

WP/19/173

IMF Working Paper

The Economic Impact of Healthcare Quality

by Anne-Line Koch Helsø, Nicola Pierri, and Adelina Yanyue Wang

***IMF Working Papers* describe research in progress by the author(s) and are published to elicit comments and to encourage debate.** The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

I N T E R N A T I O N A L M O N E T A R Y F U N D

IMF Working Paper

Research Department

The Economic Impact of Healthcare Quality

Prepared by Anne-Line Koch Helsø, Nicola Pierri, and Adelina Yanyue Wang*

Authorized for distribution by Maria Soledad Martinez-Peria

August 2019

IMF Working Papers describe research in progress by the author(s) and are published to elicit comments and to encourage debate. The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

Abstract

We study the costs of hospitalizations on patients' earnings and labor supply, using the universe of hospital admissions in Denmark and full-population tax data. We evaluate the quality of treatment based on its ability to mitigate the labor market consequences of a given diagnosis and propose a new measure of hospital quality, the "Adjusted Earning Losses" (AEL). We find a 4 percentage points difference in lost earnings between the best and worst large Danish hospitals, all else equal. We show that AEL contains significant additional information relative to traditional measures and does not suffer from worse selection issues. We also document a large decline in the labor cost of hospitalizations over time, with large variations across diseases. We find that the average post-hospitalization reduction in labor earnings declined by 25 percent (50 percent) on the intensive (extensive) margin between 1998 and 2012.

JEL Classification Numbers: I1, D24

Keywords: Hospital quality, healthcare quality, health shocks, labor supply, hospital productivity

Author's E-Mail Address: anne.line.koch.helso@econ.ku.dk; npierri@imf.org; adywang@stanford.edu

* We thank Tore Olsen for essential advice about data sources in the early stages of this project. We thank, for their insightful comments, Nick Bloom, Tim Bresnahan, Raj Chetty, Giovanni Dell'Ariceia, Mark Duggan, Liran Einav, Matt Gentzkow, Atul Gupta, Brady Horn, Caroline Hoxby, Thomas H. Jørgensen, Jean-William Laliberté, Maria Soledad Martinez Peria, Grant Miller, Torben Heien Nielsen, Petra Persson, Luigi Pistaferri, Bertel Schjerning, Pietro Tebaldi, Philippe Wingender, and seminar participants at Stanford University, University of Calgary, IMF, and 2019 ASHEcon conference. Adelina Yanyue Wang gratefully acknowledges financial support from the Integrated Benefits Institute (IBI) through a Thomas Parry Research Fellowship award. All errors remain our sole responsibility. The views expressed by the authors do not necessarily reflect those of the International Monetary Fund (IMF), IMF Executive Board, IMF management, IBI, or any other Institution.

1 Introduction

Adverse health shocks are extremely costly, both to the individual and to society at large. The average OECD spending on health as a share of GDP was 8.9% in 2016, and a staggering 17.2% in the United States. Direct medical costs, however, only make up a fraction of the total costs of morbidities: health impacts economic growth [Well, 2007], poor health reduces the capacity to work [Currie and Madrian, 1999], hospital admissions are followed by severe and long-lasting worsening of economic conditions [Dobkin et al., 2018], and earnings losses might comprise the majority of total life-cycle costs from bad health experienced by working-age individuals [De Nardi et al., 2017].

Not surprisingly, there is considerable interest in measuring and improving the quality and efficiency of healthcare. However, the extensive literature that measures and quantifies the variation in treatment quality is based on medical outcomes, such as mortality and readmission rates, rather than the expensive labor market costs associated with adverse health shocks.

We propose a novel approach to measure the quality of hospital treatment based on its ability to mitigate the labor market consequences of a given diagnosis. The intuition supporting this approach is that good healthcare should mitigate the detrimental effects of illnesses and injuries on patients' health capital [Grossman, 1972] and, consequently, on their labor market outcomes. All else equal, the better the treatment, the smaller the drop in patients' labor earnings and participation. We consider the heterogeneity in treatment quality across hospitals and call the resulting measure "Adjusted Earning Losses" (AEL hereafter).¹ We also analyze the evolution of quality over time.

To operationalize this approach, we link data on the universe of hospital admissions in Denmark with data on the labor market outcomes of the entire working-age population from 1995 to 2015. We focus on non-elderly patients (25 to 59 years old) with their first non-pregnancy related inpatient admission in four years (index-admission). This unique dataset allows us to measure the negative effects of hospitalization on the patient's earnings, which we will use to infer the quality of treatment received by the patient.

¹We call the measure "adjusted" earning losses because our specifications control for differences in patient mix along both the medical and the socio-economic dimensions.

This approach is valuable for two reasons. From a statistical perspective, it allows the researcher to use large scale datasets on observable outcomes to gather information about shocks to individuals' health status, which is an unobservable variable. In fact, AEL is positively correlated with traditional quality indexes based on mortality or readmission rates but it contains significantly new information since most of its heterogeneity is not captured by other metrics. It also sheds light on a dimension of healthcare (and health) which traditional measures have largely ignored: its role in preserving people's ability to produce earnings. This dimension is important because the ability to work is surely part of what the concept of "health" refers to.²

We find sizable heterogeneity in the labor cost of a hospitalization across different hospitals: AEL indicates a 4 percentage points difference in lost earnings between the best and worst hospitals, all else equal. Being admitted to a hospital with below-the-median AEL leads to an additional loss of 0.7% of pre-hospitalization yearly earnings with respect to being admitted to a median quality hospital.

AEL differs in several respects to the methodologies employed by the previous literature. Risk-adjusted mortality and readmission rates, which are the most widely used metrics by the existing literature, are not only based on very coarse and extreme outcomes, but they are also mainly informative for very severe and acute conditions. As we consider earnings losses, we're able to compute a quality measure for virtually any condition, including also the less severe diseases. For instance, there are only 6 (11) medical conditions³ for which we observe at least 200 deaths within 30 (360) days from discharge for non-elderly index inpatient admissions between 1998 and 2012 in the whole country. These conditions represent half of the causalities but only 4.2% (6.5%) of admissions.

We also implement an event-study design using one variable related to patients' health status and labor market outcomes that is not used in the construction of AEL: the amount of transfers from the government. The results show that patients admitted to high- and low-AEL hospitals display exactly the same pre-hospitalization trends but patients in the latter group receive significantly larger post-hospitalization support from the government,

²We searched for "health meaning" on Google dictionary: the second example provided was related to labor supply (i.e. "bad health forced him to retire").

³Namely, Acute cerebrovascular disease, AMI, Cancer of bronchus or lung, Cardiac arrest and ventricular fibril, Intracranial injury, and unclassified events.

suggesting their labor earning capacities have worsened more. While existing literature on transfer programs such as sickness benefit and disability insurance tend to focus on how the generosity of the payout affects labor earnings, our result shows that investment in improving healthcare quality may lead to both higher labor earnings (and thus more government revenue) and lower government spending. The result is especially relevant as governments across the world grapple with increasingly high expenditures on such transfer programs and declining labor force participation.

We exploit the flexibility of our approach to investigate within-hospital quality heterogeneity by estimating a different quality measure for each broad diagnostic class (or “chapter”). We find that hospitals that are better at treating one class of morbidities are, on average, better at treating other conditions as well. However, this correlation is, on average, only 14%: therefore, there exist significant within-hospital across-diseases heterogeneity.⁴

The economic impact of physical and mental impairments, which is mostly ignored by traditional measures, can be a sizable part of the risk associated with such negative events. Two additional advantages of AEL is that it has a direct economic interpretation in terms of forgone earnings and it can be estimated by combining hospital discharge data with information routinely collected by tax agencies in many countries. Therefore, it does not require “ad hoc” data collection, while the measures of health based on patients’ surveys, changes in functional and cognitive status, or tracking of specific biological metrics do. The main disadvantage of our measure is that it only focuses on the working-age population, and as such it is uninformative of the treatment quality of children and seniors.

Estimation of AEL might suffer from the same selection biases affecting traditional quality measures. In particular, previous literature on hospital quality (see Gowrisankaran and Town [1999], Geweke et al. [2003], Doyle [2011], and Doyle Jr et al. [2015]) has raised two main concerns related to unobservable patient severity. First, one could worry that the best hospitals treat a larger share of the unobservably sickest patients, which would bias

⁴Similarly, Hull [2018] studies short-term mortality and shows that there exist substantial amount of “essential heterogeneity” in hospitals’ ability to treat different patients admitted for the same (emergency) condition. Our results are different because we assume that the quality of treatment (within a hospital) is homogeneous for all patients with the same condition.

their estimated quality downwards (and vice versa for the worst hospitals). Second, better hospitals might be located closer to unobservably “better” (e.g. with higher human capital) individuals, and therefore admit a large share of them. These individuals might be more resilient to health shocks, which would also invalidate our estimation procedure. Although all of our specifications include detailed controls for socio-economic characteristics,⁵ the main diagnosis, other co-morbidities, and past medical history, which should minimize the magnitude of unobservable patient severity, these are still relevant worries. To deal with both concerns, we adopt two complementary empirical strategies to test whether being admitted to a higher quality hospital (according to our estimates) leads to a positive causal effect on post-admission earnings. We take the empirical framework from the literature assessing teachers’ value added measures [Chetty et al., 2014, Deming, 2014].

First, we instrument the quality of the hospital where a patient is admitted with the neighborhood (or “Church District”) where she lives (similarly to Geweke et al. [2003]). In fact, Danes are more likely to go to the closest hospital (neighborhoods explain more than 70% of the variation in hospital selection). The IV estimates confirm that being admitted to a better hospital decreases the post-hospitalization forgone earnings as much as predicted by the main OLS model estimates, thus supporting the causal interpretation of the quality measure.

Second, we follow Doyle [2011] and focus on patients admitted for acute non-deferrable conditions outside their home location (that is, in a hospital which is not “common” for patients from the same neighborhood). These patients are more likely to be admitted because of unforeseen health shock while they were outside their daily routine and less likely to have consciously chosen a specific hospital. Studying this sub-sample of admissions, we find that patients coming from the same location are better off being admitted to a higher quality hospital. We also show that home location does not predict post-admissions earnings dynamics when patients are treated “away” from their local hospital, suggesting that unobserved heterogeneity clustered at the local level is not a concern of paramount importance for AEL.

⁵It is crucial that we also control for individuals’ occupation, earnings, wealth, and education together with local labor market characteristics. Luckily, our rich set of variables in the Danish register data allows us to do that.

We perform two additional robustness exercises. We show that the inclusion of additional covariates increases the explanatory power of the specification used to estimate AEL while it does not shrink the estimated hospital heterogeneity. Therefore, AEL heterogeneity is “robust” to inclusion of a large set of controls. Finally, we investigate the correlation between different measures of quality and a measure of intensity of patients’ “purposeful” hospital selection (the share of out-of-region patients among the admissions for conditions which are both non-acute and deferrable). We find that this correlation is positive and of similar magnitude for AEL and traditional measures (and risk-adjusted mortality in particular). This suggests that AEL is unlikely to suffer from worse selection issues than other quality measures. It also suggests that empirical designs used to overcome selection issues in US-based studies [Doyle, 2011, e.g] could be applied (in those settings) to AEL as well.

We also document a significant decline in the labor cost of hospitalizations over time. For different diseases, we compare the change in employment probability and labor earnings of patients after their hospital admissions in different years. We find that there has been a significant decline in both earnings losses and drop in labor force participation following a hospital admission during the period 1998-2012. More specifically, we find that the post-hospitalization reduction in labor supply has declined by 8.8% on the extensive (participation), and 4.9% on the intensive margin (earnings conditional on working) on average, as measured by the accumulated change in the three years following the admission. This corresponds to an average reduction of extensive margin labor market costs of about 50%, and a 25% reduction of intensive-margin labor cost of a disease. We also find a huge variation in the reduction of labor market cost across different diseases, with large gains for cancer (lung, breast, colon, and rectum in particular), AMI, and acute cardiovascular diseases.

Our paper contributes to a large literature on healthcare quality measurements, for which practitioners, policymakers, and academics have produced several metrics of health and healthcare quality. Risk-adjusted mortality and readmission rates are the most commonly used measures of hospital quality. For instance, in the US, the Affordable Care Act (through the Hospital Readmissions Reduction Program) has targeted the reduction

of the readmission rates following hospital admissions for a few specific conditions⁶ as the relevant goal for quality improvement and cost reduction [Gupta, 2017]. Other hospital quality measures rely on patients' self-reported ability to perform activities of daily living (ADL), such as bathing or dressing, or patients' satisfaction with their treatments or their queuing time. More technical measures of treatment quality track specific biological indexes to study the impact of a treatment on the evolution of a specific medical condition, such as hemoglobin concentration for anemia [Miller et al., 2012] or white blood cell concentration for HIV [Hamilton et al., 2016].

A large health-economics literature has measured hospital quality focusing on emergency room (ER) admissions for heart attacks, heart failures, and other acute and severe conditions. For instance, see Geweke et al. [2003], Doyle Jr et al. [2015], and Hull [2018]. These studies also mostly consider readmission and mortality rates as the outcome measure. Closer to this paper are works evaluating the labor market consequences of specific healthcare practices, such as Laird and Nielsen [2016] who analyze the effect of physicians' prescribing behavior (focusing on opioids, anti-inflammatories, anti-anxiety drugs, and anti-depressants) in Demanrk, Garthwaite [2012] who studies the use of Cox-2 inhibitors (pharmaceuticals treating chronic pain and inflammation) in the US, and Shapiro [2018] who focuses on advertising of antidepressant in the US. We contribute to this literature by showing that researchers can meaningfully measure the quality of healthcare treatments using data on labor market outcomes.

Another strand of the health economics literature considers the economic consequences of severe health shocks. For instance, Dobkin et al. [2018] study the economic consequences of hospital admissions in the U.S. and estimate an average annual decline in labor market earnings of about 17 percent of pre-admission earnings during the first 3 years after admission. Using Danish register data, Fadlon and Nielsen [2015] show that negative health shocks substantially harm individual labor supply and disrupt households' consumption patterns. Gilleskie [1998] show that employees' poor health might increase work absence and hinder workers' ability to adequately perform the job tasks, leading to substantial costs for employers and governments. Our work connects these two strands of literature

⁶Heart attack, heart failure, pneumonia, Chronic Obstructive Pulmonary Disorder (COPD), and Hip/Knee replacement.

by using the economic consequences following a hospitalization to infer the treatment of the received quality.

Our paper also adds to the literature that documents significant variation in healthcare utilization, spending, and outcome across different regions and hospitals. While there has been strong evidence of large geographic variation in healthcare utilization and spending (Finkelstein et al. [2016]), there has been relatively little evidence showing how such heterogeneity translates into heterogeneity in actual health outcomes. Existing literature often find small geographic variation in health outcomes, such as mortality, despite large differences in spending and treatment practices. However, our paper demonstrates that there can be significant differences in healthcare quality even in a relatively homogeneous healthcare system such as the Danish one.

Finally, this paper is also related to the literature strand studying the factors behind the long-run decline in mortality rate [Lichtenberg, 2018], the economic effects of health heterogeneity [Well, 2007, Bloom et al., 2019], and the literature on health capital [Grossman, 1972].

The paper proceeds as follows: Section 2 describes data sources, institutional setting, and presents some raw data patterns. Section 3 presents our measure of heterogeneity across hospitals. Section 4 investigates whether this heterogeneity can be interpreted as a reliable measure of hospital quality. Section 5 deals with the evolution of the labor market consequences of hospital admissions over time. Section 6 concludes.

2 Data

2.1 Data Sources

The Danish National Hospital Register ("Landspatientregisteret"/LPR) contains nationwide data on all hospital interactions since 1995. Available information includes a 4-digit ICD10 main/action diagnosis code and other related co-diagnoses, patient type (inpatient, outpatient, ER), start- and end-date of treatment and the hospital and hospital department at which they received the treatment. The data also contain information on the treatments, operations and examinations conducted. Using a unique personal identi-

fier, we are able to link the hospital data to other full-population registers. We link data from the Danish National Hospital Register (LPR) to tax- and labor market registers so that we observe a detailed work history and all interactions with the hospital system in Denmark for the full population of Danes aged 25-59 years from 1995 to 2015. Our final data set includes an extensive list of the labor market and demographic variables, including earnings, transfer income, municipality, marital status, wealth, spousal earnings and wealth, education, occupation and socioeconomic status. As our data includes the full population of working-age Danes, we're also able to construct education-specific local labor market statistics, such as unemployment, median income etc., of each municipality.

In this study, we will focus exclusively on in-patient hospital admissions.⁷ To mitigate the problem of co-morbidities affecting the treatment outcomes, we restrict our sample to include only “index”-admissions. An “index” admission is defined as the first non birth/pregnancy related inpatient hospital admission for at least 3 years. We follow the labor-market outcome of the patients 3 years prior to-, during, and 3 years following their treatment year. Our data contains a total of 1.1 million index hospitalizations. Table 1 shows some descriptive statistics of the earnings and labor force participation before, in, and after the year of an index admission for our main analysis sample. It also shows similar statistics for a sample including all observed inpatient hospital admissions (only including one hospitalization per year per person), also in 1998-2012 for individuals who are 25-59 years old once hospitalized, and for a control group (which we use in the over time analysis) who don't experience any index admission during the observed years.

The table shows that roughly 1/3rd of all observed hospitalizations for the specified ages and years are index admission observations. Those who experience an index admission are on average slightly younger (0.4 years) compared to those who experience any inpatient admission, and have slightly lower earnings conditional on being in the labor force. However, the participation rate of the index admission sample is significantly higher, with pre-admission participation being 0.915 compared to a pre-hospitalization participation rate of 0.68 for all inpatient admissions. Both groups, however, experience a drop in

⁷The number of records related to inpatients hospital admission has been relatively stable over time, as shown by Figure A.1, while the number of ER visits (outpatient procedures) has steadily decreased (sharply increased). This allows us to be more confident that the “meaning” of an inpatient record in LPR is stable across the sample period.

participation rates of roughly four percentage points in the hospitalization year, and another three percentage points in the following year. Also, mortality rates are slightly lower in the index-admission subsample. The control group includes individuals who are either of very poor health (with too many inpatient hospitalizations to have an index-admission observation), or of very good health (without any non-birth related inpatient admission during the period). The participation rate of this group is also lower (0.82) compared to the pre-hospitalization participation rate of the index-admission sample (0.915), but earnings conditional on working are similar to that of the index admission patients. As such, we lose some representativeness when we restrict to index-admission observations only.

2.2 Institutional Settings

The Danish health care system is universal and provides free and equal access to health care for all citizens. Health care expenses are primarily tax financed and the vast majority of hospitals are publicly owned. During each year from 1998-2012, private hospitals accounted for less than 2% of total hospital expenditures. Government spending on healthcare increased by 30% from 2000 to 2010, mainly driven by an increase in hospital expenditures. Hospital financing is based on a system of politically fixed budgets that account for heterogeneity in patient mix. Following a large structural reform in 2007, the 16 counties responsible for running the hospitals were merged into 5 regions.

Compared to countries such as the U.S., there are fewer but larger hospitals in Denmark. Since the 1990s, several mergers and hospital closures have resulted in a further concentration of hospitals. Often, the general practitioners act as a gatekeeper to specialized treatment. A person's residence address determines his/her "default" hospital, to which he/she is referred to by default. However, since 1993, patients were granted the right to freely choose among all somatic hospitals in Denmark, if the requested hospital has the capacity. Since 2002, an extended free choice reform has granted patients the right to choose a state financed treatment at a private hospital if waiting times at the public hospitals are too long (2 months before 2007 and 1 month after). However, approximately 91% of patients in our sample period are admitted to hospitals in the same region where

they live.

2.3 Descriptive Patterns

To motivate our analysis, and to give examples of what the data looks like, we show examples of how individuals' earnings evolve before and after a hospital admission for different types of diseases. These illustrative examples provide suggestive evidence that earnings trajectories following hospital admissions vary for different types of diseases, over time and across hospitals.

To do so, we estimate the following linear model

$$y_{i,t} = X_{i,t}\beta + \gamma_t + \sum_{-5 \leq r \leq 5} \delta_r + \epsilon_{it} \quad (1)$$

The dependent variable is a labor market outcome of individual i at year t . Due to the descriptive nature of this exercise, we include only a small set of controls, $X_{i,t}$: age, gender and education, and γ_t denotes calendar-year time fixed effects. For individuals who experience an index hospital admission, the subscript r denotes the year relative to the hospital admission year where $r = 0$ ⁸. We add a 10% random subsample of the population who don't experience an index admission from 1995-2015 as a control group, and normalize $\delta = 0$ for these individuals. As such, the δ coefficients estimate the differences in labor market outcome for individuals hospitalized in year $t - r$ with respect to other (non-hospitalized) individuals with similar characteristics X . For illustrative purposes, we estimate equation (1) for four different diseases. These specific types of diseases are chosen because they are quite common and known to most people, and then they differ a lot in terms of their severity and duration. We consider intervertebral spondylosis (spinal degeneration), acute cerebrovascular disease, breast cancer and fracture of the lower limb⁹.

First, we consider both the intensive- and extensive margin labor market consequences

⁸For instance, $r = -2$ indicates the individual i experience an index admission in year $t + 2$. Note that in this section only, we restrict to index admissions where there is no other inpatient admission in year $t-5$ through t , while for all the main analysis the index admissions are defined as no other inpatient admissions in year $t-3$ through t .

⁹These disease classes are grouped according to the Clinical Classification Software, CCS. In the empirical analysis of hospital heterogeneity, we use a finer diagnosis classification.

of a hospital admission. We compute the intensive margins by using $\log(\text{earning})$ as the outcome variable - thereby we only include positive earnings observations. We compute the extensive margins by using a dummy variable for positive earnings as the outcome variable for all observations, with dead individuals included with zero earnings (excluding those who die does not change any of the figures notably).

The δ coefficients of equation 1 are plotted in Figure 1, where the left column of Figures show the intensive margin responses, and the right column show the extensive margin responses. The plots show that patients' earnings and labor market participation sharply decline around the year of an hospital admission. The patterns for the two margins are quite different: while the drop in the intensive margin tends to recover - especially for the less severe diseases - the extensive margin response continues to decline for all diseases.

In the left column showing the intensive margin decline in labor supply, we see that the sizes of the drop, as well as the degrees to which earnings rebound, vary a lot by disease. Patients with intervertebral spondylosis (spinal degeneration) experience an intensive-margin decline in earnings of 8% , which is halved to a 4% decline 3 years after the admission, and their labor force participation drops by 12%. Patients who suffer from an acute cerebrovascular disease experience an intensive-margin decrease in earnings of 12% and a drop in labor force participation of almost 20%. Breast cancer patients experience a somewhat smaller drop in labor supply with a 5% decrease in the intensive-margin, and a steady decline in labor force participation mounting in a 14% decrease five years after admission. Following a fracture of the lower limb, patients suffer an 8% drop in intensive-margin earnings, but are almost fully recovered after 3 years, and their labor force participation permanently decrease by 7%.

Has healthcare quality improved since 2000? We use our simple specification to provide evidence of advancement in the effectiveness of medial treatment in preventing patient earnings loss. We re-estimate equation (1) dividing patients according to the year of their hospital admissions: 2000-2002, 2003-2004, 2005-2007 and 2008-2010. To capture both extensive and intensive margin responses, the dependent variable is the log of labor earnings plus a constant (1 DKK), so that we're able to combine the intensive and extensive

margin labor market outcomes in one outcome measure. Because of the arbitrary choice of this constant (1 DKK), we only rely on this specific outcome measure for illustrative purposes, and we will refrain from using it in the main analysis (see Section 3.1). Results for the four different diseases are presented in Figure 2. We detect a striking heterogeneity over time for the different diseases. For instance, the earnings losses associated with intervertebral spondylosis (spinal degeneration) for patients treated in 2008-2010 is only borderline significant smaller than the loss of patients treated in 2000-2002. On the other hand, patients admitted for cancer of breast or acute cerebrovascular diseases seem to be much better off in the later period compared to the early 2000s.

Do healthcare quality vary across different hospitals? Again, we consider equation (1) to provide some suggestive answers to this question - now allowing the δ coefficients to vary according to the index admission's hospital. In Figure 3 we present results related to four randomly selected hospitals. The pre-hospitalization trends in earnings of individuals who suffer from acute cerebrovascular diseases are not significantly different across the different hospitals which they are admitted to. However, the earnings losses following the hospitalization are much more severe for patients of hospital A and D rather than B or C. In general, for all four diseases, patients of hospital A and D seem to suffer the most from their disease. However, these descriptive results do not control in detail for patient characteristics, specific diagnosis types, local labor market characteristics etc. In Section 3 we present a more in-depth investigation of heterogeneity of hospital treatment quality.

3 Adjusted Earning Losses

A worker's health status alters the physical and mental cost of working (or searching for an occupation), and is consequently a fundamental factor of labor supply and productivity, see evidence surveyed by Currie and Madrian [1999]. Therefore, the quality of healthcare received by a hospital patient should have an effect on his/her post-hospitalization labor market outcomes. Intuitively, as long as the cost of working is decreasing in health, and while productivity is increasing, then the economic costs of a hospital admission should

be lower for a patient receiving better care.¹⁰

Following this intuition, we define the quality of hospital h as the difference in earnings realized after a treatment at hospital h and the earnings which would have been realized after at treatment at some reference hospital 0:

$$q_h = E[y_h|r \geq 0] - E[y_0|r \geq 0] \quad (2)$$

where y_h refers to the labor market outcome of interest following an admission to hospital h . r denotes the time relative to the hospitalization, such that $r \geq 0$ includes the year of the admission and the following ones.

We refer to q_h as “Adjusted Earning Losses” (or AEL) and we propose a simple empirical framework for the estimation of AEL under the assumption that, after controlling for a large set of characteristics at the patient-, case-, and local-level, hospital selection is not driven by the residual unobservable heterogeneity (e.g. unobserved severity). Thus, we run different empirical checks to test the robustness of our main findings to possible violations of this identifying assumptions and we show remarkable stability.

3.1 Empirical Strategy

To estimate AEL, we consider all inpatient “index” hospital admissions in Denmark during the years 1998-2012, see Section 2.1 for more details. We divide the sample into five three-years periods and estimate each specification separately for each time period. We restrict the analysis to the 31 largest hospitals (see Appendix A for more details) and we exclude patients with “rare” diagnoses (less than 100 observed index admissions in 1998-2012). We include only Danish-born patients aged 25 - 59 at the time of hospitalization, in order to focus on individuals at the peak of their labor supply. We do not include a control group of non-hospitalized individuals.

The main outcomes of interest are the post-hospitalization earnings produced by in-

¹⁰This assumption would be violated if, for instance, sicker people decided to supply more hours worked because their value of leisure is lowered by a negative health shock. For our purposes, it is not important whether health shocks affect primarily labor demand or labor supply (e.g. productivity). Neither it is important to model the relationship between the health capital (a stock) and health status (a flow) [Grossman, 1972].

dividual i in year r (with $0 \leq r \leq 3$) relative to the hospital admission as a share of their mean earnings in the 3 years prior to the index admission (henceforth referred to as “normalized earnings”). The use of normalized earnings allows us to consider jointly the extensive (labor market participation and survival) and intensive margins of the labor market consequences of hospital admissions [Kleven et al., 2018]. We let h denote the hospital of the index admission, d the main diagnosis, and t one of the five three-years hospitalization periods. The analysis relies on the following linear model:

$$y_{i,r} = q_{h,t} + X_{i,r}\beta_t + \phi_{d,t} + \sum_{0 \leq r \leq 3} \delta_{r,dg,t} + \epsilon_{i,r} \quad (3)$$

where $X_{i,r}$ represents a vector of control variables, $q_{h,t}$ are (time-varying) hospital fixed effects and $\phi_{d,t}$ are (time-varying) main diagnosis (according to 4-digits ICD10 codes) fixed effects. The parameters $\delta_{r,dg,t}$ capture disease group specific evolution of earnings around the hospitalization, where r denotes the time relative to hospitalization and dg the diagnosis group.¹¹

The vector $X_{i,r}$ contains several covariates that capture both individual heterogeneity and local labor market conditions. The richness and the level of details of $X_{i,r}$ is one of the strengths of our study with respect to previous literature focusing on US hospitals. At the individual level we include several socio-economic pre-hospitalization characteristics: we flexibly control for earnings (percentile of the earnings distribution), earnings growth, decile of own assets, decile of spouse’s wealth and spouse’s earnings interacted with gender, gender, age, education, marital status, occupation, and socioeconomic status the year before hospitalization. We further control for the number of co-diagnoses related to the index admission and for other interactions with the public healthcare system, namely number of non-pregnancy related outpatient and ER visits and number of pregnancy related inpatient, outpatient, and ER visits the same year and 3 years before. We also include fixed effects related to all the co-morbidities registered in the discharge data and calendar year fixed effects. Finally, we include controls for education-specific labor market conditions, namely unemployment rates, non-in-labor-force rates, and earnings levels for individuals with same education level within the municipality, and the share of individuals

¹¹We group the 4-digit ICD10 codes into 271 diagnostic groups.

receiving disability benefits in the municipality.

The AEL estimates $q_{h,t}$ are obtained by applying OLS to equation (3) separately for each of the five three-years time periods. After dropping observations with missing values we are left with an average of 21 thousands observations per hospital for each time period (each admission relates to 4 observations since we assume dead patients' earnings are zero).

Our empirical procedure is consistent under the assumption that the hospital choices are uncorrelated with the unobserved individual characteristics or unobserved severity of health shocks. In Section 4 we discuss these potential biases, and we present evidence suggesting these concerns have a limited role for our main results.

To correct the estimated AEL for measurement error, we follow Chandra et al. [2016]¹² and implement an "Empirical Bayes Adjustment". That is, we shrink the size of the estimated quality toward its mean and shrink it more for observations that are more likely to be outliers: without such adjustment the variance of the estimated quality would be a biased measure of the variance of the true quality. Intuitively, the variance of the estimates could reflect the sum of the variance of the quality and the variance of the measurement error. We follow this procedure throughout the paper and apply it to all the quality measures produced.

3.2 Distribution of AEL

Figure 4 shows some statistics of the distribution of estimated AEL for each time period. We adopt the convention that larger values of AEL correspond to smaller losses and, therefore, higher quality. Since the dependent variable is normalized earnings, we interpret the quality measure as the percentage change in yearly earnings (compared to mean pre-hospitalization levels) expected to occur over 3 years from the hospital admissions.

We find substantial heterogeneity in hospital quality across all the time periods. AEL has an average (across the five periods) standard deviation of 1.05 percentage points.

¹²Specifically, we implement the algorithm described in their Appendix C. Our procedure differs only because we use municipality-clustered standard errors as an estimate of the variance of the hospital fixed effects and because we allow the mean of the "true" quality to be a function of hospital size rather than location.

An inpatient admission in the lowest-ranked hospital is associated with an additional decrease in annual labor earnings of 4.2 percentage points, on average, compared to the highest-ranked hospital. The hospital heterogeneity is larger for the early sample period (1998/2000), constant between 2001 and 2009, and lower for the latest period (2010/2012); this is partly due to the reduction in the number of hospitals since only 22 are active at the end of the sample. Hospital quality is positively correlated over time, but there is substantial time-varying heterogeneity. The correlation between AEL and its lag, on a three-years period scale, is 0.59 (approximately 0.84 yearly), see Figure 5.

Patients treated at hospitals with below-median hospital AEL in our analysis sample would have on average earned an additional 0.7% of their pre-hospitalization earnings for (at least) four years if they were treated at the hospital with median estimated quality, with an estimated loss of (at least) 6,400 DKK (approximately 960 USD).

To quantify the total impact on inpatient admissions for non-elderly patients, we perform a back-of-the-envelope calculation of the savings from increasing hospital quality following the formula:

$$Savings_t = \sum_{h:q_h < q_{med}} [(q_{med} - q_h) \times (N_{ht} \times \bar{Y}_{ht} \times 4)]$$

For all the below-the-median hospitals we calculate the increase in quality needed to achieve the median quality $q_{med} - q_h$. N_{ht} is the number of all non-pregnancy related inpatient admissions of patients aged 25-59 in year t for hospital h . \bar{Y}_{ht} is the mean earnings the year before hospitalization for all the patients included in N_{ht} at hospital h in year t . Then, we multiply this quantity for the number of years for which the patients' earnings are affected by health care quality (which is 4 years in our case). Our estimates imply that if all hospitals had at least the median quality, the Danish economy would save on average almost one billion DKK (approximately 140 million USD) in foregone earnings each year, or 60 million DKK (or 9 million USD) per year per hospital. These numbers are conservative estimates of the gross savings that would occur if the low-quality hospitals were to improve to the median quality because we focus on inpatient admissions

only, because better healthcare might lead to savings for more than four years, because we exclude smaller hospitals and patients outside the 25-59 age range, and we abstract from increases in wages over time. Moreover, a part of the loss of productivity caused by lower quality healthcare might be borne by employers rather than being reflected in employees' earnings. However, we do not have data to quantify the cost of improving the quality of the low-AEL hospitals.

3.3 Event-Study Approach

In this section, we adopt an “event-study” design to evaluate the evolution of patients' labor market outcomes before and after the index hospitalization and to assess the differences between patients admitted to high- and low-AEL hospitals. Specifically, we estimate the following equation, separately for patients admitted to hospitals in the top and bottom third of AEL ranking (within their admission time period):

$$y_{i,r} = \alpha_i + \gamma_t + \sum_{-3 \leq r \leq 3, r \neq -1} \delta_r + \epsilon_{it} \quad (4)$$

Since AEL is estimated with normalized earnings, the dependent variable of equation (4) cannot be based on earnings themselves, as these are mechanically correlated with the quality ranking. Therefore, we focus on the amount of transfers from the governments received by the patient. Transfers include sickness benefit, cash assistance, and unemployment benefit: they are therefore related to both labor outcomes and to patients' health status but they are not used to estimate AEL.¹³

The parameters δ_r are informative of the evolution of transfers from the government before and after the index hospital admission, controlling for patient fixed effects. Figure 6 plots the estimated coefficients, together with 95% confidence intervals based on

¹³The interested reader can find two graphs related to earning dynamics for high- and low- AEL hospitals in the top panels of Figure A.3. We also re-estimate the event-study specification by hospital ranking defined with Risk-Adjusted Mortality (RAM); this is a more insightful exercise since earnings are not included, neither as controls nor as an outcome variable, in the estimation of RAM (see Section 3.5 for details). Results are shown in the bottom panels of Figure A.3. We can reject the null that high-quality hospitals do not deliver better labor market outcomes when we focus on log earnings, at least at long horizons.

municipality-clustered standard errors.¹⁴ As expected, the amount of transfers increases sharply from the year of hospital admission onwards, signaling substitution from labor earnings to government provided insurance. Reassuringly, patients admitted to different hospitals follow exactly the same dynamics before hospital admission. Patients admitted to high-quality hospitals receive significantly less transfers from the year of admission onwards, which indicates that they are less in need of assistance and, therefore, experiencing a smaller deterioration of their labor market outcomes. It is noteworthy that the difference between the two groups persists over time.

3.4 Within Hospital Heterogeneity

One of the features of AEL that makes it a useful addition to the traditional measures of healthcare quality is its ability to provide information on a broad set of medical conditions. We exploit this property to investigate whether the best hospitals are better at treating all types of medical condition.

To categorize different medical conditions, we rely on the chapters of the ICD10 codes for the main diagnoses provided by the discharge data. The chapters represent the most general level of classification of ICD10, with “Neoplasms” (chapter 2) and “Diseases of the digestive system” (chapter 11) being two examples. We construct a chapter-specific AEL measure $q_{h,ch}$ by estimating equation (3) separately for patients admitted for a main diagnoses belonging to each chapter ch . We pool together observations for all time-periods to have larger sample size. For each pair of diagnostic chapters ch and ch' we compute the correlation between $q_{h,ch}$ and $q_{h,ch'}$. We find that the average correlation coefficient is 0.14. Therefore, while better hospitals are, on average, better at everything, there exist substantial amount of within-hospital quality heterogeneity. Similarly, the AEL estimated only for acute non-deferrable conditions¹⁵ have a 0.52 correlation with AEL estimated for

¹⁴Because of a well-known multicollinearity problem, estimation of equation (4) requires an additional assumption. Our choice is to set the coefficients of two consecutive years (2000 and 2001) to be equal. Alternatively, we estimate the specification including a “control” group of non-hospitalized individuals and allow all fixed effects to be freely estimated. Results are similar, see bottom panel of Figure A.2. Results are also similar if we define high- and low- quality hospitals as the top/bottom half or quartile of the AEL ranking, as shown by the top panels of Figure A.2.

¹⁵That is, conditions that have a weekend admission rate close to two sevenths and are classified as acute in the discharge data [Doyle, 2011].

all conditions.

3.5 Correlation with Traditional Measures

To understand whether AEL contains additional information with respect to traditional quality measures, we estimate risk-adjusted mortality (RAM) and risk-adjusted readmission (RAR) rates for acute non-deferrable conditions. To obtain these measures, we estimate a probit model:

$$Pr(D_i = 1) = \Phi(\gamma_{h,t} + X_i\beta_t + \phi_{d,t}) \quad (5)$$

where D_i is a dummy equal to one if and only if patient i died (RAM) or experienced another inpatient admission (RAR) within 30 days from the index hospital admission and $\Phi(\cdot)$ is a Gaussian cdf. The set of controls X_i includes the same health-related controls we use to estimate AEL but we do not include the extremely details patient- and local- level socio-economic controls of equation (3): we do this both for consistency with previous literature (e.g. [Hull, 2018]) and for computational reasons. We recover RAM and RAR as the derivative of the estimated probability of death or readmission (at the average value of the covariates) with respect to the hospital dummies and then perform the "Empirical Bayes Adjustment" procedure. Finally, we normalize all the estimates by subtracting the period-specific average.

We compare AEL to RAM and RAR. Focusing on AEL estimated for acute non-deferrable conditions, we find that it has a correlation of -0.29 with RAM (p -value = 0.001) and -0.13 with RAR (p -value = 0.118), see Figure 7. Since low mortality and readmission rate convey high quality, this pattern means that better hospitals according to traditional measures are also better, on average, according to AEL. We then use the traditional measures to predict AEL via linear regression. The correlation between the actual AEL and the predicted values is 0.32: most of the variation in AEL cannot be explained by mortality and readmission rates and, therefore, it contains a substantial amount of additional information that can be useful to policy-makers and researchers interested in hospital quality. The additional amount of information is especially relevant for non acute conditions. In fact, RAM and RAR can explain only 3 percent of the

variation of AEL when the latter is estimated with all index admissions.

We also compute a ranking based on our measure of quality. We compare our ranking to an alternative ranking provided by the Danish medical news journal “Dagens Medicin” in 2011. They rank the Danish hospitals based on three outcomes: treatment quality, patient satisfaction and reputation among hospital personnel. Treatment quality is an average of various different specialty-specific indicators, such as provided procedures, risk-adjusted mortality- and readmission rates - each weighted by their importance for each specialty. We find a Spearman rank correlation between the 2011 “Dagens Medicin”’s ranking and the AEL-based ranking of 0.27 for the 2007/2009 period and 0.15 for the 2010/2012 period.

4 Does AEL Capture Hospital Quality or Unobserved Heterogeneity?

In this section, we investigate whether the heterogeneity of the labor market consequences of hospital admissions across hospitals (documented in Section 3) can be interpreted as a reliable measure of differences in treatment quality.

When we measure hospital quality we aim to capture the changes in normalized earning caused by the admission to one institution rather than another (see equation 2). Estimation of equation (3) by OLS provides consistent estimates of such parameters under the assumption that the error term is uncorrelated with the hospital choice. That is, after controlling for observable factors at the patient- and local labor market-level, the remaining unobservable characteristics of admissions must be equally distributed across different hospitals.

A first reassuring finding comes from the event-study described in Section 3.3: patients admitted to high- and low-AEL hospitals do not display different pre-hospitalization trends in their transfers from the government. This variable is the labor- and health-related outcome in our data that we do not use, directly or indirectly, for the construction of AEL and it therefore provides a useful check. However, these patients could still be different in ways that do not manifest themselves before the time of hospital admission.

The literature on the measurement of hospital performance has highlighted two main concerns for this potential endogeneity problem. The *first concern* is selection on unobservable patient severity. Facing a waiting list trade-off, Danish patients with non-urgent treatment requirements are free to choose any hospital in the country with few exceptions. Therefore, the best hospitals might treat a disproportionate share of the sickest patients, all else equal, which would cause a downward bias of their estimated quality.

The *second concern* is that, since hospitals are not randomly allocated across a country's territory, some hospitals might be closer to a specific subset of patients, and therefore admit more of them. Some hospitals might be located closer to unobservably "better" (or more resilient) patients, causing an upwards bias of their estimated quality. For instance, Doyle Jr et al. [2015] write that "*patients who live relatively close to "high-tech" hospitals could be different from those who do not, in ways that are difficult to control*". A concrete example could be differences in other health-care related amenities, such as local GP quality.

In order to reduce the amount of residual heterogeneity between patients as much as possible, we include numerous different controls in all of our specifications, including a detailed set of controls for the main diagnosis, co-morbidities, and past medical history. Nonetheless, patients, nurses, and doctors are likely to have additional information about the patient, which is unobserved to the econometrician, and which affects the hospital choice of the patients.

Both endogeneity concerns are likely to be less harmful in our study compared to previous, mainly US focused, literature. This is primarily because the greater level of detail in our data set allows us to control for a wide range of patient characteristics, which consequently should limit the amount of unobserved heterogeneity. One could also argue that the concerns are smaller because of the specific institutional settings in Denmark, where fewer and larger hospitals are also more likely to serve a more representative mix of patients, and where the relative travel distance to the closest vs. second closest hospital is generally larger. Also, the Danish population is more homogeneous, with lower income inequality and poverty, lower cultural and ethnic diversity, etc. In fact, we find that each hospital in our main sample serves at least one neighborhood in each decile of the income

distribution.¹⁶ Nonetheless, selection concerns are still extremely relevant for our analysis.

The causal interpretation of the quality estimates implies that being admitted to an hospital leads to a change in normalized earnings (with respect to the baseline hospital) exactly equal to the AEL. We can therefore learn about the robustness of AEL by testing a) whether an admission to a better hospital actually leads to a smaller drop in earnings, and b) whether the size of the effect is the same as predicted by equation (3) (i.e. exactly q_h).

To do so, we exploit two complementary empirical strategies. Each one deals in turn with one of the two main concerns described above.

First, we use patients' home location as an instrument for their hospital choice. This strategy addresses the first concern, which is that the best hospitals treat patients who are unobservably more sick. That is, we relax the assumption that hospitals treat patients with (unobservably) similar characteristics (e.g. as the severity of the health shock). However, this strategy is valid only if the unobserved heterogeneity is homogeneous across neighborhoods within the country and, therefore, patients' location affect the dynamics of post-hospitalization normalized earning only through hospital choice.

Then we also consider the second concern, which could invalidate both the main specification and the IV estimates. That is, certain hospitals might be located closer to unobservably "better" (or more resilient) patients, and therefore admit more of them because of proximity. To address this concern, we focus on a small sub-sample of patients who are admitted to hospitals "far" from home for arguably random reasons. In this exercise, we find no evidence that local heterogeneity is significantly biasing our quality measures.

Furthermore, we show that AEL heterogeneity does not vanish as we add a more comprehensive set of controls (and improve the specification's explanatory power), suggesting it would be robust to the inclusion of additional unobserved patients' characteristics. Finally, we find that the correlation between hospital quality and the share of outside-the-region patients, among those who are admitted for conditions which are both deferrable and non-acute, is similar for AEL and risk-adjusted mortality. This can be thought of

¹⁶That is, each hospital is the most common hospital (for at least one diagnostic class) for patients coming from at least one neighborhood for each decile.

as a measure of “intensity” of hospital selection since these patients have to travel relatively longer distances to seek care and they can decide where to go because of lack of urgency. Therefore, it suggests that AEL is unlikely to suffer from worse selection issues than traditional quality measures.

4.1 Instrumental Variables

Let \widehat{q}_h be the AEL for hospital h estimated by applying OLS to equation (3). We consider the linear specification:

$$y_{i,r} = \alpha \cdot \widehat{q}_{h,t} + X_{i,r}\beta_t + \phi_{d,t} + \sum_{0 \leq r \leq 3} \delta_{r,dg,t} + \eta_{i,r} \quad (6)$$

It is worth comparing equations (3) (our main specification) and (6), which are identical except for $q_{h,t}$. In the former, $q_{h,t}$ is a hospital fixed effect parameter (heterogeneous across time periods t). In the latter $\widehat{q}_{h,t}$ (the estimated AEL) enters as an observed hospital characteristic, while α is the parameter of interest (we pool together data from the five time periods). The set of controls included in (3) and (6) are the same, see section 3 for details.

Equation (6) is akin to the specifications used by previous literature to assess biases in teachers’ value added measures, see equation (3) in Deming [2014] and section I.C. in Chetty et al. [2014].

A valid instrument for hospital choice would allow us to identify the effect of being admitted to a hospital with estimated quality \widehat{q}_h . Recovering $\alpha = 0$ would then imply that the measure of quality is uninformative, as it only captures heterogeneity in patient mix between hospitals (e.g. differences in case severity). $\alpha > 0$ would imply that hospitals that are better according to our measure are indeed decreasing the cost of hospital admissions, although less than expected. $\alpha = 1$ is the value consistent with our main equation (3) and would imply that OLS estimates are not significantly biased by selection issues. In fact, estimation of (6) by OLS mechanically leads to $\alpha = 1$. In Appendix B, we present results from a simple simulated model to further clarify this exercise.

Location is the main driver of hospital choice in Denmark. In fact, more than 70%

of patients in our sample are admitted to the “common” hospital, which is the hospital admitting the largest share of patients in a given neighborhood with the same medical condition. Therefore, we follow previous literature which use patient address as an instrument for hospital choice, see e.g. Geweke et al. [2003], Gowrisankaran and Town [1999].

We estimate equation (6) using patient neighborhood (Church District) as an instrument for AEL of the hospital she was admitted to, and we present results in column (1) of Table 2. We find that α is neither statistically nor economically different from 1. That is, being admitted to a higher AEL-hospital decreases the economic impact of the hospital admission exactly as much as predicted by the main specification (3). This suggests that selection issues do not affect the quality estimates described in section 3. This is in contrast with general findings of the literature analyzing hospital selection issues related to mortality rates in the USA, such as Gowrisankaran and Town [1999].

This empirical strategy is valid under the assumption that patients’ location affect their post-hospitalization earning dynamics only through hospital choice. Although we include a large set of controls at both local- and the individual level, the residual unobserved heterogeneity might invalidate this IV approach. For instance, unobservably “better” patients might be clustered around better hospitals, or local labor markets might have unobservably different characteristics for different hospitals. Furthermore, other local healthcare characteristics, such as GP quality and rehabilitation procedures, could also influence the outcome of patients at a given hospital. The same holds true for differences in case-worker procedures, e.g. requirements for disability pension eligibility, which might also vary across municipalities and across time. Therefore, next section implement two simple empirical exercises to test for biases generated by local heterogeneity.

4.2 Acute Non-Deferrable Admissions at Non-Common Hospitals

The presence of local heterogeneity might lead OLS and IV estimate to be inconsistent for the true hospital quality. To understand the extent to which patient location reveals information about their resilience to health shocks, and therefore whether local heterogeneity should be a main concern, we focus on a specific subset of patients which are

admitted to hospitals outside “far” from home for arguably random reasons.

We narrow our attention to the subsample of patients who are treated outside their local common hospital and who are admitted with an acute non-deferrable condition. That is, classified as acute by the discharge records and with a weekend-admissions share close (within 0.02) to two-sevenths. Because of the nature of their diagnosis, they are more likely to be randomly admitted to an “uncommon” hospital, as the unexpected health shock is likely to have occurred while they were away from home, and not because they conscientiously chose a hospital further away. This strategy helps us overcome problems with selection- and differences in other local health care characteristics, but it also forces us to drop the vast majority of observations and only consider a small range of acute conditions and few patients. This exercise is inspired by the identification strategy in Doyle [2011], who consider the hospital admissions of tourists in Florida experiencing an acute health shock during their vacation. However, we cannot focus on actual tourists coming from outside Denmark because we do not observe their labor market outcomes.

We consider the following specification:

$$y_{i,r} = \alpha \cdot \widehat{q}_{h,t} + X_{i,r}\beta_t + \tau_n + \phi_{d,t} + \sum_{0 \leq r \leq 3} \delta_{r,dg,t} + \epsilon_{i,r} \quad (7)$$

where n is the neighborhood where patient i lives. Equation (7) differs from (6) as we also include of a set of neighborhood fixed effects (τ_n). The parameter α reveals how (estimated) hospital quality affects the normalized earnings of patients *conditional* on their home location. We estimate (6) by OLS using the acute admissions at non-common hospitals sample. Results are reported in column (2) of Table 2. We find $\alpha \approx 1.2$: being admitted to a better hospital significantly decrease the earnings cost of the hospitalization. The estimated magnitude is larger than 1 (what predicted by the main equation 3). However, we cannot reject the null of $\alpha = 1$ at 10% confidence level (p – value = 0.21).

To investigate further whether local heterogeneity is affecting the quality estimate of section 3, we consider the alternative specification:

$$y_{i,r} = \alpha \cdot \widehat{q}_{h,t} + X_{i,r}\beta_t + \theta \cdot \widehat{q}_{n,dg,t} + \phi_{d,t} + \sum_{0 \leq r \leq 3} \delta_{r,dg,t} + \epsilon_{i,r} \quad (8)$$

where $\widehat{q}_{n,dg,t}$ is the estimated quality of the “common” hospital for patients of neighborhood (Church District) n and diagnostic group dg in period t . The intuition behind this approach is that if part of what AEL captures is local heterogeneity (e.g. patients’ unobserved characteristics or the presence of better GPs) rather than actual hospital quality, then patients coming from “better” neighborhoods should do better, even when they are treated outside the local hospital, all else equal. We focus again on acute admissions at non-common hospitals. OLS estimates of equation (8) are presented in column (3) of Table 2. The estimated α is 1.39, while θ is not different from zero: it is only the hospital where they are admitted—and *not* where they live—that matters for the economic impact of the hospitalization. Consequently, we find no evidence that AEL estimates are affected by patients heterogeneity clustered at the local level.

This exercise shows that there is no correlation between the estimated quality of a hospital and the earning dynamics of patients coming from the same area, when these patients are admitted to different hospitals for reasons arguably orthogonal to treatment quality itself. If the hospital quality was biased by local unobserved heterogeneity, this correlation should be positive.

4.3 AEL Heterogeneity and Observable Covariates

Although we include a rich set of observable covariates in each specification, most of the variance of normalized earnings is left unexplained. Therefore, a main concern for a “quality” interpretation of AEL is that its heterogeneity is derived solely by the fact that we are not able to include all potential factors affecting earning dynamics. That is, if we were able to include all possible controls, all hospitals would equally good for patients’ labor market outcomes.

An alternative approach to assess potential omitted-variables biases, within a linear regression framework, is to evaluate *jointly* the changes in OLS coefficients and R-square as one includes additional explanatory variables [Altonji et al., 2005, Oster, 2019, Finkelstein et al., 2019]. While we cannot estimate AEL from a regression that explain 100% of the variation in earning dynamics, we can measure how AEL heterogeneity is affected by the inclusion of the rich set of covariates. If AEL heterogeneity shrinks monotonically

as controls' explanatory power increases, then we might be worried it would vanish if we could run such "perfect" regression.

We define six ordered sets (or groups) of controls. Each set is a superset of the previous one. The sixth set includes all controls of equation (3), the first set includes only the main disease fixed effects, and each set in between add some additional covariates.¹⁷ We then re-estimate equation (3) for each group. For simplicity, we perform this exercise pooling together all years.

Figure 8 plots the R-square of each regression, together with two measures of AEL heterogeneity: its standard deviation and its range. By construction, each regression has higher explanatory power than the previous one. AEL heterogeneity is higher in the regressions one and two, when controls have little explanatory power. Then, as soon as we introduce controls for comorbidities, education, and pre-hospitalization occupation of the patient, both range and standard deviation stabilize around 4 and 1 percentage points. The inclusion of additional controls, such as the percentile of income, or patients and spouses wealth, increases substantially the R-square (which goes up by more than a third) but does not diminish the estimated heterogeneity in AEL.

This simple exercises cannot eliminate the concern that unobserved heterogeneity is affecting hospital selection. However, it does mitigate the concern that AEL heterogeneity captures heterogeneity of patients' unobserved characteristics rather than hospitals.¹⁸

4.4 Discussion

We find that that selection issues are unlikely to significantly bias our quality measures, as long as unobserved heterogeneity is homogeneous across different neighborhoods in the country. The estimates of equation (8) show that local heterogeneity does not seem to significantly affect our results, mitigating the main concerns and supporting a quality interpretation of AEL heterogeneity. We also find that AEL heterogeneity is "robust" to

¹⁷The second group adds controls for pre-hospitalization income (linear and square) and fixed effects for age, gender, marital status, and year. The third group adds detailed controls for comorbidities, education, and pre-hospitalization occupation.

¹⁸Previous literature [Altonji et al., 2005, Oster, 2019, Finkelstein et al., 2019] provides formal test to evaluate coefficients' stability as controls are added. However, our case is not covered by any of these papers.

the inclusion of a large set of controls.

Despite these reassuring results, several caveats apply. First, we focus (most of) our robustness analysis on the most important selection concerns underlined by previous literature. However, additional issues might be also present. For instance, the quality of the local hospital might affect patients' decision to seek care (or its timing). Moreover, our ability to compare patients with the same medical conditions crucially depends on hospitals' ability to provide appropriate diagnostic codes, which could be correlated with their quality. Second, while we argue that the impact of selection issues on our estimates is not of paramount importance, these issues might still play a (likely minor) role. Third, our instrumental variable approach recovers the effect of hospitalization for the "marginal" patients whose hospital selection depends on the home location [Imbens and Angrist, 1994, Heckman and Vytlacil, 2005]. The very sickest patients might go to the best hospitals regardless of where they are located. Therefore, our estimates, based on a homogeneous-treatment assumption, might fail to capture the ability of some hospitals to treat these very sick patients.

We finally ask: if biases caused by non-random hospital selection are present, are they more likely to affect AEL or the traditional measures of quality? We correlate AEL, RAM, and RAR (computed with acute non-deferrable conditions) with the share of out-of-region patients among the admissions for conditions which are both non-acute (according to the discharge records) and "deferrable" (weekend admission rate different than two-sevenths). The latter can be considered a measure of "intensity" of purposeful selection, since these patients are probably travelling longer distance to be admitted. AEL has a positive correlation with this share (0.13, p -value = 0.13) while RAM has a negative correlation (-0.16, p -value = 0.08) and RAR virtually no-correlation (-0.04, p -value = 0.64).¹⁹ According to this simple exercise, we do not expect AEL to be more prone to selection biases than traditional metrics, and mortality rates in particular. We leave to future research the adoption of alternative empirical designs to further test labor market outcome-based measures of healthcare quality.

¹⁹Notice that high AEL and low RAM and RAR indicate higher quality.

5 The Labor Market Costs of Hospital Admissions Over Time

5.1 Empirical Strategy

To estimate the variation in the labor market costs of hospital admission over time, we consider all inpatient “index” hospital admissions in Denmark during the years 1998-2012 (see section 2.1), together with a 10% control group of individuals without any index admissions during the same period. For each disease group, we consider the following equation:

$$y_{i,t,yg} = \alpha_i + X_{i,t}\beta + \gamma_t + \sum_{-3 \leq r \leq 3, r \neq -1} \delta_{r \times yg} + \epsilon_{i,r,t} \quad (9)$$

where α denotes individual fixed effects and $X_{i,t}$ is a vector of control variables. These includes age fixed effects, municipality- and education-specific unemployment rates and log median earnings in year t , and the share of individuals on disability benefits in the municipality in year t . We also include dummies for whether the patient is age 40-50 or above age 50 when hospitalized, a dummy for whether a hospitalized individuals had manual labor in time $r = -1$, fixed effects for the number of out-patient admissions and ER visits during the previous three years, fixed effects for the number of co-diagnoses, and fixed effects for the ICD10 chapter of the primary co-diagnosis related to the index-admission. γ_t denotes calendar-year fixed effects. $\delta_{r \times yg}$ are the parameters of interest, and these denote the “time relative to hospitalization” fixed effects interacted with the year-group of treatment: 1998-2000, 2001-2003, 2004-2006, 2007-2009 and finally 2010-2012. All year groups have relative-time reference group $r = -1$. We identify the change in labor market consequences of a specific disease as the difference in $\delta_{r \times yg}$ for $r > 0$ across different year groups.

As we include a control group, we cannot use normalized earnings as the outcome measure as in Section 3. Instead, we consider two different outcome measures: 1) a dummy indicator for earnings larger than 50,000 DKK, and 2) log earnings, conditional on earnings being larger than 50,000 DKK. While the first measure captures the extensive margin labor

participation consequences, the second captures the intensive-margin changes in earnings. Thus, the two measures should be interpreted together. If the extensive-margin drop in labor participation following a hospitalization decreases over time, this could potentially cause an increase in the intensive margin consequences over time, as better treatment could result in more individuals remaining in the labor force, but with lower earnings.

We categorize the different diseases by a clinical classification code, which groups our more than 9000 different 4-digit ICD10 codes into 271 diagnostic groups. Out of these, we consider the 150 groups with most observations, making up 97.5% of the total index admission observations, and each covering more than 850 index-observations.

A measure of the extensive-margin cost of a disease is given by $\delta_{r=3 \times yg}$ from equation 9, where the outcome variable is a dummy for whether earnings exceeds 50,000 DKK. We consider the relative time dummy for the 3rd year following the index admission, as the extensive-margin cost of a disease accumulates/continues to decline in the years following a hospital admission, as shown in Figure 1. As such, we pick the last year available within our 3-year index admission definition to capture the full accumulated effect of the different diseases.

Our measure of the intensive-margin cost of a disease is given by $\delta_{r=1 \times yg}$ as measured in equation 9, where the outcome variable is log earnings, and we only include earnings observations larger than 50,000 DKK. We consider the first year following the index admission, as this is when patients often experience the largest intensive-margin earnings loss, as earnings tend to recover for the less durable severities (as we document in section 2.3, Figure 1).

5.2 Results

Figure 9 shows the estimated $\delta_{r=3 \times yg}$ coefficients for all 5 year groups, covering the period 1998-2012. The diagnostic groups are sorted by their estimated severity in year 2010-2012, $\delta_{r=3 \times yg=2010-12}$. These coefficients measure the severity of the specific diseases, and the differences across the year groups capture the development of the disease's extensive-margin labor market consequences over time. A general finding across diseases is that there is a decline in the estimated severity over time, as the negative labor participation

consequences of the disease decrease for the later year groups. For most diseases, we even observe a steady decrease in the estimated severities, despite the relatively short time intervals of 3 years. It also seems that the decrease accelerated over time.

According to our estimation, acute cerebrovascular disease is the most severe in terms of the labor participation drop following a hospitalization. In 1998-2000, 19% of patients were dropped out of the labor force 3 years after their hospitalization. This number, however, decreased to only 6% in 2010-2012. Spondylosis (spinal degeneration) is also extremely costly for labor participation, where it caused 13% to drop out of the labor force in 1998-2012. This number dropped to 11% in 2010-12. As such, the reduction in labor participation cost is much smaller for Spondylosis compared to acute cerebrovascular disease. Other very severe diseases include poisoning by other medications and drugs (including e.g. aspirin overdose), poisoning by psychotropic agents, alcohol-related mental disorders and multiple sclerosis, causing roughly 10% of patients to drop out of the labor force.

However, since we also keep the individuals who die, giving them zero earnings, some of the change is driven by a decrease in mortality rates. Figure 13 repeats the same analysis, but exclude dead individuals. Here we find that 14% of the surviving patients dropped out of the labor force 3 years after their hospitalization for acute cerebrovascular disease in 1998-2000, compared to 9% in 2010-2012. These findings suggest that changes in mortality rates drive more than half of the estimated change in labor participation drop for this disease. Mortality explains a much smaller share for the less severe diseases.

Figure 10 shows the estimated intensive-margin severity measures given by $\delta_{r=1 \times yg}$ in equation 9 for all 5 year groups, covering the period 1998-2012 and including all of the 150 most prevalent diagnostic groups. These are sorted by the estimated intensive-margin severity in year 2010-2012, $\delta_{r=1 \times yg=2010-12}$. Compared to the extensive margin results presented in Figure 9, the intensive-margin costs of the different diseases are generally smaller. Also, while we do observe an overall decrease in disease severity over the years, the decrease is not as steady for the intensive margin as compared to the extensive margin. This, however, is not surprising, as a decrease in the extensive margin labor cost is likely to come at a “cost” with respect to the intensive margin labor cost. It seems reasonable

to assume that individuals who would have previously dropped out in the labor force, but now remain, do so with declining earnings.

Figure 12 shows the scatter plot of the estimated intensive- vs. extensive margin severities for the year groups 1998-2000 and 2010-2015, together with the 45 degree line. The plot shows that the order of the estimated severities are somewhat similar for the two margins: if a disease have large extensive-margin costs, it also have large intensive-margin costs. We also see that severities have declined in both dimensions, but where the extensive margins severity tended to dominate in year 1998-2000, it seems that the severities are more equal for the two margins in 2010-2012. As such, the decline in severity over time was largest for the extensive margin. Again, this is not very surprising, and it is in fact noteworthy that we observe such large decreases in the intensive-margin costs of the diseases, despite the fact that more individuals also remain in the labor force (large increases in the extensive margin).

For the extensive margin, we use a dummy variable for whether a person have positive earnings in a given year as the outcome variable and measure the change in extensive-margin consequences of a specific disease as $\delta_{r=1 \times yg=2010-2012} - \delta_{r=1 \times yg=1998-2000}$. We consider a dead person as a person with zero earnings, so that the extensive margin can be affected by changes in mortality rates. However, results are quantitatively similar if we drop dead people from the sample.

While Figure 9 and Figure 10 plot the severity estimated in only one year ($r = 3$ and $r = 1$ respectively), we know that the labor market outcome consequences of diseases lasts for several years. As such, when we measure the “accumulated” or total costs of a disease, it is reasonable to summarize the costs over several years following the hospitalization. We will compute the total change in the labor costs of a disease as the accumulated labor-costs for each of the 3 years following a hospitalization:

$$\text{Accumulated Change} = \sum_{r=1}^3 \delta_{r,yg=2010-2012} - \sum_{r=1}^3 \delta_{r,yg=1998-2000} \quad (10)$$

where the δ coefficients are given by equation 9. Figure 11 displays the estimated accumulated changes in severity over time, as measured by both the intensive and extensive margin, for all of the 150 most prevalent diseases. The plot confirms the previous

finding that the change in estimated severity declined more on the extensive participation margin compared to the intensive log earnings margin. For the diseases with the most significant changes in severity over time, the decline in the accumulated extensive margin severity varied from roughly 7-27 percentage points, while the decline in accumulated intensive-margin severity were more constant around 4-6 percentage points.

According to our estimates, lung cancer was the disease for which the biggest improvement happened. During the three years following a hospitalization for lung cancer, the accumulated decline in labor force participation decreased by 27 percentage points from 1998-2000 to 2010-2012 on the extensive margin. In general, we see that many diseases related to cancer have especially large declines in their labor cost over time, with secondary malignancies, breast- and colon cancer topping the list. The accumulated labor costs of acute cerebrovascular disease also declined by 20 percentage points on the extensive margin, and 10% on the intensive margin. For acute myocardial infarction (AMI), the change was 15 percentage points on the extensive margin and 8% on the extensive margin. Diseases that have seen only little or no significant decline in labor market costs are benign neoplasm of uterus, varicose veins of lower extremity and abdominal hernia.

Weighted by the number of patients during all the observed years 1998-2012, we estimate that the average accumulated 3-year decline in participation rates following a hospitalization to have decreased by 8.8 percentage points from 1998-2000 to 2010-2012. Similarly for the intensive margin, we estimate that the 3-year accumulated earnings decline following a hospitalization has dropped by 4.9 percentage points. This corresponds to an average decline in the labor force participation costs of a hospitalization of 50%, and a 25% reduction in corresponding intensive-margin earnings costs. The corresponding total change, computed as the sum of the intensive- and extensive margin changes, is 13.6 percentage points.

5.3 Discussion

The sizeable reduction of the labor market costs of hospital admissions suggests a strong improvement in the quality of healthcare treatments between 1998 and 2012. This overall improvement is likely to be the result of both advancements in global medical technology

(e.g. new drugs or medical devices) and improvements of the Danish hospitals' system.

However, part of the observed changes are probably caused by other factors than improved treatment quality. For instance, healthier behaviors of the general population (e.g. less smoking or a better diet) could also have led to higher resilience to diseases.

Moreover, changes in the labor market may also affect the evolution over time of the economic losses associated with hospital admissions. In the Appendix C, we investigate whether the introduction of the “flex-job” scheme in 1998 played a role in this improvement and find no evidence for that. We also propose a simple exercise to measure the lower-bound of the importance of healthcare-related factors in explaining the decline of the economic costs of hospitalizations. Given the assumption that factors not related to the improved treatment quality should depend only on the severity of the disease, and not the nature of the disease itself, a conservative estimate implies that at least 40% of the estimated change are due to improved healthcare treatment over time.

6 Conclusion

We introduce a novel measure of the quality of a hospital treatment based on its ability to mitigate the labor market consequences of a given diagnosis and we call it “Adjusted Earning Losses” (AEL).

To estimate AEL, we link the universe of hospital admissions in Denmark to tax data on the labor market outcomes of the entire working-age population in Denmark from 1995-2015. This allows us to measure the labor market consequences of any hospitalization, controlling for diagnosis, co-morbidities, and a very rich set of patient- and local labor market characteristics. We focus on patients’ “normalized earnings” that allow us to consider together intensive and extensive margins [Kleven et al., 2018]. To mitigate the problem of co-morbidities affecting the treatment outcomes, we consider all patients who experience their first non-pregnancy related inpatient admission in four years.

We find significant heterogeneity in the labor cost of a hospitalization across different hospitals, and our point estimates indicate that there is a 4 percentage point difference in lost earnings between the best and worst hospital, all else equal. AEL is positively correlated with traditional quality measures, such as risk-adjusted mortality and readmissions

rates but it contains substantial additional information.

We use AEL to investigate whether better hospitals are better at dealing with all medical conditions. We find a positive but low correlation between a hospital's abilities to treat patients admitted with different diagnoses, indicating substantial within-hospital heterogeneity.

As for traditional mortality and readmission based measures, AEL can be biased if unobserved patient characteristics are different across hospitals. Two main concerns that have been mentioned in the literature are 1) patients who are unobservably sicker select into the better hospitals, and thereby generate a downward bias on the estimated quality of the good hospitals, and 2) unobserved characteristics of patients are not evenly distributed across geographic regions and could bias the estimated quality of the hospitals located in unobservably "good"/"bad" neighborhoods. While we expect these concerns to be mitigated by the fact that we control for a broad range of patient characteristics (arguably broader than in previous papers), we perform two complementary empirical strategies to test whether being admitted to a higher quality hospital (according to our estimates) leads to a positive causal effect on post-admission earnings. Both tests suggest that the above mentioned concerns are not biasing our estimated quality measure in any significant way.

We also document a significant decline in the labor cost of hospitalizations over time. For different diseases, we compare the change in employment probability and labor earnings of patients after their hospital admissions in different years. We find that there has been a significant decline in both earnings and labor participation drops following a hospital admission during the period 1998-2012. More specifically, we find that the post-hospitalization reduction in labor supply has declined by 8.8% on the extensive (participation), and 4.9% on the intensive margin (earnings conditional on working) on average, as measured by the accumulated change in the three years following the admission. Thus, the total change in estimated labor costs amounts to 13.6 percentage points. This corresponds to an average reduction of extensive margin labor market costs of about 50%, and a 25% reduction of the intensive-margin labor cost of a disease.

References

- Altonji, J. G., Elder, T. E., and Taber, C. R. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy*, 113(1):151–184, 2005.
- Bloom, D. E., Canning, D., Kotschy, R., Prettner, K., and Schünemann, J. Health and economic growth: reconciling the micro and macro evidence. 2019.
- Chandra, A., Finkelstein, A., Sacarny, A., and Syverson, C. Health care exceptionalism? performance and allocation in the us health care sector. *American Economic Review*, 106(8):2110–44, August 2016.
- Chetty, R., Friedman, J. N., and Rockoff, J. E. Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9): 2593–2632, 2014.
- Currie, J. and Madrian, B. C. Health, health insurance and the labor market. *Handbook of labor economics*, 3:3309–3416, 1999.
- De Nardi, M., Pashchenko, S., and Porapakarm, P. The lifetime costs of bad health. Technical report, National Bureau of Economic Research, 2017.
- Deming, D. J. Using school choice lotteries to test measures of school effectiveness. *American Economic Review*, 104(5):406–11, 2014.
- Dobkin, C., Finkelstein, A., Kluender, R., and Notowidigdo, M. J. The economic consequences of hospital admissions. *American Economic Review*, 108(2):308–52, 2018.
- Doyle, J. J. Returns to local-area health care spending: Evidence from health shocks to patients far from home. *American Economic Journal: Applied Economics*, 3(3):221–43, 2011.
- Doyle Jr, J. J., Graves, J. A., Gruber, J., and Kleiner, S. A. Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy*, 123 (1):170–214, 2015.
- Fadlon, I. and Nielsen, T. H. Family labor supply responses to severe health shocks. *NBER Working Paper*, 2015.
- Finkelstein, A., Gentzkow, M., and Williams, H. Sources of geographic variation in health care: Evidence from patient migration*. *The Quarterly Journal of Economics*, 131(4): 1681–1726, 2016. doi: 10.1093/qje/qjw023. URL <http://dx.doi.org/10.1093/qje/qjw023>.
- Finkelstein, A., Gentzkow, M., and Williams, H. L. Place-based drivers of mortality: Evidence from migration. *NBER WP*, 2019.

- Garthwaite, C. L. The economic benefits of pharmaceutical innovations: The case of cox-2 inhibitors. *American Economic Journal: Applied Economics*, 4(3):116–37, July 2012. doi: 10.1257/app.4.3.116. URL <http://www.aeaweb.org/articles?id=10.1257/app.4.3.116>.
- Geweke, J., Gowrisankaran, G., and Town, R. J. Bayesian inference for hospital quality in a selection model. *Econometrica*, 71(4):1215–1238, 2003.
- Gilleskie, D. B. A dynamic stochastic model of medical care use and work absence. *Econometrica*, pages 1–45, 1998.
- Gowrisankaran, G. and Town, R. J. Estimating the quality of care in hospitals using instrumental variables. *Journal of Health Economics*, 1999.
- Grossman, M. On the concept of health capital and the demand for health. *Journal of Political economy*, 80(2):223–255, 1972.
- Gupta, A. Impacts of performance pay for hospitals: The readmissions reduction program. *WP*, 2017.
- Hamilton, B. H., Hincapié, A., Miller, R. A., and Papageorge, N. W. Innovation and diffusion of medical treatment. *Technical report, Working Paper*, 2016.
- Heckman, J. J. and Vytlacil, E. Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica*, 73(3):669–738, 2005.
- Hull, P. Estimating hospital quality with quasi-experimental data. *WP*, 2018.
- Imbens, G. W. and Angrist, J. D. Identification and estimation of local average treatment effects. *Econometrica (1986-1998)*, 62(2):467, 1994.
- Kleven, H., Landais, C., and Søgaaard, J. E. Children and gender inequality: Evidence from denmark. *NBER working paper*, 2018.
- Laird, J. and Nielsen, T. The effects of physician prescribing behaviors on prescription drug use and labor supply: Evidence from movers in denmark. *Harvard University Working Papers*, 2016.
- Lichtenberg, F. R. The long-run impact of new medical ideas on cancer survival and mortality. *NBER Working Paper*, 2018.
- Miller, G., Luo, R., Zhang, L., Sylvia, S., Shi, Y., Foo, P., Zhao, Q., Martorell, R., Medina, A., and Rozelle, S. Effectiveness of provider incentives for anaemia reduction in rural china: a cluster randomised trial. *BMJ*, 345:e4809, 2012.
- Oster, E. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204, 2019.
- Shapiro, B. Promoting wellness or waste? evidence from antidepressant advertising. *Becker Friedman Institute for Research in Economics Working Paper*, (2018-14), 2018.

Well, D. N. Accounting for the effect of health on economic growth. *The quarterly journal of economics*, 122(3):1265–1306, 2007.

Tables

Table 1: Descriptive Statistics

	All Admissions			Index Admissions			Control
	r=-1	r=0	r=1	r=-1	r=0	r=1	-
Mean earnings	204,742	201,737	196,661	327,639	313,750	307,559	301,963
Mean log earn earn > 50,000	12.48	12.49	12.53	12.68	12.69	12.70	12.70
Share w/ earn > 50,000	0.678	0.642	0.611	0.915	0.872	0.840	0.82
Share alive	1,000	0,984	0,966	1,000	0,989	0,981	0.97
Mean age	42.83	43.84	44.84	42.42	43.42	44.40	42.38
# Observations	3,104,532	3,104,632	3,099,490	1,098,162	1,113,321	1,111,576	3,312,609

The table shows descriptive statistics for three different samples. r specifies the year of the observation relative to the hospitalization year. We observe a total of 3.1 million non-birth related index admission observations of patients born in Denmark aged 25-59 at the time of hospitalization in year 1998-2012. This is the sample which we refer to as “all admissions”. Counting only one admission per person per year, we have approximately 2 million observations. The index admission sample is the sample we use for our main analysis, and includes the subsample of patients who experience the first observed non-pregnancy related inpatient hospital admission for at least 3 years. The control group contains a 10% random sample of everyone who didn’t experience an index admission, with similar birth-year and age restrictions as in the index-admission sample.

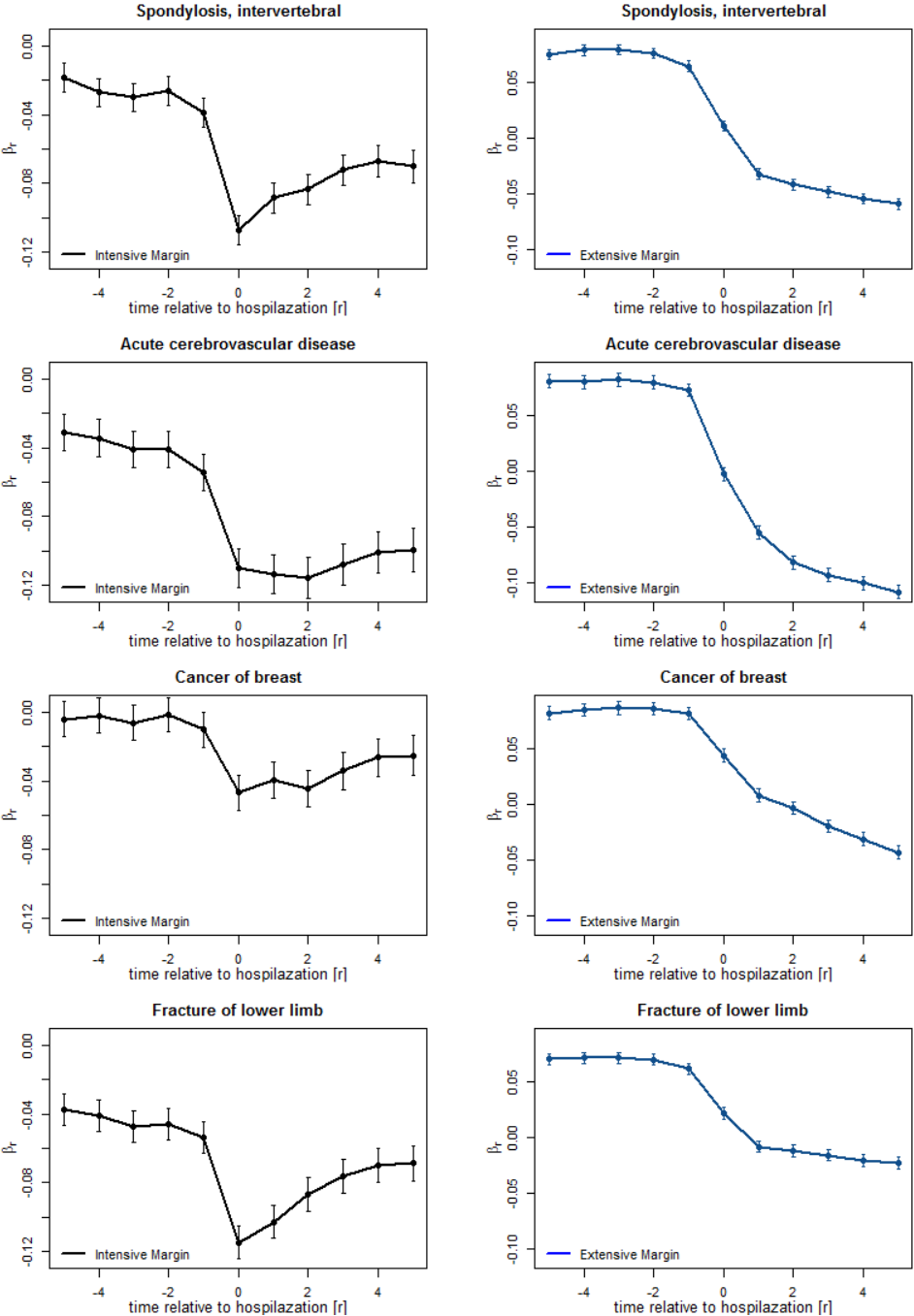
Table 2: Robustness of AEL

VARIABLES	Normalized earnings	Normalized earnings	Normalized earnings
Estimator	IV	OLS	OLS
	(1)	(2)	(3)
AEL of Admission	1.003***	1.199***	1.387**
Hospital	(0.0263)	(0.250)	(0.277)
AEL of “Common”			0.206
Hospital			(0.277)
Church District FEs		×	
Controls	×	×	×
Observations	3,113,831	59,912	52,186
Sample	All (top 31 hospitals)	Non-deferrable acute not at “Common” hospital	Non-deferrable acute not at “Common” hospital
R-squared	0.165	0.243	0.242

Column (1): results of estimating equation $y_{i,r} = \alpha \cdot \widehat{q}_{h,t} + X_{i,r}\beta + \eta_{i,r}$, where y are the normalized earnings of patient i in year (with respect to hospitalization) r . Labor earnings are normalized by the average labor earnings during the three years before the hospital admission. $\widehat{q}_{h,t}$ is the estimated AEL of the hospital h in period t , see equation (3). $\widehat{q}_{h,t}$ is treated as an endogenous variable and it is instrumented with a vector of dummy variables indicating the neighborhood (Church District) where patient i lives during the admission year. For sample selection criteria and details about the vector of controls, see sections 3 and 4 and, in particular, equation (6). Column (2): results of estimating equation $y_{i,r} = \alpha \cdot \widehat{q}_{h,t} + \tau_n + X_{i,r}\beta + \epsilon_{i,r}$. It differs from column (1) because of the inclusion of a full set of fixed effects to control for the neighborhood where patient i lived during the hospitalization year. Sample includes only patients admitted for non-deferrable acute conditions to any hospital different from the most common one among patients coming from the same neighborhood and having similar diagnosis. Estimation performed via OLS. For details, see section 4 and equation (7) in particular. Column (3): results of estimating equation $y_{i,r} = \alpha \cdot \widehat{q}_{h,t} + \theta \cdot \widehat{q}_{n,dg,t} + X_{i,r}\beta + \epsilon_{i,r}$. It differs from column (2) because the neighborhood fixed effects are substituted with $\widehat{q}_{n,dg,t}$, which is the estimated quality of the most common hospital for patients coming from the same neighborhood as i and having similar diagnosis. Observations in column (3) are less than column (2) because the quality is estimated only for the top 31 hospitals. Estimation performed via OLS. For details, see section 4 and equation (8) in particular. Standard errors in parentheses are clustered at the municipality level.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

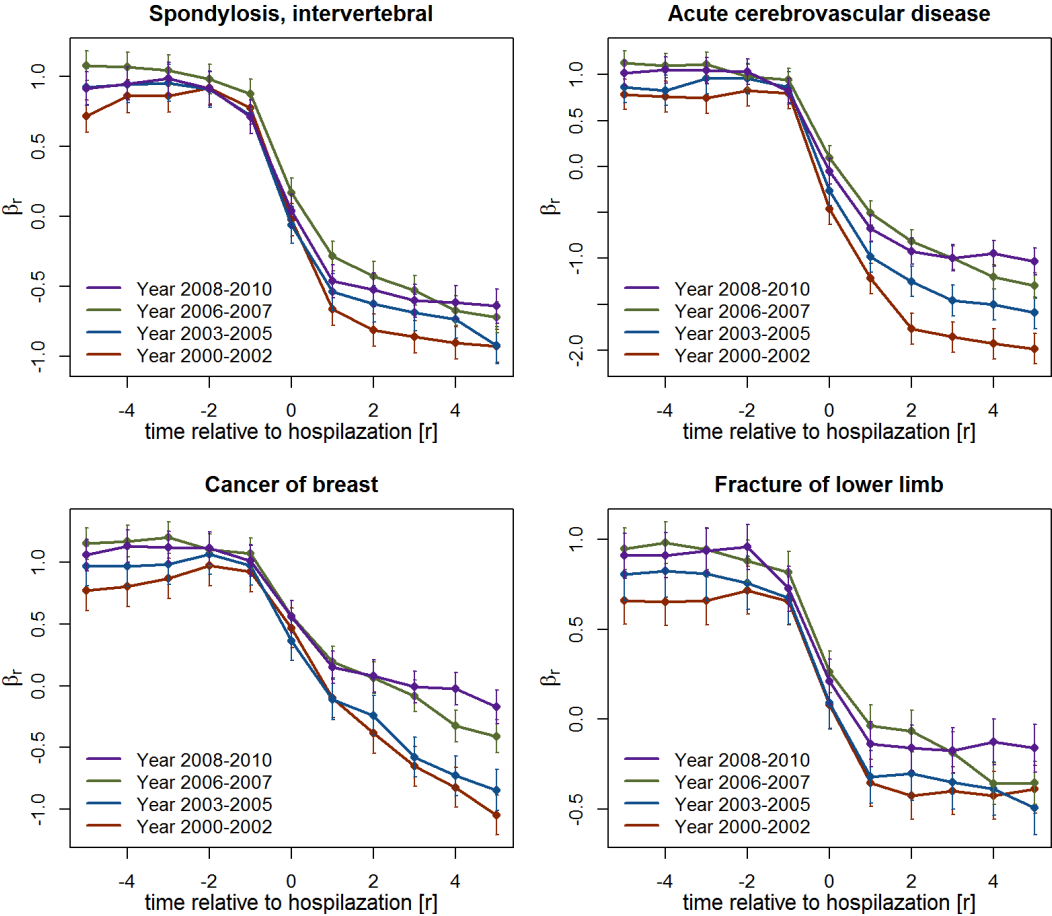
Figures

Figure 1: Labor supply response to an index hospital admission



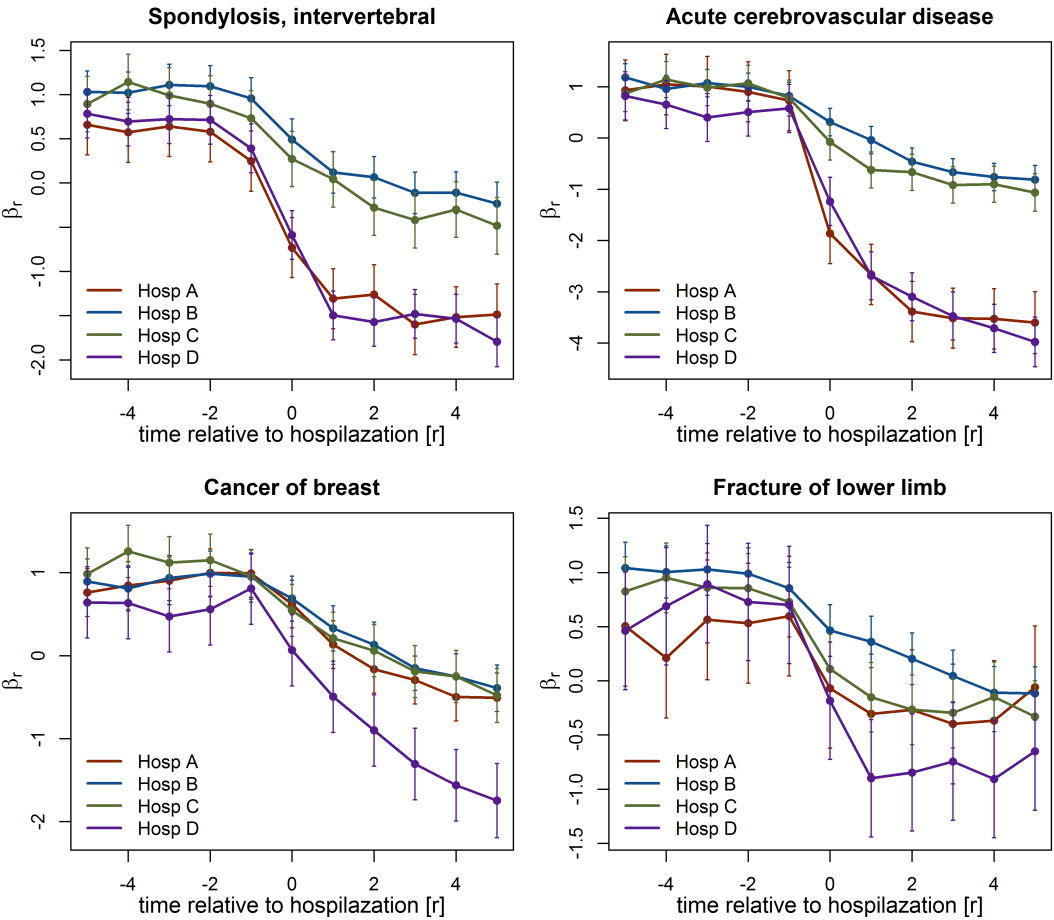
The intensive margin plots show the β_r coefficients from equation 1 with outcome variable $\log(\text{earnings})$ (excluding non-positive earnings observations). The extensive margin plot show the β_r coefficients from equation 1 with the dummy outcome variable for positive earnings. The β_r coefficients estimate the differences in labor market outcome for individuals hospitalized in year $t - r$ with respect to other (non-hospitalized) individuals with similar characteristics. Hospital admission year is 0. See section 2.3 for details.

Figure 2: Labor supply response to an index hospital admission: heterogeneity over time



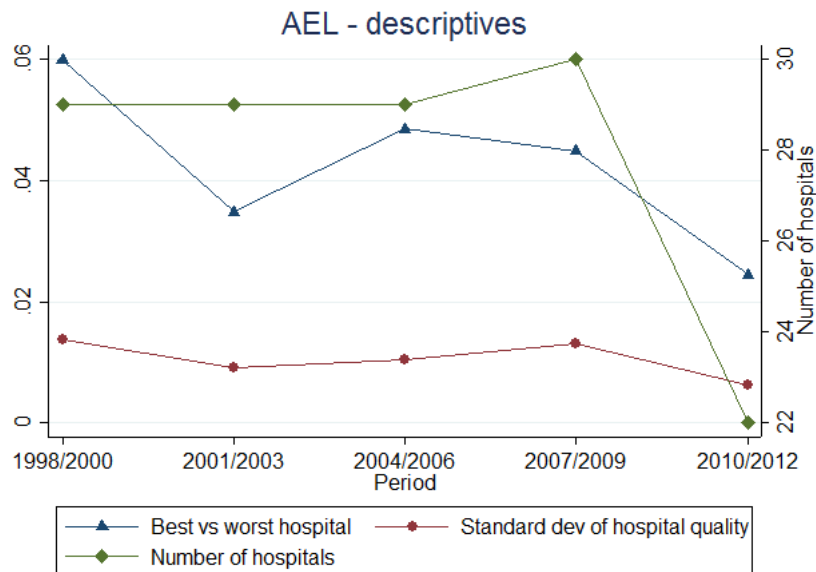
Event Study of log earnings before and after an “index”-hospital admission. Hospitalization year is 0. Each line refers to patients hospitalized in a given period. See section 2.3 for details.

Figure 3: Labor supply response to an index hospital admission: heterogeneity across hospitals



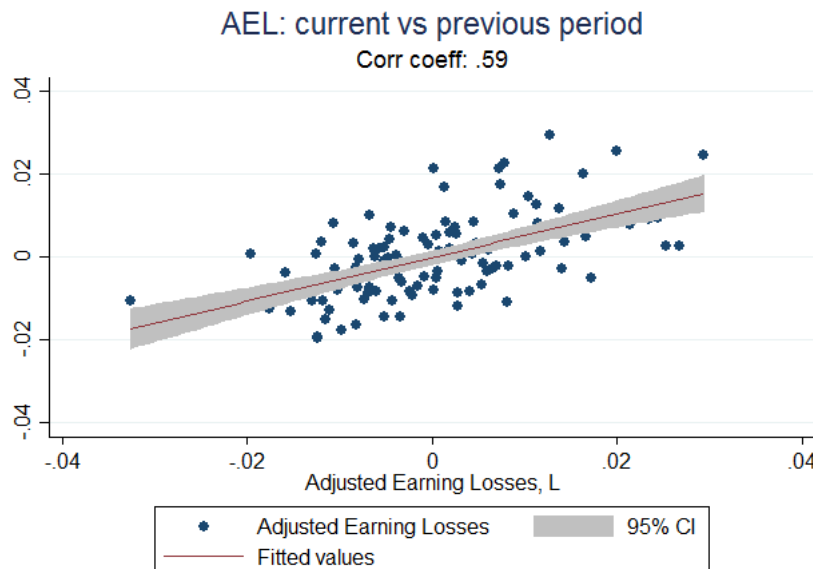
Event Study of log earnings before and after an “index”-hospital admission. Hospitalization year is 0. Each line refers to patients hospitalized in a given hospitals. See section 2.3 for details.

Figure 4: Descriptive statistics of AEL



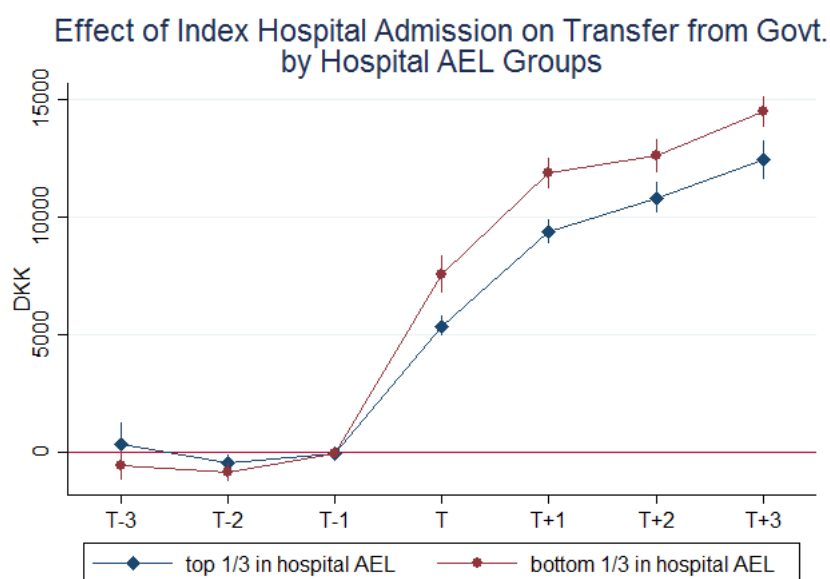
Descriptive statistics of AEL estimates for each period. Left axis is in percentage points and it refers to standard deviation and range of hospital quality. Right axis is in units and it refers to the number of hospitals. See Section 3.1 and Section 3.2 for details.

Figure 5: AEL: present versus previous period



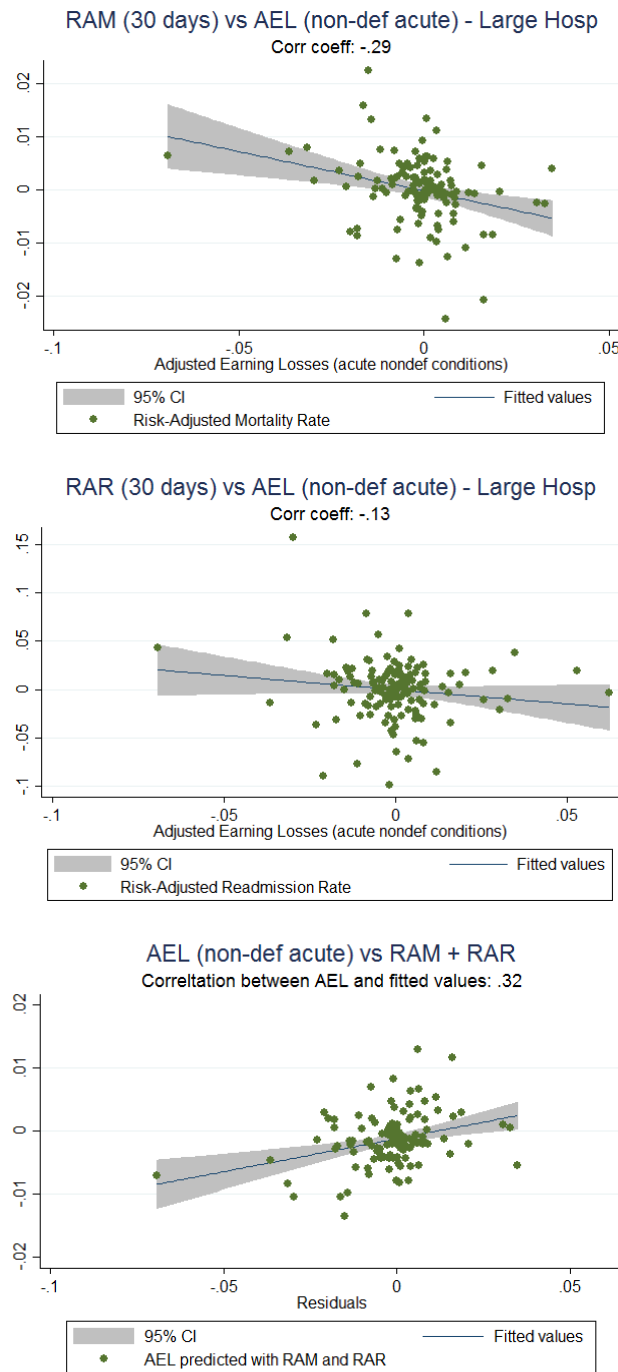
Scatterplot and line of best-fit together with 95% confidence interval for AEL and its lag. We pool together all years and AEL values are normalized by subtracting the period-specific average. See Section 3.1 and Section 3.2 for details.

Figure 6: Event Study: AEL and Transfers from Government to Patients



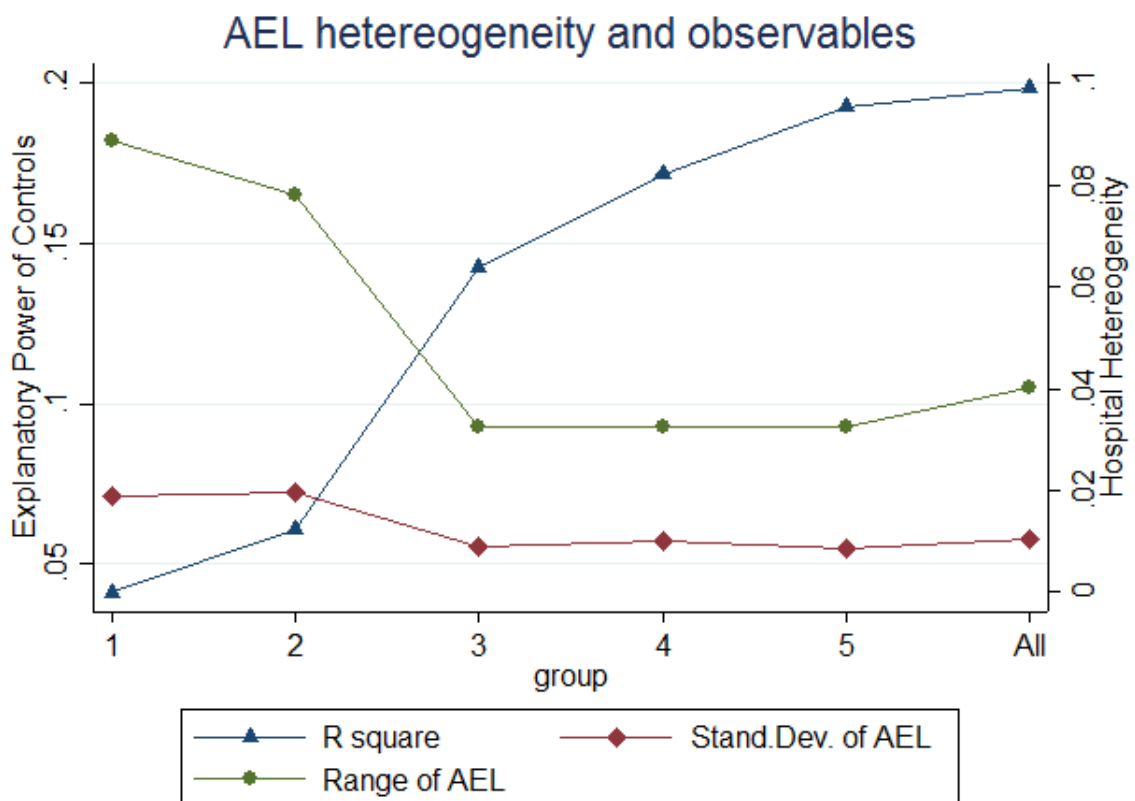
Estimated impact of an index hospital admission on the transfers from government received by the patient. The effect is estimated separately for patients admitted to hospitals in the top and bottom third of quality as measured by AEL. The estimating equation includes individual and time fixed effects, together with fixed effects for the time relative to the year of hospital admission (reported in the figure). The coefficient relative to the year before hospital admission is normalized to zero. See Section 3.3 for more details.

Figure 7: AEL and Traditional Measures



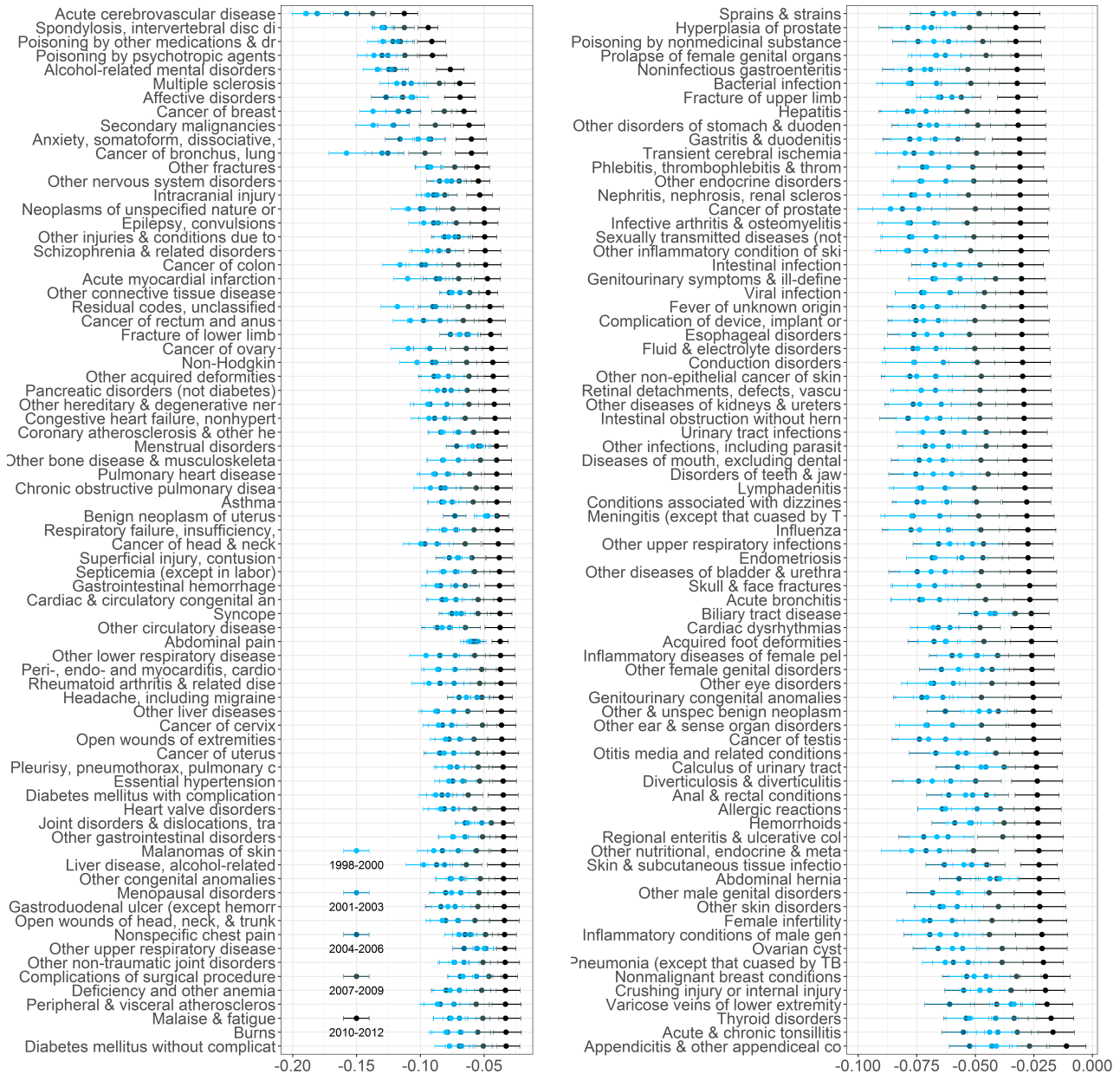
Scatterplot and line of best-fit together with 95% confidence interval for AEL and traditional quality measures: 30-days Risk-Adjusted Mortality (RAM, upper panel), 30-days Risk-Adjusted Readmission rates (RAR, middle panel), and the predicted values obtained by regressing AEL on RAM and RAR (bottom panel). We pool all years together and all quality measures are normalized by subtracting the period-specific average. See Section 3.1 and Section 3.5 for details.

Figure 8: AEL Heterogeneity and Explanatory Power of Observable Covariates



Range and standard deviation of AEL estimates together with the R-square of the regression used to estimate hospital quality (equation 3). Each “group” includes a larger set of controls, with group 1 including only main disease fixed effects and group "All" including all the controls. See Section 4.3 for more details.

Figure 9: Estimated labor force participation loss three years after a hospitalization by disease and year of treatment



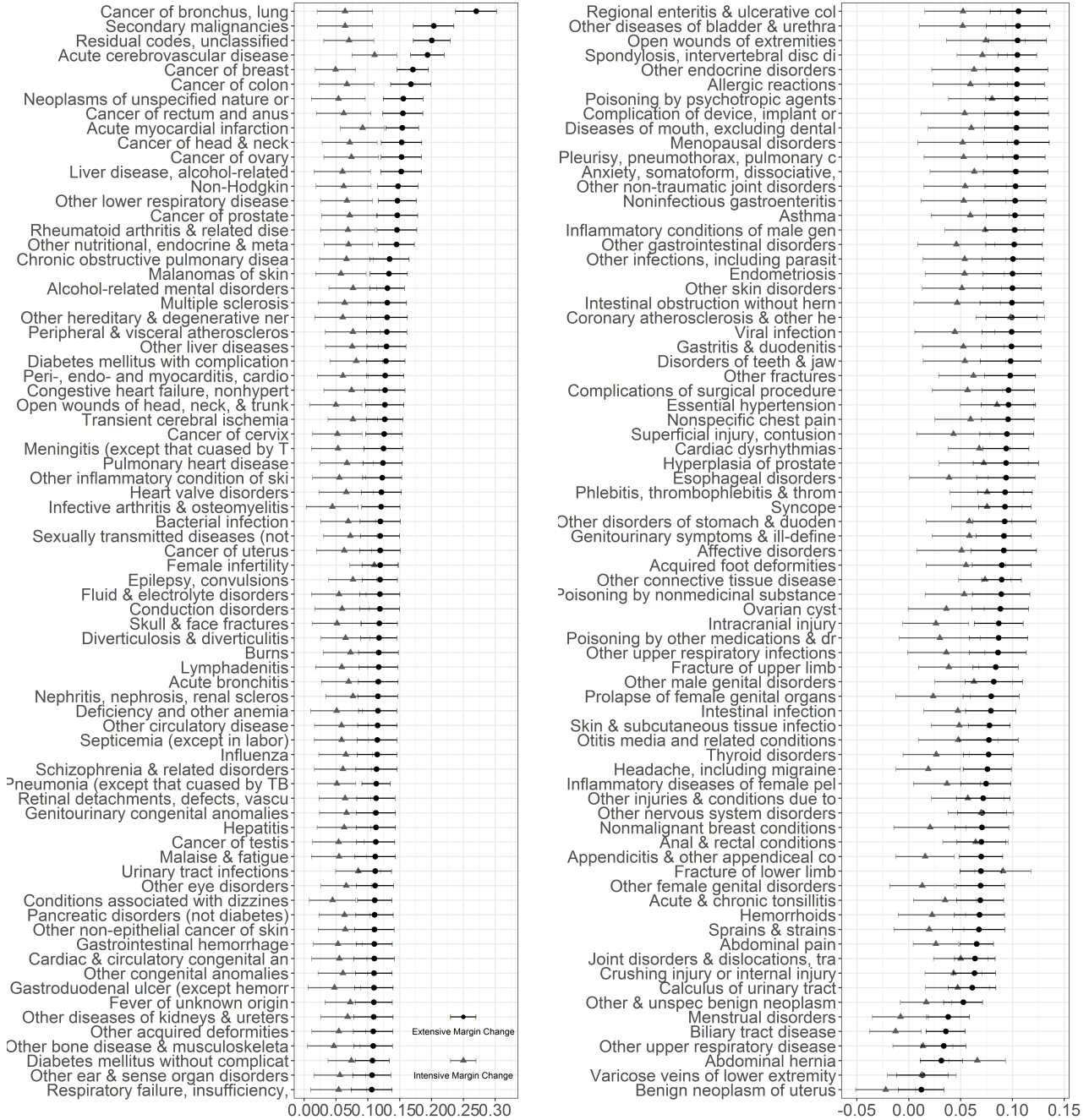
The figure shows, for each of the 150 most frequently observed disease groups, the difference in participation rates 3 years after an index admission. This difference is estimated in equation 9, where the outcome variable is a dummy for whether if earnings in a given year exceeds 50,000 DKK, and equals the estimated parameter of $\delta_{3 \times yg}$.

Figure 10: Estimated earnings loss in year r=1 by disease and year of treatment



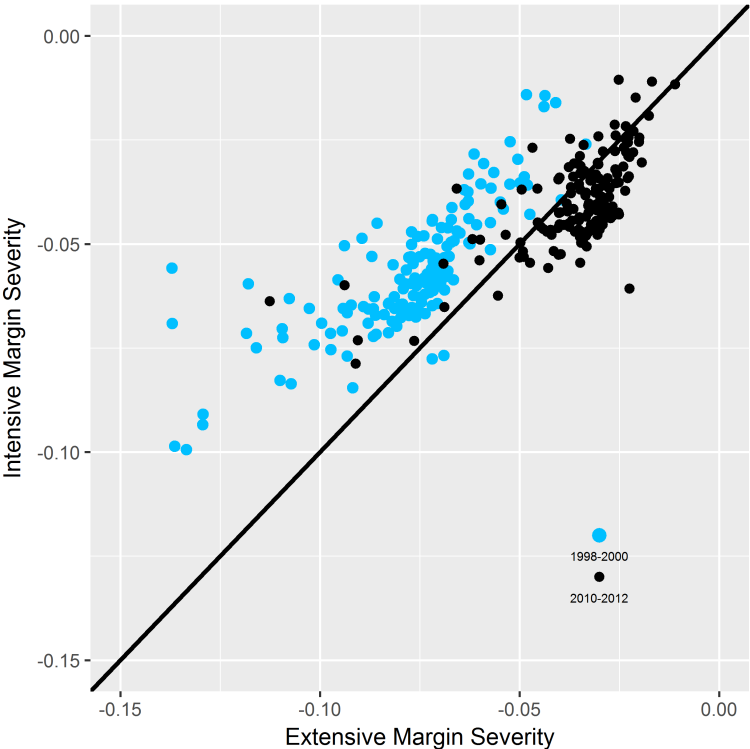
The figure shows, for each of the 150 most frequently observed disease groups, the difference in participation rates 1 year after the index admission. This difference is estimated in equation 9, where the outcome variable is log earnings, conditional on earnings exceeding 50,000 DKK, and equals the estimated parameter of $\delta_{1 \times yg}$.

Figure 11: Change in accumulated labor market cost of a hospitalization by disease group, 1998-2001 to 2010-2012, $r \in \{1, 2, 3\}$



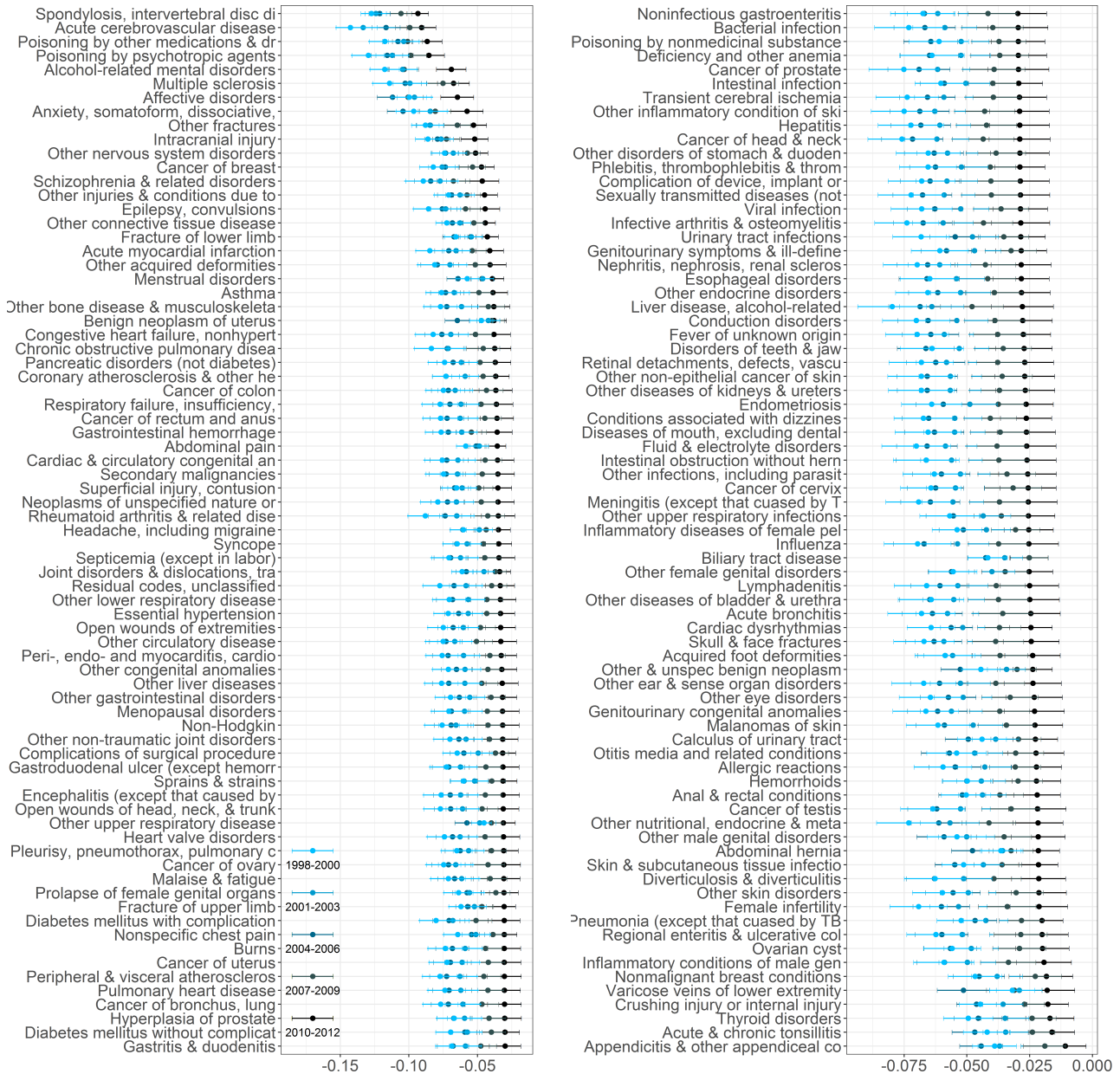
The figure shows the estimated accumulated changes in severity over time, both for the intensive and extensive margin, for all of the 150 most prevalent diseases. We measure the accumulated change as $\sum_{r=1}^3 \delta_{r, yg=2010-2012} - \sum_{r=1}^3 \delta_{r, yg=1998-2001}$, where the δ coefficients are estimated in equation 9. The black dots denote the change in β_5 over time with 95% confidence intervals, and the gray triangles and lines denote the difference as measured by the intensive margin (log earnings conditional on working).

Figure 12: Histogram of estimated pathology-specific improvements, 1998-2000 to 2008-2010



The scatter plot plots the estimated severity measured by its intensive-margin labor costs vs. its extensive margin labor cost. The y-axis denotes the intensive cost and is measured in equation 9, where the outcome variable is log earnings, conditional on earnings exceeding 50,000 DKK, and equals the estimated parameter of $\delta_{r=1 \times yg=2010-2012}$. The x-axis denotes the extensive cost of the disease and is measured in equation 9, where the outcome variable is a dummy for whether if earnings in a given year exceeds 50,000 DKK, and equals the estimated parameter of $\delta_{r=3 \times yg=2010-2012}$

Figure 13: Estimated labor force participation loss in year $r=3$ by disease and year of treatment, excluding deaths



The figure shows, for each of the 150 most frequently observed disease groups, the difference in participation rates 3 years after an index admission. This difference is estimated in equation 9, where the outcome variable is a dummy for whether if earnings in a given year exceeds 50,000 DKK, and equals the estimated parameter of $\delta_{r=3 \times yg}$

A Alternative Hospital Sample

The discharge data contains more than 100 hospital codes. Many of these codes related to hospices or small clinics specialized on a narrow set of procedures. However, for simplicity, we would like to estimate AEL on a set of similarly-sized hospitals which treats patients admitted with broad set of conditions.

We therefore focus the main analysis of Section 3 on the 31 largest hospital in Denmark. At most 30 of these are operating at any point in time, and only 22 at the end of the sample. These hospitals account for almost 90% of inpatient admissions.

We have no specific reason to believe that the meaningfulness (or usefulness) of AEL hinges on the sub-set of hospitals it is applied to. Nonetheless, we want to understand whether the main results of the paper are sensitive to the sample selection criteria. Therefore, we expand the sample to consider a larger set of 60 hospitals, which account for 98% of inpatient admissions. Both the main sample and this expanded ones are chosen by eyeballing the CDF of hospital admissions (see Figure A.4) and trying to create a more homogeneous set of hospitals.

Unsurprisingly, we find a (slightly) larger heterogeneity in hospital quality across all the time periods, as documented by Figure A.5. The average standard deviation of AEL is 1.2 percentage point. An inpatient admission in the lowest-ranked hospital is associated with an additional decrease in annual labor earnings of 5.1 percentage points, on average, compared to the highest-ranked hospital. The reduction in the number of active hospitals is stronger, since only 26 are active at the end of the sample against the 50 of the first period. The correlation between AEL and its lag (on a three-years period scale) is 0.68 in this sample, see Figure A.6.

We compare AEL to RAM and RAR and we find a stronger correlation than in the smaller sample of large hospitals. Focusing on AEL estimated for acute non-deferrable conditions, we find that it has a correlation of -0.51 with RAM (p -value = 0.000) and -0.14 with RAR (p -value = 0.047), see Figure A.7. The correlation between the actual AEL and the predicted values (obtained regressing AEL on RAM and RAR) is 0.52: half of the variation in AEL cannot be explained by mortality and readmission rates.

We also repeat the robustness exercises of section 4 and find similar results, as reported in Table A.1 and Figure A.8.

B Location IV: Simulation

We simulate a simple model of endogenous hospital selection and heterogeneous hospital quality to clarify the exercise presented in section 4 (equation 6).

We consider a population of N patients, who are equally divided between neighborhoods A and B . There are two hospitals in this county, hospital 1 with quality $q_1 > 0$ and hospital 0 with quality normalized to 0. The two neighborhoods differ only because A is closer to hospital 1, while B is closer to hospital 0. The normalized earnings for patient i admitted to hospital h is:

$$Y_i = c_0 + q_1 \cdot \mathbf{1}_{h=1}(i) + \epsilon_i \quad (11)$$

where $\mathbf{1}_{h=1}(i)$ is a dummy variable indicating whether the patient i is admitted to hospital 1 (takes value 0 if admitted to hospital 0), c_0 is a constant, and ϵ_i is a normally distributed random shock.

We assume the hospital selection process takes the form:

$$\mathbf{1}_{h=1}(i) = 1 \iff \Phi(\eta_i + c_1 \cdot \mathbf{1}_{n=A}(i) + G \cdot \epsilon_i) > \frac{1}{2}$$

where $\Phi(\cdot)$ is the CDF of a standard normal distribution, $\mathbf{1}_{n=A}(i)$ is a dummy variable indicating whether patient i lives in neighborhood A , η_i is a random shock affecting hospital choice (uncorrelated with ϵ_i). $c_1 > 0$ means that patients from neighborhood A are more likely to go to hospital 1.

The parameter G disciplines endogenous hospital selection, that is whether hospital 1 admits “worse” or “better” patients. If $G = 0$ the patients admitted to both hospitals are equal. If $G < 0$ then the best hospital admits sicker patients. If $G > 0$ then it admits more resilient patients.

As in section 3, the researcher estimates q_h by applying OLS to equation (11). Let the resulting AEL estimate be $\hat{q}_h = \hat{q}_1 \cdot \mathbf{1}_{h=1}(i)$. Then, as in section 4, she uses patient location as an instrument for hospital quality, in order to test the robustness of her AEL measure. That is, she consider the linear model:

$$Y_i = a_0 + \alpha \cdot \hat{q}_h + \zeta_i$$

She instruments the variable \hat{q}_h with the dummy $\mathbf{1}_{n=A}(i)$. Let the resulting estimate for α be α_{IV} , while the OLS estimate is α_{OLS} . How can we use these estimates to get information on the underlying selection process?

We set $q_1 = 1$ and choose arbitrary values for the other parameters.²⁰ We let the selection parameter G take value on a grid from -0.3 to 0.3 . We simulate the model 200 times for each value of G , compute \hat{q}_1 , α_{IV} , and α_{OLS} and plot the sample average.

Results are presented in Figure A.9. The solid line indicates the average AEL estimate \hat{q}_1 . As expected, OLS estimates of AEL recover the true parameter only in absence of endogenous selection (i.e. $G = 0$). When $G < 0$, the quality of hospital 1 is underestimated because it treats sicker patients, while when $G > 0$ it is overestimated. The stars indicates α_{OLS} , which is mechanically equal to 1 for all simulations.

The dotted line represents the average α_{IV} . As expected, when there is no endogenous selection, and the estimated quality is unbiased for the true quality, the parameter is equal to 1. That is, being admitted to a hospital with quality $\hat{q}_1 = 1$ has actually had an effect of size \hat{q}_1 . When $G > 0$ (and therefore $\hat{q}_1 > q_1$), the average α_{IV} is below 1; in fact, being admitted to the good hospital has an effect smaller than what is predicted by the OLS quality estimates. The opposite is true when $G < 0$.

²⁰Namely, $c_0 = 1$, $c_1 = .5$, $N = 20,000$, ϵ and η distributed as *iid* standard normal.

C Labor Market Reforms and the Decline of the Economic Costs of Hospital Admissions

A potential driver of the observed reduction in labor market consequences of hospitalization could be the introduction of the flex-job scheme in 1998. The flex job scheme provides wage subsidies for individuals with reduced working capacity, to enable partly disabled individuals to enter the labor force. As the scheme grew in popularity during the 2000’s, it could explain part of the decrease in participation rates over time. As we’re able to identify everyone who participate in the flex-job program in our data, we’re able to run a “counterfactual” robustness check to investigate whether the flex-job scheme itself is driving the observed improvement: we assume that all individuals who participated in the flex job scheme would have gone one full disability benefits instead, and have 0 earnings. Thereby we assume the “worst case” scenario of what would have happened, had the flex-job scheme not been introduced. Under this counterfactual scenario, however, we do not find very different results. In fact, we estimate slightly larger declines in both the intensive- and extensive labor market responses following a hospitalization. As such, we do not expect the introduction of this particular scheme to drive our results. This is probably due to the fact that the share of index-admission patients participating in the flex-job scheme is quite low. For index-admission patients hospitalized in 2000, 0.4% participated in the program in 1999 and 1% in 2001. For index-admission patients who were hospitalized in 2011, 3.3% participated in 2010 vs. 3.7% in 2012.

However, there have also been other labor market regulations targeted the sick, and several labor market reforms specifically aimed at decreasing sick leave were implemented during the period which we consider. Improved health and cultural changes (e.g. the common perception of how sick one should be to not work, or the pressure from the employers to return to work faster) could also explain some of the change.

While these factors might vary across disease severity - e.g. reforms targeting long-term sick leave, we expect them to have the same effect across different diagnostic groups, conditional on the disease severity. If we assume that the health care quality of all diseases have improved or remained constant over time, then we can compute a “back-of-the-envelope” lower bound estimate of how much of the improvement is due to increased treatment quality. Let the change in labor costs of a diagnostic group, $\Delta LC(dg)$ equal the sum of the change in treatment quality of the diagnostic group $\Delta TQ(dg)$ and other “labor market” factors, which are only specific to the severity of the disease $LM(severity(dg))$:

$$\Delta LC(dg) = \Delta TQ(dg) + \Delta LM(severity(dg)) \quad (12)$$

For each defined “severity” class, we consider the diagnostic group which had the lowest decline in estimated severity over time. Now, as a lower bound (we assume $TQ \geq 0$), we assume that the treatment quality of that given diagnosis did not improve at all ($\Delta TQ(dg) = 0$). Thus, we attribute the entire observed change in severity over time to other factors, ΔLM . As we assume that the contribution from ΔLM is severity dependent, but don’t vary for different diagnostic groups *within* a severity group, we’re able to identify $\Delta LM(severity)$ as:

$$\Delta\text{LM}(\text{severity}) = \min_{dg \in \text{severity}} \Delta\text{LC}(dg) \quad (13)$$

Of course, the coarseness of the split in severities will affect how much of the change in labor cost which we attribute to improvements of treatment quality, with a finer split leading to smaller changes in treatment quality. We consider the “total” labor cost to be the sum of the extensive- and intensive margin labor costs, and the severity to be the sum of intensive- and extensive margin labor costs in 1998-2000. According to our estimation, the total 3-year accumulated labor costs declined by 13.6 percentage points over the period. Relying on the back-of-the envelope strategy presented in equations 12 and 13 and assuming a split of severities in 3 equally large groups, we find that 8.8 percentage points come from health care improvements. For 5 groups of severities, this number drops to 7.7 percentage points, and 10 groups gives us 5.3 percentage points. As such, a conservative estimate would imply that improvements in healthcare quality alone have caused the labor costs of a hospitalization to decrease with at least 5.3 percentage points.

D Additional Tables

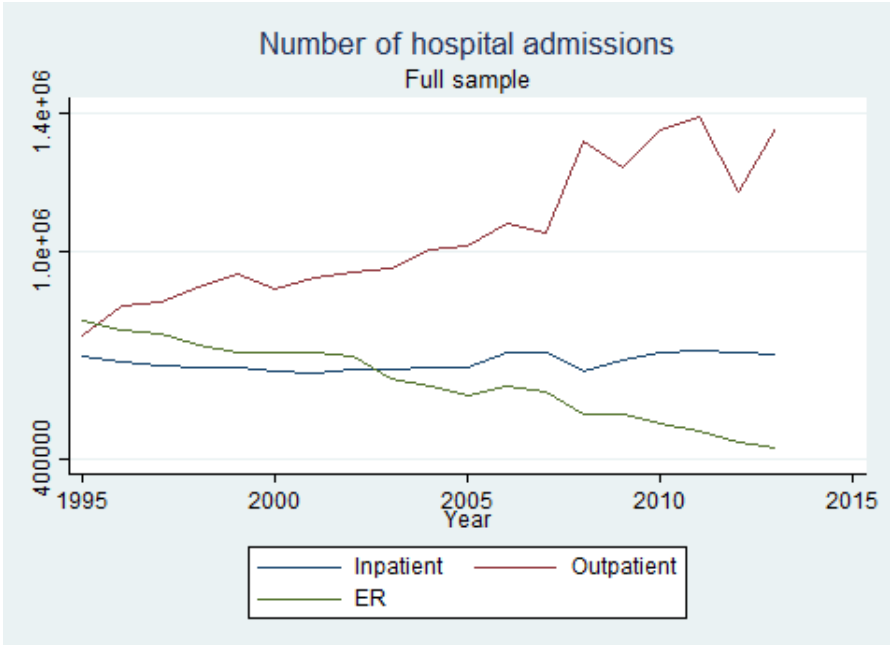
Table A.1: Robustness of AEL - Alternative Sample

VARIABLES	Normalized earnings	Normalized earnings	Normalized earnings
Estimator	IV	OLS	OLS
	(1)	(2)	(3)
AEL of Admission	1.012***	1.199***	1.374**
Hospital	(0.0223)	(0.223)	(0.223)
AEL of “Common”			-0.116
Hospital			(0.212)
Church District FEs		×	
Controls	×	×	×
Observations	3,727,345	71,396	69,471
Sample	All (top 60 hospitals)	Non-deferrable acute not at “Common” hospital	Non-deferrable acute not at “Common” hospital
R-squared	0.165	0.237	0.185

Column (1): results of estimating equation $y_{i,r} = \alpha \cdot \widehat{q_{h,t}} + X_{i,r}\beta + \eta_{i,r}$, where y are the normalized earnings of patient i in year (with respect to hospitalization) r . Labor earnings are normalized by the average labor earnings during the three years before the hospital admission. $\widehat{q_{h,t}}$ is the estimated AEL of the hospital h in period t , see equation (3). $\widehat{q_{h,t}}$ is treated as an endogenous variable and it is instrumented with a vector of dummy variables indicating the neighborhood (Church District) where patient i lives during the admission year. For sample selection criteria and details about the vector of controls, see sections 3 and 4 and, in particular, equation (6). Column (2): results of estimating equation $y_{i,r} = \alpha \cdot \widehat{q_{h,t}} + \tau_n + X_{i,r}\beta + \epsilon_{i,r}$. It differs from column (1) because of the inclusion of a full set of fixed effects to control for the neighborhood where patient i lived during the hospitalization year. Sample includes only patients admitted for non-deferrable acute conditions to any hospital different from the most common one among patients coming from the same neighborhood and having similar diagnosis. Estimation performed via OLS. For details, see section 4 and equation (7) in particular. Column (3): results of estimating equation $y_{i,r} = \alpha \cdot \widehat{q_{h,t}} + \theta \cdot \widehat{q_{n,dg,t}} + X_{i,r}\beta + \epsilon_{i,r}$. It differs from column (2) because the neighborhood fixed effects are substituted with $\widehat{q_{n,dg,t}}$, which is the estimated quality of the most common hospital for patients coming from the same neighborhood as i and having similar diagnosis. Estimation performed via OLS. For details, see section 4 and equation (8) in particular. Standard errors in parentheses are clustered at the municipality level.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

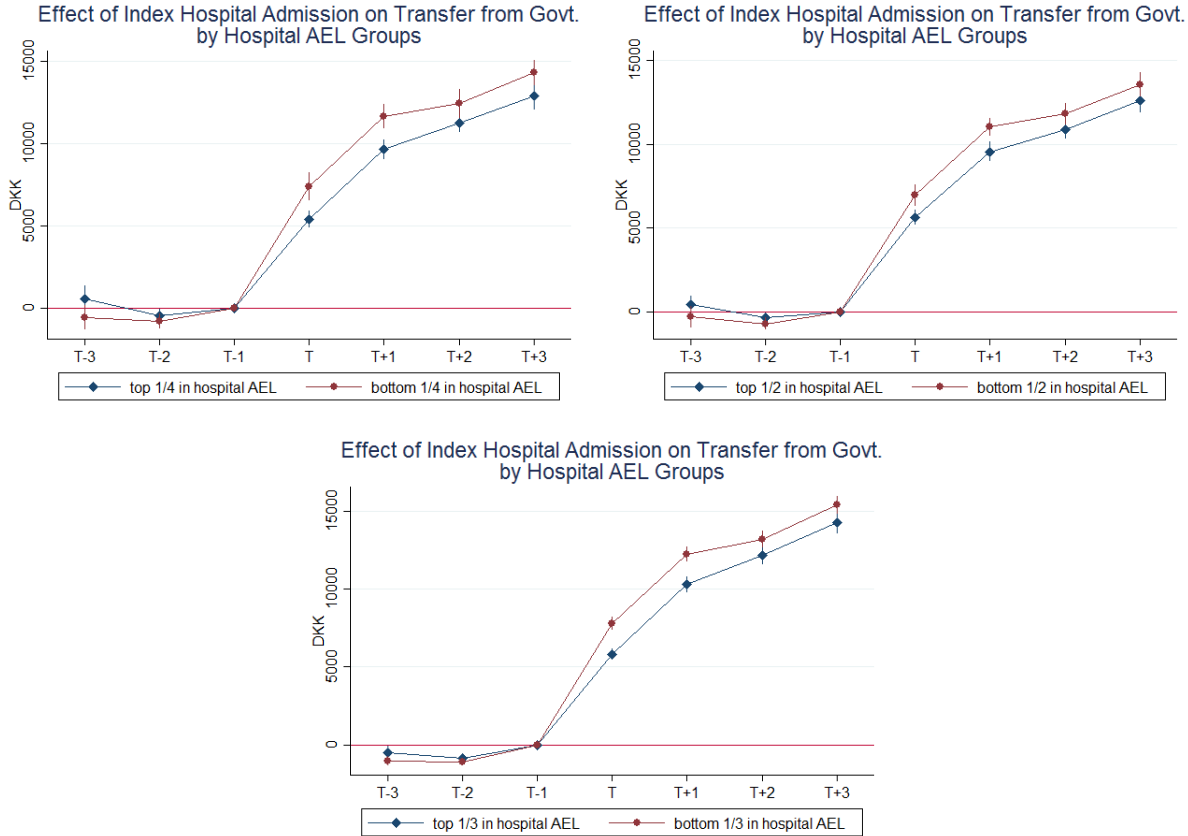
E Additional Figures

Figure A.1: Number of hospital records by type



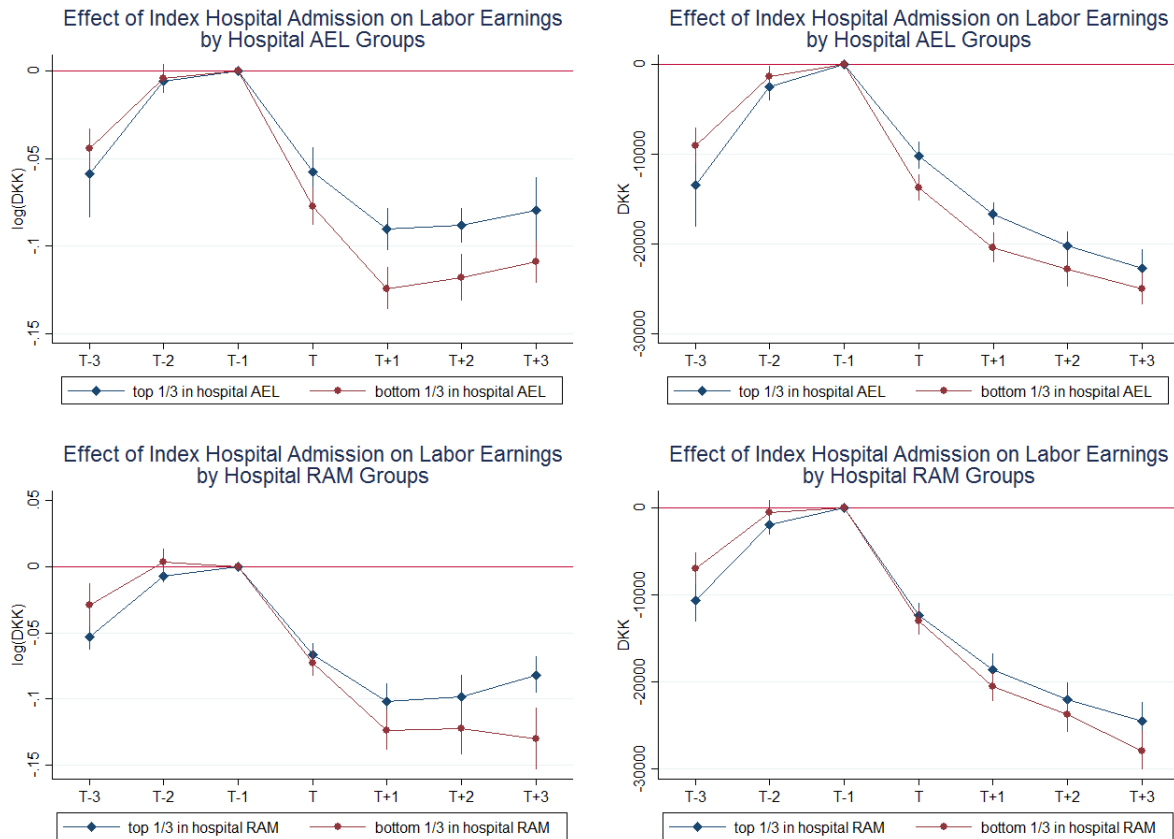
Number of non-pregnancy related hospital records in the Danish National Hospital Register by type of hospital admission (raw data).

Figure A.2: Event Study: AEL and Transfers from Government to Patients - Robustness



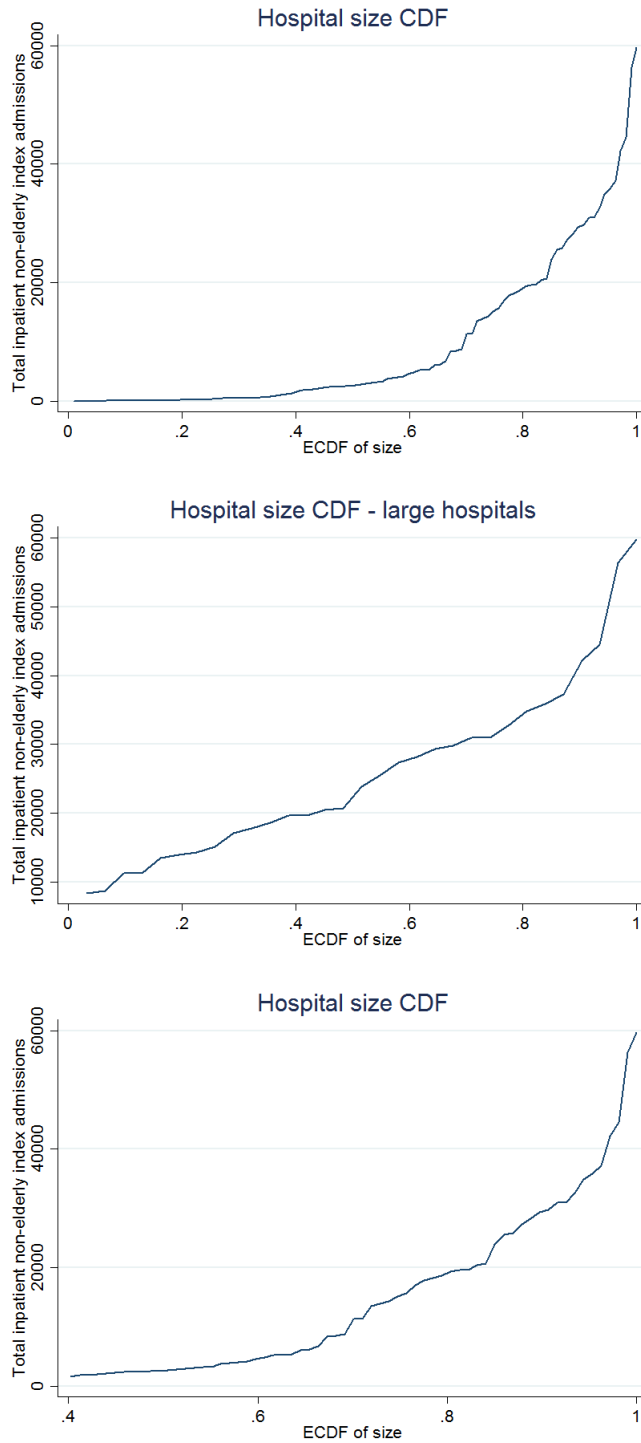
Estimated impact of an index hospital admission on the transfers from government received by the patient. The effect is estimated separately for patients admitted to hospitals in the top and bottom 25% (top-left panel), half (top-right panel), or third (bottom panel) of quality as measured by AEL. The estimating equation includes individual and time fixed effects, together with fixed effects for the time relative to the year of hospital admission (reported in the figure together with 95% confidence intervals). The coefficient relative to the year before hospital admission is normalized to zero. The bottom panel differs from Figure 6 because the results are obtained by including a control group of individuals without index hospital admissions. The estimating equation is the same as in Figure 6 for the top panels. See Section 3.3 and footnotes therein for more details.

Figure A.3: Event Study: Hospital Quality (AEL or RAM) and Labor Earnings



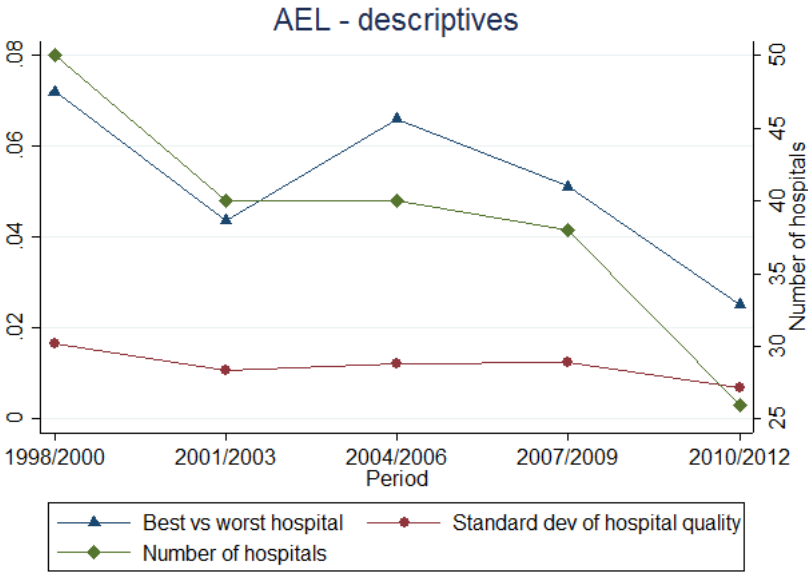
Estimated impact of an index hospital admission on Labor Earnings in logs (left panels) or levels (right panels). The effect is estimated separately for patients admitted to hospitals in the top and bottom or third of quality as measured by either AEL (top panels) or RAM (bottom panels). The estimating equation includes individual and time fixed effects, together with fixed effects for the time relative to the year of hospital admission (reported in the figure together with 95% confidence intervals). The coefficient relative to the year before hospital admission is normalized to zero. To limit the heterogeneity in patient mix between different hospital groups, we focus on individuals that have pre-hospitalization earnings in the middle of the year-specific distribution (4th to 7th decile). The patterns are similar as long as we exclude patients who are in the top and bottom decile of pre-hospitalization earnings. See Section 3.3 and footnotes therein for more details.

Figure A.4: Cumulative Distribution Function of Hospital Size



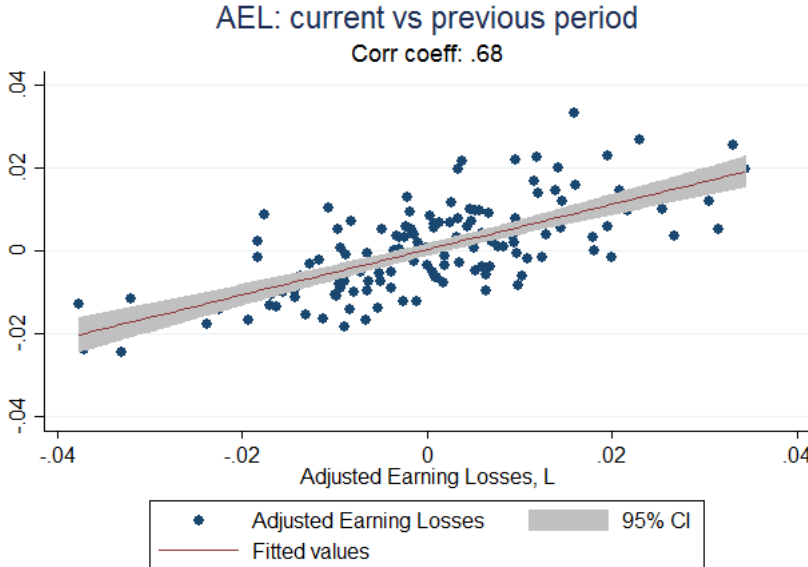
CDF of number of non-elderly inpatient index admissions for hospitals in the discharge data. All years are pooled together. The upper panel refers to all hospital codes, the middle panel only to the largest 31 hospitals (analysis in Section 3), and the lower panel to the largest 60 hospitals (analysis in Appendix A).

Figure A.5: Descriptive statistics of AEL - Alternative Hospital Sample (Appendix A)



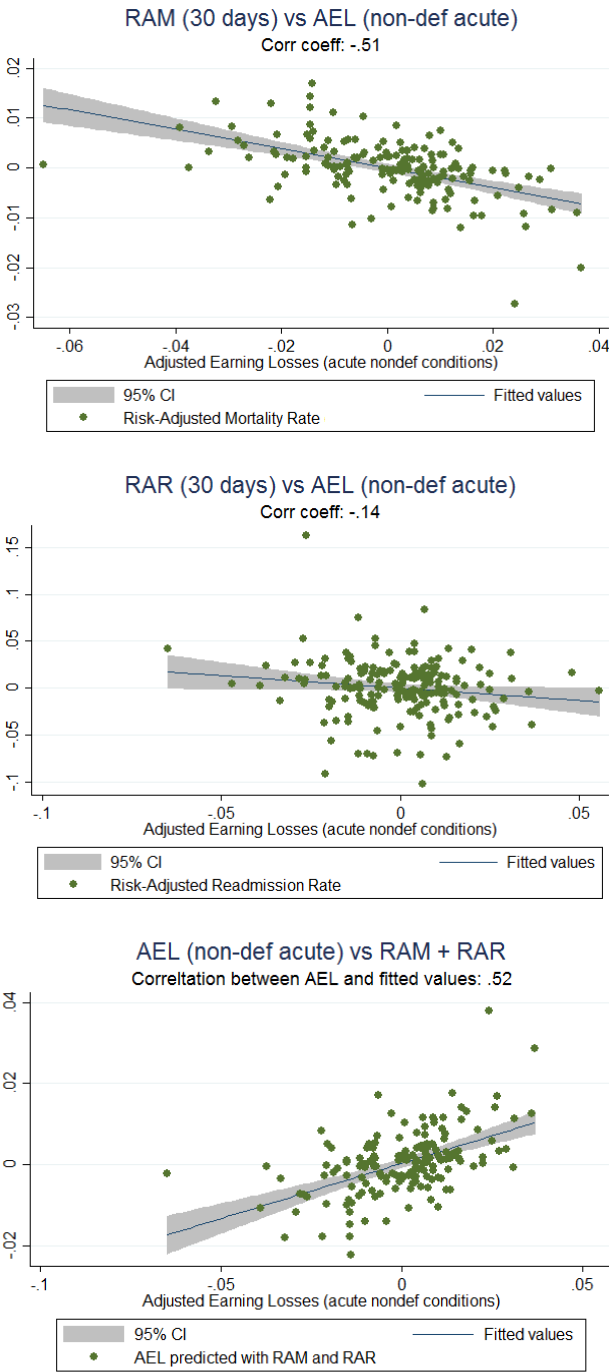
Descriptive statistics of AEL estimates for each period. Left axis is in percentage points and it refers to standard deviation and range of hospital quality. Right axis is in units and it refers to the number of hospitals. See Section 3.1 and Section 3.2 for details.

Figure A.6: AEL: present versus previous period - Alternative Hospital Sample (Appendix A)



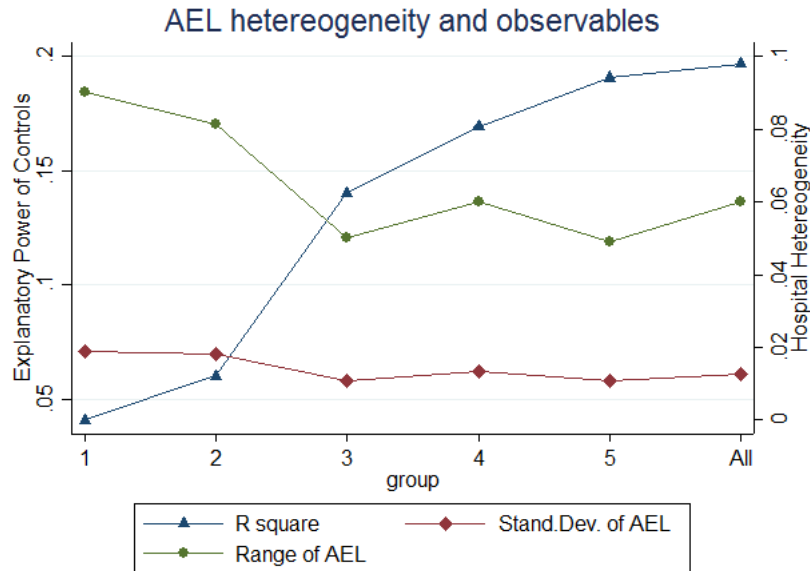
Scatterplot and line of best-fit together with 95% confidence interval for AEL and its lag. We pool together all years and AEL values are normalized by subtracting the period-specific average. See Section 3.1 and Section 3.2 for details.

Figure A.7: AEL and Traditional Measures - Alternative Hospital Sample (Appendix A)



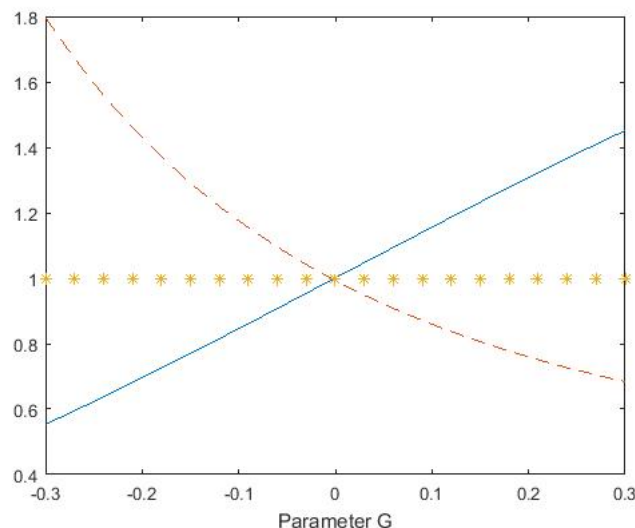
Scatterplot and line of best-fit together with 95% confidence interval for AEL and traditional quality measures: 30-days Risk-Adjusted Mortality (RAM, upper panel), 30-days Risk-Adjusted Readmission rates (RAR, middle panel), and the predicted values obtained by regressing AEL on RAM and RAR. We pool all years together and all quality measures are normalized by subtracting the period-specific average. See Section 3.1 and Section 3.5 for details.

Figure A.8: AEL Heterogeneity and Explanatory Power of Observable Covariates - Alternative Hospital Sample (Appendix A)



Range and standard deviation of AEL estimates together with the R-square of the regression used to estimate hospital quality (equation 3). Each “group” includes a larger set of controls, with group 1 including only main disease fixed effects and group “All” including all the controls. See Section 4.3 for more details.

Figure A.9: Simulated IV model



Results of the simulated IV model described in Appendix B. The solid line indicates the average estimated quality \hat{q}_1 , the dotted line represents the average α_{IV} , and the stars represent the average α_{OLS} . Averages are taken over 200 simulations of the model for each value of the endogenous selection parameter G (on the x-axis).