



HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

THE UNIVERSITY *of York*

WP 22/12

Estimating the health value added by nursing homes

Marlies Bár; Pieter Bakx; Bram Wouterse and Eddy van Doorslaer

July 2022

<http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/>

Estimating the health value added by nursing homes

Marlies Bär^{*1,2}, Pieter Bakx^{1,2}, Bram Wouterse^{1,2}, and Eddy van Doorslaer^{1,2,3}

¹*Erasmus School of Health Policy & Management, Erasmus University Rotterdam*

²*Erasmus Centre for Health Economics Rotterdam*

³*Erasmus School of Economics, Erasmus University Rotterdam*

June 2022

Abstract

Measuring performance in healthcare remains a challenge. The use of health outcomes rather than structure and process indicators is considered as the way forward, but outcome-based results risk being biased by selection. Accounting for such selection bias is more difficult in settings with small-sized providers and low chances of resident health improvement, like in the case of nursing homes. In this paper we (i) measure the health outcomes of Dutch nursing homes in terms of mortality and avoidable hospitalizations among residents, (ii) we adopt a novel approach to test for selection bias and (iii) we examine the relationship between outcomes and other nursing home quality indicators and characteristics. Using administrative data from more than 110,000 residents, we estimate the performance of the 849 largest nursing homes in the Netherlands in the period 2015-2019. Controlling for an extensive set of observable case-mix variables, we first test for the presence of selection bias using a distance-based instrumental variable. We do not find any evidence for such a structural bias. While the wide confidence intervals of the estimates display considerable imprecision, our results do reveal substantial differences between top and bottom performing nursing homes. Because the outcome-based estimates turn out to be only weakly correlated with other quality indicators, we conclude that our mortality and avoidable hospitalization-based indicators provide important complementary information. When small sample issues and case-mix differences are adequately accounted for, outcome-based indicators can provide useful policy guidance for quality improvement in nursing homes.

Keywords: Nursing homes, Value-added, Performance, Health outcomes, Administrative data, Instrumental variable

JEL Classification: C26; I10; J14

*Corresponding author. Contact: bar@eshpm.eur.nl; P.O.Box 1738, 3000 DR Rotterdam, The Netherlands

1 Introduction

Continued increasing demand and limited supply in nursing home care might reduce incentives for nursing homes to improve quality (Ching et al., 2015; Nyman, 1988). To stimulate quality improvements it is therefore important to inform consumers and policy makers by evaluating their performance on a regular basis. While outcome-based measures are commonly applied to enhance performance in other types of healthcare institutions, like in pay-for-performance schemes for hospitals, nursing home care is still primarily evaluated on the basis of structure (e.g. staffing) or processes of care (e.g. use of psychotropic drugs) in most countries (Barber et al., 2021). Since the relationship between structure, processes and outcomes of care is far from straightforward (Donabedian, 2003), it is meaningful to complement such indicators with outcome-based measures. The challenges in doing so for nursing homes are that appropriate outcomes are less easily defined, one has to rely on self-reported measures and small sample sizes, and that - similar to other sectors - it is uncertain whether quality differences persist after correction for observable case-mix differences.

In this paper, we study whether residents' health outcomes may be used to evaluate performance of nursing homes. We examine i) how much variation in health outcomes there is across nursing homes; ii) whether this can be attributed to differences in performance rather than to differences in unobserved resident characteristics; and iii) the association of structure and process-based quality indicators with those. We use administrative data from over 110,000 nursing home admissions in the Netherlands linked to data on mortality and avoidable hospitalizations and background characteristics, to estimate the health-value added of each of the 849 largest Dutch nursing homes. The identification of these nursing-home-specific effects is complicated by the fact that residents with high or low unobserved health might self-select into particular homes. We address this econometric challenge by testing whether our case-mix corrected estimates can accurately predict the outcomes for (quasi-)randomly admitted residents (i.e. those admitted to the nursing home closest to their prior residence¹). Finally, we correlate the outcome-based performance estimates to other quality indicators to verify whether structure and process indicators can explain variation in outcomes to improve understanding of the potential mechanisms involved.

Our paper makes the following contributions to the literature. First, it extends the economic value-added literature by demonstrating that by estimating a forecast coefficient when exploiting exogenous variation in provider choice, the value-added framework can still be meaningfully employed to test for the presence of selection bias, even when sufficient power to include individual instruments for each provider is lacking.² We apply the value-added framework to evaluate the presence of selection bias in performance on outcomes of relatively small entities like nursing homes. The existing value-added literature mainly focuses on larger organizations, like schools, hospitals, skilled nursing facilities or insurance plans, for which there is sufficient exogenous variation to estimate the causal impact and bias for each entity separately (Abaluck et al., 2021; Angrist et al., 2016; Chetty et al., 2014; Deming, 2014; Einav et al., 2022; Helsø et al., 2019; Kane and Staiger, 2008). As in Abaluck et al. (2021), we use the estimated forecast coefficient to examine whether our case-mix corrected

¹Geographical distance is an important determinant of nursing home choice and unlikely to be related to outcomes (other than through nursing home choice). Since Newhouse and McClellan (1998), this instrument has been used in many other settings in health and nursing home care (Cornell et al., 2019; Geweke et al., 2003; Gowrisankaran and Town, 1999; Grabowski et al., 2013; Helsø et al., 2019; Huang and Bowblis, 2018).

²This alternative test also stems from the education literature (Angrist et al., 2016; Chetty et al., 2014; Deming, 2014; Kane and Staiger, 2008), and has more recently been applied to the healthcare sector (Abaluck et al., 2021; Helsø et al., 2019).

outcome scores accurately predict causal variation in individual-level health outcomes.

Second, we contribute to the broader (health) economics literature by examining the predictive validity of case-mixed corrected outcome indicators like mortality and (avoidable) hospital (re)admission rates as measures of quality in the long-term care sector. Prior evidence from the hospital sector is not equivocal: some studies suggest that unobservable patient differences may generate misleading quality estimates (Gowrisankaran and Town, 1999; Hull, 2020), others that risk-adjusted outcomes do provide useful quality information (Doyle et al., 2019). These results cannot directly be transferred to the long-term care setting, because of its focus on preventing health deterioration rather than on improving health. The same holds for studies of Skilled Nursing Facilities in the U.S. (Einav et al., 2022; Rahman et al., 2016), where the hospital plays a more prominent role in choosing a facility, and many admissions are short-stays aimed at a discharges back to the community.

Third, we contribute to a better understanding of health outcomes across nursing homes by taking unobserved selection into account. The causal nursing home literature so far has only considered impacts on outcomes of one – often binary – characteristic at a time, like staffing levels, the presence of dementia special care units or ownership (Friedrich and Hackmann, 2021; Grabowski et al., 2013; Gupta et al., 2021; Huang and Bowblis, 2018; Joyce et al., 2018; Lin, 2014). However, since these characteristics are often strongly correlated, it is difficult to isolate the impact of a single characteristic, even with exogenous variation at the individual level (Konetzka et al., 2019). In contrast, we analyze the total variation in outcomes that can causally be attributed to provider differences. Prior research documenting overall differences in outcomes across nursing homes either does not take selection on unobservables into account (see for example Arling et al. (2007); Wouterse et al. (2022)), or it focuses on short stays in (U.S.) Skilled Nursing Facilities (see for example Einav et al. (2022); Rahman et al. (2016)). It is essential to know the extent to which selection bias drives total observed variation in outcomes across nursing homes (without attributing it to one characteristic), e.g. for providing valid quality information to consumers, for making fair comparisons of nursing homes' relative performance or for assessing returns on public healthcare spending (OECD and EuropeanCommission, 2013).

We find meaningful variation in mortality and hospitalization rates across Dutch nursing homes. We show the value of estimating performance using administrative data which allows for controlling for a large range of resident characteristics. After extensive case-mix correction, we find that the five percent best-performing nursing homes have a 7 and 14 percentage points lower mortality and avoidable hospitalization rate compared to the worst performing ones. The results from our selection bias test demonstrate that this variation in outcomes is not attributable to unobservable heterogeneity in resident characteristics. Our findings suggest that outcomes are weakly correlated with only a small subset of process and structure indicators.

2 Background

2.1 Nursing homes in the Netherlands

Nursing homes may serve two groups. First, they serve residents - or clients³ - who need long-term institutional care and who, once admitted, typically stay there for the remainder of their life. Second, they may serve clients who are discharged from the hospital for a (limited) period of rehabilitation care or post-acute care.⁴ In the Netherlands and elsewhere, there is a clear distinction between these two. In this study, we focus on the first group; long-term institutional stays.⁵

For this group, the Netherlands has comprehensive social long-term care (LTC) insurance that pays for 99.9% of total nursing home care expenditures (CBS, 2017). Nursing home care, including costs for room and board, is covered by the insurance for the entire population. Nursing home recipients pay a relatively low co-payment that covers 11% of total expenditures (Rijksoverheid, 2017). The co-payment depends on the recipient's income and wealth but not on the type of care received or the nursing home chosen (Tenand et al., 2021). This makes the Dutch nursing home care accessible.

Elders need to apply for eligibility for a nursing home admission, which is granted if someone needs supervision or care around the clock. This eligibility decision is made by an independent government agency (CIZ). CIZ also decides on the care package which indicates the intensity of nursing home care that the recipient is eligible for.⁶

Elders who are eligible for a nursing home admission may choose any nursing home with availability for the desired care intensity package. The waiting time in each of the regions in the Netherlands (during our study period) is limited: virtually all eligible elders can move to a nursing home within the 6-week period that is set as the norm by the government (NZa, 2021). However, some elders choose to delay their admission until their preferred nursing home has an opening and are then put on a nursing home-specific waiting list while they temporarily live in another nursing home or receive substitute home care.

All providers are private entities that are not allowed to make profit (Barber et al., 2021).⁷ Nursing homes receive a per diem price per client up to a budget ceiling that are negotiated with regional single-payers who contract long-term care providers. These prices are specific for each care intensity package and are constrained by a maximum price set by the government (Barber et al., 2021).

There are several measures in place to stimulate the provision of high-quality care. First, since 2017, nursing home budgets are supplemented by a subsidy for quality improvements. To receive this additional subsidy, nursing homes submit a quality improvement plan. Second, the Healthcare Inspectorate monitors quality of care, e.g. through unannounced visits. Its quality reports are **published**. Third, nursing homes are required to

³The terms nursing home residents and clients are used interchangeably throughout this paper.

⁴In the US, this care may be provided in skilled nursing facilities.

⁵Some Dutch nursing homes also offer day-care for elders who live at home or (short-term) rehabilitation care. Elders receiving these types of nursing home care are not included in this study.

⁶Residents with lower care intensity (ZZP 4) need intensive support and extensive care, with dementia care (ZZP 5) need a protective living facility with intensive dementia care, with higher care intensity (ZZP 6) need a protective living facility with intensive support and care, with highest care intensity (ZZP 7 and 8) need a protective living facility with very intensive care and treatment or support (CIZ, nd). A resident's care intensity package may change during his/her stay in a nursing home.

⁷There is a small but increasing number of for-profit nursing homes (Bos et al., 2020; Hussem et al., 2020). These nursing homes are not included in the analysis of this paper.

provide information about processes of care to the government, which is published [online](#). Finally, nursing homes are obliged to facilitate residents and their relatives to report their satisfaction with the nursing home. Almost all providers do this through a public website called [Zorgkaart Nederland](#). These online ratings are intended to assist (relatives of) future nursing home residents in selecting a nursing home.

2.2 Measuring nursing home performance

Nursing home quality is multidimensional and can be classified into three dimensions, namely structure, processes, and outcomes (Donabedian, 2003). In most countries, nursing home quality measures focus on the structure and process dimension of quality of care (Barber et al., 2021). Yet, Mor et al. (2003) and Werner et al. (2013) show that nursing homes that perform well on structure and process-based quality measures do not necessarily improve (health-related) outcomes of their residents. To provide a comprehensive set of quality information it is thus worthwhile to complement the widely used structure and process-based measures with information on outcomes. This subsection discusses prior work on outcome measurement in the nursing home sector and highlights two themes: the use of mortality and avoidable hospitalizations as outcomes measures and why there may be a selection bias in performance indicators using such outcomes.

Using mortality and avoidable hospitalizations as outcomes measures

According to Gupta et al. (2021) and McClellan and Staiger (1999) mortality has become the "gold-standard" for measuring quality in the health economics literature. Several extensive literature reviews indicate that reduced risk of mortality is associated with higher well-being of older persons (Chida and Steptoe, 2008; Martín-María et al., 2017), which makes it a good candidate for measuring nursing home outcomes. Likewise, a hospital stay can not only be costly but, more importantly, is also found to be traumatic, uncomfortable and disorienting for nursing home residents (Grabowski et al., 2007; Ouslander et al., 2000). We believe that both mortality and potentially avoidable hospitalizations are undesirable and that nursing homes with lower mortality and lower avoidable hospitalizations – all else equal – are performing better.⁸

Research shows that variation in such health outcomes can at least to some extent be attributed to factors influenced by the nursing home. For example, Cornell et al. (2019) show that residents admitted to Skilled Nursing Facilities with higher STAR ratings have lower mortality and fewer hospitalizations. Other channels through which nursing homes are found to affect outcomes – like hospitalizations and mortality – are staffing levels, private equity ownership, nonprofit status and the presence of a dementia special care unit (Grabowski et al., 2013; Gupta et al., 2021; Joyce et al., 2018; Friedrich and Hackmann, 2021). Therefore, we would expect at least some variation in terms of these outcomes across nursing homes.

The outcomes that we measure are restricted to the health domain. Ideally, we would measure outcomes that go beyond health, like individual level wellbeing and quality of life. However, routinely measuring these on a large scale in such a vulnerable population is not

⁸As some hospitalizations may be unavoidable, we focus on those that are potentially avoidable. We do not focus on potentially avoidable causes of death as most classifications of avoidable causes of death are based on premature mortality, defined as dying before the age of 70 (OECD, 2009). Since our sample is restricted to those aged 70 and older, applying such a classification would be inappropriate. Additionally, we do not expect that euthanasia has a large contribution to differences in mortality across nursing homes since euthanasia occurred only 286 times (i.e. 1 percent, according to the number of deaths in our sample in the same year) in total in nursing homes in 2017 (Heins et al., 2019)

feasible. The main advantages of using mortality and avoidable hospitalizations as outcomes are that they are not self-reported, available for the full population and not prone to measurement error. Furthermore, the econometric issues that we deal with apply to all outcome measures, making this study a relevant illustration of nursing home performance measurement problems more generally. As discussed in the previous two paragraphs, mortality and avoidable hospitalizations likely capture sufficiently relevant aspects of nursing home performance to be indicative of other types of relevant outcomes.

Selection bias in the nursing home setting

Variation in outcomes may be driven by selection bias. There are several reasons why non-random selection could occur in the nursing home setting. First, some nursing homes may selectively attract a certain type of clients. For-profit nursing homes might, for example, have an incentive to attract more profitable or less costly clients, especially when they are close to their full capacity (Gandhi, 2020; He and Konetzka, 2015). Second, different types of individuals may choose a nursing home based on different criteria, which could cause performance measures to be either positively or negatively biased. On the one hand, elders (or their family members) who consider themselves more likely to be more severely ill, and more dependent on care services, may be more inclined to choose a nursing home that has a reputation to deliver higher quality care. On the other hand, elders who are more severely ill may be less able to “shop around” for quality care, especially after a sudden impairment and end up choosing a nursing home with no waiting list, instead of one with higher perceived quality (Castle, 2003; Schmitz and Stroka-Wetsch, 2020). Also, prospective residents who are more responsive to quality might be wealthier and better educated (Bensnes and Huitfeldt, 2021) or have better informal networks. Elders with such an advantageous environment can generally be expected to have a better (unobserved) health status, and they may also be more responsive to quality indicators when choosing other types of healthcare providers (Bensnes and Huitfeldt, 2021; Cornell et al., 2019). In sum, some degree of nursing home selection may be expected, but it seems hard to predict the direction of any bias a priori.

The literature on selection bias in nursing home outcome measures is limited and, like most research on nursing home quality (Lippi Bruni et al., 2019), generally focuses on Skilled Nursing Facilities (SNFs) in the United States.⁹ Arling et al. (2007) demonstrate that shifts occur in SNF quality rankings when more observable differences in client characteristics are accounted for. This indicates that there may also be some selection on observable characteristics of SNF clients. However, how much selection does remain when many observable characteristics are already accounted for? Rahman et al. (2016) report wide variation in risk-adjusted re-hospitalization rates across SNFs (i.e. 15 percentage points between the five percent best and worst performing facilities). They show that these rates are an accurate prediction of the re-hospitalization risk of individuals admitted to these SNFs a few years later, suggesting that variation in risk-adjusted re-hospitalization rates is not driven by selection on unobservables. In contrast, Einav et al. (2022), do find evidence of selection bias driven by unobservably healthier clients being more likely to be admitted to SNFs that generate larger health improvements.

It is still an open question how any of the prior results on selection bias from the U.S. SNF may have relevance for the Dutch nursing home setting. Although both types of facilities offer institutional care mainly to older clients who cannot live at home yet or anymore,

⁹When looking at the hospital setting, the evidence on this topic is mixed: where some studies find evidence for non-random selection in hospital outcomes (Geweke et al., 2003; Gowrisankaran and Town, 1999; Hull, 2020), others find that selection bias only plays a minor role (Doyle et al., 2019; Helsø et al., 2019).

the Dutch system, like in many other developed countries, almost exclusively concerns on long-stays rather than short-stays and has a much more comprehensive social system for long term care (Barber et al., 2021). On the one hand, the role of selection likely plays a more prominent role in the U.S. setting due to financial incentives to admit short-stay non-Medicaid patients (see also Gandhi (2020)). Moreover, the focus on nonprofit nursing homes - which forms the largest part of all nursing homes in the Netherlands (Bos et al., 2020) - may induce a smaller role of non-random selection since nonprofit nursing homes may be less inclined to selectively attract healthier clients. On the other hand, the choice process may be more selective for long-stays since it requires the decision on where to reside until death (Bom, 2021), compared to where to stay for 26 days - the average length of stay in post-acute care in the U.S. (Cornell et al., 2019).

3 Data

3.1 Sample selection

3.1.1 Nursing home residents

We use administrative data provided by Statistics Netherlands encompassing the full Dutch population (more detailed information about the data sources can be found in Appendix A.1). Our sample consists of individuals who were admitted to a nursing home for the first time between January 2015 and July 2019.¹⁰ We use individual-level information on provider codes and addresses from the municipal registry, to link 87 percent of the 2015-2017 population to the nursing homes that they were admitted to.¹¹ We complement this sample by individuals who entered a nursing home in 2018 to July 2019 that we could match to nursing homes using information on addresses only. As a first step to make the resident populations across nursing homes more comparable, we dropped 9,438 (7%) admissions of residents whose age was younger than 70 at the time of admission¹², followed by 3,627 (3%) admissions for individuals for whom we have missing data on background characteristics. Our final study sample includes 119,699 nursing home residents in the mortality analyses, and 83,056 residents in the hospitalization analyses: data on hospitalizations is only available until December 2017.¹³

3.1.2 Nursing homes

Our data contains an anonymized provider code, but not the location. We do observe where individuals live and therefore identify nursing home locations by an address on which at least 5 individuals receive care within the same time period provided by the same provider (based on the provider code). Nursing home facilities belonging to the same chain organization can use the same provider code, but are distinguished using the address information. We use

¹⁰We focus on people with care packages 4-8, which are for long-term nursing home stays. That is, we exclude people who are eligible for palliative care (care package 10) or geriatric rehabilitation (care package 9).

¹¹In 31,688 cases, address information was missing and the provider code belonged to multiple nursing home facilities within an organization. For this group, we imputed to which nursing home facility the resident was admitted using admission data of the nearest neighbour with the same provider code and non-missing address information.

¹²Although this is a significant share of our sample, being admitted to a nursing home before the age of 70 is a rare event (i.e. between 0 to 0.5 percent depending on the age-group) in the Netherlands.

¹³We did not exclude residents who switched nursing homes during the study period because this may underestimate variation in performances, as they may switch from low to high quality homes.

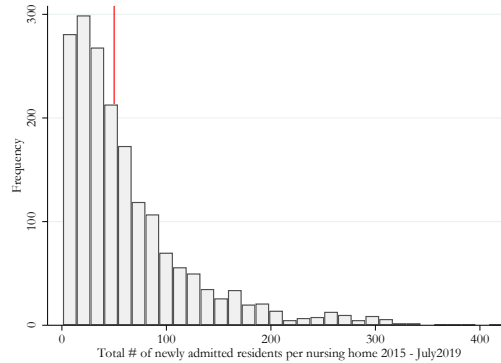


Figure 1: Total number of newly admitted admissions per nursing home

This figure shows the total number of admissions per nursing home on the x-axis and the frequency, representing the number of nursing homes, on the y-axis. All nursing homes to the right of the red vertical line are included in the main analyses.

the provider codes combined with postal codes to link the information on quality indicators. Descriptive statistics about the quality indicators of the included nursing homes can be found in Appendix Table B1.

We include the 849 largest nursing home facilities, with at least 50 new admissions during the entire study period, in our main analysis.¹⁴ Figure 1 shows the variation in size, in terms of the number of newly admitted residents across nursing homes. To ensure sufficient statistical power, we do not attempt to estimate performance for the nursing homes with fewer admissions (to the left side of the red line in Figure 1). The 21 percent of residents who are admitted to one of these 1,008 smaller nursing home facilities are included in the reference category in the analyses.

3.2 Health outcome measures

We focus on two outcomes, namely mortality and avoidable hospitalization. We define hospitalizations as potentially avoidable if they are related to a main diagnosis that could have been prevented or treated in the nursing home. For example, hospitalizations resulting from falls in nursing homes may be preventable by hip protectors (Vu et al., 2006) or by adaptations to the environment like an optimized light design for residents with cataract or height adjusted chairs (Becker and Rapp, 2010). The diagnoses that we classify as such (see also B2 in the Appendix) are based on two studies on ambulatory care-sensitive hospitalizations for the elders population (Carter, 2003; Walker et al., 2009), to which we add hospitalizations due to falls and fractures, wounds and rehabilitation as potentially preventable or treatable in a nursing home- thus being potentially avoidable. More than half of the avoidable hospitalizations are due to falls and fractures, about 8 percent to pneumonia, 6 percent to asthma and COPD, 5 percent to rehabilitation and 4 percent to kidney or urinary tract infections (Appendix Table B2).

¹⁴Individuals who are admitted to nursing homes with fewer admissions are generally in better health (i.e. lower care needs, lower healthcare expenditures - results available upon request), which may have implications for our results. However, we follow (Einav et al., 2022) by excluding the smaller nursing homes to obtain more reliable estimates.

We construct binary outcomes (equal to one if the individual died or had an avoidable hospitalization within 180 days after admission) as our main dependent variables to limit the influence of right censoring.^{15,16} The 180-day cut-off is somewhat arbitrary, but in line with the nursing home literature (Cornell et al., 2019; Intrator et al., 2004; Vossius et al., 2018). Additionally, Kaplan-Meier survival curves in Figures C1a and C1a in the Appendix confirm that most variation in time until death and until an avoidable hospitalization occurs in the first half year after nursing home admission. The robustness checks examine the sensitivity of our results to the use of different cut-offs. In our sample, the average 180-day mortality and avoidable hospitalization rates are 21 and 13 percent respectively (Table 1).^{17,18}

3.3 Case-mix controls

We control for observable differences in nursing home residents’ characteristics by including an extensive set of case-mix controls in our analyses: age at admission; gender; whether someone lives in a rural municipality, defined by an average of at least one thousand addresses per square kilometer; yearly disposable household income, standardized by household size; wealth from assets and savings; and a comprehensive set of proxies for health: whether the person visited the hospital within 30 days prior to nursing home admission (also to account for potential hospital re-admissions); the Charlson comorbidity index based on 12 comorbidities like dementia, cancer and pulmonary diseases¹⁹; the number of different types of medicine consumed; and healthcare expenditures from the year before nursing home admission. Furthermore, we include care needs as measured by the care intensity package as determined by the independent eligibility assessment agency. An overview and more extensive explanation of all covariates can be found in Appendix Table B4.

Table 1 shows how these covariates vary across the health outcomes of individuals. Older male residents, those receiving a higher care intensity package, those who use more medication, those with a higher Charlson comorbidity index or those who visited the hospital within 30 days before nursing home admission, have a higher probability of dying within the next half year. On the other hand, residents who experienced an avoidable hospitalization within 180 days after admission are, on average, younger and enter the nursing home receiving lower care intensity. This implies that healthier individuals (i.e. younger with lower care needs) are more likely to be admitted to a hospital. Nonetheless, as we control for differences in underlying health across nursing homes, we interpret a higher risk of avoidable hospitalizations an undesirable outcome.

¹⁵While mortality is already a binary event by nature, we could measure the hospital outcome as the number of avoidable hospitalizations. However, as this may be influenced by re-admissions, we use the count in our robustness tests only.

¹⁶As data on mortality is available up until 2019 and only 7 percent of individuals left the nursing home before their death, we are not concerned about right-censoring in this outcome measure. Hospitalizations are not accounted for censoring from deaths.

¹⁷The 180-day mortality rate is similar to that in skilled nursing facilities in the U.S. (Cornell et al., 2019) and slightly higher than in care homes in Norway (Vossius et al., 2018).

¹⁸Table B3 shows that, of the 23,165 residents who had at least one hospitalization within half a year after nursing home admission, 37 percent experienced a hospitalization that was potentially avoidable. Both this percentage and the 180-day hospitalization rate are higher in comparison to other studies (Carter, 2003; Intrator et al., 2004; Walker et al., 2009), likely resulting from the inclusion of falls and fractures as an avoidable cause.

¹⁹The Charlson comorbidity index is an indicator for disease burden and/or a predictor of mortality (Sundararajan et al., 2004). We use the updated version constructed by (Quan et al., 2011) which reflects a weighted score based on 12 comorbidities, among which dementia, diabetes and cancer.

Table 1: Case-mix controls by outcome

	Died within 180 days after nursing home admission					Had an avoidable hospitalization within 180 days after nursing home admission				
	All	No	Yes	Difference		All	No	Yes	Difference	
		(79.3%)	(20.7%)	Yes (-)	No	se	(87.1%)	(12.9%)	Yes (-)	No
Women (%)	0.649	0.674	0.553	-0.121***	(0.003)	0.65	0.649	0.663	0.014***	(0.005)
Age	85.027	84.786	85.951	1.166***	(0.046)	84.948	85.054	84.229	-0.826***	(0.066)
Care intensity ¹ (%)										
Lower	0.205	0.227	0.122	-0.104***	(0.003)	0.202	0.197	0.239	0.043***	(0.004)
Dementia	0.49	0.498	0.457	-0.041***	(0.004)	0.459	0.475	0.35	-0.125***	(0.005)
Higher	0.294	0.264	0.407	0.143***	(0.003)	0.326	0.315	0.4	0.085***	(0.005)
Highest	0.011	0.011	0.014	0.003***	(0.001)	0.013	0.013	0.011	-0.003**	(0.001)
Healthcare expenditures	1,347	13,437	13,553	117	(165)	10,290	9,819	13,477	3,658***	(173)
Wealth	83,185	81,992	87,760	5,768.**	(2793)	85,895	87,366	75,954	-11,412**	(4,445)
Std. household income	21,914	21,901	21,966	65	(75)	21,811	21,816	21,777	-39	(107)
Number of medicine	8.766	8.514	9.731	1.217***	(0.035)	8.82	8.637	10.057	1.42***	(0.051)
Charlson score	0.672	0.537	1.192	0.656***	(0.010)	0.703	0.687	0.816	0.129***	(0.016)
Hospital in last month (%)	0.169	0.142	0.272	0.130***	(0.003)	0.191	0.184	0.244	0.061***	(0.004)
Rural (%)	0.297	0.297	0.299	0.002	(0.003)	0.302	0.305	0.284	-0.021***	(0.005)
Year (%)										
2015	0.25	0.248	0.258	0.010***	(0.003)	0.361	0.361	0.361	0.000	(0.005)
2016	0.31	0.301	0.344	0.043***	(0.003)	0.447	0.441	0.485	0.044***	(0.005)
2017	0.26	0.259	0.266	0.008**	(0.003)	0.192	0.198	0.154	-0.044***	(0.004)
2018	0.115	0.122	0.087	-0.036***	(0.002)					
2019	0.064	0.069	0.044	-0.025***	(0.002)					
N	119,699	94,933	24,766			83,056	72,354	10,702		

This table presents the averages or shares (%) of each case-mix control variable by the mortality and avoidable hospitalization outcome including differences between those for whom the outcome equals one and zero; Age, care intensity, rural and year at moment of nursing home admission; healthcare expenditures, wealth, std. household income, number of medicine and Charlson score from the (calendar) year before admission; Standard errors (se) between brackets.

* Difference is statistically significant at 10 percent; ** at 5 percent; *** at 1 percent.

¹ Lower - intensive support and extensive care; Dementia - protective living facility with intensive dementia care; Higher - protective living facility with intensive support and care; Highest - protective living facility with very intensive care and treatment or support.

4 Empirical strategy

4.1 Observed performance

To quantify the effect of a nursing home j on the probability of an adverse health outcome, we use a linear value-added framework:

$$E(Y_i|H_i = j) = \gamma X_i + \delta_j + \varrho_i, \quad (1)$$

where Y_i is the outcome for individual i conditional on being admitted to nursing home j . The expected outcome depends on an individual's observed characteristics X_i , which include proxies for prior health, an unobserved individual component ϱ_i , and a nursing home specific effect δ_j .²⁰ δ_j is the nursing home level estimate of interest: the value-added of the nursing home, i.e. the nursing home's impact on the outcome under the condition of exogenous nursing home choice. We assume that the nursing home impact is additive and homogeneous across residents.

We estimate a linear probability model using an ordinary least squares regression (with robust standard errors) to obtain each nursing home's performance on the two health outcomes.²¹ The estimation equation is as follows:

$$Y_i = \alpha_0 + \gamma X_i + \sum_{j=1}^J \delta_j H_{ij} + \varrho_i, \quad (2)$$

where Y_i is a zero-mean dichotomous outcome variable for individual i - e.g. mortality - and X_i are the individual level case-mix controls. α_0 represents the reference category which includes all individuals that were admitted to one of the smaller nursing homes. H_{ij} is a dummy variable that equals one if individual i is admitted to nursing home j ($j = 1, 2, \dots, J$). The estimated parameter $\hat{\delta}_j$ reflects the nursing home j 's effect on the outcome - or the nursing home's value added.

Our estimates $\hat{\delta}_j$, especially those for small nursing homes, are surrounded by sampling imprecision. Like Angrist et al. (2017); Abaluck et al. (2021); Chetty and Hendren (2018); Kane and Staiger (2008), we therefore account for the statistical noise in our value-added estimates by applying a standard empirical Bayes correction (Morris, 1983). The empirical Bayes estimator of δ_j is weighted average of the precisely estimated grand mean (the average outcome across all nursing homes) and the imprecisely estimated nursing-home-specific OLS estimate $\hat{\delta}_j$, where the weight of the latter is proportional to its estimation error. As the estimation error is greater for small homes, the shrinkage is larger for these homes than for large homes.²² We use these empirical Bayes estimates when evaluating total variation in

²⁰Our value-added model deviates from the classical ones in the sense that we include proxies for individual's health as right-hand side variables instead of the individual's outcome Y_i prior to admission. The latter is simply not possible given the nature of our outcome variables.

²¹Estimating the same specification with a logit or random effects generates estimates that highly correlate (> 0.99) with the ones from the OLS procedure.

²²We use the following estimator $\delta_j^{EB} = \tau_j \hat{\delta}_j + (1 - \tau_j) \bar{\delta}_j$, where $\hat{\delta}_j$ is obtained by estimating Equation (2), $\bar{\delta}_j$ is equal to the average of $\hat{\delta}_j$ across all nursing homes, and the shrinkage factor τ_j is equal to $\frac{\sigma_{\delta}^2}{\sigma_{\delta}^2 + se_{\delta_j}^2}$ with σ_{δ}^2 being the between nursing home variation minus the average noise and $se_{\delta_j}^2$ being equal to the within nursing home variation. Under the assumption that the nursing-home specific effects are independent, this estimator is equivalent to an empirical Bayes estimate of the nursing home specific effects given that both the prior and likelihood function come from a normal distribution (Angrist et al., 2017; Chetty and Hendren, 2018; Kane and Staiger, 2008; Morris, 1983).

performance in Section 5.1 and when correlating performance on other quality indicators in Section 6.

4.2 Testing for selection bias

Individuals are not randomly assigned to nursing homes, but are to a large extent free to choose the home that they prefer. We might therefore be worried about selection bias. The question we want to address is, after we correct for observable differences in individuals' characteristics X_i , are the estimates of observed performance $\hat{\delta}_j$ biased by unobserved individual differences? This is the case if there is a correlation between and individual's unobserved health and the performance - i.e. value added - of the nursing home he or she goes to (i.e. a correlation between ϱ_i and H_{ij} in Equation 2). The bias can be either positive or negative, as preferences for nursing home quality can be both positively or negatively correlated to (unobserved) individual health (see Section 2.2).

The standard way of dealing with selection bias is to focus on plausibly exogenous variation in nursing home choice and exploit this variation, using instrumental variable analysis, to obtain (causal) estimates if δ_j . In our case, this would entail instrumenting each of the $J = 849$ nursing home dummies in Equation (2), which requires at least 849 instrumental variables to obtain a just or over-identified model. Although this can be done in some settings (see [Gowrisankaran and Town \(1999\)](#); [Hull \(2020\)](#) using such an approach for hospital care), in the setting of nursing home care, with many small-sized providers, this is not feasible because the lack of power likely causes a many weak instruments problem ([Angrist et al., 2016](#)).²³

Instead of trying to obtain a 'causal' IV-estimate for each nursing home, we test ex-post whether the observed performance measures $\hat{\delta}_j$ - estimated by Equation (2) - are biased (see also [Abaluck et al. \(2021\)](#); [Angrist et al. \(2016\)](#); [Chetty et al. \(2014\)](#); [Deming \(2014\)](#); [Helsø et al. \(2019\)](#)). The intuition behind the test is as follows. Suppose we already estimated the case-mix corrected - or value-added - scores from Equation (2) and then afterwards could randomly assign a new group of individuals over the J nursing homes. If the estimated performance scores $\hat{\delta}_j$ would be unbiased, then these scores would perfectly predict the (average) outcomes for the randomly assigned group. We could run the following regression on the sample of randomly assigned individuals:

$$Y_i = \gamma X_i + \lambda \hat{\delta}_{ij} + \varepsilon_i, \quad (3)$$

with $\hat{\delta}_{ij}$ the estimated performance score of the nursing home to which individual i has been assigned to. This regression would provide a simple test of (average) selection bias based on the forecast coefficient λ : if the values of $\hat{\delta}_{ij}$ represent (on average) the true causal effects of nursing homes on the outcome, then λ should be equal to one.²⁴ If, on the other hand, the estimates suffer from selection bias then λ will be either smaller or larger than one. If there is a positive correlation between unobserved health and nursing home quality (healthier clients are more likely to choose better nursing homes) then λ will be larger than one. We

²³In a recent working paper, [Einav et al. \(2022\)](#) estimate the added health value of skilled nursing facilities in the U.S., using a control-function approach (which is quite similar in spirit to an IV-approach) to correct for potential selection on unobserved health. The average number of treated patients in these facilities, aimed at rehabilitation, is substantially higher than in the permanent residential homes we investigate. Also, [Einav et al. \(2022\)](#) seem to use a relatively restrictive model for patient choice.

²⁴Although $\lambda = 1$ implies that there is no bias on average across nursing homes, this does not rule out the possibility that the scores of some specific homes are biased (see [Angrist et al. \(2017\)](#); [Hull \(2020\)](#))

then overestimate nursing homes’ performance, in the sense that the observed performance is better than true performance for high quality homes and lower than true performance for low quality homes. If the correlation between unobserved health and nursing home quality is negative, then the observed performance is an underestimation of the true effect.

4.3 Instrumental variable approach

In practice, we cannot randomly assign a group of clients over the different homes, and thus have to rely on quasi-exogenous variation instead. If there is a subgroup within our population for which it is credible that nursing home choice is not related to expected outcomes, then we can use this group to perform a test similar to that for the imaginary randomly assigned group in Equation (3).

The source of variation we exploit is geographical distance from a client’s home to a nursing home. Distance is an important driver of nursing home choice. Both earlier and more recent literature report distance to be a strong, if not the dominant, driver of nursing home choice (Castle, 2003; Gadbois et al., 2017; Hackmann, 2019; Schmitz and Stroka-Wetsch, 2020; Shugarman and Brown, 2006). As a result, location-based instruments are used widely to predict provider choice both in and beyond the nursing home literature (Einav et al., 2022; Gandhi, 2020; Gowrisankaran and Town, 1999; Grabowski et al., 2013; Hull, 2020; Newhouse and McClellan, 1998). Moreover, a location-based instrument is unlikely to be related to the unobserved component of the individual’s outcome as regional differences in health are expected to be small in our setting, especially since we control for an extensive set of health proxies at the individual level.

To implement the forecast test using quasi-exogenous variation in nursing home choice based on distance, we perform an IV using two-stage least squares (similar to Abaluck et al. (2021); Deming (2014); Helsø et al. (2019)). In the first stage, we predict the observed performance score $\hat{\delta}_{ij}$ of the nursing home j that individual i actually goes to using the observed performance score $\hat{\delta}_{ij}^{Closest}$ of the nursing home that is closest to individual’s i former residence:

$$\hat{\delta}_{ij} = \beta_0 + \mu X_i + \theta \hat{\delta}_{ij}^{Closest} + \varphi_i. \quad (4)$$

In the second stage, we use the first stage predictions of the performance score $\tilde{\delta}_{ij}$ to examine the effect of the nursing home performance scores on the outcomes (only) for individuals who move to a nursing home because it is the closest to their prior home. We do this by regressing individuals’ outcome on the first-stage prediction of the performance score:

$$Y_i = \alpha_0 + \gamma X_i + \lambda \tilde{\delta}_{ij} + \varepsilon_i, \quad (5)$$

If our instrument is valid (the performance of a nursing home is uncorrelated with the unobserved health of the clients that live closest to it)²⁵, the interpretation of the forecast

²⁵We further have to assume that the effect of $\hat{\delta}_j$ is homogeneous (i.e. constant for all individuals). Lambda is also affected by how compliers are distributed over the range of values that the treatment variable takes. If this distribution is not the same as in the full sample and if there are heterogeneous treatment effects, a deviation in lambda from one might not only be caused by selection bias (i.e. selection of low-mortality patients into low-mortality nursing homes) but also by systematic variation in the distribution of compliers over heterogeneous treatment effects. Our LATE would then deviate from one not because of a bias, but due to the high mortality estimates getting a lower weight due to lower compliance on that side of the distribution. We argue that this is unlikely to be an issue in our setting as almost every nursing home – irrespective of its performance – is the closest one to at least 20 people in our sample and is (in 99% of

coefficient $\hat{\lambda}$ is the same as in Equation (3); if $\hat{\delta}_j$ is an unbiased estimate of the true effect of a nursing home on clients' outcomes, it should (on average) perfectly predict the outcomes for clients who go a particular home solely because it's the closest to their prior residence. A forecast coefficient that is not equal to one then signals that observed performance is, on average, biased.

4.4 Instrumental variable assumptions

We first reflect on the assumptions that must hold to interpret $\hat{\lambda}$ as the impact of nursing home performance on a random individual's outcome. Two of these assumptions of the IV approach are that the instrumental variable – in our case performance of the closest nursing home – is (i) relevant and (ii) valid. A third condition is monotonicity.²⁶ We discuss the weak monotonicity assumption, which is sufficient for causal interpretation (Frandsen et al., 2019), in Appendix A.2. In the following subsections we pay more attention to why the relevance and validity assumption are likely to hold in this setting.

4.4.1 Relevance

The instrument is relevant if it has strong predictive power for nursing home choice. Prior studies argue and show that travel distance is the most important determinant of nursing home choice (Castle, 2003; Gadbois et al., 2017; Gandhi, 2020; Hackmann, 2019; Schmitz and Stroka-Wetsch, 2020; Shugarman and Brown, 2006). In our setting, we therefore expect that, all else equal, individuals prefer a nursing home that is closer to their prior home. Figure 2a confirms that most residents choose a nursing home that is close to their prior home: more than 60 percent of our sample is admitted to a nursing home within 5 kilometers from his or her prior home. Figure 2b shows that 21 percent chooses the nursing home that is closest to their former home. This suggests that the instrument is likely to be relevant.

The results from the first stage regression (Equation (4)) confirm that the instrument is strong. Table 2 shows that the first stage coefficient is economically and statistically significant. The partial F-statistics, which are equal to 7,189 and 18,545, both affirm the relevance of our instrument for both outcome variables (Staiger and Stock, 1997).²⁷

4.4.2 Validity

The instrument is valid if the unobserved health and other characteristics of individuals (ε_i in Equation (5)) are not correlated with the performance of the closest nursing home. That is, performance of the closest nursing home should only be related to an individual's outcome through choice. The validity assumption could be violated if unobservably (un)healthy clients are systematically located closer to the same nursing homes – keeping all individual observable characteristics fixed. This might be the case if better nursing homes are more likely to be located closer to prior homes of individuals with better underlying health (Helsø et al., 2019). This could be an important concern as previous research shows that, at least in the US, high quality nursing homes are more likely to be located in – or closer to – wealthier

the cases) chosen by at least 5 percent of those. In other words, the distribution of the compliers over the treatment variable is likely the same as the distribution of the entire study population.

²⁶This condition has received limited attention in other studies using the value-added framework, partly due to the nature of their instruments (Angrist et al., 2017; Chetty et al., 2014).

²⁷The reported F-statistics are extremely high. This is not surprising since the instrument directly links to the endogenous variable and our sample size is relatively large.

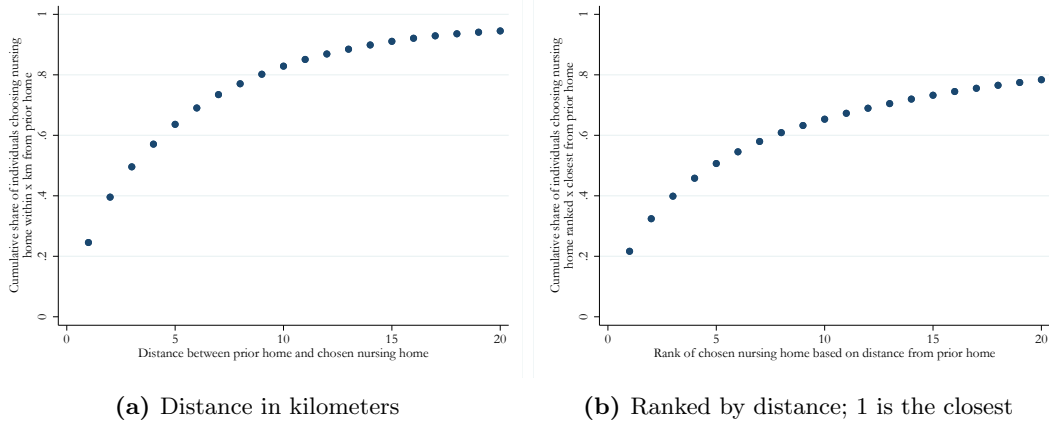


Figure 2: Relationship between nursing home choice and distance

Figure (a) shows the share of individuals being admitted to a nursing home within x kilometers from the individual's prior home. The x -axis represents the kilometers (rounded down) between one's prior home to admission and the chosen nursing home. Figure (b) shows the share of individuals being admitted to the closest nursing home from the individual's prior home. The x -axis indicates whether the chosen nursing home was the closest (at $x = 1$) or the second closest ($x = 2$) and so forth. In both figures the y -axis states the cumulative share of the full sample.

Table 2: First stage results

		Endogenous variable:	
		Performance of the chosen nursing home ($\tilde{\delta}_i$)	
		Mortality	Avoidable hospitalization
Instrumental variable:			
Performance of the closest nursing home			
$(\hat{\delta}_{ij}^{Closest})$			
	Coefficient (θ)	0.315*** (0.004)	0.497*** (0.004)
	(Partial) F-statistic	7,189	18,545
	Covariates	Yes	Yes
	N	94,905	65,265

This table reports the estimated coefficient, standard error and partial F-statistic from the first stage (Equation (4)) which is a linear regression of endogenous performance of the chosen nursing on performance of the chosen nursing home.

Standard errors between brackets. *** Statistically significantly different from zero at 1 percent; ** at 5 percent; * at 10 percent.

The sample is restricted to those that are admitted to one of the 849 largest nursing homes.

areas (Konetzka et al., 2015).²⁸

We expect the influence of this issue to be limited in our setting since we include individual level and precisely measured covariates like income and wealth as covariates when estimating performance.²⁹ Furthermore, (Cornell et al., 2019; Rahman et al., 2016) show that including zip code fixed effects limits the influence of regional differences that may be related to unobserved health and living close to a well-performing nursing home. In two robustness tests in Section 5.4, we therefore include neighbourhood characteristics as controls and neighbourhood fixed effects. We find that these additions do not change our main results.³⁰ This could either imply that the location of well performing nursing homes is not related to the health of neighbouring individuals in the Netherlands, or that our extensive set of covariates – encompassing individual level socio-economic indicators and proxies for the individuals’ health – already controls for regional variation in health.

5 Results

5.1 Observed performance

Figure 3 presents the observed performance estimates ($\hat{\delta}_j$ from Equation (2) after shrinkage) on the y-axis for all 849 largest nursing homes in groups of ten, ranked by their performance.³¹ There is a 7 percentage point difference in performance on mortality and a 14 percentage point difference in avoidable hospitalizations between the five percent best-performing and five percent worst-performing nursing homes. However, the wide 95% confidence intervals show that the individual estimates are imprecisely estimated, related to the relatively small size of most homes. The imprecision of our estimates does not facilitate the interpretation of observed differences, i.e. it remains hard to ascertain whether these are driven by true differences in performance or by imprecision.

5.2 Test for selection bias

As discussed throughout the paper, the variation observed in the figures above may be driven by unobserved heterogeneity in resident characteristics across nursing homes. In this subsection we present the results for our test for such a selection bias. More specifically, after obtaining predicted performance in the first stage (see also Equation (4) and Table 2), we obtain an estimate for the forecast coefficient through Equation (5). We test whether

²⁸In spite of this, distance-related instrumental variables are frequently used to correct for non-random selection into hospitals and nursing homes (Cornell et al., 2019; Geweke et al., 2003; Gowrisankaran and Town, 1999; Grabowski et al., 2013; Helsø et al., 2019; Huang and Bowblis, 2018; Newhouse and McClellan, 1998).

²⁹The likelihood and severity of violations of the validity assumption is small in the Dutch context for several other reasons. First, financial constraints do not play a role when choosing a nursing home, because the co-payment is the same in all nursing homes. This means that selection related to socioeconomic status is likely much more limited than in the US and many other countries. Second, elders are unlikely to select their place of residence (where they lived prior to nursing home admission) according to the performance of the nursing homes since most elders have lived in the same neighborhood for many years before they enter a nursing home (Diepstraten et al., 2020).

³⁰We do not include neighbourhood fixed effects in our main specification because the small number of people per neighbourhood moving to a nursing home minimizes the within-neighbourhood variation in which of the nursing homes is the closest one. This significantly reduces the power of the first stage, which in turn decreases the precision of our forecast coefficient.

³¹Estimates are published in groups of ten as, for privacy reasons, the results for individual nursing homes cannot be published.

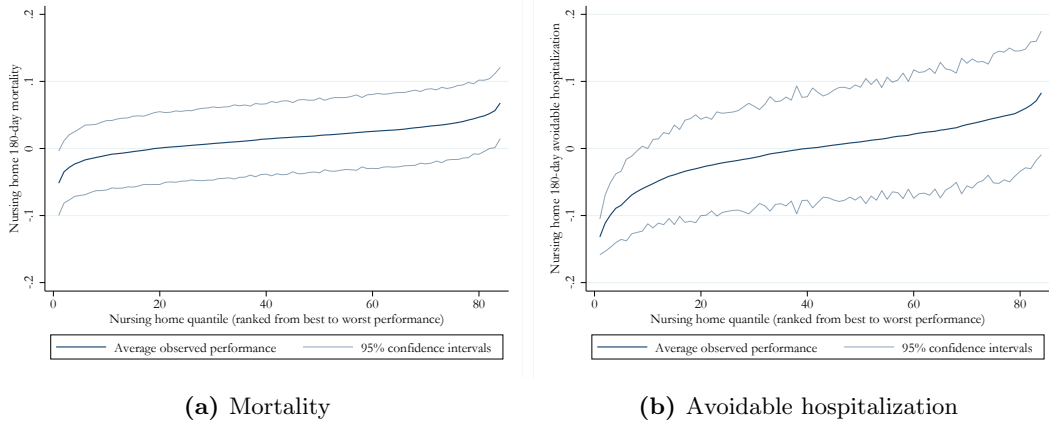


Figure 3: Observed performance

This figure displays the nursing home performances on 180-day mortality (a) and avoidable hospitalizations (b). We present estimated performance (by Equation (2)) after empirical Bayes shrinkage. Nursing homes are ranked on their performances and subsequently divided into 84 equally sized groups of 10 to 11 nursing homes. The x-axis represent these nursing home groups. The y-axis indicates average observed performance of each of these groups and its confidence intervals, which are calculated based on the standard error from a randomly chosen nursing home within each group.

the estimated forecast coefficient $\hat{\lambda}$ is equal to one. If we fail to reject this test, we interpret our estimated performance estimates as unbiased on average.

The forecast coefficients $\hat{\lambda}$ are economically and statistically not significantly different from one (Table 3). The estimated forecast coefficients deviate only minimally from one: choosing a nursing home with an above average mortality of 2 percentage points instead of one of 1 percentage point, increases the risk of dying by 1.07 percentage points.³² These minor deviations from one are for both outcomes not statistically significant different ($p = .408$ and $p = .104$). This implies that observed performance is, on average, unbiased and is likely to predict true differences in performance across nursing homes.

5.3 Subgroup analysis

One of the assumptions of the value-added model is that nursing home performance scores are homogeneous across residents (see Section 4.3). When estimating the relevance of observed performance for clients with different care needs, we shed light on whether this assumption is plausible: i.e. is observed performance representative for all clients. For this, we use the forecast coefficient estimated through Equation (5) replacing predicted performance by observed performance of the chosen nursing home to compare deviations of $\hat{\lambda}$ from one for the different subgroups. This test provides insights on whether observed performance is more informative for specific groups, which also is a relevant question on itself.

The regression results in Table B5 in the Appendix show that the estimates on both outcomes

³²To compare, the absolute deviation of the forecast coefficient from one lies within the ones found in studies on hospital performance including extensive controls by [Helsø et al. \(2019\)](#) ($\hat{\lambda} = 0.956$) and [Hull \(2020\)](#) ($\hat{\lambda} = 1.086$).

Table 3: Forecast coefficient

Predicted performance of the chosen nursing home ($\tilde{\delta}_{ij}$)	Individual level outcome (Y_i)	
	Mortality	Avoidable hospitalization
Forecast coefficient ($\hat{\lambda}$)	1.067*** (0.085)	1.070*** (0.042)
Forecast bias test ($\hat{\lambda} = 1$)		
χ^2 statistic	0.68	2.64
p-value	0.408	0.104
Covariates	Yes	Yes
N	94,905	65,265

This table reports the regression results from the second stage (from Equation (5)), which estimates the impact of predicted performance of the chosen nursing home on the individual level outcome, either mortality or avoidable hospitalization. The test statistic report the χ^2 statistic and the p-value when testing $\hat{\lambda} = 1$. Standard errors are reported between parentheses and p-values between (squared) brackets.

Standard errors between brackets. *** Statistically significantly different from zero at 1 percent; ** at 5 percent; * at 10 percent.

The sample is restricted to those that are admitted to one of the 849 largest nursing homes.

are predictive for residents of all care needs, but for some more accurately than for others. We find that the variation in observed performance on mortality is (on average) slightly overestimated ($\hat{\lambda} < 1$) for individuals with lower and the highest care needs. Although these results suggest that the estimates on outcomes are somewhat heterogeneous across care need groups, this does not affect the IV result from Section 5.2.³³

5.4 Robustness

We examine the robustness of our results with two additional sets of checks: (i) by including larger sets of controls; and, (ii) by using different definitions for our health outcome measures. We inspect how they correlate with our baseline estimates and whether the result of no structural bias (from Section 5.2) is robust to these adjustments.

First, in Section 4.4 we reflected on the validity of our instrumental variables. We mentioned that any systematic differences in unobserved health that are related to the location of someone's prior home are threats to this validity. Therefore, in two robustness checks, we include either neighbourhood characteristics that might be related to someone's health as additional control variables (i.e. average property (house) value, average household income and the share of households living below the poverty threshold at the neighbourhood level as measures of neighbourhood living standards) or include neighbourhood fixed effects.

³³From additional instrumental variable analyses by subgroup, we observe that performances on both outcomes are unbiased for clients of all care needs, except for performance on avoidable hospitalizations for people with higher care needs. Nevertheless, in this case the 1 lies only just outside of the 95% confidence bounds of $\hat{\lambda}$. These results are available upon request.

Columns 1-2 in Table 4 demonstrate that our results are to a large extent robust to the inclusion of these covariates and neighbourhood fixed effects: performance estimates from both models are highly correlated with our baseline estimates. Additionally, at least with 95% certainty, we cannot reject that the forecast coefficients $\hat{\lambda}$ are equal to one when including neighbourhood controls or fixed effects.³⁴

Second, by estimating different specifications of our outcome measures, we verify whether they are sensitive to how they are defined. If performance varies across different stages of admissions, e.g. between the first 90 and 180 days, having a strict cut-off within the outcome measure may not be appropriate. Nevertheless, we find that performance on our main outcomes is highly correlated – with correlations of at least 0.8 – with those of the other specifications in columns 3-5 in Table 4. This also holds for using only falls and fractures instead of all avoidable hospitalizations as an outcome. However, we find a statistically significant bias in observed performance on one-year mortality, discouraging the usage of this outcome as a quality measure.

³⁴One thing to note is that the forecast coefficient in the model including neighbourhood fixed effects (Column 2) is imprecisely estimated. The fact that $\hat{\lambda}$ is only weakly to not statistically significant from zero for both outcomes implies that the within neighbourhood variation in nursing home performance is not predictive for an individual's health outcome, likely because there are relatively few nursing homes per neighbourhood.

Table 4: Robustness tests

	Mortality				
	Neighbourhood controls		Other outcome specifications		
	Characteristics (1)	Fixed effects (2)	90 days (3)	365 days (4)	# days alive (5)
A. Correlation baseline estimates	.999	.881	.830	.792	-.915
B. Forecast coefficient ($\hat{\lambda}$) – performance chosen NH ($\tilde{\delta}_{ij}$)	1.067*** (.086)	1.143* (.632)	1.102*** (.087)	1.221*** (.083)	1.091*** (.081)
C. Forecast bias test ($\hat{\lambda} = 1$)					
χ^2 statistic	.65	.05	1.46	7.32***	1.34
p-value	.420	.815	.227	.007	.247
Covariates	Yes	Yes	Yes	Yes	Yes
N (individuals)	83,501	94,605	94,905	88,619	94,905
	Avoidable hospitalization				
	Neighbourhood controls		Other outcome specifications		
	Characteristics (1)	Fixed effects (2)	90 days (3)	Falls and fractures (4)	# av. hosp. (5)
A. Correlation baseline estimates	.9996	.805	.898	.770	.953
B. Forecast coefficient ($\hat{\lambda}$) – performance chosen NH ($\tilde{\delta}_{ij}$)	1.076*** (.044)	.683 (.684)	1.065*** (.053)	1.041*** (.062)	1.042*** (.049)
C. Forecast bias test ($\hat{\lambda} = 1$)					
χ^2 statistic	2.85*	.22	1.46	.44	.70
p-value	.091	.639	.227	.510	.402
Covariates	Yes	Yes	Yes	Yes	Yes
N (individuals)	64,004	64,992	65,265	65,265	65,265

Panel A shows how our baseline performance estimates correlate to the ones specified in each of the columns. Panel B reports the forecast coefficient, which is equal to the coefficient of predicted performance of the chosen nursing home (by the first-stage regression) in a regression with the individual level outcome as a dependent variable. Panel C tests whether the forecast coefficient $\hat{\lambda}$ is different from zero.

Column 1 includes average property (house) value, average household income and the share of households living below the poverty threshold at the neighbourhood level as additional control variables. Column 2 includes neighbourhood fixed effects. In the remaining columns, the outcome variables are specified respectively as 3) the occurrence of an adverse health outcome within 90 days; 4) within 365 days for mortality and being hospitalized due to a fall or fracture for avoidable hospitalization; 5) the number of days alive within 180 days after admission for mortality and the number of avoidable hospitalizations within 180 days.

Standard errors between brackets. *** Statistically significantly different from zero at 1 percent; ** at 5 percent; * at 10 percent.

Column 1 excludes individuals from very small neighbourhoods, as there is no data available for those. Estimates in Panel A are correlated to estimates obtained from this sample (excl. small neighbourhoods), but with the baseline model (no neighbourhood characteristics). Column 2 excludes individuals for whom the neighbourhood could not be identified. Column 4 for the mortality outcome excludes those admitted to a nursing home from January 2019 onward.

6 Correlations with quality indicators

In this section we examine to what extent quality indicators on other dimensions – like process and structure – can explain observed variation in outcomes. We explore the association of observed performance to publicly available measures of nursing home quality that are often used in comparisons (Castle and Ferguson, 2010; Spilsbury et al., 2011).³⁵ We use these results to evaluate whether performance on outcomes could complement the available indicators based on the other dimensions of quality.

6.1 Process and outcome quality indicators

The nursing home mortality scores are positively correlated with high levels of psychotropic medicine use which is a process-based indicator of low quality that is reported by the Dutch Health and Youth Care Inspectorate (Table 5). 180-day mortality in nursing homes in which all clients use psychotropic medicines is 2 percentage points higher compared to nursing homes in which none of its clients uses psychotropic medicine. Although the coefficient is rather small, the sign of the correlation is in line with the medical literature (Bronskill et al., 2009). Psychotropic medicine use may be related to mortality through side effects that may be more harmful to an older population, like diarrhea (Lindsey, 2009) and delirium. On the other hand, the relationships may also be confounded by (unobserved) other types of nursing home quality. Phillips et al. (2018) argue for example that the number of registered nurse hours is one of the main drivers of antipsychotic medication use among nursing home residents, which may in turn affect mortality through other channels or processes.

Moreover, we find that nursing homes with high rates of avoidable hospitalizations (low quality) have lower pressure sores rates (high quality). At first sight, this correlation may appear to be opposite to what one would expect. However, the negative association with pressure sores may well be a result of residents spending a relatively long time in their nursing home beds, which increases the risk of pressure sores. However, at the same time, spending a lot of time in bed could prevent nursing home residents from falling, which is one of the main contributors of avoidable hospitalizations. In that case, the negative associations are plausible, although it may raise questions about the interpretation of the avoidable hospitalization outcome. Other possible explanations are that pressure sores may be under-reported in the bottom performing nursing homes (Kaltenthaler et al., 2001), or that the (uncorrected) variation in pressure sores is driven by case-mix differences which causes the relationship with performance to be negative if those in worse health (pressure sores \uparrow) are more likely to be admitted to better performing nursing homes (avoidable hospitalizations \downarrow).

6.2 Structure quality indicators

Table 6 presents associations between nursing home characteristics – or structure characteristics – and their mortality and avoidable hospitalizations outcomes. Most estimated coefficients are relatively close to zero. This may mean that, although the multivariate regression includes various observed characteristics³⁶, the results may be confounded by

³⁵Almost every characteristic that we consider is an average over multiple years between 2015 and 2018. The descriptive statistics and a more elaborate description of these characteristics can be found in Appendix Table B1.

³⁶When examining correlations of the same characteristics in bi-variate regressions, we find very similar results. The only difference is that staff absenteeism becomes statistically significant at 10 percent when excluding the other characteristics as covariates.

Table 5: Bi-variate regression with quality measures

<i>Regression:</i>	Mortality		Avoidable hospitalizations		
	(1) Coefficient	(2) Constant	(3) Coefficient	(4) Constant	(5) N
I. Psych. medicine use (↓= <i>better</i>)	0.020*** (0.005)	0.007*** (0.002)	-0.015 (0.010)	0.001 (0.004)	705
II. Physical restraint use (↓= <i>better</i>)	0.008* (0.004)	0.012*** (0.001)	-0.002 (0.009)	-0.004 (0.003)	719
III. Pressure sores (↓= <i>better</i>)	-0.005 (0.019)	0.015*** (0.001)	-0.098** (0.039)	0.000 (0.002)	693
IV. Online rating (1 = <i>worst</i> ; 10 = <i>best</i>)	-0.000 (0.001)	0.015** (0.007)	0.001 (0.002)	-0.010 (0.015)	709

Regression results of eight (2×4 (I-IV)) separate bi-variate regressions at the nursing home level with either performance on mortality (column 1-2) or avoidable hospitalization (column 3-4) as dependent variables. It uses the performance estimates obtained in Equation (2) after shrinkage. Nursing home characteristics are also at the nursing home facility level. Descriptive statistics of nursing home characteristics can be found in Appendix Table B1

Standard errors between brackets. *** Statistically significantly different from zero at 1 percent; ** at 5 percent; * at 10 percent.

(unobserved) other types of nursing home quality, such as managerial quality, which may have offsetting effects: it may reduce the number of staff or higher educated nurses by empowering nurse aids, which in turn could improve outcomes (Barry et al., 2005). Yet, the very weak correlations may also imply that existing structure-based quality indicators do not accurately capture variation in performance on the health outcomes that we measure.

Nevertheless, there are some structure quality indicators that show a somewhat stronger association with performance on outcomes. For instance, we find that a relatively long waiting list is (weakly) negatively associated with nursing home mortality: having a one standard deviation larger waiting list to client ratio (of 12 percent) is associated with a 0.28 percentage points lower mortality rate. Caution is warranted as this association may be due to reverse causality; when nursing home mortality is low, turnover of clients is also low, which may in turn result in longer waiting lists. On the other hand, even in a situation in which mortality rates are not publicly available – as in the Netherlands – there may also be some perception of quality that makes nursing homes with lower mortality more popular.

Our results do not provide evidence for strong relationships between various staffing indicators and performance on health outcomes. This may seem surprising since some studies report that adverse outcomes are related to, for example, lower (registered) nurse employment (Friedrich and Hackmann, 2021; Lin, 2014) and higher nurse turnover (Antwi and Bowblis, 2018). However, the findings from a literature review (see Spilsbury et al. (2011)) suggest that the evidence on this topic has been contradictory. Our results do indicate that nursing homes with larger shares of specialists – like geriatricians and psychologists – relative to total staffing are likely to have lower mortality. Although this relationship is statistically significant, in economic terms it is relatively weak.

Finally, the reported coefficients in Table 6 suggest that the size of nursing home organiz-

ations is linked to performance on avoidable hospitalizations. Keeping the other observed characteristics fixed, an organization with 6 additional facilities (equal to one standard deviation) is associated with a 0.6 percentage points higher avoidable hospitalization rate. This implies that nursing homes that belong to a larger (chain) organization score worse on avoidable hospitalization performance. This finding is in line with quantitative evidence from the United States ([Grabowski et al., 2016](#); [You et al., 2016](#)), who argue that this relationship could be explained by chain targeted nursing homes being of lower quality because of, for example, a poor financial situation, both before and after acquisition. Qualitative evidence suggests that differences between low and high hospitalization nursing homes are related to how the staff approaches the decision to hospitalize ([Cohen et al., 2017](#)). Nursing homes with low rates generally make this decision case-by-case, whereas those with higher rates are more likely to approach it as an algorithmic process. The decision process may well be related to whether the nursing home belongs to a non-chain organization since they are characterized by having more autonomous staff ([Kruzich, 2005](#)), being more flexible in care provision ([Lucas et al., 2007](#)) and since staff may have a more personal relationship with the residents.

Table 6: Multi-variate regression with nursing home characteristics

	Mortality (1)	Avoidable hospitalization (2)
Facility level		
Number of people on waiting list	-0.023* (0.013)	0.008 (0.020)
Number of clients	-0.000 (0.002)	-0.005 (0.004)
Organisation level		
Number of facilities	0.000 (0.000)	0.001** (0.000)
Operating profit margin	-0.000 (0.000)	0.001 (0.001)
Solvency ratio	0.000 (0.000)	0.000 (0.000)
Liquidity	0.000 (0.000)	-0.000 (0.000)
FTE per client	0.009 (0.006)	0.005 (0.017)
Percentage high educated nurses	0.004 (0.008)	-0.000 (0.012)
Percentage specialists	-0.070** (0.028)	-0.137 (0.125)
Staff turnover	-0.027 (0.018)	-0.010 (0.047)
Staff absenteeism	-0.039 (0.115)	0.287 (0.251)
Expenditures on external staff	0.036 (0.029)	-0.003 (0.063)
Constant	0.014 (0.014)	-0.004 (0.025)
R-squared	0.036	0.053
N (facilities)	540	540
n (clusters = organizations)	177	177

Regression results of two multivariate regressions with either nursing home specific performance on mortality (column 1) or avoidable hospitalization (column 2) as dependent variables. It uses the performance estimates obtained in Equation (2) after shrinkage. Nursing home characteristics are either at the nursing home facility level or at the organisation level, which are copied to all facilities within the same organisation. Descriptive statistics of nursing home characteristics can be found in Appendix Table B1

Standard errors between brackets. *** Statistically significantly different from zero at 1 percent; ** at 5 percent; * at 10 percent.

7 Conclusion and discussion

As the quality of care in nursing homes has been the subject of vigorous public debate for decades, the sector might benefit from improved performance measurement based on health outcomes to complement the often-used structure and process measures (Barber et al., 2021; OECD, 2005; OECD and EuropeanCommission, 2013). However, quality estimates tend to rely on self-reported outcomes, be based on a small sample of residents, and be hampered by selection bias. The question is whether these challenges can be addressed and how outcome information can be used best to evaluate performance.

We have addressed the following three questions in this paper. First, how large is the variation in health outcomes across nursing homes? Second, to what extent can this variation be attributed to differences in performance of the nursing homes rather than to unobservable differences in case-mix? Third, is there any relationship with quality indicators based on structure and processes? If such existing indicators to a large extent explain variation in health outcomes, then complementing information on structure and processes with information on outcomes may be of less importance.

We use detailed administrative data to estimate variation in performance on outcomes – mortality and avoidable hospitalization risk – when correcting for observable case-mix differences. In addition, we apply a novel test developed in the value-added literature to examine the role of selection bias in nursing home outcomes. Finally, we examine how these outcomes relate to nursing home characteristics on other dimensions.

After controlling for differences in case-mix, we find substantial heterogeneity in clients' health outcomes across Dutch nursing homes. Due to the small population sizes, the estimates are relatively imprecise, but we can statistically distinguish top performers from bottom performers. We find that the probability of dying or being hospitalized within 180 days after admission is 7 to 14 percentage points higher in the five percent worst performing nursing homes compared to the best. The variation in the avoidable hospitalization outcome is comparable to the variation in rehospitalization rates of Skilled Nursing Facilities in the United States reported by (Rahman et al., 2016), 2016. Moreover, we do not find that unobserved heterogeneity in client characteristics due to non-random selection into nursing homes leads to biased performance estimates. The correlation with other indicators of provider quality is limited, indicating that outcome-based estimates supplement existing process and structure indicators.

Although our findings suggests that nursing homes vary in terms of outcomes, the imprecision in the point estimates is large compared to the observed differences. This means that even when using detailed data and noise-reducing methods like empirical Bayes, it remains difficult to measure variation in outcomes of small-scale providers. As a consequence, differentiating between nursing homes based on performance on outcomes, for example when benchmarking and in pay-for-performance schemes, should be executed with caution, especially when considering the non-extremes.

Our results imply that the observed variation in nursing home outcomes unbiasedly predicts variation in causal mortality and avoidable hospitalization performance. This is important, as it means that case-mix adjustment based on observable characteristics is sufficient for measuring nursing home performance based on outcomes, at least on average. However, selection bias may still be an issue in other settings: this study only includes non-profit nursing homes and things may be different in other institutional settings, e.g. for-profit nursing homes may have stronger incentives to attract healthier clients (Gandhi, 2020).

Also the exclusion of nursing homes with fewer than 50 admissions is important: since our descriptive results show that there is at least some selection into larger vs smaller nursing homes based on observables, there might also be selection based on unobserved characteristics into these smaller homes.

While our results may be seen as reassuring with respect to selection at the aggregate level, they do not imply that observed performance is unbiased for every nursing home separately. Any selection bias due to unobserved heterogeneity may cancel out if it happens to be in the negative direction for some nursing homes and in the positive one for others. While the average bias could then be zero, it may still result into misclassification of certain high performers as some of the lowest, as was observed in the case of hospitals (Geweke et al., 2003; Gowrisankaran and Town, 1999; Hull, 2020). However, given that estimating a bias for individual entities requires sufficient statistical power, this investigation was not feasible in our setting with small providers. The question of whether case-mix adjusted performance on outcomes can be used to promote quality improvements through pay-for-performance incentives remains thus an open question.

All in all, our results suggest that in designing policies to improve the quality of nursing home care, such as public reporting of quality, the dashboards should be expanded with outcome measures. This is especially important in the long-term care sector for which expenditures are expected to grow but for which information on outcomes is limited. Having such additional information is useful, both for nursing homes themselves to identify where improvements may be achievable, and for users aiming to make more informed choices. As our findings suggest that there is no detectable selection bias at the aggregate level, directing users to the nursing homes with the best observable case-mix adjusted outcomes could generate positive benefits in terms of health.

References

- Abaluck, J., Bravo, M. C., Hull, P., and Starc, A. (2021). Mortality Effects and Choice Across Private Health Insurance Plans. *The Quarterly Journal of Economics*, 136(3):1557–1610.
- Angrist, J., Hull, P., Pathak, P., and Walters, C. (2016). Interpreting tests of school value-added. *American Economic Review*, 106(5):388–92.
- Angrist, J. D., Hull, P. D., Pathak, P. A., and Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. *The Quarterly Journal of Economics*, 132(2):871–919.
- Antwi, Y. A. and Bowblis, J. R. (2018). The impact of nurse turnover on quality of care and mortality in nursing homes: Evidence from the great recession. *American Journal of Health Economics*, 4(2):131–163.
- Arling, G., Lewis, T., Kane, R. L., Mueller, C., and Flood, S. (2007). Improving quality assessment through multilevel modeling: The case of nursing home compare. *Health Services Research*, 42(3):1177–1199.
- Barber, S. L., van Gool, K., Wise, S., Wood, M., Or, Z., Penneau, A., Milstein, R., Ikegami, N., Kwon, S., Bakx, P., et al. (2021). Pricing long-term care for older persons. World Health Organization.

- Barry, T., Brannon, D., and Mor, V. (2005). Nurse Aide Empowerment Strategies and Staff Stability: Effects on Nursing Home Resident Outcomes. *The Gerontologist*, 45(3):309–317.
- Becker, C. and Rapp, K. (2010). Fall prevention in nursing homes. *Clinics in Geriatric Medicine*, 26(4):693–704. Falls and Their Prevention.
- Bensnes, S. and Huitfeldt, I. (2021). Rumor has it: How do patients respond to patient-generated physician ratings? *Journal of Health Economics*, 76:102415.
- Bom, J. (2021). Formal and informal long-term care in an ageing society. Doctoral thesis, Erasmus University Rotterdam, the Netherlands.
- Bos, A., Kruse, F. M., and Jeurissen, P. P. T. (2020). For-profit nursing homes in the netherlands: What factors explain their rise? *International Journal of Health Services*, 50(4):431–443.
- Bronskill, S. E., Rochon, P. A., Gill, S. S., Herrmann, N., Hillmer, M. P., Bell, C. M., Anderson, G. M., and Stukel, T. A. (2009). The relationship between variations in antipsychotic prescribing across nursing homes and short-term mortality: quality of care implications. *Medical Care*, 47(2):1000–1008.
- Carter, M. W. (2003). Factors associated with ambulatory care - sensitive hospitalizations among nursing home residents. *Journal of Aging and Health*, 15(2):295–331.
- Castle, N. G. (2003). Searching for and Selecting a Nursing Facility. *Medical Care Research and Review*, 60(2):223–247.
- Castle, N. G. and Ferguson, J. C. (2010). What is nursing home quality and how is it measured? *Gerontologist*, 50(4):426–442.
- CBS (2017). Eigen betalingen zorg. Statistics Netherlands. Available at <https://www.cbs.nl/nl-nl/maatwerk/2017/07/eigen-betalingen-zorg>.
- Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.
- Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228.
- Chida, Y. and Steptoe, A. (2008). Positive Psychological Well-Being and Mortality: A Quantitative Review of Prospective Observational Studies:. *Psychosomatic Medicine*, 70(7):741–756.
- Ching, A. T., Hayashi, F., and Wang, H. (2015). Quantifying the impacts of limited supply: the case of nursing homes. *International Economic Review*, 56(4):1291–1322.
- CIZ (n.d.). Zorgprofielen. Retrieved at April 29, 2020, from <https://www.ciz.nl/zorgprofessional/over-de-wlz/zorgprofielen>.
- Cohen, A. B., Knobf, M. T., and Fried, T. R. (2017). Avoiding Hospitalizations From Nursing Homes for Potentially Burdensome Care: Results of a Qualitative Study. *JAMA Internal Medicine*, 177(1):137–139.

- Cornell, P. Y., Grabowski, D. C., Norton, E., and Rahman, M. (2019). Do report cards predict future quality? the case of skilled nursing facilities. *Journal of Health Economics*, 66:208–221.
- Deming, D. J. (2014). Using school choice lotteries to test measures of school effectiveness. *American Economic Review*, 104(5):406–411.
- Diepstraten, M., Douven, R., and Wouterse, B. (2020). Can your house keep you out of a nursing home? *Health Economics*, 29(5):540–553.
- Donabedian, A. (2003). *An Introduction to Quality Assurance in Health Care*. Oxford University Press.
- Doyle, J. J., Graves, J. A., and Gruber, J. (2019). Evaluating measures of hospital quality: Evidence from ambulance referral patterns. *The Review of Economics and Statistics*, 101(5):841–852.
- Einav, L., Finkelstein, F., and Mahoney, N. (2022). Producing health: Measuring value added of nursing homes. Available at <https://leinav.people.stanford.edu/research>.
- Frandsen, B. R., Lefgren, L. J., and Leslie, E. C. (2019). Judging judge fixed effects. NBER working paper 25528, National Bureau of Economic Research.
- Friedrich, B. U. and Hackmann, M. B. (2021). The Returns to Nursing: Evidence from a Parental-Leave Program. *The Review of Economic Studies*, 88:2308—2343.
- Gadbois, E. A., Tyler, D. A., and Mor, V. (2017). Selecting a Skilled Nursing Facility for Postacute Care: Individual and Family Perspectives. *Journal of the American Geriatrics Society*, 65(11):2459–2465.
- Gandhi, A. (2020). Picking Your Patients: Selective Admissions in the Nursing Home Industry. Available at SSRN: <https://dx.doi.org/10.2139/ssrn.3613950>.
- Geweke, J., Gowrisankaran, G., and Town, R. J. (2003). Bayesian inference for hospital quality in a selection model. *Econometrica*, 71(4):1215–1238.
- Gowrisankaran, G. and Town, R. J. (1999). Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics*, 18:747–767.
- Grabowski, D. C., Feng, Z., Hirth, R., Rahman, M., and Mor, V. (2013). Effect of nursing home ownership on the quality of post-acute care: An instrumental variables approach. *Journal of Health Economics*, 32(1):12–21.
- Grabowski, D. C., Hirth, R. A., Intrator, O., Li, Y., Richardson, J., Stevenson, D. G., Zheng, Q., and Banaszak-Holl, J. (2016). Low-quality nursing homes were more likely than other nursing homes to be bought or sold by chains in 1993–2010. *Health Affairs*, 35(5):907–914.
- Grabowski, D. C., O’Malley, A. J., and Barhydt, N. R. (2007). The costs and potential savings associated with nursing home hospitalizations. *Health Affairs*, 26(6):1753–1761.
- Gupta, A., Howell, S. T., Yannelis, C., and Gupta, A. (2021). Does private equity investment in healthcare benefit patients? evidence from nursing homes. NBER working paper 28474, National Bureau of Economic Research.

- Hackmann, M. B. (2019). Incentivizing better quality of care: The role of medicaid and competition in the nursing home industry. *American Economic Review*, 109(5):1684–1716.
- He, D. and Konetzka, R. T. (2015). Public Reporting and Demand Rationing: Evidence from the Nursing Home Industry. *Health Economics*, 24(11):1437–1451.
- Heins, M., Donker, G., Versteeg, S., and Korevaar, J. (2019). Ontwikkelingen in het aantal euthanasiegevallen en achterliggende factoren. Nivel, Utrecht, the Netherlands.
- Helsø, A. L., Pierri, N., and Wang, A. (2019). The economic impact of healthcare quality. IMF Working Paper No. 19/173.
- Huang, S. S. and Bowblis, J. R. (2018). The principal–agent problem and owner-managers: An instrumental variables application to nursing home quality. *Health Economics*, 27(11):1653–1669.
- Hull, P. (2020). Estimating quality with quasi-experimental data. Available at <https://about.peterhull.net/wp>.
- Hussem, A., Bakx, P., and Tenand, M. (2020). Publieke middelen voor private ouderenzorg. Netspar Design paper 143, Network for Studies on Pensions, Aging and Retirement.
- Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.
- Intrator, O., Zinn, J., and Mor, V. (2004). Nursing home characteristics and potentially preventable hospitalizations of long-stay residents. *Journal of the American Geriatrics Society*, 52(10):1730–1736.
- Joyce, N. R., McGuire, T. G., Bartels, S. J., Mitchell, S. J., and Grabowski, D. C. (2018). The impact of dementia special care units on quality of care: An instrumental variables analysis. *Health Services Research*, 53(5):3657–3679.
- Kaltenthaler, E., Withfield, M., Walters, S., Akehurst, R., and Paisley, S. (2001). Uk, usa and canada: how do their pressure ulcer prevalence and incidence data compare? *Journal of Wound Care*, 10(1):530–535.
- Kane, T. J. and Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. NBER working paper 14607, National Bureau of Economic Research.
- Konetzka, R. T., Yang, F., and Werner, R. M. (2019). Use of instrumental variables for endogenous treatment at the provider level. *Health Economics*, 28(5):710–716.
- Konetzka, T. R., Grabowski, D. C., Perraiillon, M. C., and Werner, R. M. (2015). Nursing home 5-star rating system exacerbates disparities in quality, by payer source. *Health Affairs*, 34(5):819–827.
- Kruzich, J. M. (2005). Ownership, chain affiliation, and administrator decision-making autonomy in long-term care facilities. *Administration in Social Work*, 29(1):5–24.
- Lin, H. (2014). Revisiting the relationship between nurse staffing and quality of care in nursing homes: An instrumental variables approach. *Journal of Health Economics*, 27:13–24.

- Lindsey, P. L. (2009). Psychotropic medication use among older adults: what all nurses need to know. *Journal of gerontological nursing*, 35(9):28–38.
- Lippi Bruni, M., Mammi, I., and Verzulli, R. (2019). Quality in nursing homes. Oxford University Press.
- Lucas, J. A., Levin, C. A., Lowe, T. J., Robertson, B., Akincigil, A., Sambamoorthi, U., Bilder, S., Paek, E. K., and Crystal, S. (2007). The relationship between organizational factors and resident satisfaction with nursing home care and life. *Journal of Aging & Social Policy*, 19(2):125–151.
- Martín-María, N., Miret, M., Caballero, F. F., Rico-Urbe, L. A., Steptoe, A., Chatterji, S., and Ayuso-Mateos, J. L. (2017). The Impact of Subjective Well-being on Mortality: A Meta-Analysis of Longitudinal Studies in the General Population. *Psychosomatic Medicine*, 79(5):565–575.
- McClellan, M. and Staiger, D. (1999). The quality of health care providers. NBER working paper 7327, National Bureau of Economic Research.
- Mor, V., Berg, K., Angelelli, J., Gifford, D., Morris, J., and Moore, T. (2003). The quality of quality measurement in U.S. nursing homes. *Gerontologist*, 43(SPEC. ISS. 2):37–46.
- Morris, C. N. (1983). Parametric empirical bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78(381):47–55.
- Newhouse, J. P. and McClellan, M. (1998). Econometrics in outcomes research: The use of instrumental variables. *Annual Review of Public Health*, 19:17–34.
- Nyman, J. A. (1988). Excess demand, the percentage of medicaid patients, and the quality of nursing home care. *The Journal of Human Resources*, 23(1):76–92.
- NZa (2021). Monitor ouderenzorg - regionale verschillen (deel 1). Nederlandse Zorgautoriteit. Available at https://puc.overheid.nl/nza/doc/PUC_636235_22/1/.
- OECD (2005). Ensuring quality long-term care for older people. Policy brief, Organisation for Economic Co-operation and Development. Available at <https://www.oecd.org/els/health-systems/Ensuring-quality-long-term-care-for-older-people.pdf>.
- OECD (2009). *Premature mortality, in Health at a Glance 2009: OECD Indicators*. OECD Publishing, Paris.
- OECD and European Commission (2013). *A Good Life in Old Age? Monitoring and Improving Quality in Long-term Care*. OECD Health Policy Studies, OECD Publishing.
- Ouslander, J. G., Weinberg, A. D., and Phillips, V. (2000). Inappropriate hospitalization of nursing facility residents: A symptom of a sick system of care for frail older people. *Journal of the American Geriatrics Society*, 48(2):230–231.
- Phillips, L. J., Birtley, N. M., Petroski, G. F., Siem, C., and Rantz, M. (2018). An observational study of antipsychotic medication use among long-stay nursing home residents without qualifying diagnoses. *Journal of Psychiatric and Mental Health Nursing*, 25(8):463–474.

- Quan, H., Li, B., Couris, C. M., Fushimi, K., Graham, P., Hider, P., Januel, J.-M., and Sundararajan, V. (2011). Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. *American Journal of Epidemiology*, 173(6):676–682.
- Rahman, M., Grabowski, D. C., Mor, V., and Norton, E. C. (2016). Is a skilled nursing facility’s rehospitalization rate a valid quality measure? *Health Services Research*, 51(6):2158–2175.
- Rijksoverheid (2017). Rijksbegroting. Government of the Netherlands. Available at https://www.rijksbegroting.nl/2017/verantwoording/jaarverslag,kst244411_29.html.
- Schmitz, H. and Stroka-Wetsch, M. A. (2020). Determinants of nursing home choice: Does reported quality matter? *Health Economics*, 29(7):766–777.
- Shugarman, L. R. and Brown, J. A. (2006). Nursing home selection: How do consumers choose? volume i: Findings from focus groups of consumers and information intermediaries. Rand corporation, Office of Disability, Aging and Long-Term Care Policy Office of the Assistant Secretary for Planning and Evaluation U.S. Department of Health and Human Services, Washington D.C.
- Spilsbury, K., Hewitt, C., Stirk, L., and Bowman, C. (2011). The relationship between nurse staffing and quality of care in nursing homes: A systematic review. *International Journal of Nursing Studies*, 48(6):732–750.
- Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3).
- Statistics Netherlands (n.d.). Definitions: Degree of urbanisation. Retrieved at October 26, 2021, from <https://www.cbs.nl/en-gb/onze-diensten/methods/definitions/degree-of-urbanisation>.
- Sundararajan, V., Henderson, T., Perry, C., Muggivan, A., Quan, H., and Ghali, W. A. (2004). New icd-10 version of the charlson comorbidity index predicted in-hospital mortality. *Journal of Clinical Epidemiology*, 57:1288–1294.
- Tenand, M., Bakx, P., and Wouterse, B. (2021). The impact of co-payments for nursing home care on use, health, and welfare. EsCHER Working Paper Series No. 2021002, Erasmus University Rotterdam. Available from: <https://www.eur.nl/en/research/escher/research/working-papers>.
- Vossius, C., Selbæk, G., Šaltytė Benth, J., and Bergh, S. (2018). Mortality in nursing home residents: a longitudinal study over three years. *PloS one*, 13(9):e0203480.
- Vu, M. Q., Weintraub, N., and Rubenstein, L. Z. (2006). Falls in the nursing home: Are they preventable? *Journal of the American Medical Directors Association*, 7(3, Supplement):S53–S58. CME Digest.
- Walker, J. D., Teare, G. F., Hogan, D. B., Lewis, S., and Maxwell, C. J. (2009). Identifying potentially avoidable hospital admissions from canadian long-term care facilities. *Medical Care*, 47(2):250–254.
- Werner, R. M., Rita, T., and Kim, M. (2013). Quality improvement under nursing home compare: The association between changes in process and outcome measures. *Medical Care*, 51(7):582–588.

- Wouterse, B., Bakx, P., and Wong, A. (2022). Measuring nursing home performance using administrative data. Forthcoming in *Medical Care Research and Review*.
- You, K., Li, Y., Intrator, O., Stevenson, D., Hirth, R., Grabowski, D., and Banaszak-Holl, J. (2016). Do nursing home chain size and proprietary status affect experiences with care? *Medical care*, 54(3):229.

A Appendix

A.1 Data sources

We use administrative data provided by Statistics Netherlands encompassing the full Dutch population. The data on individuals' nursing home care use and care needs, determined by the independent eligibility assessor, comes from the Dutch Central Administrative Office (CAK). We combine the anonymized provider codes and admission dates included in the nursing home use data with data on addresses from the mandatory municipal registry to link individuals to their chosen nursing home.

We obtain mortality data from the municipal registries, which includes the exact date of death. Data for our other outcome variable on hospital visits comes from NZa (Nederlandse Zorgautoriteit) and includes both ICD-10 and DBC (in Dutch: Diagnose Behandeling Combinatie) codes.

We also obtain age and sex from the municipal registries, and supplement this with the following data sources to define our other covariates: medicine consumption by ATC-code from SGZ (Gezondheid en Zorg); income and wealth data from tax registries; and healthcare expenditures from insurance data provided by Vektis.

A.2 Monotonicity assumption

An IV estimate can only be interpreted as a local average treatment effect (LATE) under the monotonicity assumption. Because our (non-binary) instrument has multiple values, we can interpret λ as a weighted sum of multiple LATEs defined by different pairs of values of the instrument (Imbens and Wooldridge, 2009). We can only interpret λ as LATE if the value of the treatment variable ($\hat{\delta}_{i,j}$) is increasing (or decreasing) with every increment in the value of the instrument, and if the monotonicity assumption holds for any two values of the instrument.

In our case, monotonicity means that if a person chooses a low mortality nursing home far away because the closest nursing home has slightly higher-than-average mortality – e.g. 1 percentage point above average –, this person should also choose a low mortality nursing home if the closest nursing home has even higher mortality — e.g. 2 percentage points above the average. The same person should also choose a low mortality nursing home if it was to be the closest one.

While we cannot test this directly, we can do the following: first, we plot the average performance estimate of the chosen nursing home (the endogenous treatment variable) against the instrument. Figures C2a and C2a shows that the endogenous performance estimates on average monotonically increase with the instrument. Second, we follow Frandsen et al. (2019) and run the first-stage regressions for a wide range of subgroups. We find positive and statistically significant coefficients for all of them. This is suggestive evidence in favor of the weak monotonicity assumption, which is sufficient for causal interpretation (Frandsen et al., 2019).

B Appendix - Tables

Table B1: Descriptive statistics nursing home quality indicators

Variable	Description	Mean	Std. dev.	N
Facility level				
Psychotropic medicine use	Share of clients using psychotropic medicine	0.370	0.166	705
Physical restraint use	Share of clients to whom physical restraints are applied	0.303	0.184	719
Pressure sores	Share of clients with pressure sores	0.042	0.041	693
Online rating	Online rating according to Zorgkaart Nederland (1-10)	7.761	0.826	709
Number of people on waiting list	Number of people actively or passively waiting per nursing home	0.092	0.119	540
Number of clients	Number of clients	100.263	54.119	540
Organisation level				
Number of facilities	Number of facilities belonging to organization (= 1 if no chain)	6.921	6.104	177
Operating profit margin	Operating profit divided by total revenue multiplied by 100	1.376	2.603	177
Solvability	Equity-to-asset ratio: equity divided by total revenue multiplied by 100	38.290	15.299	177
Liquidity	Current ratio: current assets divided by current liabilities multiplied by 100	177.882	112.499	177
FTE per client	Care-related FTE per nursing home client	0.697	0.174	177
Percentage high educated nurses	Percentage of employed nurses that is high educated (in terms of FTE)	0.331	0.239	177
Percentage specialists	Percentage of care-related FTE that is specialized staff (i.e. geriatricians, medical doctors, psychologists and nurse practitioners)	0.026	0.031	177
Staff turnover	Staff turnover measured by outflow divided by average staffing (in terms of FTE)	0.143	0.063	177
Staff absenteeism	Percentage staff absenteeism	0.063	0.011	177
Expenditures on external staff	Percentage of salary expenses on external staffing	0.066	0.045	177

This table lists the nursing home quality indicators used in Section 6 in this paper. It reports a more elaborate description of each of the indicators, the mean, standard deviation and number of unique observations. Some indicators are reported at the facility level, while others are reported at the organisation level.

Table B2: Avoidable hospitalization diagnoses

Diagnoses	Percent
Anaemia	2.50
Angina pectoris	2.24
Asthma & COPD	6.32
Dehydration	0.44
Diabetes	3.59
Falls & fractures	52.82
Gastroenteritis	1.51
Grand mal seizure disorders	2.37
Hypertension	0.98
Infections skin	2.66
Infections other	1.25
Kidney or urinary tract infections	4.14
Pneumonia	8.16
Rehabilitation	5.02
Sepsis	3.14
Wounds	2.85
Total	(n = 10,702) 100

This table lists the diagnoses of hospitalizations that we identify as potentially avoidable, which is based on [Carter \(2003\)](#); [Walker et al. \(2009\)](#).

Table B3: Hospitalization by avoidable and type of stay

Type	Avoidable hospitalization		Total (%)
	No (%)	Yes (%)	
Outpatient stay	78.2	63.6	72.9
Inpatient stay - one-day	4.0	2.3	3.4
Inpatient stay - overnight	17.8	34.0	23.7
Total	(n = 18,968)	(n = 10,702)	(n = 29,670)

This table reports whether the hospitalization within 180 days after nursing home admission was an outpatient, a one-day inpatient stay or an overnight inpatient stay. It includes individuals who had at least one hospitalization within 180 days after nursing home admission and separates by those of which at least one of them was potentially avoidable or not.

Table B4: Variable specifications

	Variable	Explanation
Health outcomes Y_{ij}	Mortality	Binary outcome variable that equals one if resident i died within 180 days after nursing home admission. Data on the date of death is obtained from death registries.
	Avoidable hospitalization	Binary outcome variable that equals one if resident i had a potentially avoidable hospitalization. Diagnoses for which a hospitalization is perceived as potentially avoidable are listed in Table B2.
Case-mix controls X_i	Women	Equals one if resident i is a woman
	Age	Age of at admission
	Care intensity	A care intensity package (in Dutch: Zorgzwaartepakket (ZZP)) is a proxy for the intensity of nursing home care that the recipient needs according to an independent care assessor from the Care Assessment Centre (CIZ). Residents with lower care (ZZP 4) intensity need intensive support and extensive care, with dementia care (ZZP 5) need a protective living facility with intensive dementia care, with higher care intensity need a protective living facility with intensive support and care, with the highest care intensity (ZZP 7 and 8) need a protective living facility with very intensive care and treatment or support (CIZ, nd).
	Healthcare expenditures	Yearly healthcare expenditures in logarithmic form from the calendar year before admission, obtained from claims data.
	Wealth	Wealth from assets and savings (excluding home equity) of the household in the calendar year prior to admission. Variable enters the control function in logarithmic form.
	Standardized household income	Yearly disposable household income in the calendar year before admission from tax registries, standardized for household size. Variable enters the control function in logarithmic form.
	Number of medicine	Is equal to the number of different types of medicine resident i received in the calendar year before admission. Types are distinguished based on ATC3 codes.
Continued on next page		

Table B4 – continued from previous page

	Variable	Explanation
	Charlson score	The Charlson comorbidity index is a weighted score based on 12 different comorbidities: congestive heart failure, dementia, pulmonary disease, connective tissue disorder, liver disease, diabetes, diabetes complications, paraplegia, renal disease, cancer, metastatic cancer, severe liver disease and HIV (Quan et al., 2011). Diagnoses are obtained from hospital visits in the 365 days before nursing home admission.
	Hospital in last month	An indicator for whether the individual had an hospitalization within the last 30 days before the individual were admitted to a nursing home.
	Rural	This binary variable indicates whether the resident's prior home was in a rural (equals 1) or urban (equals 0) municipality. We categorize a municipality as urban if the area is at least moderately urbanised (> 1,000 addresses per square kilometer) (Statistics Netherlands, nd).
	Year	Indicates in which year the individual is admitted to a nursing home.

Table B5: Subgroup analysis

		Outcome: Mortality			
		Lower	By care intensity		Highest
			Dementia	Higher	
Performance chosen NH ($\hat{\delta}_{ij}$)					
Coefficient		0.774*** (0.053)	0.991*** (0.038)	1.135*** (0.055)	0.669** (0.279)
Constant		-0.461*** (0.068)	-0.521*** (0.052)	-0.310*** (0.062)	-0.935** (0.392)
controls		Yes	Yes	Yes	Yes
R-squared		0.065	0.066	0.093	0.078
N		17,923	46,144	29,703	1,135
		Outcome: Avoidable hospitalization			
		Lower	By care intensity		Highest
			Dementia	Higher	
Performance chosen NH ($\hat{\delta}_{ij}$)					
Coefficient		1.004*** (0.053)	0.850*** (0.033)	1.155*** (0.043)	1.200*** (0.216)
Constant		0.135 (0.089)	-0.061 (0.049)	0.429*** (0.058)	0.353 (0.345)
controls		Yes	Yes	Yes	Yes
R-squared		0.054	0.029	0.057	0.044
N		11,811	29,869	22,693	892

For this exercise we distinguish between four groups: (1) clients with a lower level of care intensity; (2) clients with a higher level of care intensity; (3) clients with the highest level of care intensity; and (4) clients with dementia care needs. These are identified by the care intensity package, as explained in Footnote ⁶ and Appendix Table B4. For each of these groups, we regress the individual level outcome – mortality or avoidable hospitalization – on the observed performance score of the nursing home to which the individual is admitted ($\hat{\delta}_{ij}$).

Standard errors between brackets. *** Statistically significantly different from zero at 1 percent; ** at 5 percent; * at 10 percent.

C Appendix - Figures

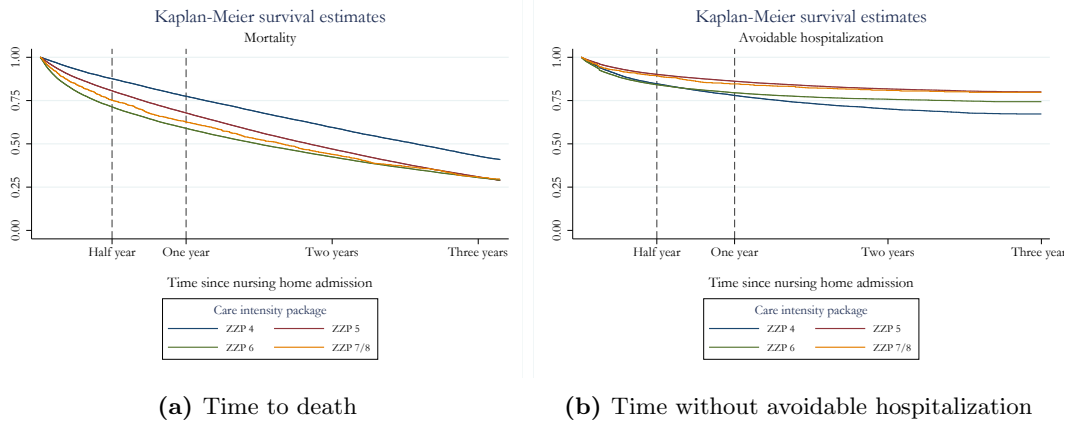


Figure C1: Kaplan-Meier survival curves by care intensity

This figure shows Kaplan-Meier survival curves for time to death (mortality) and time to an avoidable hospitalization by care intensity package. Time to death for those who enter the nursing home with a higher care intensity is, on average, shorter than for those with a lower care intensity. On the other hand, those who have relatively high or dementia care needs, stay longer without having an avoidable hospitalization. This can partly be explained by this group dying earlier.

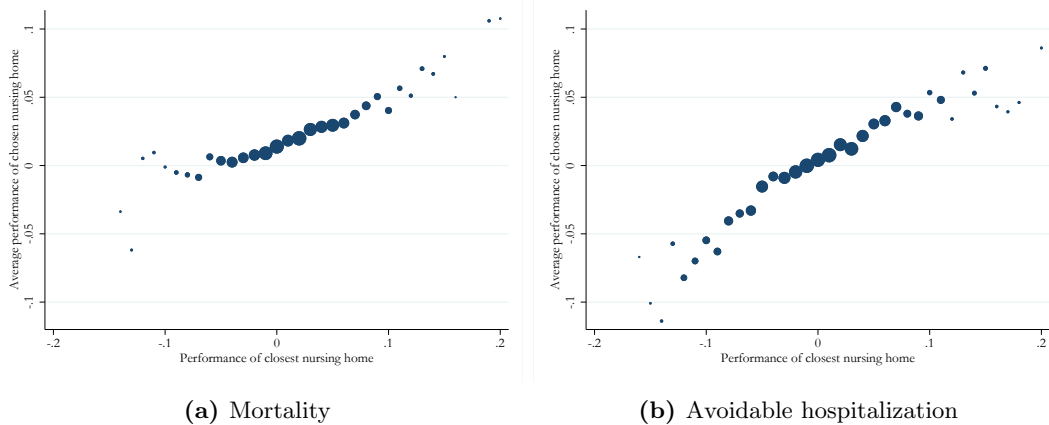


Figure C2

This figure (nonparametrically) shows the relationship between performance of the chosen nursing home (i.e. the endogenous variable) and performance of the closest nursing home (i.e. the instrumental variable) for both outcomes: mortality (a) and avoidable hospitalization (b). The size of the data points reflect the group size on which the average is based.