DEPARTMENT OF ECONOMICS
JOHANNES KEPLER UNIVERSITY OF LINZ

Evaluating hospital performance

by

Gerald J. PRUCKNER
Thomas SCHOBER

Working Paper No. 2503
March 2025

Johannes Kepler University of Linz
Department of Economics
Altenberger Strasse 69
A-4040 Linz - Auhof, Austria
www.econ.jku.at

gerald.pruckner@jku.at

Evaluating hospital performance

Gerald J. Pruckner[a] and Thomas Schober[b]

[a] *Department of Economics, Johannes Kepler University Linz, Austria*
[b] *New Zealand Policy Research Institute, Auckland University of Technology, New Zealand*

March 9, 2025

Abstract

Patient selection remains a major challenge in evaluating hospital performance. We exploit the quasi-random assignment of patients to hospitals, based on a rotation schedule between hospitals in the Upper Austrian capital of Linz. In the instrumental variable (IV) framework, we use high-quality administrative data and estimate hospital performance on patient outcomes such as mortality and readmission. We contrast these results with those of traditional risk adjustment models based on patient observables.

We find that the assessment of hospital performance is sensitive to the inclusion of patient observables and that increasing the number of socioeconomic covariates to better control for patient risk profiles does not always help bring risk-adjusted estimates closer to IV estimates. Our results suggest that common risk adjustment does not adequately control for patient differences between hospitals and that hospital quality indicators based on common administrative data should be interpreted with caution. The trend toward personalized medicine may support the process of collecting more clinical information at the individual level, thus allowing for better quality comparisons between hospitals.

JEL Classification: I11, I18, H42, L13
Keywords: Hospital performance, hospitalization, health risk adjustment, admission schedule

Corresponding author: Thomas Schober, Auckland University of Technology, 42 Wakefield Street, Auckland 1142; thomas.schober@aut.ac.nz. For helpful discussions and comments, we would like to thank participants at the research seminar of the Department of Economics in Linz and the CINCH Research Center in Essen, the World Congress of the International Health Economics Association, the Annual Meeting of the Austrian Health Economics Association, and the New Zealand Association of Economists Conference. We gratefully acknowledge financial support from the Anniversary Fund of the Oesterreichische Nationalbank (OeNB) under project no. 18460.

1 Introduction

Hospitals are major healthcare providers, accounting for two-fifths of all healthcare spending in OECD countries (OECD, 2023), and play a central role in policy efforts to improve healthcare. Many countries are shifting towards a greater focus on patient outcomes rather than inputs. In some countries, hospital quality indicators are publicly reported to inform patients and increase incentives for providers to improve quality of care. A prominent example is the US Medicare website medicare.gov, which allows a search for hospitals and lists several indicators such as mortality rates for heart attack patients. Pay-for-performance initiatives are designed to directly reward the quality of care. For example, the Advancing Quality program provides financial incentives linked to hospital performance in England (Kristensen et al., 2014) and the Hospital Readmission Reduction Program penalizes hospitals with higher-than-expected readmission rates for certain conditions in the US (Gupta, 2021).

Patient selection is a major challenge in comparing hospitals. Patients and healthcare professionals involved in their care may choose hospitals based on their capabilities. Therefore, the highest-quality hospitals may treat the sickest patients, which can lead to worse average patient outcomes than in other hospitals. Commonly used hospital quality indicators rely on risk-adjustment methods to account for differences in patients' health status and characteristics and compare outcomes such as mortality and readmissions of a standardized patient population. However, these methods are often criticized for their inability to fully control for the differences in patients between hospitals (Lilford and Pronovost, 2010; Goodacre et al., 2015; Baker and Chassin, 2017; Doyle et al., 2019). Important characteristics such as the severity of illness are often not (well) observed in the available data. The differences in clinical coding and admission practices may affect the observable variables. Therefore, using these observed characteristics may increase the bias that risk adjustment intends to reduce (Mohammed et al., 2009).

This study assesses hospital performance using exogenous variation in hospital admissions to account for patient selection. This variation is induced by the unique feature of inpatient care in the Austrian province of Upper Austria. In the capital city of Linz, hospitals agreed on a rotation schedule in which one or two hospital(s) is (are) primarily responsible for inpatient admissions each day. For patients requiring acute care, the schedule creates a quasi-random allocation system for different hospitals. We use this variation in an instrumental variable (IV) framework to estimate hospital performance based on patient outcomes, using high-quality administrative data. We contrast these results with those of a "traditional" risk adjustment approach that accounts for patient differences through a set of observable characteristics. We assess whether the risk adjustment adequately accounts for patient selection.

In practice, the implementation of risk-adjustment methods varies widely (Pross et al., 2017; Cacace et al., 2019; Goodacre et al., 2015), with differences in both the choice of outcome variables and variables used to account for observable patient characteristics. This can be attributed to the availability of data and discretionary choices. An early introduction in England, called the hospital standardized mortality ratio, focused on in-hospital mortality (Jarman et al., 1999), and was later adopted in several other countries (CIHI, 2007; Heijink et al., 2008; Miyata et al., 2008). Other hospital performance indicators include post-discharge mortality (Pouw et al., 2013; Hosein et al., 2014; Ridgeway et al., 2019) or hospital readmission rates (Roberts et al., 2018).

Regarding the variables used for risk adjustment, quality indicators are usually adjusted for patient age, sex, and comorbidities for which data are readily available in routine hospital data (Cacace et al., 2019). To better account for case severity, additional characteristics such as the length of hospital stay are sometimes added (for example, Jarman et al., 2010). An important concern is that several variables available in routine data are endogenous and result from hospital performance. A recent topic of discussion has been whether social factors should be included as they are typically correlated with patient outcomes. Omitting these factors could lead healthcare providers to avoid enrolling vulnerable patients. Conversely, adjusting for social risk could allow hospitals to justify providing lower-quality care for these patients (Braithwaite, 2018; Berry and Chien, 2016).

This study contributes to the literature on the evaluation of hospital performance. Doyle et al. (2015) and Doyle et al. (2019) indicate that patients treated in hospitals that incur higher costs and score higher on commonly used quality measures have, on average, better health outcomes. These studies have also used instrumental variables to account for patient selection. In this case, the exogenous variation comes from ambulance companies with different propensities to transport patients to particular hospitals. Although our identification strategy is similar, we aim to assess the extent to which (different versions of) risk adjustment adequately account for patient selection. Understanding this is crucial, as risk-adjusted quality indicators are widely used in the literature to analyse various aspects of hospital performance, including racial disparities in healthcare (Chandra et al., 2022) and it's relationship with profitability (Garthwaite et al., 2022).

Evidence suggests that different risk-adjustment approaches can lead to substantially different conclusions. Shahian et al. (2010) compare in-hospital mortality in Massachusetts acute care hospitals using four varying risk adjustment methods, including selection of hospital cases and patient covariates. These results are often inconsistent, with some hospitals having lower than expected mortality for one variant and higher than expected mortality for another variant. Austin et al. (2015) indicate a similar lack of agreement between hospital rating systems for the US as a whole. We also analyze the variability of risk adjustment approaches by using different sets of covariates to account for patient characteristics. More importantly, by taking advantage of the random assignment of patients to hospitals, our instrumental variable estimates provide a credible benchmark for comparison.[1]

We find a significant disagreement when comparing hospital performances using risk adjustment and the IV approach. Consistent with previous findings, we find that hospital performance assessment is sensitive to the inclusion of patient observables. However, increasing the number of covariates to better control for patient characteristics does not always bring the risk-adjusted estimates closer to the IV estimates. Our results suggest that common risk adjustment does not adequately control for patient differences between hospitals and that hospital quality indicators based on common administrative data should be interpreted with caution.

The remainder of this study is organized as follows. Section 2 describes the institutional setting, including the admission schedule and data. Section 3 outlines the IV and risk adjustment approaches. Section 4 presents the results and Section 5 concludes the paper.

2 Background and data

2.1 Institutional setting

Austria has a Bismarckian social security system, with compulsory public health insurance. Membership in social insurance schemes is determined by occupation and place of residence, meaning that patients cannot choose their schemes freely. Most population is covered by the Austrian Health Insurance Fund, which is administered by nine regional health insurance organizations. Regional funds cover active and retired private sector employees (including the unemployed) and their co-insured

[1] A body of related literature documents the wide variation in the diagnostic behavior of healthcare providers, both between (Song et al., 2010; Welch et al., 2011; Finkelstein et al., 2017) and within (Fadlon and Van Parys, 2020; Gowrisankaran et al., 2023) regions. Hospitals with diagnosis-related group (DRG) payment systems may also distort the coding of patients in response to financial incentives (Cook and Averett, 2020; Di Giacomo et al., 2017; Jürges and Köberlein, 2015).

dependents.[2] Insured individuals have access to a wide range of services, including outpatient visits to general practitioners (GPs) and specialists, hospital care, and prescription medicines. Healthcare costs are covered by public health insurance with little or no co-payment, such as a prescription fee of 6.10 € (2019).

Public and private non-profit hospitals largely provide inpatient care. Expenditure on hospital care is covered by social security contributions and taxes at various federal levels. Hospital services are paid for using a DRG system. Similar to other DRG systems, hospital cases are classified into diagnostic case groups, which serve as a proxy for the cost of cases based on diagnosis and treatment (Bachner et al., 2018).

Quality control and benchmarking of Austrian hospitals remains in its infancy. The Austrian Inpatient Quality Indicators have been developed to compare the quality of inpatient care. These indicators include patient outcomes such as mortality and complication rates, but transparency is limited because most data are not publicly reported at the hospital level (Bachner et al., 2018; Wörndle et al., 2023).

2.2 Data

The Regional Health Insurance Fund of Upper Austria provides detailed individual-level data on healthcare utilization, covering more than 1.5 million people in the province of Upper Austria. We analyze data on hospital admissions from 2010 to 2019, during which more than 1 million inpatient admissions occurred in the four hospitals participating in the admission schedule. Inpatient data include the date of admission, length of hospital stay, and principal and secondary diagnoses according to the International Classification of Diseases (ICD-10). We use principal diagnoses to select disease groups for the empirical analysis and secondary diagnoses to measure comorbidities. We also use the indicator of ambulance use on the day of admission to identify emergency cases. The data also include basic demographic characteristics such as age and sex.

Hospital patients can be linked to the Austrian Social Security Database (ASSD), a linked employer-employee dataset containing the labor market history of all private-sector employees in Austria (Zweimüller et al., 2009). The data source includes the socioeconomic characteristics of the patients. We use the patients' income earned in the year prior to hospital admission to calculate their daily wages. For pensioners, we use their last recorded wage, whereas for dependents with no income, we use

[2]Special social insurance institutions provide compulsory health insurance for certain occupational groups, such as farmers, civil servants, or the self-employed.

the wage of the primary insured as a proxy for socioeconomic status. The data also include a broad occupational indicator of whether the patient has ever held a white-collar job and an indicator of whether the patient has an academic degree to reflect educational background. The ASSD also provides information on mortality, allowing us to observe deaths both inside and outside the hospitals.

2.3 Admission schedule

Our empirical analysis focuses on four acute care hospitals in the Upper Austrian capital of Linz, one of which is publicly owned, and three of which are owned by Catholic orders. All hospitals are non-profit and operate under the same DRG system. Hospitals coordinate their inpatient admissions via so-called admission days (Ordensklinikum Linz, 2019). Each day, one or two hospitals are primarily responsible for the admission of acute patients according to a defined admission schedule. The schedule covers 21 days, as listed in Table 1.[3] Hospital admissions change every day, except Sunday, when the Saturday hospital remains in charge. The two smaller hospitals share the same admissions but split the female and male patients. The admission schedule is balanced so that each hospital is in charge for seven days during the three weeks.

The aim of the admission schedule is to systematically distribute patients between hospitals and facilitate easy planning. For example, hospital operators can reduce staffing levels on non-admission days and concentrate staff in relevant departments on admission days. Hospital staff handling admissions are not allowed to turn patients away on admission days. On non-admission days, the patients are referred to the relevant hospital. Emergency service control centers are informed of the schedule and ambulances are instructed to follow the admission schedule and visit the designated hospitals. This ensures that patients are admitted to hospitals that are prepared to provide the necessary care.

Table 1: Admission schedule

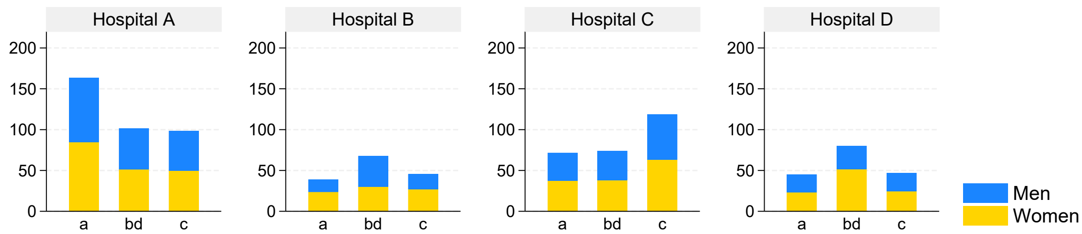
	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Week 1	c	a	c	a	bd	c	c
Week 2	a	bd	a	bd	c	a	a
Week 3	bd	c	bd	c	a	bd	bd

Notes: On admission day a, hospital A is primarily responsible for inpatient admissions. On admission day c, it is hospital C. On admission day bd, hospital B and hospital D share the responsibility but split female and male patients. The schedule repeats every three weeks.

[3] In the following, admission days are indicated in lower case and hospitals in upper case.

The admission schedule implies that, for acute patients, the date affects the hospital to which they are admitted. However, there are some important exceptions, even for acute patients. Maternity and pediatric units always accept patients; therefore, cases related to childbirth and childcare do not follow this rule. Among the four hospitals, only hospital A has an accident unit; therefore, accident patients are always admitted there. The admission plan does not apply to patients with stroke. Patients are admitted if a hospital has a stroke unit or at least one neurology department (hospitals A, B, and C). Other exceptions are patients whose non-admission would mean an acute danger to life or serious damage to health, to be decided on a case-by-case basis by doctors on duty. Finally, patients who have recently undergone surgery at a hospital and those returning due to unforeseen post-operative complications are exempt from the admission schedule.

Figure 1: Hospital admissions



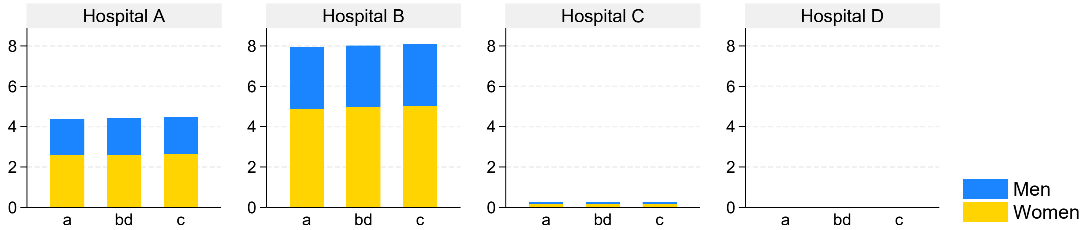
Notes: This figure shows the average number of admissions in hospitals A, B, C, and D on different admission days a, bd, and c.

Figure 1 shows how the admission schedule affects admissions in practice. It shows the average number of admissions per day for the four general acute care hospitals, separately for the three admission days. On admission day a, hospital A admits an average of more than 160 patients covered by the Upper Austrian Regional Health Insurance Fund. This is comparable to 100 patients on other days. A similar pattern is observed across all hospitals, with admissions peaking when it is the hospital's designated turn according to the admission schedule.

For two hospitals with the same admission day, Figure 1 confirms that the patient's sex plays a crucial role. On admission day bd, hospital B admits on average, twice as many men as on other days (38 compared with 15 and 19), while there is only a small increase in the number of female patients. Conversely, significantly more women are admitted to hospital D on this day.

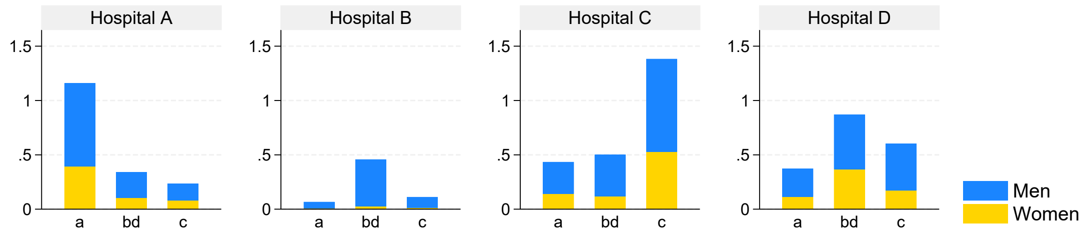
Each hospital admits a certain number of patients, even on days when they are not responsible according to schedule. This is owing to the large number of planned admissions and acute exceptions discussed above. Figure 2 illustrates this by showing the admissions of patients with senile cataract (ICD-10 diagnosis H25).

Figure 2: Hospital admissions – senile cataract



Notes: This figure shows the average number of patients admitted with senile cataract to hospitals A, B, C, and D on different admission days a, bd, and c.

Figure 3: Hospital admissions – acute myocardial infarction



Notes: This figure shows the average number of patients admitted with acute myocardial infarction to hospitals A, B, C, and D on different admission days a, bd, and c.

Cataracts are a leading cause of visual impairment and blindness, and affect millions worldwide, particularly as they age (Liu et al., 2017). Patients with this diagnosis are almost exclusively treated in hospitals A and B, which receive five and eight patients per day, respectively. Notably, there are no peaks on admission days, as care is planned and the admission schedule does not apply.

In contrast, Figure 3 shows admissions for acute myocardial infarction (AMI) (I21), commonly known as a heart attack. Each hospital admitted significantly more patients on days primarily scheduled for acute care. This is because the treatment for AMI is highly time-sensitive and typically unplanned.

However, even for diagnoses of such acute conditions, there are variations, meaning that not all patients are admitted to the appropriate hospital as planned. One reason for this pattern is the measurement error. Admission days start and end at 7am; however, because the data do not include the exact time of admission, cases admitted between midnight and 7am are assigned to the wrong admission day (except for Sunday mornings). Second, not all conditions that typically require immediate care are acute. Some might be related to continuing care after transfer from another hospital, planned readmissions, or conditions discovered or occurring during hospital visits for other purposes. Third, because the admission schedule is not legally binding, we expect hospitals to routinely deviate from it. For example, a patient with a history of care at a particular hospital may be readmitted to that hospital regardless of the admission schedule.

It is because of such deviations why we use an instrumental variable framework to estimate hospital performance, where admission days explain patient admissions partly. In this framework, admissions that do not follow the planned schedule reflect non-compliance, similar to randomized trials in which participants do not receive the treatment to which they are assigned.

3 Empirical strategy

We use the admission schedule in an instrumental variable framework to compare the quality of care between the hospitals. For some patients (compliers), the day on which an acute health problem occurs affects the hospital to which they are admitted. Thus, the schedule creates an exogenous variation in admissions, addressing the common concern that patients are selected into hospitals based on unobservable characteristics.

We then contrast the IV results with traditional risk adjustment, in which we control for observable characteristics to account for patient differences between hospitals. We evaluate several variants of risk adjustment that differ in patient observables to assess the extent to which these methods adequately account for patient selection.

Instrumental variables We evaluate the performance of four acute care hospitals participating in the admission schedule using the two-stage least squares method. The corresponding second-stage equation is as follows:

$$Y_i = \beta_1 B_i + \beta_2 C_i + \beta_3 D_i + X_i' \zeta + \epsilon_i, \tag{1}$$

where Y_i, the outcome Y of patient i, is regressed on the binary variables B_i, C_i, and D_i, indicating whether the patient was admitted to hospitals B, C, or D, and the control variables X_i. Hospital A serves as the base hospital. Therefore, β_1 can be interpreted as the effect of being admitted to hospital B compared with hospital A.

We examine three outcome variables: (i) in-hospital mortality, (ii) 30-day mortality after admission, and (iii) 30-day readmission after discharge. Readmission is defined as the admission to any hospital with any diagnosis. In-hospital mortality is often used as an indicator of the quality of care because it is commonly available in administrative hospital data and does not require linkage to other data sources. One concern is that hospitals may have different discharge practices, which could bias the results. For example, there may be differences in the transfer of high-risk patients to more specialized hospitals or palliative care facilities (Pouw et al., 2013; Kozar et al., 2014). However, post-discharge mortality is also influenced by factors outside the hospital's control, such as the quality of outpatient care.

Patient readmission is often used as an indicator of hospital performance, as it may indicate ineffective treatment at index admission. However, differences in survival rates between hospitals may affect the number of patients at risk of readmission, leading to potential bias when comparing readmission rates (Laudicella et al., 2018).

Because hospital admissions are endogenous, we instrument the hospital indicators using *admission days* derived from the admission schedule. These variables indicate whether a given hospital was responsible for inpatient admissions on the day the patient was admitted. Two hospitals have the same days of admission. As they agreed to split the patients by sex, we interact their admission day with the sex of the patient in the three first-stage equations.

$$B_i = \gamma_1 c_i + \gamma_2 bd_i + \gamma_3 c_i \times f_i + \gamma_4 bd_i \times f_i + X_i'\kappa + \eta_i, \tag{2}$$

$$C_i = \alpha_1 c_i + \alpha_2 bd_i + \alpha_3 c_i \times f_i + \alpha_4 bd_i \times f_i + X_i'\psi + \mu_i, \tag{3}$$

$$D_i = \delta_1 c_i + \delta_2 bd_i + \delta_3 c_i \times f_i + \delta_4 bd_i \times f_i + X_i'\phi + \xi_i, \tag{4}$$

where the instrumental variables c_i and bd_i indicate whether hospital C or hospitals B and D had an admission day on the day a patient was admitted and f_i is an indicator of whether the patient is female. The interaction of f_i with admission days is used as an additional instrument, whereas the direct effect of females is included in the control vector X_i. We also control for the patients' principal diagnosis, age, and year of admission.

Instrument validity The underlying assumption is that admission days only affect outcomes by influencing the hospital to which a patient is admitted and that there is no direct effect of the admission day on hospital outcomes. A problem could arise if the admission days were correlated with weekdays, such that a particular hospital was always responsible for admissions on Saturdays. An extensive literature suggests that outcomes are worse for patients admitted on weekends, which has been linked to a poorer quality of care, sicker patients, and inconsistencies in coding between weekends and weekdays (Black, 2016). However, as discussed in Section 2.3, the admission schedule follows a rotating pattern, so that after three weeks, each hospital was responsible for admissions exactly once on each day of the week.

A related issue is potential differences in patient selection in the observed patient pool. Hospitals may differ in the extent to which they admit patients as inpatients or treat them as outpatients, depending on the severity of the underlying condition. This could result in different patients being admitted depending on which hospital is currently responsible for inpatient admissions. Similarly, there could be differences in the extent to which non-participating hospitals are used as substitutes for care at participating hospitals. For example, in out-of-town emergencies, ambulances may change their tendency to transfer patients to hospitals in cities other than Linz. To investigate this issue, we analyze whether observable patient characteristics differed between admission days.

Note that non-compliance with the admission schedule is not a fundamental problem. As discussed in Section 2.3, a large proportion of hospital care involves planned visits that do not follow the admission schedule. Even during emergencies, if a patient seeks care in a hospital that is not scheduled to receive patients on a particular day, there is no legal obligation to turn the patient away. Finally, there may be medical reasons for deviations from the schedule; for example, if a patient has been treated in a particular hospital in the past. Therefore, our estimates must be interpreted as local average treatment effects (LATE) for patients for whom the admission schedule is important–the so-called compliers.

Risk Adjustment We contrast the IV results with a risk-adjustment approach in which we control for observable characteristics to account for patient differences between hospitals. Here, we estimated Equation 1 by using ordinary least squares (OLS). We estimate different models, labeled RA1 to RA4, in which the control variables are added stepwise.

RA1 includes the same control variables as the IV outcomes: patient diagnosis, age, sex, and year of admission. In RA2, we add the Charlson Comorbidity Index to account for the severity of the comorbidities.[4] RA3 adds control variables for socioeconomic status. We include in the regressions variables indicating whether the patient has a university degree, whether they have ever held a white-collar job, and the wage level at which they worked, represented by quintile-based indicators for the five wage groups. RA4 adds indicators for emergency transport and length

[4]The Charlson Comorbidity Index is a scoring system that assesses a patient's comorbidities and was developed to classify those that may affect mortality risk (Charlson et al., 1987). We use the implementation for ICD-10 diagnostic codes by Quan et al. (2005), which is based on 17 differently weighted conditions. For a recent review of their clinimetric properties, see (Charlson et al., 2022).

of hospital stay. Here, we group the length of stay into categories of 1 day, 2 days, 3 to 9 days, 10 to 15 days, 16 to 21 days, and more than 22 days, following the Canadian risk adjustment practice (CIHI, 2016). Adding different sets of control variables allows us to explore the sensitivity of the results and test how the estimates of hospital performance using risk adjustment compare with the IV results.

Selection of diagnoses As discussed in Section 2.3, the admission schedule is relevant only in certain hospital cases. It cannot be used to evaluate typically planned hospital care (e.g., for cataract patients). Following Doyle et al. (2015), we restrict the analysis to diagnoses with high weekend admission rates. The rationale is that diagnoses that occur frequently on Saturdays and Sundays are indicative of 'non-discretionary' conditions requiring immediate care, as opposed to discretionary admissions, which decrease significantly at weekends.

For diseases evenly distributed throughout the week, the weekend admission rate is two sevenths (0.256). In our sample, patients with AMI have a weekend admission rate of 0.216, indicating that 21.6 % of patients are admitted on Saturdays or Sundays. In contrast, less than 1 % of senile cataract patients are admitted on weekends. For our primary analysis, we focus on diagnoses with at least 50 admissions in each of the four hospitals participating in the admission schedule, and then select the 50 three-digit ICD-10 diagnoses with the highest weekend admission rates. This results in a threshold of 0.213 for the weekend admission rate and includes conditions such as AMI, acute appendicitis, cerebral infarction, and pneumonia. Table A.1 in the Appendix lists the corresponding diagnoses. We also perform a sensitivity analysis using the top 75 diagnoses. In addition, injury-related diagnoses are excluded because only one of the included hospitals has a department specializing in the treatment of injured patients.

To examine whether and how the selection of diagnoses affects the empirical results, we present the results based on an alternative approach that selects diagnoses for which the admission schedule is statistically relevant. We run separate first-stage regressions for each 3-digit ICD-10 code and apply under-identification tests to assess whether the instrumental variables are correlated with endogenous regressors (see, e.g., Windmeijer (2024)). In this sensitivity check, we focus on the 219 diagnoses in our sample for which the Kleibergen-Paap Lagrange multiplier test was statistically significant at the 5 % level.[5] There is considerable overlap in diagnoses between the two selection approaches, as diagnoses for which the instrument is statistically relevant have higher weekend admission rates.

[5]See Appendix Table A.2 for the 30 most common diagnoses that pass this threshold.

4 Results

4.1 Hospital descriptives

Table 2 compares the patient characteristics in the four hospitals. With 54.4 to 117.7 inpatient admissions per day, the hospitals vary considerably in size.[6] There is also variation in patient characteristics, with the proportion of female patients ranging from 51 % to 56 %, and the average patient age from 60.5 to 63.6 years.

Table 2: Patient characteristics by hospital

	Hospital A	Hospital B	Hospital C	Hospital D	Total
Admissions per day	117.7	54.4	80.5	57.5	310.1
Female	0.51	0.53	0.53	0.56	0.53
Age	61.5	62.8	60.5	63.6	61.8
Patient health					
Charlson Comorbidity Index	0.80	0.56	0.55	0.93	0.72
Length of stay	5.30	4.37	5.15	5.49	5.13
Emergency transport patients	0.02	0.02	0.02	0.02	0.02
Socio-economic characteristics					
White-collar worker	0.56	0.56	0.58	0.56	0.57
Academic degree	0.03	0.03	0.03	0.03	0.03
Wage	85.40	82.91	86.29	84.36	85.00
N	429876	198543	293956	210119	1132494

Notes: This table shows admissions and patient characteristics in hospitals A, B, C, and D. N is the number of observations.

Differences are also observed in the health status of the patients. The Charlson Comorbidity Index ranges from 0.55 to 0.93, and the average length of hospital stay varies by more than one day. The correlations between patient characteristics are plausible. For example, hospital D has, on average, the oldest patients who have more comorbidities and stay the longest. The proportion of emergency transport patients is low and similar for all hospitals, and there are only small differences in the socioeconomic characteristics of the patients; the proportion of patients with (former) white-collar employment range from 56 % to 58 %. All hospitals have similar proportions of patients with academic degrees and similar wages.

4.2 First stage results

The graphical representation of the admission figures in Section 2.3 suggests that the admission schedule affects the admission patterns in the four participating hospitals.

[6]The total number of hospital beds is in a comparable range between 337 and 886 in 2015 (Oö. Gesundheitsfonds, 2016).

First-stage regressions formally test this relationship for selected diagnoses with high weekend admission rates.

Table 3: First stage results

	Hospital B	Hospital C	Hospital D
c	0.147***	0.419***	0.038***
	(0.003)	(0.004)	(0.003)
bd	0.494***	−0.023***	0.109***
	(0.004)	(0.003)	(0.003)
c × female	−0.029***	0.065***	0.018***
	(0.004)	(0.006)	(0.004)
bd × female	−0.344***	0.009*	0.386***
	(0.005)	(0.004)	(0.005)
N	122,286	122,286	122,286
Mean of dpt.	0.209	0.241	0.207
Partial R^2	0.188	0.265	0.175
Sanderson-Windmeijer F	5951.358	17132.041	6366.754
Cragg-Donald F: 2939.9			
Kleibergen-Paap F: 2267.2			

Notes: The regressions also control for patient sex, age, and principal diagnosis at the 3-digit level. The bottom panel of the table shows the number of observations, the mean of the dependent variable and tests for weak identification, including the Sanderson-Windmeijer F statistic, the Cragg-Donald F statistic, and the Kleibergen-Paap Wald rk F statistic (Sanderson and Windmeijer, 2016; Cragg and Donald, 1993; Kleibergen and Paap, 2006). Robust standard errors in parentheses. * $p < 0.05$. ** $p < 0.01$, *** $p < 0.001$.

Table 3 shows the estimation results for Equations 2–4, highlighting that admission days have a strong impact on the probability of being admitted to a particular hospital. For example, if hospital C is primarily responsible for inpatient admissions (on *c* days), the probability of being admitted to this hospital increases by 42 (men) and 48 (women) percentage points (pp) compared with when hospital A is responsible. This is a large effect given that the average probability of being admitted to this hospital is 24 %.

Similar effects can be observed in other hospitals, although the effects vary much more by sex. On a *bd* day, the increase in the probability of admission is much greater for men (49 pp) than for women (15 pp) in hospital B, and the opposite is true in hospital D. This result follows from the design of the schedule, which separates women from men on this day.

The results also confirm a considerable amount of non-compliance, in which the coefficients for admission days are not equal to 1. For the reasons discussed in Section 2.3, there are "deviations" from the admission schedule. A few patients are admitted to any hospital regardless of the current admission day. However, a partial R^2 between 0.18 and 0.27 for our instrumental variables indicates that the schedule explains a considerable amount of variation in hospital admissions. Instruments are statistically significant. The joint F-tests for the instruments in the equations are between 5,951 and 17,132. The Kleibergen-Paap Wald rk F, which tests for weak identification of the three endogenous variables, is 2,267.

4.3 Hospital effects

Table 4 shows the effect of hospital admission on the 30-day mortality. The results for the second stage of the instrumental variable model are shown in column 1, and the results for the different risk-adjustment models are in columns 2-5. Figure 4 illustrates the results.

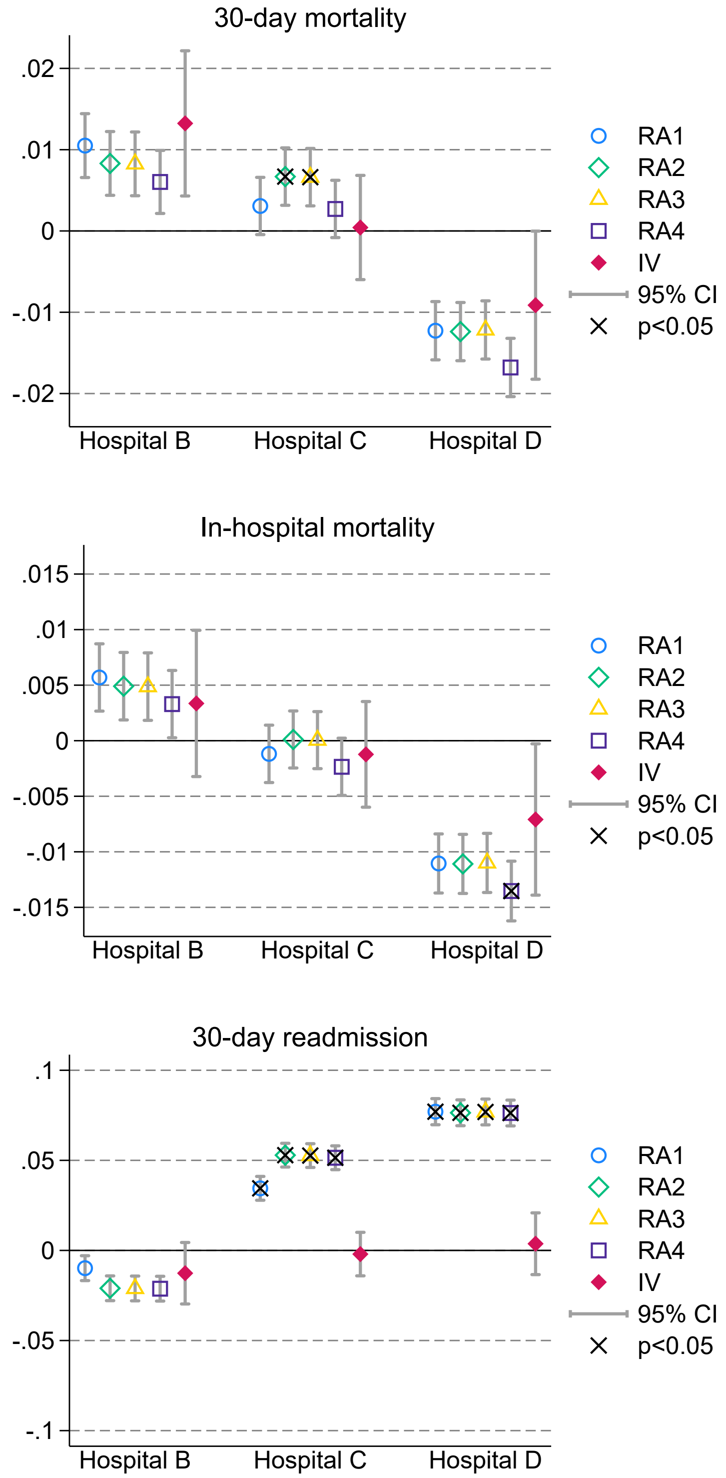
The IV estimates indicate a 1.3 pp higher 30-day mortality rate for admission to hospital B. With an average mortality rate of 5.6 %, this is 23 % higher than that of hospital A. The risk-adjustment estimates are consistent with these results and range between 0.6 and 1.1 pp. While the results of both the IV estimation and risk adjustment models suggest a lower 30-day mortality for hospital D, we find no statistically significant effect for hospital C according to the IV specification. Two risk-adjustment models (RA2 and RA3) show an increase in mortality, and the estimates are statistically different from the IV estimates.[7]

The effects of in-hospital mortality in Table 5 and the middle panel of Figure 4 are similar. All specifications suggest a higher mortality rate for hospital B and a lower mortality rate for hospital D. However, the IV effect for hospital B is not statistically significant, and that for hospital D is quantitatively smaller than the risk-adjusted estimates.

The greatest discrepancy between the two estimation methods is found for hospital readmissions. While the IV results suggest that the probability of readmission is similar across hospitals, risk adjustment shows significant differences. For hospitals C and D, the IV estimates are sufficiently precise to rule out the large effects of risk adjustment. Interestingly, increasing the number of covariates from RA1 to RA4 to better control for patient differences does not help bring risk adjustment in line with the IV estimates. Although the point estimates and their significance levels for mortality rates are similar for the IV- and risk-adjusted specifications, the two estimation methods yield significantly different results for readmission rates.

[7] We test the difference by "stacking" the regressions and estimating OLS and 2SLS simultaneously, as described in Hull (2022).

Figure 4: Effects of admission hospital on patient outcomes



Notes: This figure shows the effects of the admission hospital on 30-day mortality (upper panel), in-hospital mortality (middle panel) and 30-day readmission (bottom panel). The full estimation output is included in Tables 4 to 6.

Table 4: Effects of admission hospital on 30-day mortality

	(1) IV	(2) RA1	(3) RA2	(4) RA3	(5) RA4
Hospital B	0.013*	0.011*	0.008*	0.008*	0.006*
	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)
Hospital C	0.000	0.003	0.007*°	0.007*°	0.003
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Hospital D	−0.009*	−0.012*	−0.012*	−0.012*	−0.017*
	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)
N	122,286	122,286	122,286	122,286	122,286
Mean of dept.	0.062	0.062	0.062	0.062	0.062
Mean of hospital A	0.056	0.056	0.056	0.056	0.056

Notes: All regressions control for primary disease, year of admission, patient sex, and age. RA2-RA4 progressively add additional control variables. RA2 adds indicators for comorbidities based on the Charlson Comorbidity Index. RA3 adds indicators for wage group, education, and occupation. RA4 adds indicators for emergency transport and length of hospital stay. Robust standard errors in parentheses. * indicates that the estimate is statistically significant at the 5 % level. ° indicates that the estimate is statistically different from the corresponding IV estimate at the 5 % level.

4.4 Sensitivity analysis

We conduct two sensitivity analyses using different samples. First, departing from the baseline specification, we extend the included diagnoses to the top 75 three-digit ICD-10 codes with the highest weekend admission rates. Compared to the main sample, this changes the number of hospital cases in the analysis from 122,286 to 184,944. The results of the first stage in the Appendix Table A.3 confirm the large and significant effects of the admissions schedule on the admissions pattern for this sample.

Similar to the baseline results, Figure 5 shows the greatest disagreement between the estimation methods when hospital readmissions are considered. The risk-adjustment approach showes large and statistically significant effects of hospitals on the readmission rates. However, consistent with the baseline results, the IV coefficients for the top 75 diagnoses remains insignificant and the estimates are sufficiently precise to reject all but one (RA1) of the risk adjustment estimates. The IV and risk adjustment effects on 30-day and in-hospital mortality are quantitatively similar, and in most cases, do not differ significantly from each other.

The second sensitivity check, described in Section 3, uses a larger set of diagnoses for which the instruments are statistically relevant, following an under-identification test for correlations between the instruments and the endogenous regressors. This increased the number of hospital cases over the period analyzed to 486,794. The first-stage results in the Appendix Table A.7 confirm that the admissions schedule

Table 5: Effects of admission hospital on in-hospital mortality

	(1) IV	(2) RA1	(3) RA2	(4) RA3	(5) RA4
Hospital B	0.003	0.006*	0.005*	0.005*	0.003*
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Hospital C	−0.001	−0.001	0.000	0.000	−0.002
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Hospital D	−0.007*	−0.011*	−0.011*	−0.011*	−0.014*°
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
N	122,286	122,286	122,286	122,286	122,286
Mean of dept.	0.032	0.032	0.032	0.032	0.032
Mean of hospital A	0.030	0.030	0.030	0.030	0.030

Notes: All regressions control for primary disease, year of admission, patient sex, and age. RA2-RA4 progressively add additional control variables. RA2 adds indicators for comorbidities based on the Charlson Comorbidity Index. RA3 adds indicators for wage group, education, and occupation. RA4 adds indicators for emergency transport and length of hospital stay. Robust standard errors in parentheses. * indicates that the estimate is statistically significant at the 5 % level. ° indicates that the estimate is statistically different from the corresponding IV estimate at the 5 % level.

has a significant effect on hospital choice for this set of diagnoses.

Again, we observe substantial differences in the assessment of hospital quality between the two estimation approaches (Figure 6). In all but one case, the IV- and risk-adjusted estimates of the effect of hospitals on readmissions differ significantly. The results are also consistent with the baseline specification, although the IV estimator, similar to the different versions of risk adjustment, provides a significantly higher readmission rate for hospital C, whereas these effects remain insignificant for Hospitals B and D in the IV variant. As before, there is only a small difference between the IV and risk-adjusted mortality estimates for hospitals B and D. In contrast, the IV- and risk-adjusted mortality effects for hospital C differ significantly, both statistically and quantitatively. Overall, this sample exhibits the largest differences between the estimation methods.

5 Conclusion

In this study, we examine hospital performance by exploiting exogenous variations in hospital admissions to control for patient selection. In the capital of Upper Austria, Linz, an admission schedule regulates which hospital is responsible for admitting acute cases on that day. Using an instrumental variables approach and high-quality administrative data, we estimate hospital performance in terms of mortality and readmission rates, and compare the results with traditional risk adjustment approaches to control for patient selection.

Table 6: Effects of admission hospital on 30-day readmission

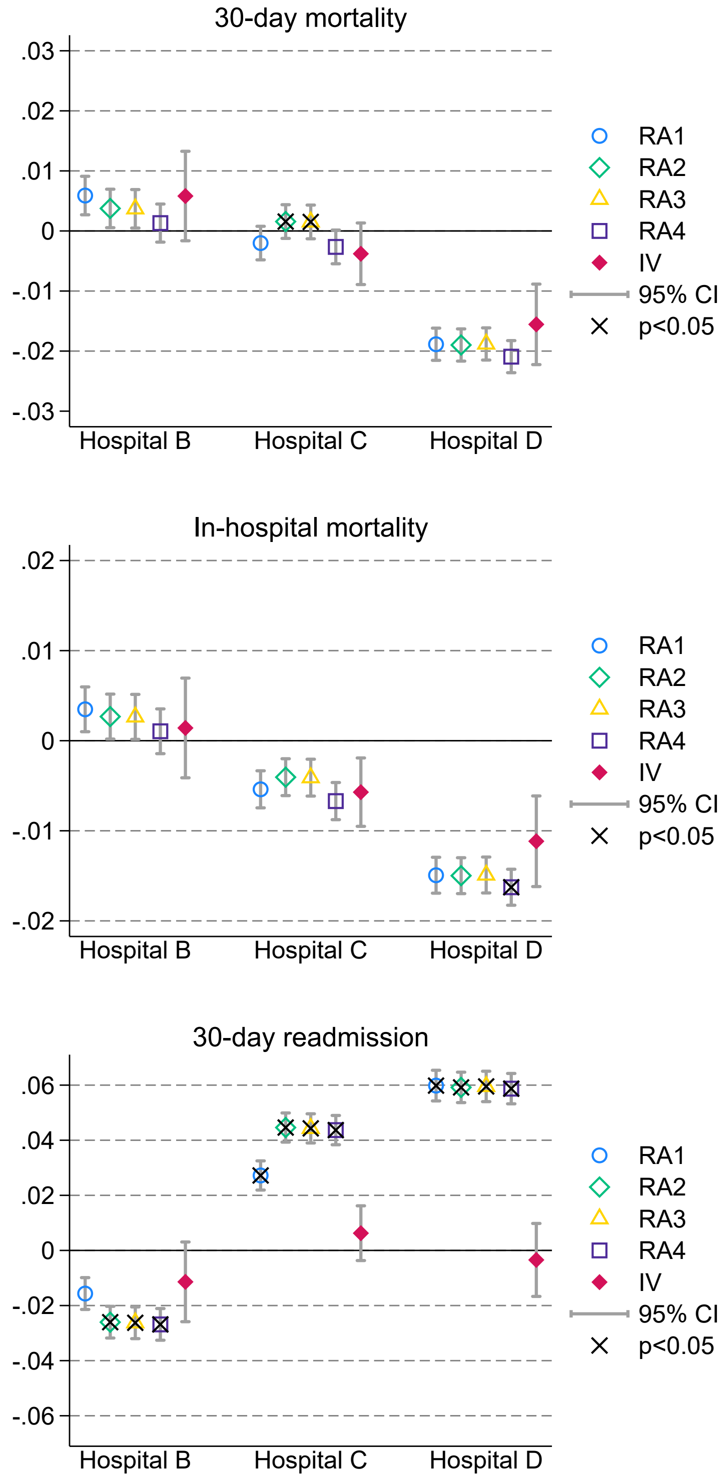
	(1) IV	(2) RA1	(3) RA2	(4) RA3	(5) RA4
Hospital B	−0.013	−0.010*	−0.021*	−0.021*	−0.021*
	(0.009)	(0.004)	(0.004)	(0.004)	(0.004)
Hospital C	−0.002	0.034*°	0.053*°	0.053*°	0.051*°
	(0.006)	(0.003)	(0.003)	(0.003)	(0.003)
Hospital D	0.004	0.077*°	0.076*°	0.077*°	0.076*°
	(0.009)	(0.004)	(0.004)	(0.004)	(0.004)
N	121,260	121,260	121,260	121,260	121,260
Mean of dept.	0.26	0.26	0.26	0.26	0.26
Mean of hospital A	0.23	0.23	0.23	0.23	0.23

Notes: All regressions control for primary disease, year of admission, patient sex, and age. RA2-RA4 progressively add additional control variables. RA2 adds indicators for comorbidities based on the Charlson Comorbidity Index. RA3 adds indicators for wage group, education, and occupation. RA4 adds indicators for emergency transport and length of hospital stay. Robust standard errors in parentheses. * indicates that the estimate is statistically significant at the 5 % level. ° indicates that the estimate is statistically different from the corresponding IV estimate at the 5 % level.

We find significant differences in the estimation of hospital performance depending on whether we use the IV approach or classical models of risk selection based on observable patient characteristics. Consistent with existing literature, we show that the assessment of hospital performance is sensitive to patient characteristics. However, increasing the number of patient characteristics in the risk adjustment does not always lead to the results of these models converging with the IV estimators. Overall, our results suggest that risk adjustment with observable characteristics from administrative data does not adequately control for differences between patients and their diseases. Therefore, hospital quality indicators derived from administrative data should be considered carefully.

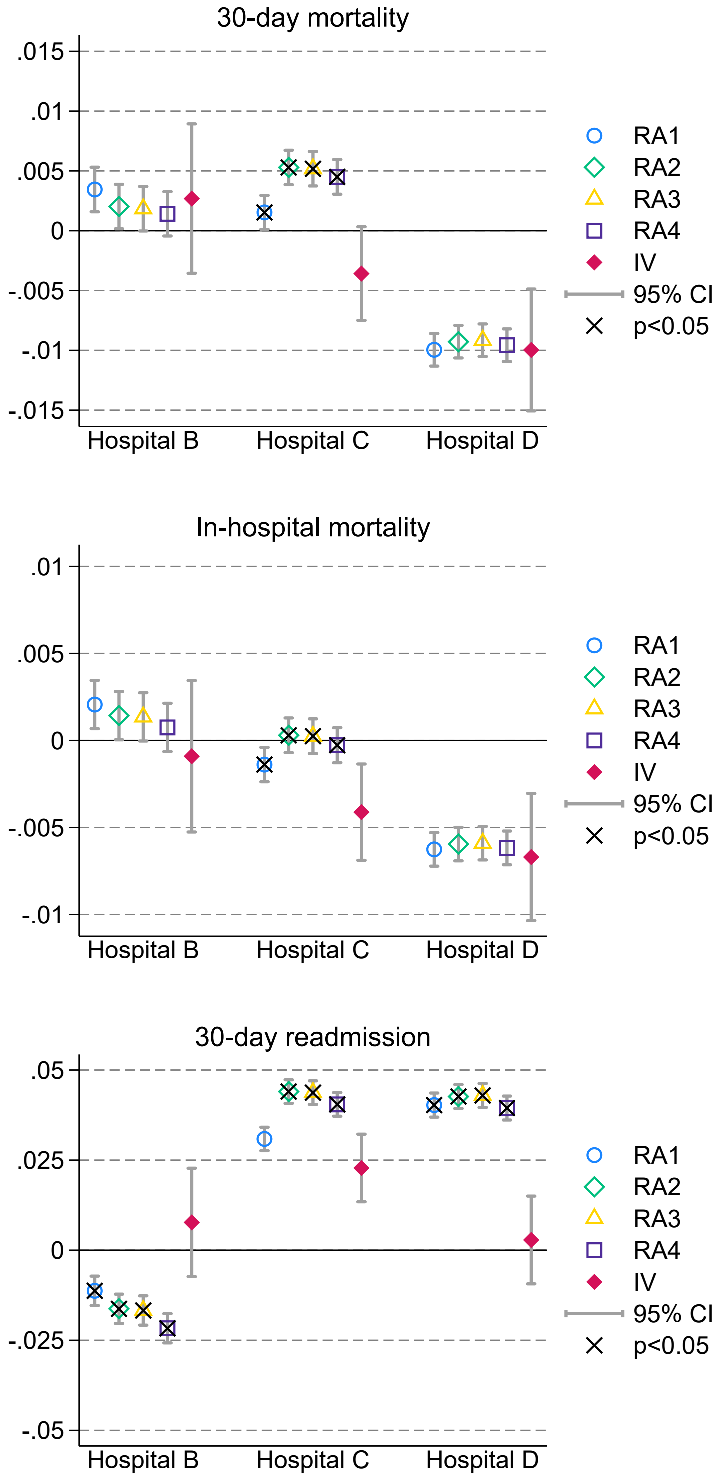
Our results support further development of process-oriented hospital indicators. If socioeconomic characteristics such as age and sex, as well as diagnoses, are not sufficient to adequately control the severity of illness, the collection of additional personal data seems unavoidable. Information collected during the treatment that differs between patients with identical diagnoses is important for improved risk adjustment in the context of hospital evaluations. Examples of personal background information include the individual medical services provided, complications that occurred, combination of medications administered, and the time between the onset of a medical problem and the necessary treatment. Personalized medicine is increasingly supporting the process of collecting information at an individual level and, by combining all available measures, should enable better quality comparisons between hospitals.

Figure 5: Effects of admission hospital – top 75 diagnoses



Notes: This figure shows the effects of the admission hospital on 30-day mortality (upper panel), in-hospital mortality (middle panel) and 30-day readmission (bottom panel) based on a sample of the 75 diagnoses with the highest weekend admission rates. The full estimation output is included in Tables A.4 to A.6.

Figure 6: Effects of admission hospital – relevant diagnoses



Notes: This figure shows the effects of the admission hospital on 30-day mortality (upper panel), in-hospital mortality (middle panel), and 30-day readmission (bottom panel) based on the sample including all statistically relevant diagnoses. The full estimation output is included in Tables A.8 to A.10.

References

Austin, J Matthew, Ashish K Jha, Patrick S Romano, Sara J Singer, Timothy J Vogus, Robert M Wachter and Peter J Pronovost (2015), 'National hospital ratings systems share few common scores and may generate confusion instead of clarity', *Health Affairs* **34**(3), 423–430.

Bachner, Florian, Julia Bobek, Katharina Habimana, Joy Ladurner, Lena Lepuschutz, Herwig Ostermann, Lukas Rainer, Andrea E Schmidt, Martin Zuba, Wilm Quentin and Julia Winkelmann (2018), Austria: Health System Review 2018, Vol. 20, European Observatory on Health Systems and Policies, pp. 1–254.

Baker, David W and Mark R Chassin (2017), 'Holding providers accountable for health care outcomes', *Annals of Internal Medicine* **167**(6), 418–423.

Berry, Jay G and Alyna T Chien (2016), 'To risk adjust or not to risk adjust: should that be the question?', *JAMA Pediatrics* **170**(4), 319–320.

Black, Nick (2016), 'Higher mortality in weekend admissions to the hospital: true, false, or uncertain?', *JAMA* **316**(24), 2593–2594.

Braithwaite, R Scott (2018), 'Risk adjustment for quality measures is neither binary nor mandatory', *JAMA* **319**(20), 2077–2078.

Cacace, Mirella, Max Geraedts and Elke Berger (2019), Public reporting as a quality strategy, *in* R.Busse, N.Klazinga, D.Panteli and W.Quentin, eds, 'Improving healthcare quality in Europe: Characteristics, effectiveness and implementation of different strategies', Health Policy Series: 53, World Health Organization. Regional Office for Europe.

Chandra, Amitabh, Pragya Kakani and Adam Sacarny (2022), 'Hospital allocation and racial disparities in health care', *Review of Economics and Statistics* pp. 1–39.

Charlson, Mary E, D Pompei, KE Ales and CR MacKenzie (1987), 'A new method of classifying prognostic comorbidity in longitudinal studies: development and validation', *Journal of Chronic Diseases* **40**(5), 373–83.

Charlson, Mary E, Danilo Carrozzinoc, Jenny Guidi and Chiara Patierno (2022), 'Charlson comorbidity index: a critical review of clinimetric properties', *Psychotherapy and Psychosomatics* **91**(1), 8–35.

CIHI (2007), *HSMR: A new approach for measuring hospital mortality trends in Canada*, Canadian Institute for Health Information, Ottawa, Ontario.

CIHI (2016), *Hospital Standardized Mortality Ratio - Technical Notes November 2016*, Canadian Institute for Health Information, Ottawa, Ontario.

Cook, Amanda and Susan Averett (2020), 'Do hospitals respond to changing incentive structures? evidence from medicare's 2007 drg restructuring', *Journal of Health Economics* **73**, 102319.

Cragg, John G and Stephen G Donald (1993), 'Testing identifiability and specification in instrumental variable models', *Econometric Theory* **9**(2), 222–240.

Di Giacomo, Marina, Massimiliano Piacenza, Luigi Siciliani and Gilberto Turati (2017), 'Do public hospitals respond to changes in drg price regulation? the case of birth deliveries in the i talian nhs', *Health Economics* **26**, 23–37.

Doyle, Joseph, John Graves and Jonathan Gruber (2019), 'Evaluating measures of hospital quality: Evidence from ambulance referral patterns', *Review of Economics and Statistics* **101**(5), 841–852.

Doyle, Joseph, John Graves, Jonathan Gruber and Samuel Kleiner (2015), 'Measuring returns to hospital care: Evidence from ambulance referral patterns', *Journal of Political Economy* **123**(1), 170–214.

Fadlon, Itzik and Jessica Van Parys (2020), 'Primary care physician practice styles and patient care: Evidence from physician exits in medicare', *Journal of Health Economics* **71**, 102304.

Finkelstein, Amy, Matthew Gentzkow, Peter Hull and Heidi Williams (2017), 'Adjusting risk adjustment–accounting for variation in diagnostic intensity', *The New England Journal of Medicine* **376**(7), 608.

Garthwaite, Craig, Christopher Ody and Amanda Starc (2022), 'Endogenous quality investments in the us hospital market', *Journal of Health Economics* **84**, 102636.

Goodacre, Steve, Mike Campbell and Angela Carter (2015), 'What do hospital mortality rates tell us about quality of care?', *Emergency Medicine Journal* **32**(3), 244–247.

Gowrisankaran, Gautam, Keith Joiner and Pierre Thomas Léger (2023), 'Physician practice style and healthcare costs: Evidence from emergency departments', *Management Science* **69**(6), 3202–3219.

Gupta, Atul (2021), ‘Impacts of performance pay for hospitals: The readmissions reduction program’, *American Economic Review* **111**(4), 1241–1283.

Heijink, Richard, Xander Koolman, Daniel Pieter, André van der Veen, Brian Jarman and Gert Westert (2008), ‘Measuring and explaining mortality in dutch hospitals; the hospital standardized mortality rate between 2003 and 2005’, *BMC Health Services Research* **8**(1), 73.

Hosein, Shaun F, Derek J Roberts, Tanvir Chowdhury Turin, David Zygun, William A Ghali and Henry T Stelfox (2014), ‘A meta-analysis to derive literature-based benchmarks for readmission and hospital mortality after patient discharge from intensive care’, *Critical Care* **18**, 1364–8535.

Hull, Peter (2022), “‘stacking” regressions’, Twitter Thread, link available at https://about.peterhull.net/metrix.

Jarman, B, D Pieter, AA Van Der Veen, RB Kool, P Aylin, A Bottle, GP Westert and S Jones (2010), ‘The hospital standardised mortality ratio: a powerful tool for dutch hospitals to assess their quality of care?’, *BMJ Quality & Safety* **19**(1), 9–13.

Jarman, Brian, Simon Gault, Bernadette Alves, Amy Hider, Susan Dolan, Adrian Cook, Brian Hurwitz and Lisa I Iezzoni (1999), ‘Explaining differences in english hospital death rates using routinely collected data’, *BMJ* **318**(7197), 1515–1520.

Jürges, Hendrik and Juliane Köberlein (2015), ‘What explains drg upcoding in neonatology? the roles of financial incentives and infant health’, *Journal of Health Economics* **43**, 13–26.

Kleibergen, Frank and Richard Paap (2006), ‘Generalized reduced rank tests using the singular value decomposition’, *Journal of Econometrics* **133**(1), 97–126.

Kozar, Rosemary A, John B Holcomb, Wei Xiong and Avery B Nathens (2014), ‘Are all deaths recorded equally? the impact of hospice care on risk-adjusted mortality’, *Journal of Trauma and Acute Care Surgery* **76**(3), 634–641.

Kristensen, Søren Rud, Rachel Meacock, Alex J Turner, Ruth Boaden, Ruth McDonald, Martin Roland and Matthew Sutton (2014), ‘Long-term effect of hospital pay for performance on mortality in england’, *New England Journal of Medicine* **371**(6), 540–548.

Laudicella, Mauro, Stephen Martin, Paolo Li Donni and Peter C Smith (2018), ‘Do reduced hospital mortality rates lead to increased utilization of inpatient emergency care? a population-based cohort study’, *Health Services Research* **53**(4), 2324–2345.

Lilford, Richard and Peter Pronovost (2010), ‘Using hospital mortality rates to judge hospital performance: A bad idea that just won’t go away’, *BMJ* **340**, c2016.

Liu, Yu-Chi, Mark Wilkins, Terry Kim, Boris Malyugin and Jodhbir S Mehta (2017), ‘Cataracts’, *The Lancet* **390**(10094), 600–612.

Miyata, Hiroaki, Hideki Hashimoto, Hiromasa Horiguchi, Shinya Matsuda, Noboru Motomura and Shinichi Takamoto (2008), ‘Performance of in-hospital mortality prediction models for acute hospitalization: hospital standardized mortality ratio in japan’, *BMC Health Services Research* **8**(1), 229.

Mohammed, Mohammed A, Jonathan J Deeks, Alan Girling, Gavin Rudge, Martin Carmalt, Andrew J Stevens and Richard J Lilford (2009), ‘Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of english hospitals’, *BMJ* **338**, b780.

OECD (2023), *Health at a Glance 2023: OECD Indicators*, OECD Publishing, Paris.

Ordensklinikum Linz (2019), ‘Aufnahmetage’, `https://www.ordensklinikum.at/de/patienten/aufnahmetage/`. [accessed 2019-September-11].

Oö. Gesundheitsfonds (2016), *Geschäftsbericht 2015 - Tätigkeit und Gebarung*, Oö. Gesundheitsfonds, Linz, Austria.

Pouw, Maurice E, LM Peelen, KGM Moons, CJ Kalkman and HF Lingsma (2013), ‘Including post-discharge mortality in calculation of hospital standardised mortality ratios: retrospective analysis of hospital episode statistics’, *BMJ* **347**.

Pross, Christoph, Alexander Geissler and Reinhard Busse (2017), ‘Measuring, reporting, and rewarding quality of care in 5 nations: 5 policy levers to enhance hospital quality accountability’, *The Milbank Quarterly* **95**(1), 136–183.

Quan, Hude, Vijaya Sundararajan, Patricia Halfon, Andrew Fong, Bernard Burnand, Jean-Christophe Luthi, L Duncan Saunders, Cynthia A Beck, Thomas E Feasby and William A Ghali (2005), ‘Coding algorithms for defining comorbidities in icd-9-cm and icd-10 administrative data’, *Medical Care* **43**(11), 1130–1139.

Ridgeway, Greg, Mette Nørgaard, Thomas Bøjer Rasmussen, William D Finkle, Lars Pedersen, Hans Erik Bøtker and Henrik Toft Sørensen (2019), 'Benchmarking danish hospitals on mortality and readmission rates after cardiovascular admission', *Clinical Epidemiology* **11**, 67–80.

Roberts, Eric T, Alan M Zaslavsky, Michael L Barnett, Bruce E Landon, Lin Ding and Michael J McWilliams (2018), 'Assessment of the effect of adjustment for patient characteristics on hospital readmission rates: Implications for pay for performance', *JAMA International Medicine* **178**(11), 1498–1507.

Sanderson, Eleanor and Frank Windmeijer (2016), 'A weak instrument f-test in linear iv models with multiple endogenous variables', *Journal of Econometrics* **190**(2), 212–221.

Shahian, David M, Robert E Wolf, Lisa I Iezzoni, Leslie Kirle and Sharon-Lise T Normand (2010), 'Variability in the measurement of hospital-wide mortality rates', *New England Journal of Medicine* **363**(26), 2530–2539.

Song, Yunjie, Jonathan Skinner, Julie Bynum, Jason Sutherland, John E Wennberg and Elliott S Fisher (2010), 'Regional variations in diagnostic practices', *New England Journal of Medicine* **363**(1), 45–53.

Welch, H Gilbert, Sandra M Sharp, Dan J Gottlieb, Jonathan S Skinner and John E Wennberg (2011), 'Geographic variation in diagnosis frequency and risk of death among medicare beneficiaries', *JAMA* **305**(11), 1113–1118.

Windmeijer, Frank (2024), 'Testing underidentification in linear models, with applications to dynamic panel and asset pricing models', *Journal of Econometrics* **240**(2), 105104.
URL: *https://www.sciencedirect.com/science/article/pii/S030440762100097X*

Wörndle, Michael, Silvia Jordan and Albrecht Becker (2023), 'Quality measurement in the Austrian health care system', https://www.uninetz.at/en/contributions/. UniNEtZ, Allianz Nachhaltige Universitäten in Österreich [accessed 2024-February-26].

Zweimüller, Josef, Rudolf Winter-Ebmer, Rafael Lalive, Andreas Kuhn, Jean-Philippe Wuellrich, Oliver Ruf and Simon Büchi (2009), 'Austrian social security database', The Austrian Center for Labor Economics and the Analysis of the Welfare State, Johannes Kepler University Linz, Austria.

A Appendix

Table A.1: Weekend admission rate of diseases

Rank	ICD	Description	N	W. rate	Excl.
1	F10	Mental and behavioural disorders due to use of alcohol	1827	0.332	
2	E16	Other disorders of pancreatic internal secretion	475	0.301	
3	J03	Acute tonsillitis	1607	0.296	
4	J11	Influenza, virus not identified	518	0.276	
5	J10	Influenza due to other identified influenza virus	538	0.264	
6	R04	Haemorrhage from respiratory passages	2122	0.262	
7	J06	Acute upper respiratory infections of multiple and unspecified sites	738	0.262	
8	T50	Poisoning by diuretics and other and unspecified drugs/substances	1029	0.259	x
9	J81	Pulmonary oedema	361	0.258	
10	R40	Somnolence, stupor and coma	436	0.257	
11	J69	Pneumonitis due to solids and liquids	1463	0.248	
12	R50	Fever of other and unknown origin	1120	0.246	
13	J96	Respiratory failure, not elsewhere classified	944	0.246	
14	S32	Fracture of lumbar spine and pelvis	1853	0.243	x
15	J15	Bacterial pneumonia, not elsewhere classified	2592	0.242	
16	K35	Acute appendicitis	2520	0.240	
17	A09	Other gastroenteritis and colitis	4981	0.240	
18	B99	Other and unspecified infectious diseases	3099	0.239	
19	H81	Disorders of vestibular function	5280	0.239	
20	K56	Paralytic ileus and intestinal obstruction without hernia	3635	0.237	
21	F41	Other anxiety disorders	588	0.235	
22	S72	Fracture of femur	3384	0.234	x
23	A08	Viral and other specified intestinal infections	607	0.234	
24	G40	Epilepsy	3553	0.234	
25	R11	Nausea and vomiting	1002	0.233	
26	A41	Other sepsis	2388	0.232	
27	I63	Cerebral infarction	7272	0.231	
28	J45	Asthma	1308	0.231	
29	K52	Other noninfective gastroenteritis and colitis	4754	0.229	
30	I46	Cardiac arrest	464	0.228	
31	R55	Syncope and collapse	5503	0.227	
32	E86	Volume depletion	1419	0.226	
33	I95	Hypotension	1344	0.226	
34	K59	Other functional intestinal disorders	2544	0.226	
35	J18	Pneumonia, organism unspecified	15770	0.223	
36	K85	Acute pancreatitis	2203	0.223	
37	R73	Elevated blood glucose level	595	0.222	
38	G45	Transient cerebral ischaemic attacks and related syndromes	3030	0.219	
39	A04	Other bacterial intestinal infections	1063	0.218	
40	K92	Other diseases of digestive system	2726	0.218	
41	A46	Erysipelas	4062	0.217	
42	I61	Intracerebral haemorrhage	1080	0.217	
43	I21	Acute myocardial infarction	8507	0.216	
44	J40	Bronchitis, not specified as acute or chronic	1285	0.216	
45	J44	Other chronic obstructive pulmonary disease	10616	0.216	
46	J20	Acute bronchitis	1180	0.216	
47	N12	Tubulo-interstitial nephritis, not specified as acute or chronic	380	0.216	
48	T63	Toxic effect of contact with venomous animals	436	0.216	x
49	E87	Other disorders of fluid, electrolyte and acid-base balance	2336	0.215	
50	R57	Shock, not elsewhere classified	451	0.213	

Notes: This table lists the 50 diagnoses with the highest weekend admission rate. It shows the rank, the 3-digit ICD-10 code and description, the number of observations and the weekend admission rate. We also indicate which diagnoses are excluded from the main analysis because they are injury-related.

Table A.2: Hospital diagnoses for which the admission schedule is relevant

Code	Description	Count
I25	Chronic ischaemic heart disease	29676
C34	Malignant neoplasm of bronchus and lung	22503
J18	Pneumonia, organism unspecified	15771
I48	Atrial fibrillation and flutter	13723
I50	Heart failure	11937
M54	Dorsalgia	11725
N39	Other disorders of urinary system	11146
J44	Other chronic obstructive pulmonary disease	10616
I10	Essential (primary) hypertension	10127
N20	Calculus of kidney and ureter	9388
C78	Secondary malignant neoplasm of respiratory and digestive organs	9139
K80	Cholelithiasis	8833
I21	Acute myocardial infarction	8507
C25	Malignant neoplasm of pancreas	8376
K40	Inguinal hernia	7974
I63	Cerebral infarction	7272
N13	Obstructive and reflux uropathy	6840
I70	Atherosclerosis	6664
K29	Gastritis and duodenitis	6354
K57	Diverticular disease of intestine	6000
R55	Syncope and collapse	5503
E11	Type 2 diabetes mellitus	5390
H81	Disorders of vestibular function	5280
A09	Other gastroenteritis and colitis of infectious and unspecified origin	4982
N18	Chronic kidney disease	4907
K52	Other noninfective gastroenteritis and colitis	4754
R07	Pain in throat and chest	4732
I35	Nonrheumatic aortic valve disorders	4707
T81	Complications of procedures, not elsewhere classified	4235
R10	Abdominal and pelvic pain	4216

Notes: This table lists the 30 most common ICD-3 diagnoses for which the first stage underidentification test is statistically significant at the 5 % level.

Table A.3: First stage results - top 75 diagnoses

	Hospital B	Hospital C	Hospital D
c	0.119***	0.415***	0.030***
	(0.003)	(0.003)	(0.003)
bd	0.475***	−0.032***	0.103***
	(0.003)	(0.002)	(0.003)
c × female	−0.037***	0.053***	0.029***
	(0.003)	(0.005)	(0.004)
bd × female	−0.358***	0.006	0.395***
	(0.004)	(0.003)	(0.004)
N	184,944	184,944	184,944
Mean of dpt.	0.184	0.242	0.225
Partial R^2	0.177	0.257	0.170
Sanderson-Windmeijer F	9123.721	26124.308	10458.409
Cragg-Donald F: 4601.2			
Kleibergen-Paap F: 3557.7			

Notes: The regressions also control for patient sex, age and principal diagnosis at the 3-digit level. The bottom panel of the table shows the number of observations, the mean of the dependent variable and tests for weak identification, including the Sanderson-Windmeijer F statistic, the Cragg-Donald F statistic and the Kleibergen-Paap Wald rk F statistic. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.4: Effects of admission hospital on 30-day mortality - top 75 diagnoses

	(1) IV	(2) RA1	(3) RA2	(4) RA3	(5) RA4
Hospital B	0.006	0.006*	0.004*	0.004*	0.001
	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)
Hospital C	−0.004	−0.002	0.002°	0.002°	−0.003
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Hospital D	−0.016*	−0.019*	−0.019*	−0.019*	−0.021*
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
N	184,944	184,944	184,944	184,944	184,944
Mean of dept.	0.057	0.057	0.057	0.057	0.057
Mean of hospital A	0.056	0.056	0.056	0.056	0.056

Notes: All regressions control for primary disease, year of admission, patient sex and age. RA2-RA4 progressively add additional control variables. RA2 adds indicators for comorbidities based on the Charlson Comorbidity Index. RA3 adds indicators for wage group, education, and occupation. RA4 adds indicators for emergency transport and length of hospital stay. Robust standard errors in parentheses. * indicates that the estimate is statistically significant at the 5 % level. ° indicates that the estimate is statistically different from the corresponding IV estimate at the 5 % level.

Table A.5: Effects of admission hospital on in-hospital mortality - top 75 diagnoses

	(1) IV	(2) RA1	(3) RA2	(4) RA3	(5) RA4
Hospital B	0.001	0.003*	0.003*	0.003*	0.001
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Hospital C	−0.006*	−0.005*	−0.004*	−0.004*	−0.007*
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Hospital D	−0.011*	−0.015*	−0.015*	−0.015*	−0.016*°
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
N	184,944	184,944	184,944	184,944	184,944
Mean of dept.	0.030	0.030	0.030	0.030	0.030
Mean of hospital A	0.031	0.031	0.031	0.031	0.031

Notes: All regressions control for primary disease, year of admission, patient sex and age. RA2-RA4 progressively add additional control variables. RA2 adds indicators for comorbidities based on the Charlson Comorbidity Index. RA3 adds indicators for wage group, education, and occupation. RA4 adds indicators for emergency transport and length of hospital stay. Robust standard errors in parentheses. * indicates that the estimate is statistically significant at the 5 % level. ° indicates that the estimate is statistically different from the corresponding IV estimate at the 5 % level.

Table A.6: Effects of admission hospital on 30-day readmission - top 75 diagnoses

	(1) IV	(2) RA1	(3) RA2	(4) RA3	(5) RA4
Hospital B	−0.011	−0.016*	−0.026*°	−0.026*°	−0.027*°
	(0.007)	(0.003)	(0.003)	(0.003)	(0.003)
Hospital C	0.006	0.027*°	0.045*°	0.044*°	0.044*°
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
Hospital D	−0.003	0.060*°	0.059*°	0.060*°	0.059*°
	(0.007)	(0.003)	(0.003)	(0.003)	(0.003)
N	183,419	183,419	183,419	183,419	183,419
Mean of dept.	0.25	0.25	0.25	0.25	0.25
Mean of hospital A	0.23	0.23	0.23	0.23	0.23

Notes: All regressions control for primary disease, year of admission, patient sex and age. RA2-RA4 progressively add additional control variables. RA2 adds indicators for comorbidities based on the Charlson Comorbidity Index. RA3 adds indicators for wage group, education, and occupation. RA4 adds indicators for emergency transport and length of hospital stay. Robust standard errors in parentheses. * indicates that the estimate is statistically significant at the 5 % level. ° indicates that the estimate is statistically different from the corresponding IV estimate at the 5 % level.

Table A.7: First stage results - relevant diagnoses sample

	Hospital B	Hospital C	Hospital D
c	0.058***	0.264***	0.014***
	(0.001)	(0.002)	(0.002)
bd	0.273***	−0.008***	0.069***
	(0.002)	(0.002)	(0.002)
c × female	−0.008***	0.060***	0.023***
	(0.002)	(0.003)	(0.003)
bd × female	−0.199***	−0.001	0.279***
	(0.002)	(0.002)	(0.003)
N	486,794	486,794	486,794
Mean of dpt.	0.126	0.248	0.244
Partial R^2	0.085	0.112	0.075
Sanderson-Windmeijer F	10196.711	26655.103	11608.097
Cragg-Donald F: 5245.4			
Kleibergen-Paap F: 4227.8			

Notes: The regressions also control for patient sex, age and principal diagnosis at the 3-digit level. The bottom panel of the table shows the number of observations, the mean of the dependent variable and tests for weak identification, including the Sanderson-Windmeijer F statistic, the Cragg-Donald F statistic and the Kleibergen-Paap Wald rk F statistic. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.8: Effects of admission hospital on 30-day mortality - relevant diagnoses

	(1) IV	(2) RA1	(3) RA2	(4) RA3	(5) RA4
Hospital B	0.003	0.003*	0.002*	0.002	0.001
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
Hospital C	−0.004	0.002*°	0.005*°	0.005*°	0.005*°
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Hospital D	−0.010*	−0.010*	−0.009*	−0.009*	−0.010*
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
N	486,794	486,794	486,794	486,794	486,794
Mean of dept.	0.036	0.036	0.036	0.036	0.036
Mean of hospital A	0.033	0.033	0.033	0.033	0.033

Notes: All regressions control for primary disease, year of admission, patient sex and age. RA2-RA4 progressively add additional control variables. RA2 adds indicators for comorbidities based on the Charlson Comorbidity Index. RA3 adds indicators for wage group, education, and occupation. RA4 adds indicators for emergency transport and length of hospital stay. Robust standard errors in parentheses. * indicates that the estimate is statistically significant at the 5 % level. ° indicates that the estimate is statistically different from the corresponding IV estimate at the 5 % level.

Table A.9: Effects of admission hospital on in-hospital mortality - relevant diagnoses

	(1) IV	(2) RA1	(3) RA2	(4) RA3	(5) RA4
Hospital B	−0.001	0.002*	0.001*	0.001	0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Hospital C	−0.004*	−0.001*°	0.000°	0.000°	−0.000°
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hospital D	−0.007*	−0.006*	−0.006*	−0.006*	−0.006*
	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
N	486,794	486,794	486,794	486,794	486,794
Mean of dept.	0.017	0.017	0.017	0.017	0.017
Mean of hospital A	0.017	0.017	0.017	0.017	0.017

Notes: All regressions control for primary disease, year of admission, patient sex and age. RA2-RA4 progressively add additional control variables. RA2 adds indicators for comorbidities based on the Charlson Comorbidity Index. RA3 adds indicators for wage group, education, and occupation. RA4 adds indicators for emergency transport and length of hospital stay. Robust standard errors in parentheses. * indicates that the estimate is statistically significant at the 5 % level. ° indicates that the estimate is statistically different from the corresponding IV estimate at the 5 % level.

Table A.10: Effects of admission hospital on 30-day readmission - relevant diagnoses

	(1) IV	(2) RA1	(3) RA2	(4) RA3	(5) RA4
Hospital B	0.008	−0.011*°	−0.016*°	−0.017*°	−0.022*°
	(0.008)	(0.002)	(0.002)	(0.002)	(0.002)
Hospital C	0.023*	0.031*	0.044*°	0.044*°	0.040*°
	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)
Hospital D	0.003	0.040*°	0.043*°	0.043*°	0.039*°
	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)
N	483,179	483,179	483,179	483,179	483,179
Mean of dept.	0.30	0.30	0.30	0.30	0.30
Mean of hospital A	0.28	0.28	0.28	0.28	0.28

Notes: All regressions control for primary disease, year of admission, patient sex and age. RA2-RA4 progressively add additional control variables. RA2 adds indicators for comorbidities based on the Charlson Comorbidity Index. RA3 adds indicators for wage group, education, and occupation. RA4 adds indicators for emergency transport and length of hospital stay. Robust standard errors in parentheses. * indicates that the estimate is statistically significant at the 5 % level. ° indicates that the estimate is statistically different from the corresponding IV estimate at the 5 % level.