (Medicare Payment Advisory Commission 2018). Most hospital discharge planners are currently fairly hands off in directing patients to particular SNFs (Tyler et al., 2017). This nudge would ensure that every hospitalized Medicare beneficiary would be made aware of the quality ratings at the time of discharge, while patients would retain the freedom to select their preferred SNF.

In addition to reforming the hospital discharge process, new alternative payment models are another potential mechanism to encourage increased use of star ratings. ACOs and bundled payment models put hospitals at risk for post-discharge spending and outcomes. As such, these models create an incentive for hospitals to develop relationships with SNFs that can reasonably shorten the length of SNF stay while also reducing the likelihood of re-hospitalization during the episode (Mechanic 2016). Many hospital systems are forming preferred networks to partner with higher quality SNFs (MedPAC

2018). As these payment models expand and shift from voluntary to mandatory participation, our results suggest that star ratings can be a meaningful source of information for hospitals looking to identify SNF partners whose patients are less likely to be readmitted to the hospital.

In summary, a large literature has suggested that SNF star ratings and patient outcomes are correlated. Our paper is the first to find that admission to a higher star SNFleads to lower mortality, decreased hospitalization, and shorter SNF and nursing home stays. . Promoting the use of star ratings by patients and discharge planners is a strategy that Medicare might pursue as it considers ways to achieve these outcomes in post-acute care.

References

Arling, G., T. Lewis, R. L. Kane, C. Mueller and S. Flood (2007). "Improving quality assessment through multilevel modeling: the case of nursing home compare." Health services research 42(3 Pt 1): 1177-1199.

Bahar, Dany, (2014), IVREG2HDFE: Stata module to estimate an Instrumental Variable Linear Regression Model with two High Dimensional Fixed Effects, https://EconPapers.repec.org/RePEc:boc:bocode:s457841

Bowblis, J. R. and C. S. Brunt. (2013). "Medicare Skilled Nursing Facility Reimbursement and Upcoding." *Health Economics* 23(7), 821-840.

Brooks, J.M., Irwin, C.P., Hunsicker, L.G., Flanigan, M.J., Chrischilles, E.A., Pendergast, J.F. (2006). Effect of dialysis center profit-status on patient survival: a comparison of risk-adjustment and instrumental variable approaches. *Health Services Research* 41 (6), 2267–2289.

Centers for Medicare and Medicaid Services (CMS). (2017). "Design for Nursing Home Compare Five-Star Rating System: Technical Users' Guide."

Centers for Medicare and Medicaid Services (CMS). "Five-star quality rating system." (Internet). Access date 1 Feb 2017. https://www.cms.gov/medicare/provider-enrollment-and-certification/certificationandcomplianc/fsqrs.html

Deyo, R.A., Cherkin, D.C., Ciol, M.A. (1992). "Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases." *Journal of Clinical Epidemiology* 45 (6), 613–619.

Doyle, J. J., J. A. Graves and J. Gruber (2017). "Evaluating Measures of Hospital Quality." National Bureau of Economic Research, Inc, Working Paper No. 23166.

Elixhauser, A., Steiner, C., Harris, D.R., Coffey, R.M. (1998). "Comorbidity measures for use with administrative data." *Medical Care* 36 (1), 8–27.

Gowrisankaran, G. and Town, R.J. (1999). "Estimating the quality of care in hospitals using instrumental variables." *Journal of Health Economics, 18* (6):747-767.

Grabowski, D. C. and E. C. Norton (2012). Nursing Home Quality of Care. The Elgar Companion to Health Economics, Second Edition, edited by A. M. Jones. Cheltenham, UK, Edward Elgar Publishing, Inc: 307-317.

Grabowski, D. C., D. G. Stevenson, D. J. Caudry, A. J. O'Malley, L. H. Green, J. A. Doherty and R. G. Frank (2017). "The Impact of Nursing Home Pay-for-Performance on Quality and Medicare Spending: Results from the Nursing Home Value-Based Purchasing Demonstration." Health Services Research 52(4):1387-1408.

Grabowski, D. C., Z. Feng, R. Hirth, M. Rahman and V. 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.

Hirth, R. A., D. C. Grabowski, Z. Feng, M. Rahman and V. Mor (2014). "Effect of nursing home ownership on hospitalization of long-stay residents: an instrumental variables approach." International journal of health care finance and economics 14(1): 1-18.

Institute of Medicine (2013). J.P. Newhouse, A.M. Garber, R.P. Graham, M.A. McCoy, M.Mancher, & A. Kibria (Eds.). Variation in Health Care Spending: Target Decision Making, Not Geography. Washington, DC: The National Academies Press. https://doi.org/10.17226/18393.

Intrator, O., J. Hiris, K. Berg, S. C. Miller and V. Mor (2011). "The residential history file: studying nursing home residents' long-term care histories(*)." Health Serv Res 46(1 Pt 1): 120-137.

Kimball, C.C., C.I. Nichols, R.M. Nunley, J.G. Vose, J.B. Stambough (2018). "Skilled Nursing Facility Star Rating, Patient Outcomes, and Readmission Risk After Total Joint Arthroplasty." Journal of Arthroplasty 33(10):3130-3137.

Konetzka, R.T., M.C. Perraillon (2016). "Use Of Nursing Home Compare Website Appears Limited By Lack Of Awareness And Initial Mistrust Of The Data." Health Affairs 35(4):706-713.

Konetzka, R. T., D. C. Grabowski, M. C. Perraillon and R. M. Werner (2015). "Nursing home 5-star rating system exacerbates disparities in quality, by payer source." Health affairs 34(5): 819-827.

Li, Y., X. Cai, L. G. Glance, W. D. Spector and D. B. Mukamel (2009). "National release of the nursing home quality report cards: implications of statistical methodology for risk adjustment." Health services research 44(1): 79-102.

McClellan, M., McNeil, B.J., Newhouse, J.P. (1994). "Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables." JAMA: The Journal of the American Medical Association 272 (11), 859–866.

McWilliams, J. M., L. G. Gilstrap, D. G. Stevenson, M. E. Chernew, H. A. Huskamp and D. C. Grabowski (2017). "Changes in Postacute Care in the Medicare Shared Savings Program." JAMA internal medicine 177(4): 518-526.

Mechanic, R. (2016). Medicare's Bundled Payment Initiatives: Considerations for Providers. Issue Brief. Washington, D.C., American Hospital Association.

Medicare Payment Advisory Commission (2009). Payment Basics: Skilled Nursing Facility Services Payment System. Washington, DC.

Medicare Payment Advisory Commission (2016). Report to the Congress: Medicare Payment Policy. Washington, DC.

Medicare Payment Advisory Commission (2017). Report to the Congress: Medicare payment policy. Washington, DC.

Medicare Payment Advisory Commission (2018). "Encouraging Medicare beneficiaries to use higher quality post-acute care providers (Chapter 5)." Report to the Congress: Medicare payment policy. Washington, DC. Available at: http://www.medpac.gov/docs/default-source/reports/jun18_ch5_medpacreport_sec.pdf?sfvrsn=0

Mor, V. (2006). "Defining and measuring quality outcomes in long-term care." J Am Med Dir Assoc 7(8): 532-538; discussion 538-540.

Morris, J.N., Fries, B.E., Morris, S.A. (1999). "Scaling ADLs within the MDS." The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences 54 (11), M546–M553.

Mukamel, D. B., L. G. Glance, Y. Li, D. L. Weimer, W. D. Spector, J. S. Zinn and L. Mosqueda (2008). "Does risk adjustment of the CMS quality measures for nursing homes matter?" Med Care 46(5): 532-541.

Newhouse, J. P. and A. M. Garber (2013). "Geographic variation in Medicare services." The New England Journal of Medicine 368(16): 1465-1468.

37

Norton, E.C. (1992). "Incentive Regulation of Nursing Homes." Journal of Health Economics 11(2):105–128.

Norton, E.C., J. Li, A. Das, and L.M. Chen (2018). "Moneyball in Medicare." Journal of Health Economics 61:259–273.

Ogunneye, O., M.B. Rothberg, J. Friderici, M.T. Slawsky, V.T. Gadiraju, M.S. Stefan (2015). "The association between skilled nursing facility care quality and 30-day readmission rates after hospitalization for heart failure." American Journal of Medical Quality 30(3):205-13.

Rahman, M., Gozalo, P., Tyler, D., Grabowski, D. C., Trivedi, A., & Mor, V. (2014). Dual eligibility, selection of skilled nursing facility, and length of Medicare paid postacute stay. Medical Care Research and Review, 71(4), 384-401.

Rahman, M., D. C. Grabowski, V. Mor and E. C. Norton (2016). "Is a Skilled Nursing Facility's Rehospitalization Rate a Valid Quality Measure?" Health Services Research 51(6): 2158-2175.

Rahman, M., E. C. Norton and D. C. Grabowski (2016). "Do hospital-owned skilled nursing facilities provide better post-acute care quality?" Journal of Health Economics 50: 36-46.

Rahman, M., Foster, A.D., 2015. "Racial segregation and quality of care disparity in US nursing homes." Journal of Health Economics 39, 1–16.

Rahman, M., J. S. Zinn and V. Mor (2013). "The Impact of Hospital‐Based Skilled Nursing Facility Closures on Rehospitalizations." Health services research 48(2pt1): 499-518.

Ryskina, K.L., R.T. Konetzka, R.M. Werner (2018). "Association between 5-star Nursing Home Report Card Ratings and Potentially Preventable Hospitalizations." Inquiry 55:1-14.

Shugarman, L.R., and J.A. Brown (2007). "Nursing Home Selection: How Do Consumers Choose? Volume I: Findings from Focus Groups of Consumers and Information Intermediaries." Santa Monica, CA: RAND Corporation.

Sood, N., P. J. Huckfeldt, J. J. Escarce, D. C. Grabowski and J. P. Newhouse (2011). "Medicare's bundled payment pilot for acute and postacute care: analysis and recommendations on where to begin." Health Affairs 30(9): 1708-1717.

Staiger, D. and J.H. Stock. 1997. "Instrumental variables regression with weak instruments." *Econometrica* 65(3): 557–586.

Tan, H.J., Norton, E.C., Ye, Z., Hafez, K.S., Gore, J.L., Miller., D.C., 2012. "Long-term survival following partial vs radical nephrectomy among older patients with early-stage kidney cancer." *JAMA: The Journal of the American Medical Association* 307 (15), 1629–1635.

Tyler, D.A., E.A. Gadbois, J.P. McHugh, R.R. Shield, U. Winblad, V. Mor (2017). "Patients Are Not Given Quality-Of-Care Data About Skilled Nursing Facilities When Discharged From Hospitals." Health Affairs 36(8):1385-1391.

Unroe, K.T., M.A. Greiner, C. Colón-Emeric, E.D. Peterson, L.H. Curtis (2012). "Associations between published quality ratings of skilled nursing facilities and outcomes of Medicare beneficiaries with heart failure." Journal of the American Medical Directors Association 13(2):188.e1.

U.S. General Accounting Office (2002). Nursing Homes: Public Reporting of Quality Indicators Has Merit, but National Implementation Is Premature. Washington, DC: U.S. General Accounting Office.

Werner, R.M., E.C. Norton, R.T. Konetzka, D. Polsky. (2012). "Do consumers respond to publicly reported quality information? Evidence from nursing homes." Journal of Health Economics 31(1):50–61

List of Exhibits

41

42

Table 1: Outcomes (N=1,278,456)

Variable	Mean	Std. Dev.	Min	Max
Binary outcomes				
Any rehospitalization within 30 days	0.177	0.382	0	1
Any hospitalization within 180 days	0.536	0.499	0	1
Death within 30 days	0.071	0.256	0	1
Death within 180 days	0.214	0.410	0	1
Becoming long-stay nursing resident	0.155	0.362	0	1
Number of days in different settings				
Number of days at home	70.21	64.14	0	180
Number of days at home with home health	28.33	38.08	0	180
Number of days in skilled nursing facility	48.24	49.22	1	181
Number of days in hospital	5.48	10.96	0	180
Number of days in hospice	3.69	17.26	0	180
Number of days deceased	25.04	52.88	0	179

Note: We started counting the day from day 0 when everyone was in nursing home. Summation of the mean number of days is 181

Table 2: Descriptive characteristics (N=1,278,456)

Variable	Mean	Std. Dev.	Patients with rating < ZIP code level median N=404,402	Patients with rating = ZIP code level median N=503,535	Patients with rating > ZIP code level median N=372,741
Demographic characteristics					
Age	81.34	8.20	81.224	81.494	81.359
Language Spanish	0.01	0.11	0.013	0.011	0.011
Other language	0.01	0.12	0.016	0.011	0.014
Black	0.08	0.26	0.089	0.065	0.074
Hispanic	0.04	0.19	0.040	0.034	0.035
Other race	0.02	0.15	0.026	0.020	0.024
Female	0.65	0.48	0.637	0.652	0.665
Full dual-eligible	0.14	0.35	0.170	0.135	0.125
Partial dual-eligible	0.04	0.20	0.043	0.042	0.036
Married	0.36	0.48	0.348	0.368	0.362
Clinical characteristics from index hospital claims					
Deyo-Charlson comorbidity score	2.17	2.14	2.267	2.139	2.099
Elixhauser comorbidity score	3.39	1.83	3.461	3.365	3.336
Total ICU days	1.59	3.84	1.690	1.539	1.523
Total CCU days	0.62	2.30	0.642	0.592	0.630
Hospital length of stay	6.82	6.06	7.136	6.655	6.688
Clinical characteristics from MDS assessments					
Shortness of breath	0.21	0.41	0.209	0.215	0.205
ADL score (0-28, higher = worse)	17.11	4.68	17.427	16.977	16.951
CPS scale	1.34	1.44	1.454	1.310	1.266
Stroke	0.10	0.30	0.106	0.096	0.092
Lung disease	0.22	0.42	0.226	0.223	0.213
Alzheimer's disease	0.04	0.20	0.049	0.042	0.040
Non-Alzheimer's dementia	0.17	0.37	0.187	0.161	0.156
Hip fracture	0.08	0.28	0.079	0.085	0.085
Multiple sclerosis	0.00	0.05	0.003	0.003	0.002
Heart failure	0.20	0.40	0.204	0.203	0.195
Diabetes	0.30	0.46	0.311	0.295	0.289
Schizophrenia	0.01	0.08	0.008	0.005	0.005
Bipolar disease	0.01	0.10	0.012	0.009	0.009
Aphasia	0.01	0.12	0.015	0.013	0.013

Notes: ICU= intensive care unit, CCU=coronary care unity, MDS=minimum data set, ADL = activities of daily living

Table 3: Characteristics of patients with different values of IV (logged distance of the nearest 5 star SNF) compared to their respective zip code level median

	Patients with IV < ZIP code level median 190,767	Patients with IV = ZIP code level median 857,869	Patients with IV > ZIP code level median 254,848
Age	81.233	81.366	81.320
Language Spanish	0.012	0.012	0.012
Other language	0.013	0.013	0.014
Black	0.074	0.076	0.074
Hispanic	0.037	0.036	0.037
Other race	0.024	0.023	0.024
Female	0.651	0.650	0.649
Full dual-eligible	0.145	0.141	0.146
Partial dual-eligible	0.042	0.040	0.043
Married	0.360	0.362	0.361
Deyo-Charlson score	2.177	2.170	2.182
Elixhauser score	3.389	3.389	3.388
Total ICU days	1.640	1.563	1.621
Total CCU days	0.591	0.628	0.613
Hospital length of stay	6.759	6.826	6.856
Shortness of breath	0.212	0.210	0.212
ADL score (0-28, higher = worse)	17.000	17.130	17.143
CPS scale	1.345	1.336	1.357
Stroke	0.098	0.097	0.100
Lung disease	0.222	0.220	0.223
Alzheimer's disease	0.043	0.043	0.045
Non-Alzheimer's dementia	0.165	0.167	0.169
Hip fracture	0.085	0.082	0.082
Multiple sclerosis	0.003	0.003	0.003
Heart failure	0.202	0.200	0.203
Diabetes	0.302	0.297	0.299
Schizophrenia	0.006	0.006	0.006
Bipolar disease	0.010	0.010	0.009
Aphasia	0.013	0.014	0.014

Table 4: First stage: Regression of star rating of admitting SNF onto patient characteristics and distances of different star rating SNFs

Log of distance to nearest SNF rating	Median distance (in miles) to nearest SNF	(1) No fixed effects	(2) Hospital fixed effects	(3) ZIP code fixed effects	(4) Hospital and ZIP code fixed effects	(5) SNF and ZIP code fixed effects
1-star	6.8	0.129***	0.0804***	0.0683***	0.0684***	0.0698***
		[0.00221]	[0.00150]	[0.00195]	[0.00194]	[0.00290]
2-star	5.0	0.129***	0.0855***	0.0633***	0.0628***	0.0629***
		[0.00256]	[0.00173]	[0.00239]	[0.00238]	[0.00320]
3-star	4.9	0.0227***	0.0198***	0.0101***	0.0103***	0.0123***
		[0.00258]	[0.00172]	[0.00246]	[0.00244]	[0.00319]
4-star	4.1	-0.155***	-0.0866***	-0.0804***	-0.0809***	-0.0778***
		[0.00308]	[0.00214]	[0.00330]	[0.00329]	[0.00418]
5-star	5.5	-0.226***	-0.130***	-0.121***	-0.120***	-0.118***
		[0.00250]	[0.00183]	[0.00261]	[0.00260]	[0.00357]
Observations		1,278,456	1,278,456	1,278,456	1,278,456	1,277,815
R-squared		0.134	0.249	0.262	0.2993	0.836
F-statistic		3126.72	1122.40	771.17	364.99	388.97

Note: Robust *standard errors* are in square brackets. Clustering units were ZIP code in column 1, 3-4, hospital in column 2 and SNF in column 5. Each column represents a separate regression. All regressions include patient characteristics listed in table 2 and month of SNF admission dummies.

*** $p<0.01$, ** $p<0.05$, * $p<0$.

Table 5: First stage regression with star rating as categorical variable with ZIP code fixed effect model

Log of distance to nearest SNF rating	(1) Admission to 2 star SNF	(2) Admission to 3 star SNF	(3) Admission to 4 star SNF	(4) Admission to 5 star SNF
1-star	0.0141***	0.00749***	0.00595***	0.00532***
	(0.000699)	(0.000707)	(0.000780)	(0.000643)
2-star	-0.0753***	0.0234***	0.0166***	0.0104***
	(0.00113)	(0.000948)	(0.000993)	(0.000821)
3-star	0.0250***	-0.0884***	0.0293***	0.0186***
	(0.000919)	(0.00117)	(0.00113)	(0.000845)
4-star	0.0203***	0.0402***	-0.115***	0.0408***
	(0.00108)	(0.00133)	(0.00171)	(0.00129)
5-star	0.00857***	0.0184***	0.0352***	-0.0679***
	(0.000833)	(0.000938)	(0.00126)	(0.00120)
Observations	1,278,456	1,278,456	1,278,456	1,278,456
R-squared	0.164	0.156	0.170	0.211

Note: Robust *standard errors* clustering error by ZIP codes are in parentheses. Each column represent a separate regression. All regressions include patient characteristics listed in table 2, month of SNF admission dummies and patient's residential ZIP code fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Effect of SNF star rating (assumed continuous) on patient outcomes (binary)

		Any acute hospitalization within 30 days	Any hospitalization within 180 days	Death within 30 days	Death within 180 days	Became long-stay nursing-home resident
OLS	Without patient characteristics or any FE	-0.00774*** (0.000299)	-0.0119*** (0.000404)	-0.00577*** (0.000196)	-0.0152*** (0.000326)	-0.0188*** (0.000368)
	Without patient characteristics but with ZIP FE	-0.00689*** (0.000319)	-0.0107*** (0.000423)	-0.00543*** (0.000217)	-0.0145*** (0.000360)	-0.0181*** (0.000367)
	With patient characteristics and ZIP code FE	-0.00291*** (0.00031)	-0.00593*** (0.00041)	-0.00179*** (0.00021)	-0.00606*** (0.00031)	-0.00957*** (0.00032)
	With patient characteristics, SNF FE and ZIP code FE	-0.00083 (0.000650)	-0.00093 (0.000819)	-6.81E-05 (0.000430)	-0.00073 (0.000644)	-0.00109* (0.000585)
2SLS	Without patient characteristics or any FE	-0.0123*** (0.000946) [40.47]***	-0.0182*** (0.00129) [30.35]***	-0.00761*** (0.000581) [9.14]**	-0.0190*** (0.000976) [12.06]***	-0.0166*** (0.00120) [6.06]*
	Without patient characteristics but with ZIP FE	-0.00850*** (0.00276) [0.168]	-0.0106*** (0.00364) [0.929]	-0.00376** (0.00191) [0.485]	-0.0124*** (0.00302) [0.250]	-0.0112*** (0.00275) [7.85]**
	With patient characteristics and ZIP code FE	-0.00500 (0.00306) [0.27]	-0.00441 (0.00404) [0.18]	-0.00479** (0.00205) [2.52]	-0.0120*** (0.00311) [2.28]	-0.00782*** (0.00296) [0.626]
	With patient characteristics, SNF FE and ZIP code FE	-0.00604** (0.00305) [3.68]*	-0.00578 (0.00405) [1.91]	-0.00497** (0.00207) [5.25]**	-0.0120*** (0.00312) [8.94]***	-0.00706** (0.00286) [5.71]**

Note: Robust *standard errors* clustering error are in parentheses. Each coefficient and associated standard errors is derived from a separate regression. Hausman test statistics are reported in square brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Effects of star rating measured as categorical variable (1-star as baseline category) on outcomes (number of days in different settings) using 2SLS model with ZIP code fixed effects

Star rating of admitted SNF	Any acute hospitalization within 30 days	Any hospitalization within 180 days	Death within 30 days	Death within 180 days	Became long-stay nursing-home resident
2-star	-0.002	-0.00786	-0.00603	-0.00426	-0.0253***
	(0.00889)	(0.0118)	(0.00603)	(0.00914)	(0.00868)
3-star	-0.0132	-0.0199	-0.0144**	-0.0281***	-0.0321***
	(0.0107)	(0.0142)	(0.00724)	(0.0109)	(0.0104)
4-star	-0.0163	-0.0211	-0.0205***	-0.0380***	-0.0380***
	(0.0118	(0.0157)	(0.00795)	(0.012)	(0.0115)
5-star	-0.0191	-0.0215	-0.0195**	-0.0445***	-0.0430***
	(0.0135)	(0.0178)	(0.00904)	(0.0137)	(0.0132)

Note: Robust *standard errors* clustering error by ZIP codes are in parentheses. Each column represent a separate regression. All regressions include patient characteristics listed in table 2, month of SNF admission dummies and patient's residential ZIP code fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Regression of alternative outcomes (days in settings) on SNF star rating (assumed continuous; ZIP code fixed effect models)

		A: Home	B: Home health	C: Nursing home	D: Hospital	E: Hospice	F: Death
No fixed effect	OLS	2.728*** (-0.0616)	0.0119 (-0.0469)	-1.790*** (-0.046)	-0.127*** (-0.00883)	0.0157 (-0.0133)	-0.838*** (-0.0373)
	2SLS	4.898*** (-0.113)	-1.262*** (-0.0785)	-1.661*** (-0.0995)	-0.262*** (-0.0215)	-0.00812 (-0.0347)	-1.704*** (-0.0989)
Hospital fixed effects	OLS	2.112*** (0.0781)	0.0784 (0.0488)	-1.417*** (0.0653)	-0.114*** (0.00951)	0.00487 (0.0146)	-0.663*** (0.0467)
	2SLS	1.273*** (0.271)	-0.621*** (0.190)	0.617*** (0.223)	-0.229*** (0.0488)	0.0216 (0.0745)	-1.060*** (0.213)
ZIP code fixed effects	OLS	2.330*** (0.0513)	0.191*** (0.0348)	-1.714*** (0.0453)	-0.112*** (0.00895)	0.00676 (0.0138)	-0.702*** (0.0397)
	2SLS	1.732*** (0.450)	1.124*** (0.315)	-1.139*** (0.401)	-0.182** (0.0843)	-0.0951 (0.141)	-1.440*** (0.401)
ZIP code and hospital fixed effects	OLS	2.149*** (0.698)	0.162*** (0.0459)	-1.534*** (0.0543)	-0.109*** (0.00894)	0.00403 (0.0176)	-0.672*** (0.0392)
	2SLS	1.672*** (0.568)	1.143*** (0.459)	-1.070*** (0.513)	-0.189** (0.0836)	-0.0811 (0.196)	-1.476*** (0.390)

Note: Robust *standard errors* clustering error (by ZIP codes in all models except hospital fixed effect models where clustering unit is hospital) are in parentheses. Each coefficient and associated standard errors is derived from a separate regression. All regressions include patient characteristics listed in table 2 and month of SNF admission dummies. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Regression of outcomes on star rating (assumed continuous) with different model or sample specifications

	Any acute hospitalization within 30 days	Any hospitalization within 180 days	Death within 30 days	Death within 180 days	Became long-stay nursing-home resident
Hospital fixed effects N=1,276,307	-0.00707*** (0.00166)	-0.0113*** (0.00224)	-0.00309*** (0.00108)	-0.00909*** (0.00170)	0.00340** (0.00160)
Hospital & ZIP code fixed effects N=1,276,307	-0.00520* (0.00311)	-0.00480 (0.00409)	-0.00470** (0.00208)	-0.0122*** (0.00315)	-0.00725** (0.00299)
Hospitals without SNF N=901,340	-0.00811** (0.00372)	-0.00495 (0.00487)	-0.00424* (0.00247)	-0.00939** (0.00377)	-0.00794** (0.00351)
Hospitals with SNF N=369,706	-0.0166 (0.0168)	-0.00318 (0.0226)	-0.00364 (0.0113)	-0.0262 (0.0168)	-0.0114 (0.0163)
Including DRG fixed effects N=1,276,307	-0.00558* (0.00305)	-0.00524 (0.00399)	-0.00534*** (0.00203)	-0.0130*** (0.00304)	-0.00749** (0.00293)
Urban beneficiaries N=1,050,252	-0.00766** (0.00369)	-0.00706 (0.00479)	-0.00725*** (0.0024)	-0.0191*** (0.00368)	-0.00793** (0.00348)
Rural beneficiaries N=226,198	0.000665 (0.0055)	-0.00113 (0.00748)	0.000742 (0.00388)	0.0022 (0.00584)	-0.00773 (0.00558)

Note: Robust standard errors clustering error are in parentheses. Each coefficient and associated standard errors is derived from a separate regression. All regressions include patient characteristics listed in table 2 and month of SNF admission dummies. *** p<0.01, ** p<0.05, * p<0.1

51

Figure 1: Distribution of patients in different settings in 180 days following SNF admission
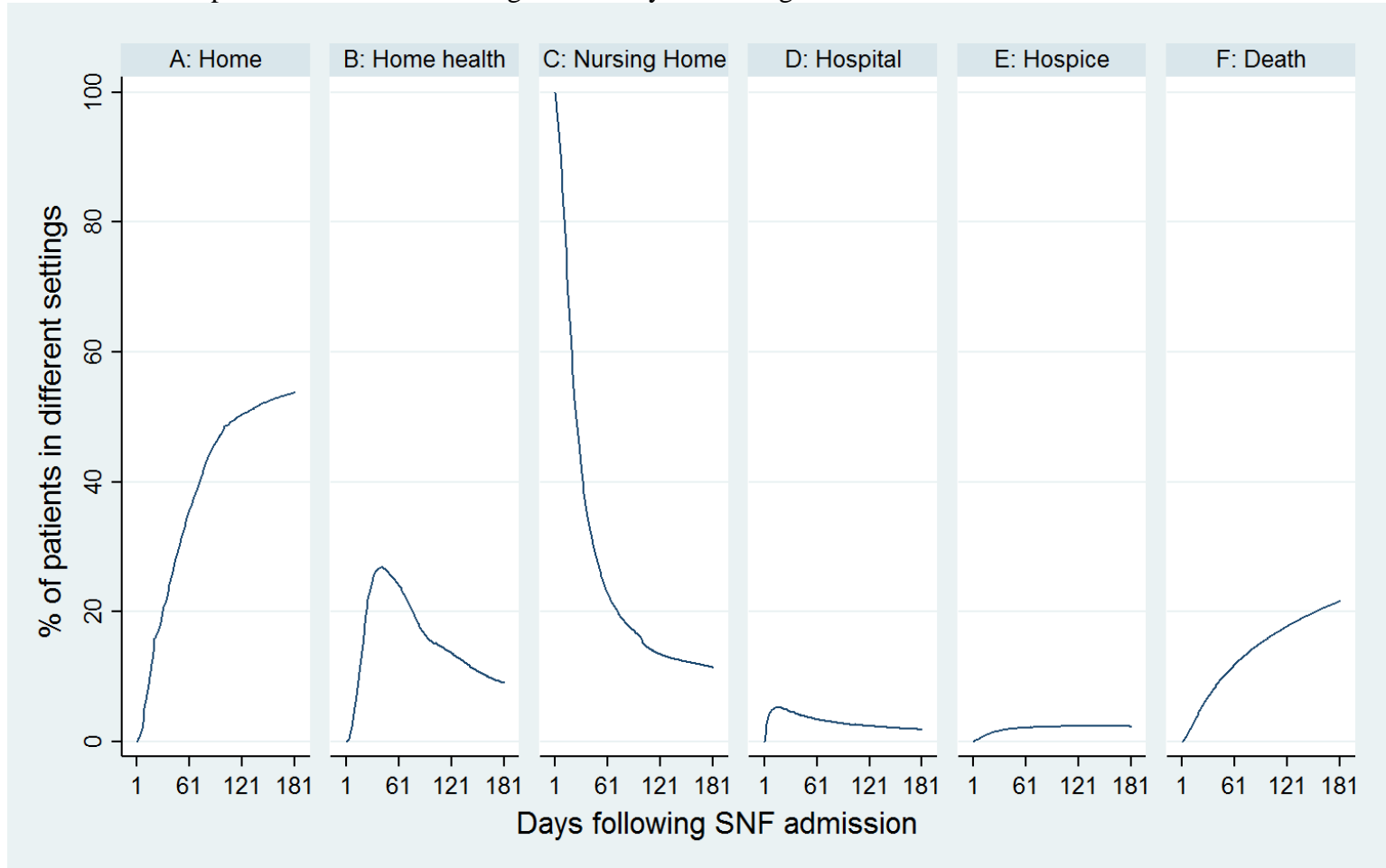
Figure 2: Distribution of patients and skilled nursing facilities by overall star-rating and month of admission

Graphs by Overall Star Rating of SNFs

53
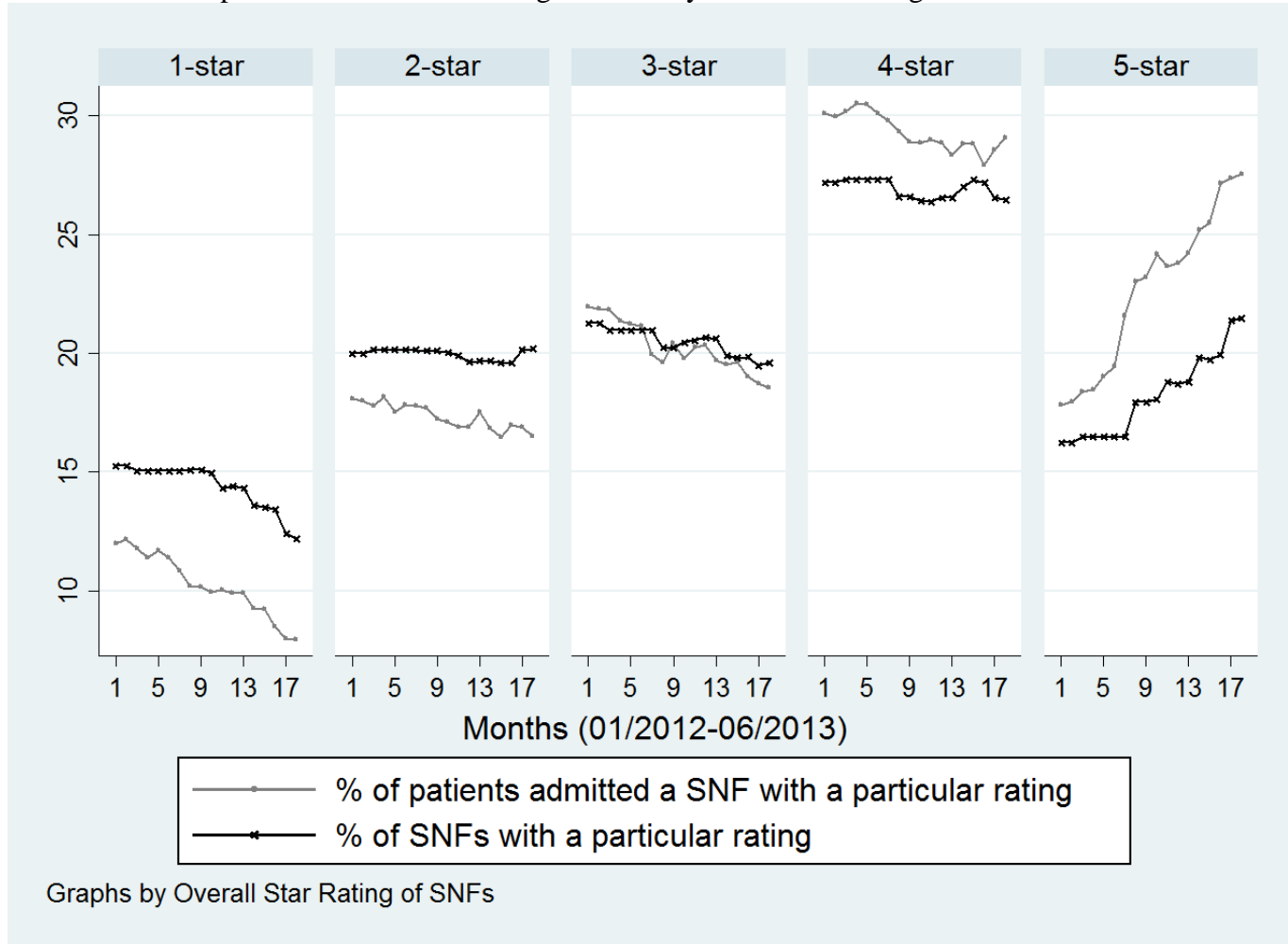
Figure 3: Test of the first stage
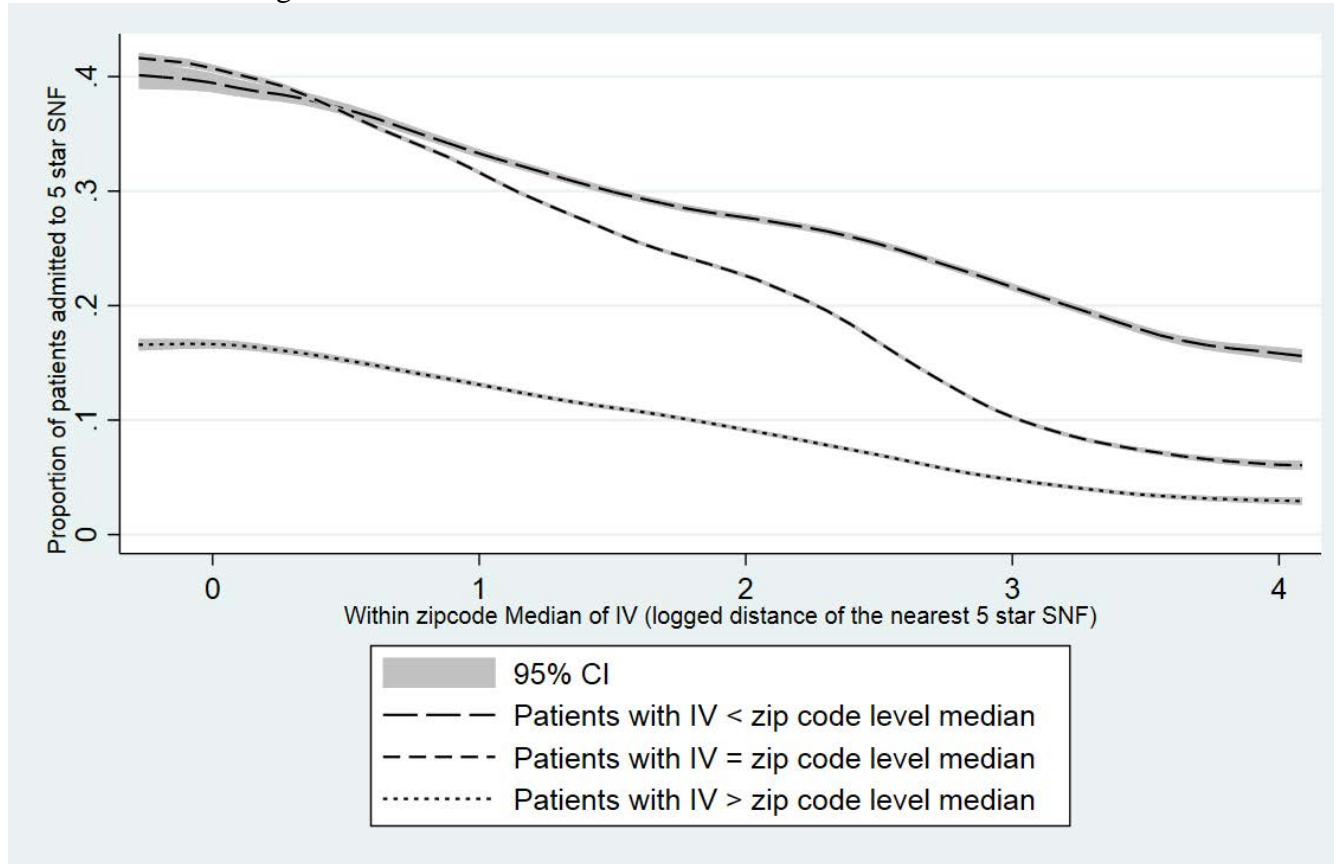
Note: figure excludes upper and lower 5% of the sample.

Figure 4: Comparison of patient characteristics (measured in terms of predicted likelihood of admission to a 5-star SNF) between patients with lower and higher values of IV within a ZIP code
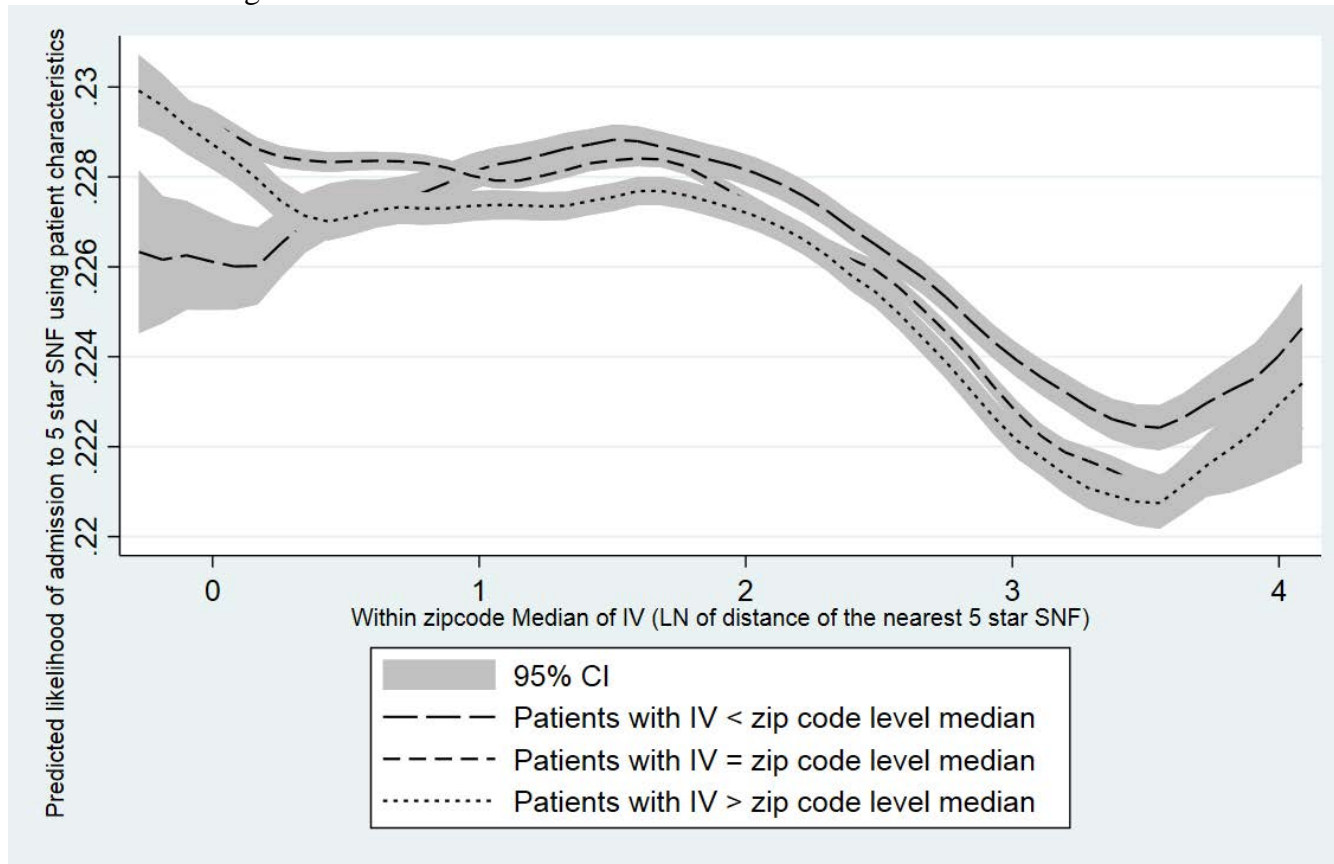
Note: figure excludes upper and lower 5% of the sample.

Figure 5: Effect of increase in overall star rating on probability of being in a location, by day.
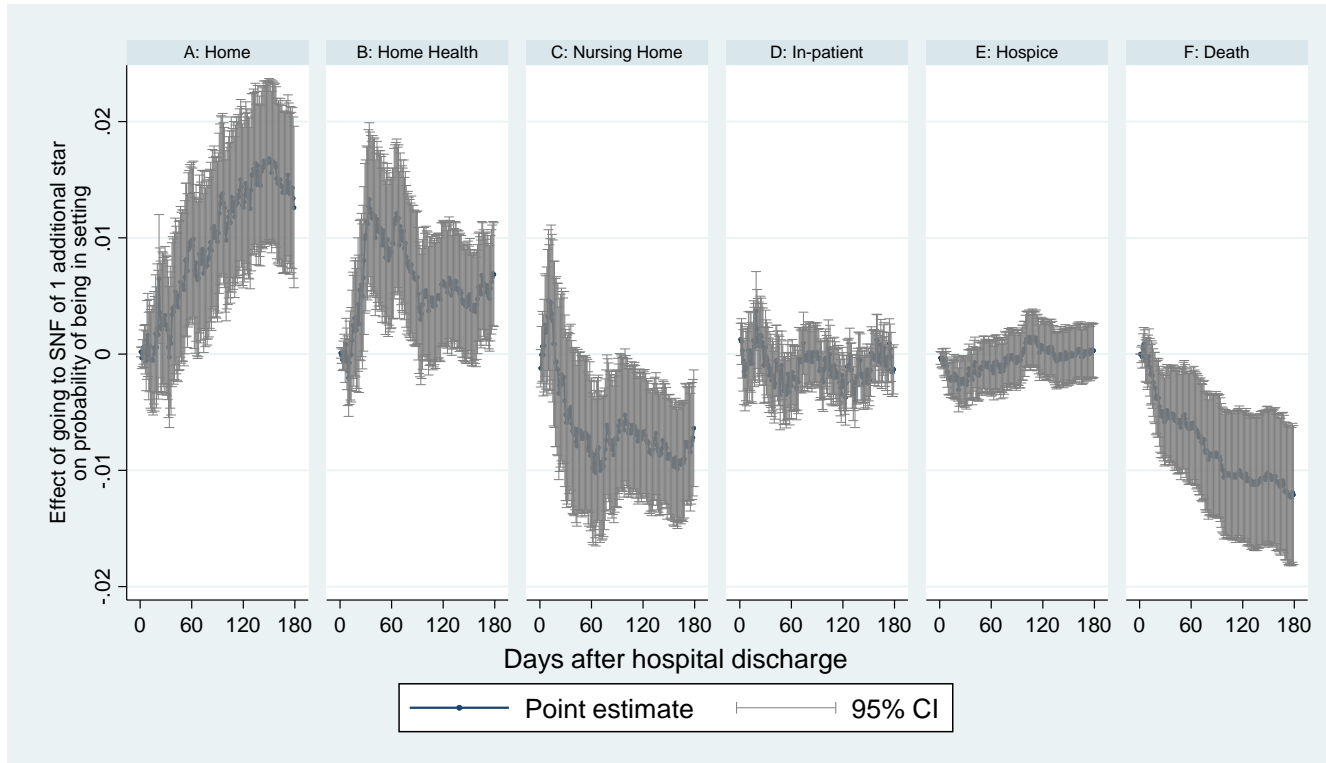
Figure A1: Number of SNFs faced by patients within different mile radius from residential zip codes

excludes outside values
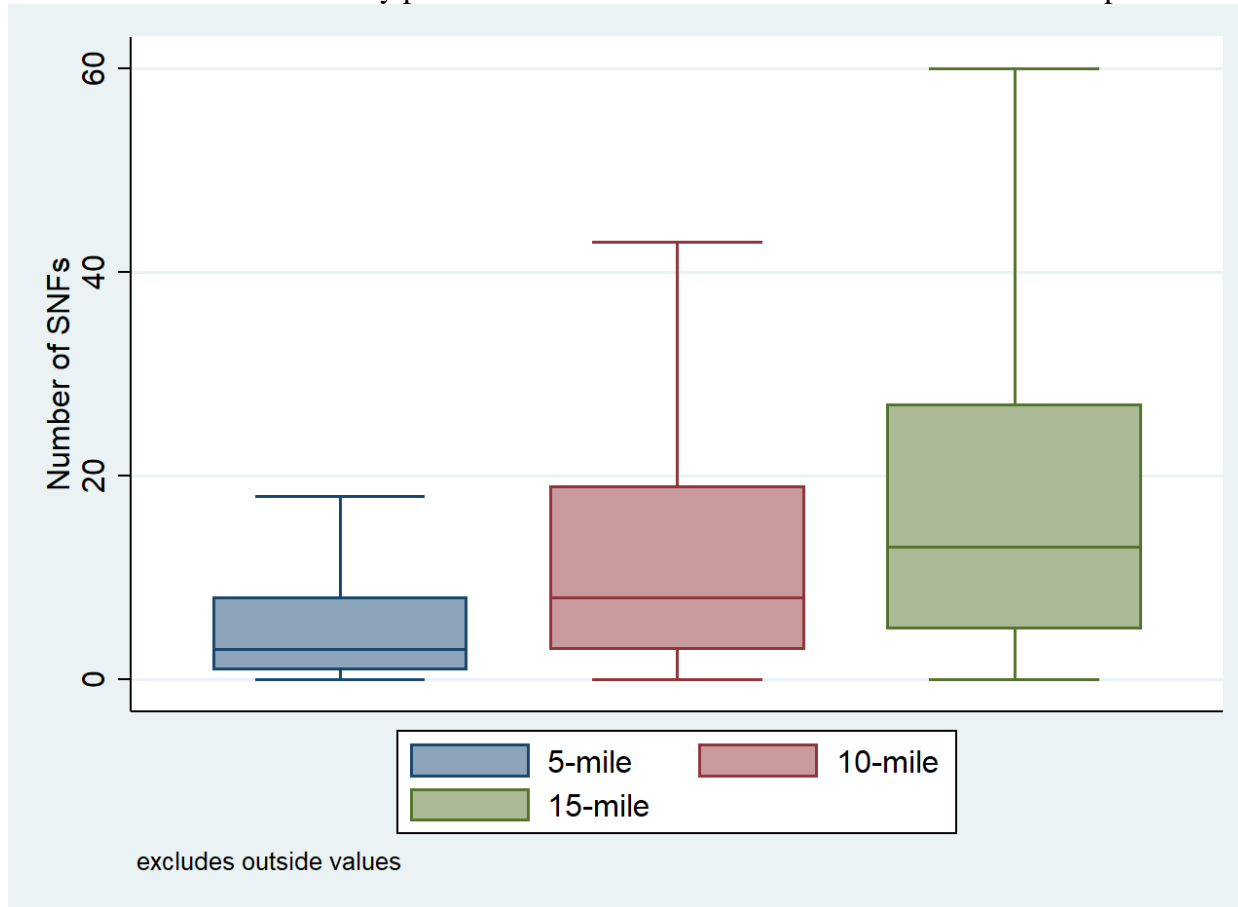
Table A1: Marginal effects of change in IV by one mile on likelihood of entering a SNF with different star-rating (in percentage point: based on regressions presented in Table 5)

	(1) Admission to 2 star SNF	(2) Admission to 3 star SNF	(3) Admission to 4 star SNF	(4) Admission to 5 star SNF
Increase in distance of the nearest 1-star SNF by one mile	0.21	0.12	0.09	0.07
Increase in distance of the nearest 2-star SNF by one mile	-1.50	0.46	0.34	0.20
Increase in distance of the nearest 3-star SNF by one mile	0.51	-1.80	0.59	0.39
Increase in distance of the nearest 4-star SNF by one mile	0.49	0.98	-2.80	1.00
Increase in distance of the nearest 5-star SNF by one mile	0.16	0.33	0.64	-1.24

Table A2: Reduced-form regression of outcomes onto instrument

Variable: log of distance to nearest SNF of star rating		Any acute hospitalization within 30 days	Any acute hospitalization within 180 days	Death within 30 days	Death within 180 days	Became long-stay nursing home resident
Without patient characteristics or FE	1 star	-0.00325***	-0.00484***	-0.000128	-0.00177***	-0.00172***
		(0.000246)	(0.000356)	(0.000163)	(0.000272)	(0.000310)
	2 star	-0.00215***	-0.00206***	0.000626***	-0.000126	0.000456
		(0.000323)	(0.000460)	(0.000222)	(0.000368)	(0.000433)
	3 star	-0.000675**	-0.000856*	0.00136***	0.00205***	-0.000185
		(0.000343)	(0.000477)	(0.000230)	(0.000383)	(0.000417)
	4 star	0.00205***	0.00259***	0.00221***	0.00404***	0.000723
		(0.000380)	(0.000527)	(0.000249)	(0.000413)	(0.000499)
	5 star	0.00194***	0.00303***	0.00209***	0.00474***	0.00573***
		(0.000291)	(0.000411)	(0.000193)	(0.000318)	(0.000379)
No patient characteristics, ZIP code FE	1 star	0.000591	0.000130	0.000600**	0.00280***	0.000660*
		(0.000401)	(0.000534)	(0.000281)	(0.000438)	(0.000395)
	2 star	0.000658	0.000843	0.00104***	0.00369***	0.00229***
		(0.000523)	(0.000691)	(0.000366)	(0.000577)	(0.000516)
	3 star	0.00194***	0.00220***	0.00183***	0.00611***	0.00276***
		(0.000519)	(0.000687)	(0.000373)	(0.000583)	(0.000516)
	4 star	0.00296***	0.00305***	0.00268***	0.00821***	0.00421***
		(0.000658)	(0.000853)	(0.000441)	(0.000701)	(0.000634)
	5 star	0.00261***	0.00280***	0.00154***	0.00613***	0.00386***
		(0.000534)	(0.000691)	(0.000364)	(0.000591)	(0.000527)
Patient characteristics and ZIP code FE	1 star	0.000744*	0.000706	0.000289	0.00253***	0.000611
		(0.000398)	(0.000529)	(0.000271)	(0.000401)	(0.000383)
	2 star	0.000514	0.00109	0.000477	0.00275***	0.00220***
		(0.000513)	(0.000676)	(0.000349)	(0.000526)	(0.000490)
	3 star	0.00162***	0.00215***	0.00110***	0.00479***	0.00252***
		(0.000508)	(0.000677)	(0.000351)	(0.000527)	(0.000492)
	4 star	0.00249***	0.00250***	0.00199***	0.00653***	0.00345***
		(0.000643)	(0.000832)	(0.000420)	(0.000636)	(0.000603)
	5 star	0.00176***	0.00163**	0.00108***	0.00478***	0.00263***
		(0.000527)	(0.000684)	(0.000344)	(0.000541)	(0.000499)
Patient characteristics, SNF and ZIP code FE	1 star	0.000764*	0.000621	0.000261	0.00249***	0.000589
		[0.000393]	[0.000524]	[0.000272]	[0.000401]	[0.000363]
	2 star	0.000774	0.000957	0.000428	0.00269***	0.00219***
		[0.000491]	[0.000668]	[0.000341]	[0.000507]	[0.000468]
	3 star	0.00168***	0.00199***	0.00117***	0.00503***	0.00229***
		[0.000505]	[0.000680]	[0.000348]	[0.000515]	[0.000480]
	4 star	0.00246***	0.00254***	0.00203***	0.00660***	0.00325***
		[0.000627]	[0.000812]	[0.000415]	[0.000633]	[0.000576]
	5 star	0.00192***	0.00184***	0.00110***	0.00491***	0.00253***
		[0.000504]	[0.000678]	[0.000340]	[0.000533]	[0.000485]

Robust standard errors, clustering error by ZIP codes, in parentheses. Each coefficient and associated standard errors is derived from a separate regression. All regressions include month of SNF admission indicators. *** p<0.01, ** p<0.05, * p<0.1

Table A3: Effects of star rating measured as categorical variable (1-star as baseline category) on outcomes (number of days in different settings) using 2SLS model with ZIP code fixed effects and SNF fixed effects

Star rating of admitted SNF	Any acute hospitalization within 30 days	Any hospitalization within 180 days	Death within 30 days	Death within 180 days	Became long-stay nursing-home resident
2-star	-0.00182	-0.00733	-0.00577	-0.00308	-0.0236***
	(0.00870)	(0.0116)	(0.00603)	(0.00896)	(0.00825)
3-star	-0.013	-0.0199	-0.0154**	-0.0299***	-0.0280***
	(0.0106)	(0.0141)	(0.00734)	(0.0108)	(0.01000)
4-star	-0.0181	-0.0229	-0.0212***	-0.0379***	-0.0338***
	(0.0117)	(0.0155)	(0.00803)	(0.0119)	(0.0110)
5-star	-0.0223*	-0.0255	-0.0201**	-0.0448***	-0.0386***
	(0.0134)	(0.0178)	(0.00915)	(0.0137)	(0.0126)

Note: Robust *standard errors* clustering error by ZIP codes are in parentheses. Each column represent a separate regression. All regressions include patient characteristics listed in table 2, month of SNF admission dummies and patient's residential ZIP code fixed effects. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A4: Placebo test. Reduced-form regression of outcomes (30-day rehospitalization, and 30-day mortality, and 180-day mortality) on values of the instrument among patients discharged from hospital directly to home.

	Variable: Log of distance to nearest SNF rating	Coefficient	St. Err.	P-value
180-day mortality N=3558538	1 star	0.00015	0.00016	0.35
	2 star	-0.00018	0.00021	0.38
	3 star	0.00006	0.00021	0.76
	4 star	0.00001	0.00024	0.97
	5 star	0.00012	0.00020	0.55
30-day mortality N=3558538	1 star	0.00004	0.00007	0.57
	2 star	-0.00004	0.00009	0.63
	3 star	-0.00013	0.00009	0.15
	4 star	-0.00009	0.00011	0.39
	5 star	0.00004	0.00009	0.69
30-day rehospitalization N=3558538	1 star	-0.00024	0.00021	0.25
	2 star	0.00011	0.00027	0.68
	3 star	0.00016	0.00027	0.54
	4 star	0.00071	0.00033	0.03
	5 star	0.00032	0.00027	0.24

Note: Sample includes Medicare fee-for-service beneficiaries discharged from an inpatient hospital stay directly to home in 2012-2013. We excluded patients with hospital or SNF stay or hospice use 90 days prior to the index hospital admission. Model covariates include sex, race, age, diagnosis categories, and year-month effects.

Table A5: Falsification check: Reduced-form (OLS) effect of instrument on probability of discharge to skilled nursing facility (N=10,873,882)

Variable: log of distance to nearest SNF of star rating	Discharge to SNF versus other setting	
	(1)	(2)
1 star	0.00432***	9.86e-05
	(0.000297)	(0.000190)
2 star	-0.00115***	-3.98e-05
	(0.000368)	(0.000246)
3 star	-0.00365***	-0.000339
	(0.000390)	(0.000248)
4 star	-0.00876***	0.000101
	(0.000439)	(0.000288)
5 star	-0.00435***	3.70e-06
	(0.000345)	(0.000237)
ZIP code fixed effects included	NO	YES

Note: Sample includes Medicare fee-for-service beneficiaries discharged from an inpatient hospital in 2012-2013. Patients were discharged to home, to a SNF, intensive rehabilitation facility, to another hospital, or died. We excluded patients with hospital, SNF stay or hospice use 90 days prior to the index hospital admission. Robust standard errors accounting for ZIP clusters shown in parentheses. All models include month fixed effects.

Table A6: Association between within SNF change in star-rating and patient characteristics

VARIABLES	coef	se	tstat	pval
Language Spanish	0.000292	(0.00571)	0.0510	0.959
Other language	-0.00335	(0.00492)	-0.680	0.496
Black	0.00220	(0.00214)	1.026	0.305
Hispanic	-0.00584*	(0.00333)	-1.753	0.0796
Other race	0.00487	(0.00372)	1.309	0.191
Female	0.000829	(0.00108)	0.764	0.445
Full dual-eligible	-0.00504***	(0.00158)	-3.197	0.00139
Partial dual-eligible	-0.00382	(0.00252)	-1.515	0.130
Married	-0.000548	(0.00110)	-0.500	0.617
Deyo-Charlson score	-0.000238	(0.000278)	-0.853	0.394
Elixhauser score	6.43e-05	(0.000328)	0.196	0.845
Total ICU days	0.000160	(0.000161)	0.998	0.318
Total CCU days	7.33e-05	(0.000234)	0.314	0.754
Hospital length of stay	8.61e-05	(0.000107)	0.806	0.420
Shortness of breath	-0.00273**	(0.00133)	-2.043	0.0411
ADL score (0-28, higher = worse)	-0.000218*	(0.000128)	-1.706	0.0880
CPS scale	0.000264	(0.000430)	0.612	0.540
Stroke	-0.000369	(0.00170)	-0.217	0.829
Lung disease	-0.000542	(0.00127)	-0.429	0.668
Alzheimer's disease	0.00157	(0.00247)	0.636	0.525
Non-Alzheimer's dementia	-0.000442	(0.00144)	-0.307	0.759
Hip fracture	0.00231	(0.00179)	1.293	0.196
Multiple sclerosis	0.00316	(0.00931)	0.340	0.734
Heart failure	9.11e-05	(0.00129)	0.0708	0.944
Diabetes	0.00210*	(0.00115)	1.831	0.0671
Schizophrenia	0.0182***	(0.00638)	2.848	0.00440
Bipolar disease	-0.00297	(0.00494)	-0.602	0.547
Aphasia	0.00402	(0.00430)	0.935	0.350
Constant	3.264***	(0.00613)	532.9	0
Observations	1,280,318			
R-squared	0.826			

Includes month-year fixed effects. Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A7: Robustness Check: 2SLS Models Using Mortality-Adjusted Outcome Measures

	30-Day Rehosp, conditinal on alive on day 30	180-Day Rehosp, conditional on being alive on day 180	Long-Stay nursing patient, conditional on being alive on day 180	Days at Home, mortality adjusted	Days in Hospital, mortality adjusted	Days in SNF, mortality adjusted	Days at Home with Home Health, mortality adjusted
No FE	-0.0110***	-0.0161***	-0.0229***	6.503***	-0.865***	-4.431***	-1.235***
	(0.000882)	(0.00144)	(0.00147)	(0.272)	(0.0582)	(0.227)	(0.201)
ZIP FE	-0.00482	-0.00272	-0.0152***	2.880***	-0.572***	-2.939***	-1.123***
	(0.00308)	(0.00465)	(0.00370)	(0.528)	(0.163)	(0.524)	(0.326)
ZIP FE and patient covariates	-0.00379	-0.000613	-0.0109***	1.797***	-0.448***	-2.029***	-1.063***
	(0.00305)	(0.00458)	(0.00343)	(0.450)	(0.158)	(0.461)	(0.324)
Observations	1,102,958	1,004,307	1,004,307	1,278,174	1,278,174	1,278,174	1,278,174

Robust standard errors clustered by ZIP code in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table A8. Previous estimates of the association of SNF star rating with patient outcomes

Study	Cohort	Outcome	Effect	Estimate of relative reduction
Kimball et al. 2018	Knee and hip arthroplasty patients	30-day readmission	5 star vs. 1 star	-30%
		90-day readmission	5 star vs. 1 star	-27%
		30-day mortality	5 star vs. 1 star	-50%
		90-day mortality	5 star vs. 1 star	-25%
Ogunneye et al. 2015	Patients with acute decompensated heart failure (ADHF)	30-day all-cause readmission	high (3-5 star) vs low (1-2)	-20%
		30-day ADHF readmission	high (3-5 star) vs low (1-2)	-20%
Unroe et al. 2012	Patients with heart failure	90-day all cause hospitalization	5 star vs. 1 star	-8%

65