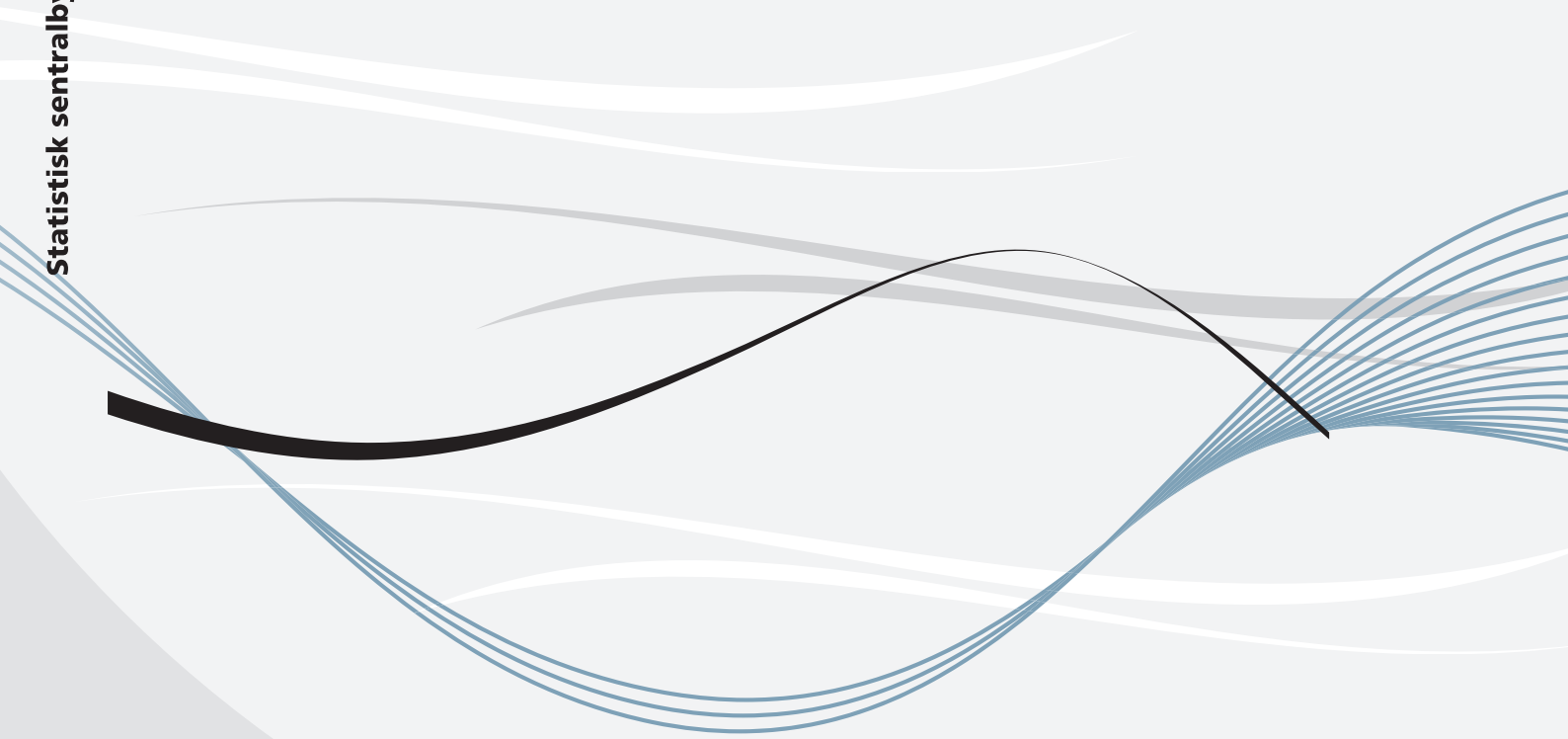
Discussion Papers
Statistics Norway
Research department
No. 890
December 2018

Statistics Norway

Statistisk sentralbyrå

Anna Godøy and Ingrid Huitfeldt

Regional variation in healthcare utilization and mortality



Discussion Papers No. 890, November 2018
Statistics Norway, Research Department

Anna Godøy and Ingrid Huitfeldt

Regional variation in healthcare utilization and mortality

Abstract:

Geographic variation in healthcare utilization has raised concerns of possible inefficiencies in healthcare supply, as differences are often not reflected in health outcomes. Using comprehensive Norwegian microdata, we exploit cross-region migration to analyze regional variation in healthcare utilization. Our results indicate that hospital region factors account for half of the total variation, while the rest reflect variation in patient demand. We find no statistically significant association between the estimated hospital region effects and overall mortality rates. However, we document a negative association with relative utilization-intensive causes of death such as cancer, suggesting high-supply regions may achieve modestly improved health outcomes.

Keywords: healthcare supply, healthcare demand, healthcare spending, regional variation, health outcomes.

JEL classification: H51, I1, I11, I13

Acknowledgements: This research has received support from the Research Council of Norway (grants #214338, #227117 and #256678). Data made available by Statistics Norway and the Norwegian Patient Registry have been essential for this project. We are grateful to Anna Aizer, Amy Finkelstein, Simen Gaure, Sverre AC Kittelsen, Samuel Kleiner and Edwin Leuven for helpful discussions, suggestions and comments at various stages of this project. Comments from conference and seminar participants at ASHEcon 2016, EALE 2016, student micro lunch at the University of Chicago, 2016, and the ESOP lunch seminar at the University of Oslo, 2016 are gratefully acknowledged.

Address: Anna Godøy, Institute for Research on Labor and Employment, University of California, Berkeley; Institute for Social Research, Norway; and Statistics Norway.
E-mail: anna.godoy@berkeley.edu

Ingrid Huitfeldt Statistics Norway and the Frisch Centre, Norway.
E-mail: ingrid.huitfeldt@gmail.com)

Discussion Papers comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no/en/forskning/discussion-papers
http://ideas.repec.org/s/ssb/dispap.html

ISSN 1892-753X (electronic)

Sammendrag

Det er betydelig regional variasjon i bruk av helsetjenester mellom sykehus. Dette kan reflektere ulike faktorer – det kan skyldes at noen sykehus tilbyr flere eller grundigere tjenester enn andre, men det kan også forklares ved at sykehusene har ulike pasientgrupper, for eksempel ved at befolkningen i høyforbruksregioner har dårligere underliggende helsetilstand, og dermed etterspør mer helsetjenester. Analysene våre skiller mellom faktorer knyttet til geografisk område (inkludert sykehuset) og faktorer knyttet til pasienten. Dette er mulig ved å studere hvordan bruk av helsetjenester endrer seg når personer flytter mellom regioner. Resultatene viser at faktorer knyttet til bosted forklarer rundt halvparten av variasjonen i bruk av helsetjenester, mens resten kan forklares av pasientenes bakgrunn. Variasjon i forbruk som ikke kan forklares av pasienthelse og -bakgrunn tyder på at noen sykehus er relativt mindre effektive enn andre. Imidlertid er det vanskelig å si noe om hva som er det optimale forbruksnivået uten å analysere hvordan forbruk påvirker helseutfall. Vi forsøker å svare på dette ved å se på sammenhengen mellom forbruk av helsetjenester og helseutfall. I gjennomsnitt finner vi ingen statistisk signifikant sammenheng mellom høyt forbruk og død, men når vi går spesifikt inn på ulike dødsårsaker ser det ut til at høyforbruksområder har lavere forekomst av kreftdød. Mer generelt finner vi at høyforbruksområder har lavere dødelighet fra dødsårsaker som er kjennetegnet ved mer bruk av helsetjenester i årene før død. Dette kan indikere at høyforbrukssykehus ikke nødvendigvis er ineffektive.

1 Introduction

Geographic variation in healthcare utilization has raised concerns of possible inefficiencies in the supply of healthcare. In particular, we may be concerned that some regions are spending too much on healthcare, given that high utilization regions tend not to achieve better health outcomes (Finkelstein et al., 2016; Skinner, 2011). In this paper, we leverage detailed microdata from Norway to answer two questions. First, to what extent is regional variation in healthcare utilization driven by place-specific factors, as opposed to variation in underlying patient health? Second, is higher regional supply of healthcare associated with better health outcomes?

We argue that both questions are central to policymakers seeking to make sense of regional variation in healthcare utilization. In principle, regional variation in healthcare utilization can be driven by variation in demand factors, such as patient health, as well as supply factors, such as physicians' practice styles. Generally, demand-driven variation is seen as less problematic - regions may have higher or lower average utilization rates depending on whether the inhabitants require more or less care. Supply driven variation on the other hand, typically signals inefficiencies.

On the one hand, variation in hospital region effects could indicate inefficiently high utilization if higher regional supply does not translate to better health outcomes. In this case, reducing healthcare utilization in high supply regions can lead to efficiency gains. If, on the other hand, high supply regions do have better health outcomes, we may instead be concerned with utilization being too low in low utilization regions, and the prescribed policy response may involve raising utilization rates in selected regions. In other words, policy recommendations are likely to depend on the answer to the second question, that is, the impact of hospital region effects on health outcomes.

Previous research from the U.S. has uncovered substantial regional variation in healthcare utilization (Finkelstein et al., 2016; Song et al., 2010; Baicker et al., 2004; Fisher et al., 2009, 2003a,b). Finkelstein et al. (2016) finds that 40-50% of this variation is attributable to patient demand factors, while the rest is explained by supply factors. The majority of existing papers, however, concludes that regional variation in healthcare

spending is primarily driven by the supply side (see eg. Cutler et al., 2018; Chandra et al., 2012; Anthony et al., 2009).

Meanwhile, it is not *a priori* clear if these findings would translate to a nationalized single payer healthcare system, where hospitals are similar in terms of payment schemes and physician incentives, and patients face no to negligible co-payments. Furthermore, existing literature from the U.S. is mainly based on the Medicare population, which includes only patients aged 65 years or older. The present paper draws on data from the entire Norwegian population and includes all major hospitals in the country over the period 2008-2013, removing concerns about selection into the sample.[1]

Identifying and estimating hospital region effects in the presence of patient heterogeneity is complicated by the fact that patient demand for healthcare is largely unobservable. Individual demographic variables such as age, gender and education, are admittedly crude proxies for underlying health status. To identify hospital region effects, we follow closely the approach of Finkelstein et al. (2016), exploiting migration of patients across hospital referral regions. Specifically, we estimate panel models of log healthcare utilization with place and patient fixed effects, controlling fully for time invariant individual heterogeneity. Similar models with two-way fixed effects have been used previously in research separating the impacts of workers and firms on wage inequality (e.g. Abowd et al., 1999, 2002; Combes et al., 2008; Card et al., 2013; Gibbons et al., 2014), as well as in papers studying exposure to neighborhoods on intergenerational mobility, schooling and mortality (e.g. Chetty and Hendren, 2018a,b; Chetty et al., 2016), teacher quality (e.g. Rothstein, 2010; Jackson, 2013; Chetty et al., 2014a,b; Mansfield, 2015), and physician practice styles (Molitor, 2018).

The model allows for movers and stayers to have systematically different utilization, and for utilization to be correlated with the movers' origin or destination choices. The key identifying assumption is that conditional on person and place, mobility is as good as random with respect to health. Our model thus mirrors a difference in differences

[1]Data contain all public hospitals as well as private providers contracting with the health authorities. Very few healthcare institutions operate as for-profit institutions without any contract with public health authorities.

design, which requires that *trends* in latent health demand do not vary systematically with the movers' origin or destination. To test this assumption empirically, we implement an event-study approach, estimating patterns of healthcare utilization around the time of migration.

In the second part of the paper, we turn to the analysis of health outcomes, estimating panel models of cause specific mortality rates as functions of the estimated hospital region effects. This analysis relates to an unsettled literature, mainly from the U.S., on the relationship between spending and health (see, e.g. Doyle et al., 2015; Joynt and Jha, 2012; Doyle Jr, 2011; Cutler et al., 2018). Our mortality analysis makes two distinct contributions to this field. First, we link mortality to the estimated patient and hospital region effects rather than average utilization. Second, we merge information on cause of death to individual utilization data in order to shed further light on the link between spending and mortality.

Interpreting the correlation between regional utilization and mortality rate is complicated by the fact that regions with sicker individuals will tend to have higher demand for healthcare, driving up average utilization rates. This form of omitted variable bias will lead to a positive correlation between utilization rates and mortality. Meanwhile, our empirical strategy exploiting interregional migration yields a set of hospital region effects that are effectively purged of patient demand factors. To be clear, the estimated hospital region effects may reflect both local variation in the supply of healthcare, as well as a number of other factors such as environmental or social factors. This can in turn complicate the analysis of health outcomes, as we cannot distinguish between the impacts of healthcare supply per se and unobserved place characteristics.

To address this issue, we leverage variation in utilization intensity across causes of death. If regional supply of healthcare shifts mortality rates, we might expect the largest effects for conditions where patients tend to use more hospital services in the time leading up to death, such as cancer. Meanwhile, effects should be smaller for causes associated with lower average utilization rates, like deaths from external causes. To be clear, we are not claiming to estimate true causal effects of spending, rather, the models should be

seen as predictive.[2]

Our results show that place factors account for roughly half of the gap between average utilization in high and low utilization regions. This result is robust to a number of sensitivity checks, including alternative hospital market definitions, using balanced samples to avoid compositional bias, and alternative model specifications. The estimated figures are similar to those found by Finkelstein et al. (2016), which is remarkable given the many institutional differences in healthcare systems, in terms of e.g. hospital financing and physician compensation.

The mortality analysis finds no significant association of hospital region effects and all-cause mortality. However, the picture changes somewhat when we distinguish between major causes of death. In particular, the models find that higher hospital region effects are associated with a statistically significant reduction in cancer deaths. More generally, higher hospital region effects tend to predict lower mortality from relatively utilization-intensive causes of death, suggesting that high supply regions may in fact achieve modestly improved health outcomes.

The rest of the paper is structured as follows. Section 2 describes the institutional setting and data. Section 3 presents the econometric models and discusses identifying assumptions. Results are presented in Section 4. Section 5 presents estimated models of cause specific mortality. Finally, Section 6 concludes.

2 Institutions and data

2.1 Institutional setting

Somatic specialist healthcare in Norway is funded primarily through taxes and transfers from the national government. The reimbursement scheme from national level to regional health authorities entails a fixed part and an activity-based part. Since 2002 four state-owned regional health authorities have had the overall responsibility for pro-

[2]Our approach estimating impacts by cause of death can only be interpreted causally if we are willing to make the strong assumption that cause of death (but not death alone) is uncorrelated with other place characteristics.

viding specialist healthcare services to their region's population. The regional health authorities own in total 24 health trusts whose task is to execute healthcare provision to their respective referral regions. A health trust may comprise several hospitals and other institutions.

Specialized healthcare is rationed by wait time, aiming at prioritizing patients based on their medical needs for healthcare. Access to hospital services is either by emergency admissions or through referrals from general practitioners (GPs) acting as gatekeepers, thus being responsible for all initial assessment, examinations and treatment of patients.

Since 2001 patients who are referred to specialist healthcare have had the right to choose the hospital at which they want to receive treatment. Patients may choose to be treated at hospitals outside of their referral region; either at another health trust within their region or in another region, but the latter is infrequently observed.[3]

Patients' healthcare expenses are mainly subsidized by national insurance schemes. Some services, such as outpatient admissions and visits to GPs are subject to small co-payment rates. In 2015, the out-of-pocket payment rate for an outpatient procedure was 320NOK (∼40USD). However, once a patient's yearly total out-of-pocket healthcare expenditures exceed about 2100NOK (∼260USD) all further expenses within that calendar year are reimbursed.

To summarize, the Norwegian hospital system is characterized by universal coverage, low co-payments, and a high degree of centralization. Hospitals face the same financial incentives, and physicians at hospitals are employed on fixed salary rather than on a fee-for-service or capitation fee basis. This should leave less scope for supply-driven demand, and similar moral hazard problems.

2.2 Data, sample and summary statistics

The empirical analysis is based on data that combine several administrative registers from Statistics Norway, the Norwegian Patient Registry (NPR), the Control and Payment of

[3] 90% of elective surgeries are performed within the patients' own region, and 22% chooses a hospital outside of their catchment region but still within their residential region (Huitfeldt, 2016). An information service called *Free Hospital Choice* facilitates the option to choose hospital by making quality indicators such as expected wait time publicly available.

Health Reimbursement (KUHR), and the Cause of Death Registry. A unique personal identifier is provided every Norwegian resident at birth or upon immigration, enabling us to match the health records with administrative data of the entire resident population of Norway.

Data provided by Statistics Norway contain birth and death dates, sex, district and municipality of residence, country of origin, education, occupation, annual earnings and welfare benefits receipt. These data are linked with patient data from NPR, containing complete patient level observations for all somatic public hospitals and private hospitals contracting with regional health authorities in Norway from 2008 onward. Records include main and secondary diagnoses (ICD10), surgical and medical procedures (NCSP/NCMP), time of deaths in/out of hospital, exact time, date and institution of admissions and discharges, time of referral, length of stay, diagnosis groups and diagnosis cost weight. Each patient discharged at a somatic hospital is assigned a diagnosis group that uniquely determines the reimbursement rate. Healthcare utilization is defined as an individual's yearly total hospital care expenditures, calculated by applying the diagnosis group system and prices (for year 2012) on each year. We finally add the KUHR database, which contains all visits in the primary care sector, as well as visits to specialists. Data include date of visit, diagnosis codes, reimbursement codes and size of patient deductible.

Our sample covers a period of six years, from 2008 to 2013. For the baseline estimation sample, two additional restrictions are imposed. First, we retain only people aged between 30 and 75. The assumptions underlying our empirical approach may be less likely to hold for younger and older persons. For younger people, we note that individuals are exempt from the legal requirement to register change of address while enrolled in education. This could potentially make mobility data less accurate for teenagers and younger adults, who may delay changing their address until after they complete schooling. Meanwhile older adults are more likely to move for health related reasons, which would undermine our identification strategy. In addition, we exclude people who move between HRRs more than once during the 6 year period. This restriction eases the event study approach as all movers will have one well-defined move year. In the robustness section we relax

this assumption, and estimate the two-way fixed effect model with no restriction on the number of moves. Note that both the restriction on age and number of moves are applied only to the baseline estimation sample used to estimate hospital region and patient fixed effects. In the subsequent analysis of mortality, all ages are retained in the analysis sample.

The resulting estimation sample contains 15,570,065 person-year observations.[4] In our empirical models, identification of hospital region effects is obtained by individuals who move between regions. Table 1 shows descriptive statistics for stayers and movers separately. Compared to stayers, movers are more likely to be male and foreign-born. Movers are also more likely to be in school - roughly 15% of the movers are enrolled in education at the first year of observation, compared to 8.4% in the stayer sample. On average, movers are younger than stayers; a majority of movers are between 30 and 44 years old. Residential origins are quite evenly distributed among movers and stayers, although slightly more of the movers compared to stayers originate from the South East region (capital area).

The average person is followed for 5.4 years in the stayer group, and 5.45 in the moving group. There are several entry and exit routes from the sample: a small share dies during the study period, 2.5% in the stayer group and 0.5% in the moving group. Individuals will also enter and exit the age-groups under study (aged 30-75), and there may be both immigration and emigration; we only observe residents. There are 116,367 unique movers, and 2,792,692 unique stayers (i.e roughly 4% movers).

The moving population has a slightly lower average annual utilization, which again is likely due to the lower share of elderly among this group. As many as 67% of the movers never visit the hospital during the study period; the share is only slightly lower in the stayer population. The distribution of utilization is right-skewed for both movers and stayers. In Appendix Table A1 we show the full distribution of utilization in logs and levels.

The main geographic unit of analysis is a hospital referral region (HRR). We will define

[4]We additionally exclude individuals who move in the first or last year of our sample, as these do not provide any useful variation.

Table 1: Descriptive statistics of estimation sample

	Stayers	Movers
Female	0.49	0.46
Norwegian-born	0.86	0.74
Enrolled in education	0.084	0.15
Age first observed		
30-44	0.44	0.69
45-59	0.33	0.22
60-75	0.24	0.089
First observed residence		
North	0.096	0.091
Mid	0.14	0.10
West	0.21	0.13
South East	0.56	0.68
Annual health care utilization (USD)		
Mean	1184.6	906.3
Standard deviation	5636.8	5296.6
Share of patient-years with zero	0.66	0.68
Average number of years observed	5.40	5.45
Share who dies during study	0.025	0.0049
Number of patient-years	15,080,854	634,012
Number of patients	2,792,692	116,367

Notes: Table shows descriptive statistics for movers vs. stayers aged 30-75 based on data for the period 2008-2013.

these regions in two different ways: (i) 28 local hospitals conditional on them having both maternity ward and emergency room; (ii) 19 health trusts with defined catchment regions. Some health trusts do not serve their own catchment region; these may have different functions or be highly specialized. For both definitions, the hospital referral regions are defined based on residential municipality. We will apply definition (i) of hospital referral regions in our baseline estimations; definition (ii) will be used in the robustness section. Using definition (i), there are on average about 1.9 institutions within each HRR.

As discussed above, patients may seek medical care outside their own region of residence. In our sample, we calculate average utilization rates for the HRRs based solely on patients' residence region, regardless of where care was actually provided. About one fifth of total expenditures occur outside of a patient's HRR of residence.

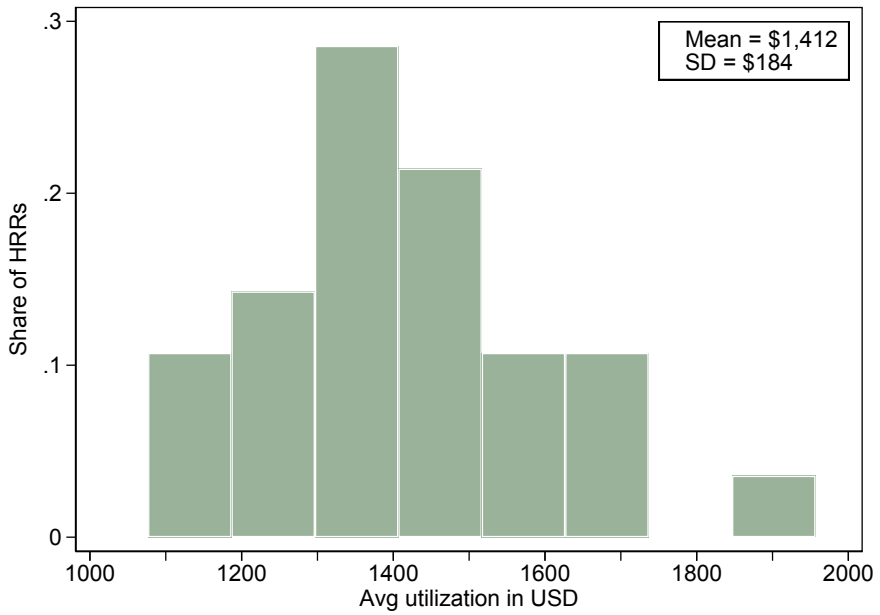


Figure 1: Distribution of utilization (in USD).
Figure shows the distribution of yearly average utilization (in USD) per patient in the 28 hospital referral regions.

Figure 1 shows the distribution of yearly average patient utilization across HRRs. The average HRR has an average utilization of 1,412USD per patient per year (standard deviation 184USD). In Figure A1 we show that the spread is substantial even after purging utilization for sex, age and educational differences. The geographic pattern of utilization can be seen in Figure 2, where colors illustrate quintiles of healthcare utilization.

3 Empirical models

We begin our empirical analysis by disentangling the components of utilization attributable to place-specific heterogeneity, e.g. hospital quality or physician knowledge; and patient-specific heterogeneity, e.g. health endowment or preferences. Next, we use the estimated place and patient components to shed light on their relative importance in explaining differences in average patient utilization across hospital regions. Importantly, our goal is not to estimate the individual health production function, nor to evaluate the impact of place on individual utilization. Rather, we aim at exploring sources of differences in average patient utilization between hospitals.

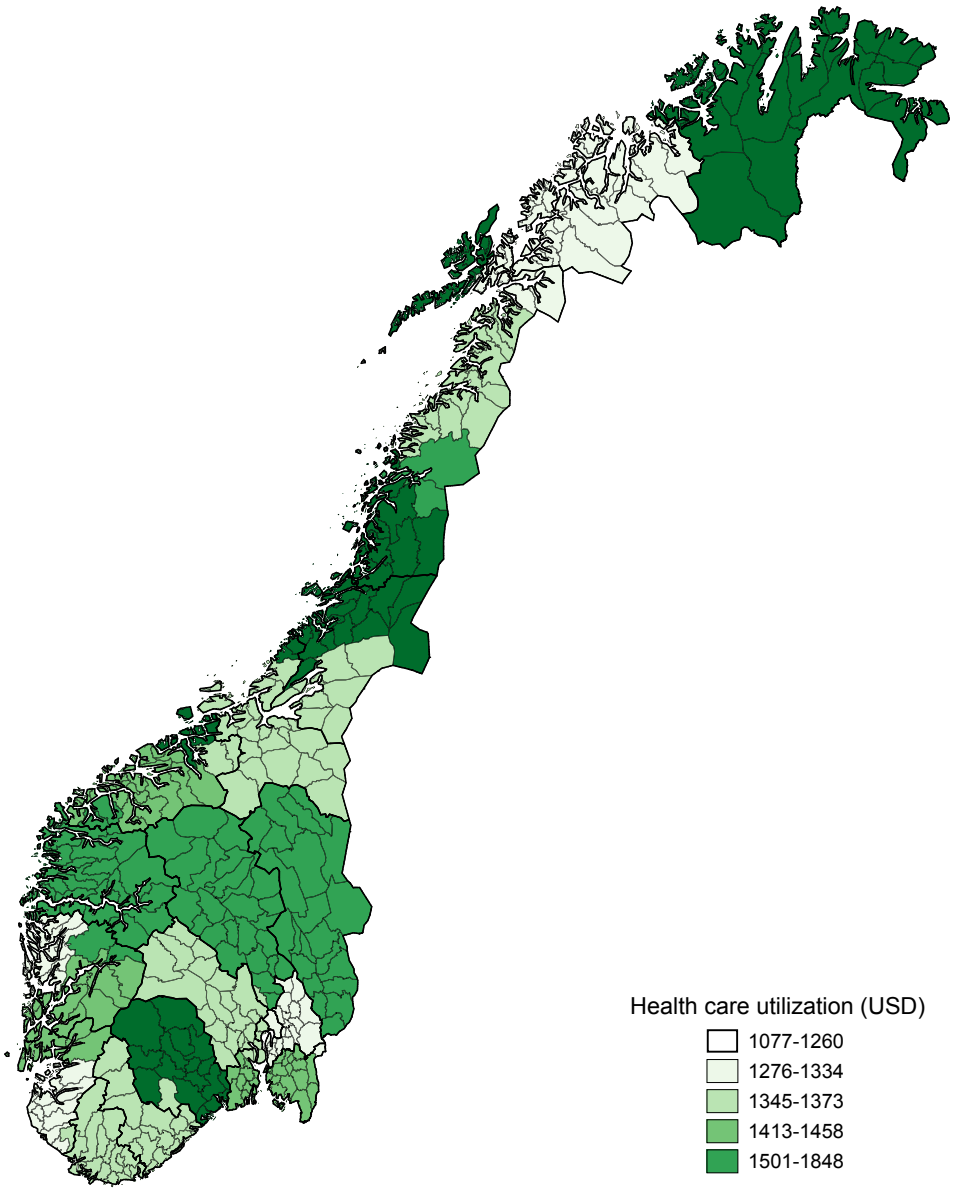


Figure 2: Map of Norway. Utilization (in USD) by hospital referral region.
Figure shows the geographic distribution of yearly average utilization per patient in the 28 hospital regions, divided into quintiles. Thick solid lines mark HRR borders; thin solid lines mark municipality borders.

3.1 Fixed effects models

The empirical specification closely follows Abowd et al. (2002, 1999); Card et al. (2013) and Finkelstein et al. (2016) where the dependent variable y_{it}, person i's log of utilization of healthcare (plus 1) in year t, is expressed as a function of individual heterogeneity, hospital region heterogeneity, and measured time-varying characteristics:

$$y_{it} = \alpha_i + \gamma_{j(i,t)} + X_{it}\lambda + \varepsilon_{it}, \tag{3.1}$$

where $i = 1, ..., N, t \in \{n_{i1}, .., n_{iT}\}$, and the function $j(i,t)$ indicates the hospital region j of individual i in year t, where $j = 1, ..., J$. There are T_i observations per individual and $N^* = \sum_i T_i$ total observations.[5] The component α_i is the individual effect, and $\gamma_{j(i,t)}$ is the hospital region effect. Time-varying covariates are included as X_{it}, and in the baseline specification this includes fixed effects for year and age (in 5-year bins) only.[6] We explore richer versions of Equation (3.1) in the robustness section.

As discussed in e.g. Bonhomme et al. (2017), Lamadon et al. (2017) and Finkelstein et al. (2016), causal interpretation of the parameters in Equation (3.1) rests on two important assumptions. First, mobility needs to be exogenous to the utilization residual, which would follow if, e.g., the assignment of individuals to hospital regions is random conditional on all observable controls and time invariant unobservables. Second, we assume a log additive functional form. This implies that individuals who move from hospital region j' to j'' will on average experience an average utilization change of $\gamma_{j''} - \gamma_{j'}$, whereas those who move in the opposite direction will experience an average change of $\gamma_{j'} - \gamma_{j''}$.

These assumptions flexibly allow for rich patterns of sorting, as the moving decision may be related to α_i or γ_j. For example, the model allows for movers and non-movers to have systematically different utilization levels, and for utilization levels to be correlated with the movers' origin or destination. Moreover, mobility may be related to characteristics of hospitals unrelated to utilization, such as geographic location, and of the individual, such as her earnings potential. We return to a thorough discussion of the validity of the identifying assumptions below.

The results from the two-way fixed effects model in Equation (3.1) form the basis for two decomposition exercises that quantify the relative impact of the estimated patient and hospital region effects. The first is an additive decomposition in means following

[5] In estimation of model (3.1) we drop the year of move, as we do not have information on the exact date of move. This exclusion avoids attributing utilization to the wrong hospital region.

[6] Note that, as the individual fixed effects absorb the cohort effect, age and year are perfectly collinear. In Table 4 we show that our specification is robust to alternative ways of including age in the model. In principle, our model could also include fixed effects for relative year of moving (where relative year for non-movers are normalized to zero). This allows the possibility that the decision to move is correlated with health shocks. In our baseline model we focus on the simplest model formulated in Equation (3.1), but the robustness section shows that inclusion of such relative year dummies does not affect our results.

Finkelstein et al. (2016), where we ask how much of the *difference* in average utilization between high utilization regions and low utilization regions can be explained by the type of patients they have, and how much is due to place-specific factors. Second, we note that the relative impacts of patient and hospital region effects may also depend on the degree of sorting, i.e. how the two components are correlated. To examine this, we implement a variance decomposition exercise where we ask how much of the *variation* in average utilization at hospitals is explained by the variance in individual factors, hospital region factors, and sorting, respectively.

As a starting point for the additive decomposition exercise, we use the estimates from Equation (3.1), and average over hospital referral regions:

$$\hat{y}_j = \bar{c}_j + \hat{\gamma}_j, \tag{3.2}$$

where $\hat{y}_j$ is the sample analog of $\bar{y}_j$, $\hat{\gamma}_j$ are the estimated hospital region effects, and we label $\bar{c}_j$ as an average patient compound effect, comprising fixed effects for patient, age and year. Hospital referral regions are then split into two groups (A, B) depending on the average utilization $\hat{y}_j$ at the hospitals. We next calculate the difference between the mean hospital region (compound patient) effect estimates in the two groups, and finally divide by the difference in average utilization. This renders a hospital region share $\frac{\hat{\gamma}_A - \hat{\gamma}_B}{\hat{y}_A - \hat{y}_B}$ and a patient compound share $\frac{\bar{c}_A - \bar{c}_B}{\hat{y}_A - \hat{y}_B}$. In one specification we split the hospitals into groups A and B by median utilization; in a second we include only the top and bottom quartiles.

The variance decomposition is more standard though we perform the exercise at the hospital level rather than at the individual level. Collapsing Equation (3.1) over hospital referral regions as in Equation (3.2) and then taking the variance yields[7]

$$var(\bar{y}_j) = var(\bar{c}_j) + var(\hat{\gamma}_j) + 2cov(\bar{c}_j, \hat{\gamma}_j). \tag{3.3}$$

In both decomposition exercises we calculate the standard errors using 500 bootstrap replications at the patient level. We additionally account for potential limited mobility

[7] In practice, we separate out the year effect from the patient compound effect in the variance decomposition exercise.

bias using a split-sample jackknife approach (Dhaene and Jochmans, 2015).[8]

3.2 Identifying assumptions

The estimated hospital region effects can only be interpreted causally if mobility is conditionally independent of latent health outcomes. To structure the discussion on endogenous mobility, we follow Card et al. (2013) and assume that the error term ε_{it} in Equation (3.1) consists of three separate random effects: a match component $\eta_{j(i,t)}$, a unit root component ν_{it}, and a transitory error ω_{it}:

$$\varepsilon_{it} = \eta_{ij(i,t)} + \nu_{it} + \omega_{it} \tag{3.4}$$

The match effect $\eta_{ij(i,t)}$ represents an idiosyncratic utilization premium or reduction obtained by individual i at hospital j, relative to the baseline level $\alpha_i + \gamma_j$. Match effects arise if e.g. some hospitals are highly specialized in treating certain types of patients. The unit root component ν_{it} captures potential drift in the individual's utilization over time, such as health deterioration. The transitory component ω_{it} represents any left-out mean-reverting factors. We assume that $\eta_{ij(i,t)}$ has mean zero for all i and for all j; and both ν_{it} and ω_{it} have mean zero for each person in the sample.

Sorting on match effects: Bias can arise if individuals sort to hospitals based on the idiosyncratic match component $\eta_{ij(i,t)}$. This form of sorting changes the interpretation of the estimated hospital region effects since different individuals have different utilization premiums at any given hospital, depending on their match component. In the limit, if all moves are due to the match component, we could expect all moves to lead to increased

[8] It is well known that incidental parameter bias caused by a large number of place-specific parameters is likely to introduce upward bias to the place component and a downward bias to the sorting component, with the size of the bias depending inversely on the degree of patient migration between hospital regions (Andrews et al., 2008). We suspect that the bias caused by limited mobility is small in our setting, as we are only estimating 28 hospital region effects, and there are several hundred movers from each hospital region. Nonetheless, to correct for potential limited mobility bias we use a split-sample jackknife approach (Dhaene and Jochmans (2015), see also, Lamadon et al. (2017); Bonhomme et al. (2017)). This estimator is based on half-sample estimation where, within each hospital referral region, migrants (and stayers) are randomly split into two approximately equal-sized subsamples. We then estimate Equation (3.1) separately in each subsample. The bias-corrected estimate is equal to twice the full-sample estimate minus the mean of the half-sample estimates.

utilization.

Drift: Endogenous mobility may arise if patients with gradually declining health systematically move to different types of hospitals. If individuals with deteriorating health systematically move to high utilization regions, we might overestimate the importance of hospital region effects, as the drift component ν_{it} will be positively correlated with the change in the hospital region effects. In other words, Equation (3.1) will be biased if *trends* in utilization vary systematically with the movers' origin or destination.

Transitory error: Shocks or fluctuations in the transitory error ω_{it} may be associated with systematic moves between higher and lower utilization regions. For example, if individuals who experience a negative health shock (i.e. high utilization) are more likely to move to higher utilization regions, estimated hospital region effects might be amplified.

3.3 Event-study framework

To assess whether these assumptions hold in our data, we introduce an event-study framework tracking individuals' utilization before and after they move. This model serves a dual purpose: having shown that endogenous mobility does not seem to be a concern, the event-study model's estimates will give a first indication of the relative importance of patient and hospital region effects in explaining variation in average utilization.

If everyone moved from low-utilization hospital region j' to high-utilization hospital region j'', we could plot average utilization by relative year to move, and then study whether the movers increase their utilization. However, in the data we observe people moving in both directions: from high to low utilization regions and the other way around. These moves could cancel each other out and produce a flat event-study figure. Moreover, the "magnitude" of the moves varies considerably: while some persons move from regions that are fairly similar, other moves are characterized by much larger differences in average healthcare utilization in the origin and destination regions. To account for this, we follow Finkelstein et al. (2016) and augment the standard event-study model to consider both

the direction and magnitude of the move. With this in mind, we define

$$\delta_i = \bar{y}_{j''(i)} - \bar{y}_{j'(i)}$$

as the difference in average log utilization in the destination ($\bar{y}_{j''(i)}$) and origin ($\bar{y}_{j'(i)}$) hospital regions. δ_i can be interpreted as a scaling factor, capturing the direction and magnitude of i's move. Appendix Figure B1 shows the distribution of δ_i. The distribution is fairly symmetrical with mean just above zero which means that slightly more people move from low to high utilization hospital regions than there are people moving from high to low utilization hospital regions.

Having defined the relevant parameters, we formulate the following event-study equation, where the scaling factor δ_i is interacted with a set of dummy variables indicating event time k (i.e. relative year of move):

$$y_{it} = \alpha_i' + \beta_{k(i,t)}\delta_i + X_{it}\lambda' + \varepsilon_{it}'. \tag{3.5}$$

Here, as before, α_i' are fixed effects capturing any time invariant characteristics of individual i, including unobserved characteristics that are correlated with the choice of origin or destination region, and X_{it} is a vector of age (in 5-year bins) and year dummies.

The primary coefficients of interest are the $\beta_{k(i,t)}$, capturing the effects of the relative year coefficients multiplied by the scaling factor δ_i. Our data allow estimation of β_k for $k =\in [-4, 4]$. The coefficients $\{\beta_{-4}, ..., \beta_4\}$ are only identified relative to each other; we use the normalization that $\beta_{-1} = 0$.

In Appendix B, we show that if the assumptions underlying the two-way fixed effects model hold, the coefficients β_k from Equation (3.5) can be related to the parameters in Equation (3.1) as follows:

$$\beta_k = \begin{cases} 0 & \text{if } k < 0 \\ \frac{\gamma_{j''(i)} - \gamma_{j'(i)}}{\bar{y}_{j''(i)} - \bar{y}_{j'(i)}} & \text{if } k > 0 \end{cases} \tag{3.6}$$

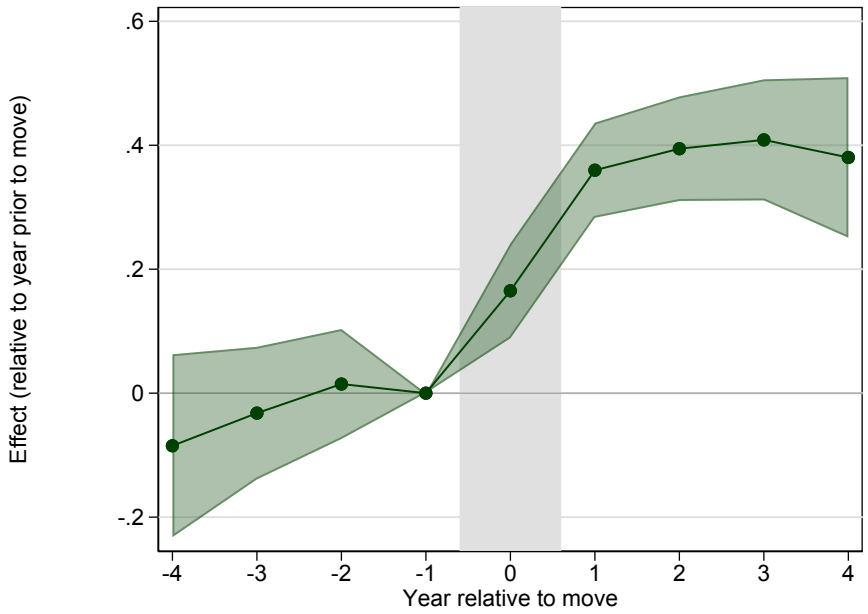


Figure 3: Event-study figure
Figure shows point estimates of β_k from Equation (3.5).

Since we do not have fully continuous data, in the calendar year of the move ($k = 0$), the coefficient should be a positive number between these two values, i.e. $\beta_0 \in \left(0, \frac{\gamma_{j''(i)} - \gamma_{j'(i)}}{\bar{y}_{j''(i)} - \bar{y}_{j'(i)}}\right)$.

The event study model also serves to give a first indication of the relative importance of hospital region effects. Intuitively, if differences in utilization are driven entirely by differences in patient factors, individual utilization is not expected to change around the year of move. On the other hand, if the variation in average utilization is driven entirely by hospital region effects, individual utilization should respond with a one-to-one change with the magnitude of the move, i.e. coefficients of 1 for $k > 0$.

4 Results

4.1 Event study results

Figure 3 plots the estimated coefficients β_k together with 95% confidence intervals. Recall that we identified three forms of potentially problematic endogenous mobility: drift, sorting on matching effects, and correlated fluctuations in the transitory error. First, the pattern of estimated β_k before and after the move gives a direct indication of the presence

of problematic drift. The figure shows no clear systematic trends in utilization prior to move, suggesting that drift in individuals' utilization is uncorrelated with the movers' origin or destination. The event study also gives an indication of whether fluctuations in the transitory error ω_{it} systematically correlate with mobility patterns. Generally, we would expect any systematic moving in response to gradual changes in health status to give rise to an upward trend in the estimated β_k in the years leading up to the move. The event study model does not find any clear evidence of this.

There are also no signs of any trends post move. A positive sloping post-trend could be the case in presence of habit formation, where, as outlined in Finkelstein et al. (2016), today's patient preferences is a function of historic utilization. If this were the case, we would expect that people moving from high to low utilization regions experienced more persistence compared to opposite moves. To investigate this more closely, we have estimated an event-study model where we separate between people moving from high to low utilization regions, and people moving from low to high utilization regions. Appendix Figure C1 indicates no such pattern; both the size of the jump and the post-trend are similar in the two cases.

Appendix Figure C1 can also be used to evaluate the assumption of no sorting on match effects. To see this, consider the case with systematic positive sorting on match effects. In the limit, all moves may lead to increased utilization. In this case, patients who move from high to a low utilization regions would still see increased utilization. Estimating the event study model on this subsample could yield event study estimates that were negative. Meanwhile, if there is no sorting on match effects the change in utilization around the time of move should be symmetrical. This is exactly what we see: individuals moving from low to high utilization regions seem to experience utilization changes that are equal in magnitude (but of different sign) to individuals moving from high to low utilization regions.[9] This provides suggestive evidence against the possibility of sorting on match effects.

[9]Recall that event time is scaled by both the magnitude and direction of move. Hence, panel (a), which plots utilization for individuals moving from high- to low utilization hospital regions, displays a *positive* jump upon move although individuals *decrease* their utilization.

To further assess the importance of match effects, we estimate a fully saturated model that includes a dummy for each individual-hospital region pair. If match effects are important, the saturated model will fit the data much better than the additively separable baseline model. Adjusted R^2 increases only marginally in the saturated model, implying that the improvement in fit is modest.[10]

The absence of match effects also provides justification for our log additive model. Note that log additivity does not, however, completely rule out complementarities, as patient and hospital region effects affect the *level* of utilization multiplicatively. This means, that the level utilization will vary more across places for sicker individuals compared to that for healthy individuals, and that more weight is put on differences in the lowest part of the utilization distribution.

As discussed in the previous section, if fluctuations in the transitory error ω_{it} systematically affect mobility patterns through gradual health deterioration, we would expect to see an increasing trend in the estimated event time coefficients β_k for $k < 0$. The estimated coefficients plotted in Figure 3 do not exhibit a clear trend, indicating that changes in health that happen over time do not systematically correlate with mobility patterns. In absence of such an increasing trend, the only remaining threat would be a health shock that induces systematic moving within the same year. Though this is in general difficult to verify, a likely implication is that such acute conditions would induce intense treatment immediately following the move. If so, this would have generated a spike in the first year after the move, and perhaps be more prominent for persons moving from low to high utilization hospital regions; we observe no such patterns in our event-study graphs.

To summarize, the estimated event study model lends support to our key identifying assumptions of conditionally exogenous mobility and log additivity. Figure 3 also gives a first indication of the relative importance of hospital region effects. The estimated relative year coefficients β_k exhibit a positive jump at the time of the move, from 0 to approximately 0.4. We can interpret this as the place factors' share of utilization, or vice

[10]Baseline model: $R^2 = 0.4657, Adj.R^2 = 0.3478$. Saturated model $R^2 = 0.4693, Adj.R^2 = 0.3494$.

versa, that approximately $1 - 0.4 = 0.6$ is the patient share. Next, we present results from the baseline twoway fixed effects model.

4.2 Fixed effects estimates

Estimation of Equation (3.1) by ordinary least squares produces coefficient estimates $\hat{\alpha}_i$, $\hat{\gamma}_{j(i,t)}$, $\hat{\lambda}$, and $\hat{\varepsilon}_{it}$. Table 4 plots the estimated hospital region effects against average log utilization. The figure shows an upward sloping, fairly linear relationship between the two variables: Hospital regions with higher average utilization tend to have higher estimated fixed effects. Looking at the estimated linear slope coefficient gives an estimate of the quantitative importance of hospital region effects in determining average hospital region utilization. To illustrate, if the geographical variation in average utilization was driven entirely by patient effects, the estimated hospital region effects should not be correlated with average hospital region utilization, yielding a slope coefficient of zero. In the opposite scenario, where geographical variation is entirely driven by place specific factors, the model should yield a slope coefficient of 1. The estimated slope coefficient of 0.49 thus indicates that variation in hospital region effects accounts for roughly half of the difference in average utilization between hospital referral regions.

We proceed by presenting results from the two decomposition exercises. The additive decomposition gives the relative shares of patient and hospital region effects, respectively, in explaining the difference in utilization between hospital regions. Table 2 shows that place factors account for 39-59% of the difference in utilization between hospital regions above and below median utilization, while the remainder is explained by patient characteristics. Results are almost equivalent when comparing hospitals with average utilization in the first quartile to the fourth quartile.

In Table 3 we present results from the variance decomposition exercise, with estimates for the variances and covariances of hospital region effects, average patient effects and sorting, as well as their respective shares of the variation in utilization. (See Table D1 for the full list of estimates). Table 3 additionally shows the split-sample jackknife results (Dhaene and Jochmans, 2015). The unadjusted estimate of the share of utilization

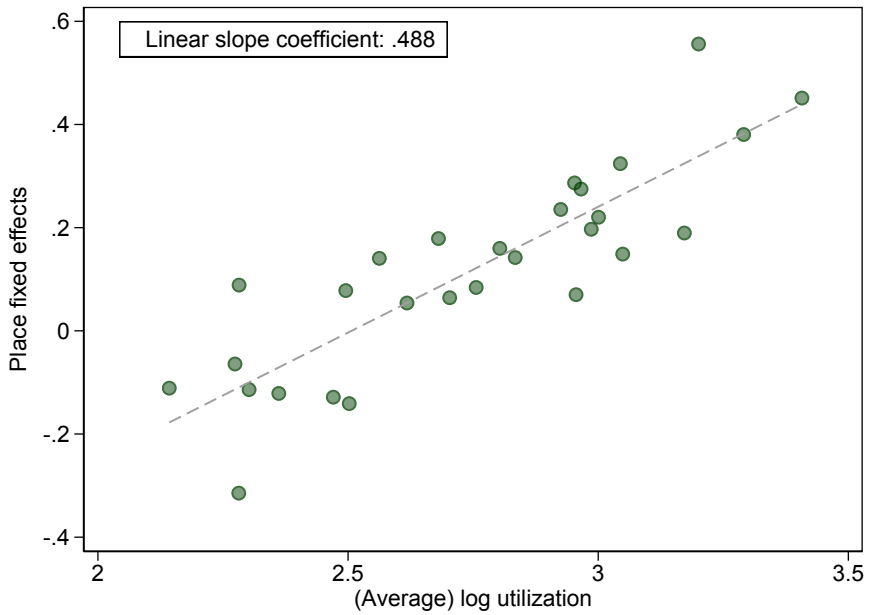


Figure 4: Hospital region effects and (average) log utilization across hospital regions.
Figure shows estimated hospital region effects and average patient utilization by hospital regions (from Equation (3.1))

Table 2: Additive Decomposition of Hospital Level Log Utilization

	(1) Above/below median	(2) Top/bottom 25%
Difference in average log utilization		
Overall	0.535	0.828
Due to hospital regions	0.263	0.392
Due to patients (id+age+year)	0.271	0.435
Share of difference due to		
Places	0.49	0.47
	(0.05)	(0.05)
Patients	0.51	0.53

Notes: Additive decomposition of log utilization based on estimation of Equation (3.1). Standard errors are calculated using 500 bootstrap replications at the patient level. R^2 is 0.478, while adjusted R^2 is 0.362.

explained by place factors is 32%, while the bias corrected measure is smaller, 26%. The confidence intervals of the non-adjusted place variance share and the bias corrected estimates are overlapping. Moreover, about 34% of the utilization variation is explained by patient specific factors in the unadjusted estimate, compared to the bias-corrected estimate 28%. The sorting component amounts to 34% (bias corrected 46%) of the variance in average log utilization.

Table 3: Variance Decomposition of Hospital Level Log Utilization

	(1) Not bias corrected	(2) Bias corrected
Variance of average log utilization	0.12	
Variance of hospital region effects	0.038	0.031
Variance of average patient (id+age) effects	0.041	0.034
Covariance of average patient (id+age) and place	0.020	0.027
Share of variance due to		
Place	0.32	0.26
	(0.06)	(0.14)
Patient	0.34	0.28
Sorting	0.34	0.46

Notes: Table shows variance decomposition at the hospital region level. Parameters estimated in Equation (3.1) are averaged within hospital referral regions. Bias corrected variances and covariances of fixed effects are estimated using a split-sample jackknife estimator (Dhaene and Jochmans, 2015). Place share is calculated using 500 replications at the patient level. See Table D1 for the full list of estimates

4.3 Robustness tests

To test the robustness of our estimates, we re-run our main model on different samples and specifications and perform the additive decomposition exercise for each model. Model (1) in Table 4 includes place and year effects only. This gives an upper bound of the hospital region effects, and emphasizes that we will overstate the hospital region effects if we naively ignore the role of sorting. In row (2) we add an extensive set of individual control variables; age, female and three categories of educational attainments, including all combinations of interactions between these controls. This significantly lowers the place share of utilization differences. Nonetheless, places still account for almost 90% of the difference in average utilization.

Model (3) reports results from a specification closer to our baseline model (which is shown in the last row for comparison). Here, place, patient and year fixed effects are included, but no age effects. This substantially decreases the place share as compared to the models with no individual fixed effect. Now, the place share amounts to about 46% of the difference in average utilization between high and low utilization regions, which is almost identical to the baseline model. Models (4) through (6) present results from other minor changes to the baseline specification, all of which yield place shares close to the

baseline model: Model (4) shows results when the baseline model includes relative year of move fixed effects. This allows the possibility that the decision to move (but not the direction) is correlated with health shocks. In row (5) we additionally add an interaction between five-year age dummies, gender, and educational attainment, and in row (6) we substitute the age dummies with squared and cubic age variables.

Our baseline sample is unbalanced as people are observed for a different number of years before and after their move. To see whether compositional changes affect our estimates, we run our model on different subsamples where we for each subsample only include movers from the same year as well as all stayers. All models give reasonably consistent estimates in the ranges of the baseline model, perhaps except from the model with 2009-movers. The additive decomposition is shown in column (7) through (10), while event-study estimates for each subsample are shown in Appendix E. Eye-balling the different panels adds confidence to our assumption that trends in utilization are not systematically related to the origin or destination of movers.

In row (11) we expand our sample to include movers who move multiple times during the time period. In model (12) we apply an alternative market definition where hospital referral regions are aggregated into 19 regions, rather than the 28 used in the baseline model (regions now represent the health trusts rather than local hospitals). Both models give similar place shares as the baseline. Event-study estimates corresponding to the higher market level definition are shown in Appendix E.

Next, we estimate the model with log utilization replaced by a binary indicator for hospital visit (row 13). If regional variation is primarily driven by the intensive margin (i.e. more services for a given patient), as opposed to the extensive margin, we would expect the binary model to display less variation in the estimated hospital region effects compared to that of our baseline model. However, the two models yield comparable hospital region shares, indicating that hospitals may also differ in the extent to which patients ever visit the hospital.

We finally estimate the model on a sample of persons aged 65 and older.[11] This

[11]Recall that our baseline sample excludes individuals who are younger than 30 or older than 70.

Table 4: Robustness tests - additive decomposition

	N	Mean of y	Diff in y	Diff in place	Place share	S.E.
(1) HRR, year	15,570,065	2.50	0.53	0.54	1.00	0.006
(2) HRR, year, age*female*educ	15,570,065	2.50	0.53	0.48	0.89	0.006
(3) HRR, patient, year	15,570,065	2.50	0.53	0.26	0.49	0.05
(4) HRR, patient, year, rel.year, age	15,570,065	2.50	0.53	0.26	0.48	0.05
(5) HRR, patient, year, rel.year, age*fem*educ	15,570,065	2.50	0.53	0.26	0.48	0.05
(6) HRR, patient, year, (age^2+age^3)*fem*educ	15,570,065	2.50	0.53	0.26	0.49	0.05
(7) Movers in 2009	15,213,062	2.50	0.54	0.36	0.67	0.12
(8) Movers in 2010	15,214,399	2.50	0.54	0.25	0.47	0.10
(9) Movers in 2011	15,210,969	2.50	0.54	0.22	0.41	0.09
(10) Movers in 2012	15,202,631	2.50	0.54	0.24	0.44	0.13
(11) Multiple moves	16,112,380	2.49	0.53	0.24	0.45	0.05
(12) Bigger HRR	15,570,065	2.50	0.48	0.23	0.48	0.04
(13) Binary utilization	5,570,065	0.356	0.075	0.036	0.48	0.05
(Baseline) HRR, patient, year, age	15,598,499	2.50	0.53	0.26	0.49	0.05

Notes: Additive variance decomposition on various samples. Standard errors are calculated using 500 bootstrap replications at the patient level.

corresponds with the "medicare sample" used by Finkelstein et al. (2016). Results are presented in Appendix F. The event study plot (Figure F1) looks less convincing for this group. The model estimates an upward path in the event time coefficients in the years leading up to the move. This indicates that the assumption of conditionally random mobility may be less likely to hold for this sample, which in turns supports the exclusion of elderly patients from the main analysis. Estimating the twoway fixed effects model on this sample yields a place share of 0.67, compared with 0.49 for the baseline sample. Taken at face value, this suggests that hospital region effects may be more important for elderly individual's utilization. However, these results should be interpreted with caution, given the mobility patterns of Figure F1.

4.4 Correlates of hospital region effects

To study the drivers of regional variation, we link the estimated hospital region effects with observable attributes of the hospital regions, and estimate simple bivariate and multivariate OLS regressions. These regressions should not be given a causal interpretation. The models presented in previous sections uncovered evidence of sorting, that is, the estimated place and patient effects are positively correlated. As a result, regions with higher estimated hospital region effects will tend to have residents who utilize more healthcare, even though the estimated hospital region effects are purged of the direct effects of patient demand. More generally, there may be unobserved local characteristics that drive both the observable attributes and the hospital region effects, giving rise to spurious correlations.

We study the correlation between hospital referral region fixed effects and the following standardized variables averaged at the hospital region level: travel time in minutes to closest hospital, travel time in minutes to primary care physician, population size, specialist nurses/midwives per capita, specialist physicians per capita, unemployment rate, disability insurance rate, local health budget as share of regional budget,[12] and primary care visits.[13] Figure 5(a) presents results from bivariate regressions, where the hospital region effects are regressed separately on each observable characteristics; while Figure 5(b) presents results from a multivariate regression where all observables are included in one regression. Standard errors are clustered on the HRR level. This results in 28 clusters which is below what is generally perceived to be the minimum number of clusters required to perform valid inference. To avoid overstating the significance of the findings, we implement the wild bootstrap for significance tests (Cameron and Miller, 2015; Roodman, 2015).

The bivariate model indicates that rural regions tend to have higher hospital region effects. The estimated hospital region effects are significantly positively associated with travel time to nearest hospital and primary care physician (GP), and negatively correlated

[12]The health budget variables are taken from Fiva and Natvik (2017)

[13]Note that the utilization measure in the main model only includes specialist care/hospital visits; it does not include primary care.

with population size.

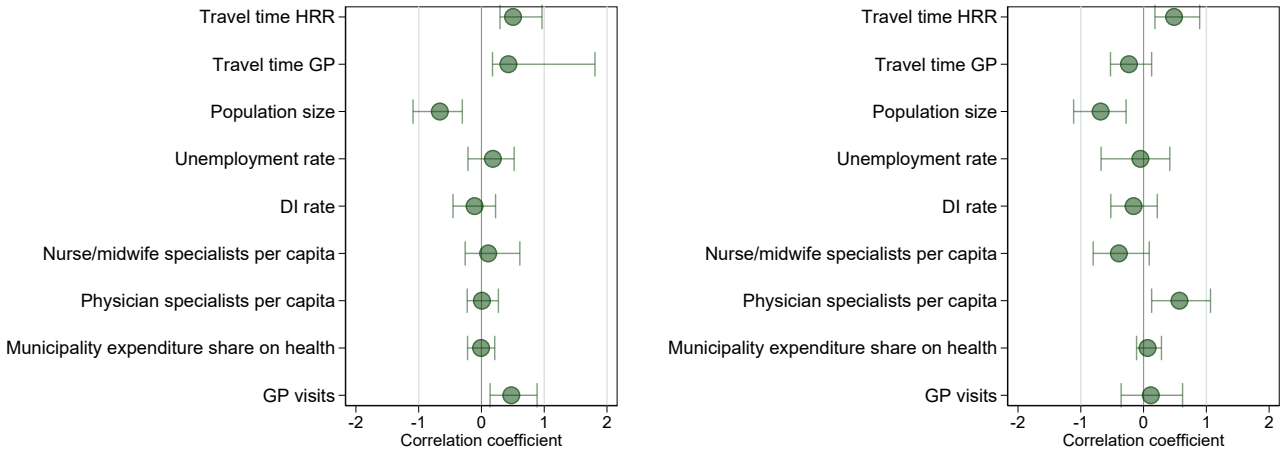
Existing literature finds that regions with more specialists, or more physicians involved in treating one patient, have higher costs and quantities of care (see, e.g. Baicker and Chandra, 2004; Currie et al., 2016). In the bivariate regressions, however, neither the number of nurse specialists/midwives per capita, nor physician specialists have any explanatory power.

Our utilization measure captures hospital utilization, which are services provided by the state/regional level. Other healthcare services, such as primary care, and services to the elderly and disabled, are organized at lower levels (municipalities). The health share of the local budget does not seem to have any economic or statistical significant association with the estimated hospital region effects. On the other hand, there is a significant, positive correlation between the hospital region effects and average number of visits to primary care physicians. This finding could potentially reflect complementarities of primary and specialist care, which is particularly interesting as primary care is the gatekeeper for specialist healthcare.

In the multivariate model, travel time to closest hospital and small population size still predict higher hospital region effects. Moreover, the number of physician specialists per capita is now positively correlated with hospital region effects. This positive association is consistent with a story where specialists are driving utilization through e.g. providing more care. However, the correlation may also reflect long term dynamics, in that regions with poorer health may have attracted more high skilled medical professionals over time in response to patient demand.

5 Health outcomes – cause-specific mortality

In the results so far, we have seen that there is substantial variation across regions in healthcare utilization that cannot be explained by observable or time-invariant patient characteristics alone. Variation that is driven by hospital region effects is potentially concerning for policymakers, as it is suggests some places provide inefficiently high or low



(a) Bivariate regressions

(b) Multivariate regression

Figure 5: Correlates of estimated hospital region effects.
Standard errors made based on the wild bootstrap using the empirical t distribution, clustered at the HRR level. Significance level in brackets: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

levels of care. The optimal utilization level is, however, difficult to pinpoint. Utilization might be high due to over-treatment, suggesting inefficiently high levels of utilization. On the other hand, high utilization levels may be efficient if this is due to higher quality of care. A natural question is therefore whether regions with high hospital region effects achieve better health outcomes.

5.1 Empirical models of mortality

Models linking health outcomes to average utilization rates are typically difficult to interpret because the causality tends to run both ways - while the regional level of care may affect the health outcomes of residents, the health status of residents would also influence the demand for healthcare and utilization patterns. The econometric model in this paper has identified and estimated hospital region effects that control for patient demand. In this section, we estimate a set of models linking the two estimated components of HRR-level utilization - the estimated patient and hospital region fixed effects, to cause specific mortality.

Our baseline empirical approach estimates a linear model of regional mortality rates. For these models, the sample is collapsed by HRR and demographic group (i.e. age and 1-year age), yielding a sample of regional average age and gender specific mortality rates

over the 2008-2013 period. Letting d_{gj} denote the mortality rate of group g in region j, we estimate the following regression equation:

$$d_{gj} = x_{gj}\beta^x + \hat{\gamma}_j\beta^\gamma + \bar{c}_j\beta^c + \varepsilon_{gj} \tag{5.1}$$

where x_{gt} is a vector of gender and age dummies. $\hat{\gamma}_j$ is the estimated γ hospital region effect of region j, while $\bar{c}_j$ is the average estimated patient effect, defined as the sum of the individual and age effects from Equation (3.1). For reference we also estimate a model linking d_{gj} to average local utilization $\hat{y}_j$. In order to ease the quantitative interpretation of our estimates, before the model is estimated average utilization $\hat{y}_j$ and the two components of utilization $\hat{\gamma}_j, \bar{c}_j$ are standardized to have mean zero and a standard deviation of one.

The primary parameter of interest in Equation (5.1) is β^γ. This parameter captures the expected change in mortality rates associated with a one standard deviation increase in hospital region effect. We want to stress that the estimated relationship should be thought of as predictive rather than causal. Hospital region effects may be correlated with other place characteristics such as climate, pollution or economic opportunity, that affect mortality independent of healthcare spending (see also Finkelstein et al., 2018). Moreover, there could be nonrandom sorting on health, e.g. if places with high healthcare spending conditional on patient demand also have healthier residents.[14] These concerns complicate the interpretation of the overall mortality model in Equation (5.1).

Meanwhile, there is considerable variation between the different causes of death in how much specialist healthcare patients utilize in their last years of life. Total healthcare utilization during the final three years before death averages 35,052 USD for patients who die of cancer, compared to 19,090 USD for patients who die of cardiovascular conditions and 15,157 USD for deaths from external causes. If higher healthcare supply as measured by larger estimated hospital region effect has a negative effect on mortality, effects should be larger for relatively utilization intensive causes of death. To study this hypothesis, we

[14]Note that our empirical model ensures that variation in healthcare spending that results directly from variation in patient demand is purged from the estimated hospital region effects.

estimate the regression model in Equation (5.1) separately for each cause of death m, and link the estimated $\hat{\beta}^{\gamma m}$'s to the average utilization intensity within each cause.[15] ICD10 codes are used to define grouped causes of death $m = 1, ..., M$.

If the supply channel is important, effects should be higher for utilization intensive causes. On the other hand, a lack of correlation between the estimated $\hat{\beta}^{\gamma m}$ and utilization intensity may indicate that the relationship between healthcare supply and mortality is driven by unobserved place heterogeneity.

In addition to linear models of aggregate death rates, we also use the underlying individual level data to estimate Cox proportional hazards model of mortality.[16] For overall mortality, the hazard function at age τ takes the following form:

$$r(\tau) = h(\tau) \exp(x_{it}\beta^x + \hat{\gamma}_j\beta^\gamma + \bar{c}_j\beta^c) \tag{5.2}$$

where x_{it} is gender and education and $h(\tau)$ is an unspecified baseline hazard. For cause-specific mortality, we estimate the corresponding competing risks models (Fine and Gray, 1999), treating deaths by causes other than m as the competing event. The models will be estimated by maximum likelihood.

All models are estimated on the full sample of stayers and movers, without age restrictions. Estimating the models on the sample of only movers would allow for the inclusion of origin fixed effects, however we are reluctant to do so for at least three reasons. First, as indicated in Table 1, the sample of movers is much smaller and deaths comparatively rare. Second, it would require strong assumptions on exogenous mortality in order for the model to be informative. In particular, the models do not include individual fixed effects, and consequently require mobility decisions to be uncorrelated with health status. This is a stronger assumption than our twoway fixed effects model of utilization rates, which requires only exogeneity with respect to *changes* in health status. Third, to the extent that healthcare supply affects mortality, we might expect the effect to occur with some lag. This kind of dynamics creates complications not present in the analysis of utilization

[15] Alternative measures such as the share admitted for at least one inpatient stay yield similar results.
[16] Duration models with a Gompertz baseline hazard provide qualitatively similar results.

Table 5: Mortality rates

	(1) All causes	(2) Cancer	(3) Heart	(4) External
Model 1: HRR average utilization				
Log utilization	-0.401	-0.282**	-0.0172	-0.00965
p-value	[0.464]	[0.00100]	[0.955]	[0.720]
Model 2: Estimated place, patient effects				
$\hat{HRR}$	-0.365	-0.216**	-0.0565	-0.00875
p-value	[0.237]	[0.0220]	[0.610]	[0.559]
$\bar{c}$	-0.0929	-0.180*	0.0728	-0.000563
p-value	[0.839]	[0.0660]	[0.696]	[0.985]
$\bar{y}$	8.480	2.199	2.700	0.517
N	5920	5920	5920	5920

Notes: Dependent variable is the death rate per 1000 inhabitants over the 2008-2013 period. Observations weighted by the population in each demographic x HRR cell. Regressions include controls for gender and 1-year age. Significance tests based on the wild bootstrap using the empirical t distribution, clustered at the HRR level. p values testing whether the effect is zero in brackets: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

patterns.

5.2 Mortality results

Table 5 presents selected estimates from linear models of death rates. The first column of Table 5 shows results on mortality from all causes. Panel one shows that average regional utilization is not significant in explaining variation in mortality. Panel two shows that this holds also when the model includes the estimated patient and hospital region effects, $\bar{c}$ and $\hat{HRR}$. That is, places that have higher healthcare utilization do not appear to have lower mortality rates.

Estimates for deaths from cancer, heart disease and external causes are shown in columns 2-4 of Table 5. The model finds a significant and negative estimate of higher $\hat{HRR}$ on deaths from cancer. A one standard deviation increase in the hospital region effect predicts 28 fewer cancer deaths per 100,000, or a 12.8% reduction relative to the mean. For the two other causes of death studied, hospital region effects have no significant association with mortality rates. To summarize, Table 5 seems to find a negative association of hospital region effects and mortality only for deaths from cancer, which is

the most treatment-intensive of the three groups.

Estimated duration models (Appendix Table E1) yield qualitatively similar results: A one standard deviation increase in hospital region effects predicts 17.4 percent reduction in all cause mortality. Competing risks models estimate a corresponding 18.8 percent reduction in the cancer mortality rate. Meanwhile, the models find no significant correlations between hospital region effects and deaths from heart disease and deaths from external causes.

To see if this pattern holds more generally we use the first letter of the ICD 10 code to classify all deaths occurring in the sample period. Equation (5.1) is then estimated separately for each group.

Figure 6 plots the estimated $\hat{\beta}^{\gamma}$ from these regressions against the average log utilization around the year of death. If higher hospital region effects have an impact on mortality, we would expect the estimated $\hat{\beta}^{\gamma}$ to fall with our measure of treatment intensity: higher hospital region effects should have larger effects for conditions that typically involve more medical interventions in the years leading up to death. Figure 6 finds that this pattern seems to be present in our data: the model finds larger, negative $\hat{\beta}^{\gamma}$ for conditions with high utilization rates, while the estimated $\hat{\beta}^{\gamma}$ are around zero for conditions that are less utilization intensive.

To summarize, while healthcare utilization is not significant in predicting average mortality rates, we find some evidence of association with cause specific mortality rates. In particular, higher hospital region effects are significant in predicting lower rates of cancer deaths. More generally, higher estimated hospital region effects are associated with reductions in mortality from causes of death that are characterized by higher healthcare utilization around the time of death. The estimated effects tend to be somewhat imprecisely estimated. Still, these results suggest that high place-specific utilization may translate to better health outcomes, meaning high utilization regions are not necessarily inefficient.

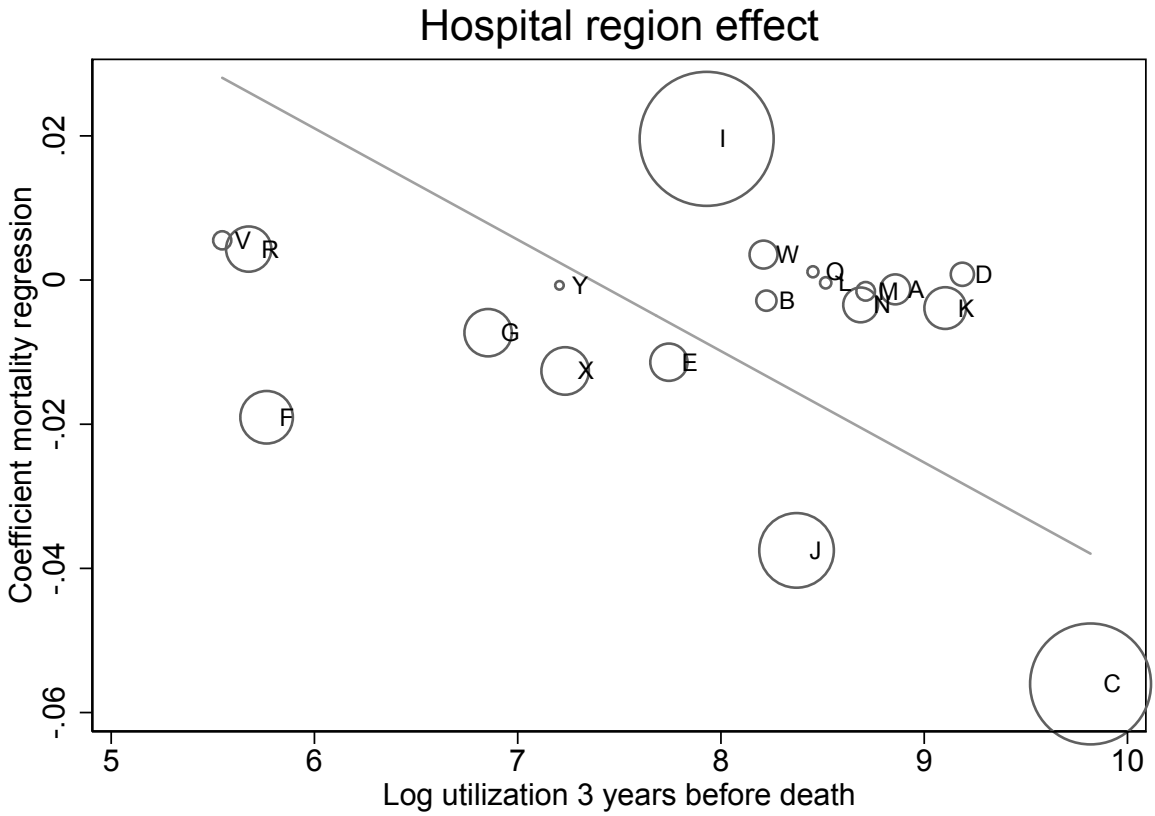


Figure 6: Hospital region effects and mortality, by cause of death.
Figure plots estimated $\hat{\beta}^{\gamma}$ and average log healthcare utilization in the three years before death, by ICD 10 chapter. Observations weighted by number of deaths in each category.

6 Conclusion

This paper analyzes regional variation in healthcare utilization, with two main objectives. First, we distinguish between two distinct sources of regional variation: patient effects, capturing variation in demand across patient population, and hospital region effects, which we can interpret as the supply of healthcare broadly defined. Following Finkelstein et al. (2016), we use migration data to decompose regional variation in healthcare utilization, finding that place-specific factors account for roughly half of the total difference between average utilization in high and low utilization regions, while the rest is explained by patient characteristics.

The second part of the analysis links the estimated hospital region effects to mortality data. The results suggest that higher hospital region effects are not significantly associated with overall mortality. Meanwhile, there seems to be a statistically significant

negative association between higher hospital region effects and mortality for utilization intensive causes of death, such as cancer. The policy implications of this result are not immediately clear. First, we should be careful in drawing policy conclusions from these models, as they are primarily predictive rather than causal. Moreover, even if we were willing to accept the estimates as causal, there could be heterogeneity in the ability of hospitals to deliver quality care (Chandra and Staiger, 2017). Finally, from a cost benefit perspective, the modest reductions in cancer mortality uncovered by our analysis may not be enough to justify higher spending.

One point of interest was how the relative importance of place would differ in a centralized system like Norway, compared to the literature which is primarily focused on the U.S. One could argue that Norway's centralized single payer system, with hospital physicians employed on a fixed salary (rather than on fee for service or capitation based contracts), should be expected to have less variation in place-specific factors in healthcare delivery compared to a more decentralized system like in the U.S. Our estimated share of healthcare utilization that can be contributed to hospital regions is slightly smaller, yet not statistically different from the effect found in Finkelstein et al. (2016), who estimate that between 50-60% of total variation reflect supply differences. Their paper uses data on elderly patients, while we look at the full population, making direct comparison difficult.[17] Still, the fact that the range of the reported estimates tend to overlap indicates that the importance of place-specific factors is not dramatically different in the two populations, despite significant institutional differences.

[17]When we restrict our estimation sample to individuals age 65 and older, we find that the share of regional variation attributable to hospital region effects increases somewhat, though the estimated event study plots for these older patients suggest potentially endogenous mobility, complicating the interpretation of this result.

References

Abowd, J. M., R. H. Creecy, F. Kramarz, et al. (2002). Computing person and firm effects using linked longitudinal employer-employee data. Technical report, Center for Economic Studies, US Census Bureau.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008). High wage workers and low wage firms: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society) 171*(3), 673–697.

Anthony, D. L., M. B. Herndon, P. M. Gallagher, A. E. Barnato, J. P. Bynum, D. J. Gottlieb, E. S. Fisher, and J. S. Skinner (2009). How much do patients' preferences contribute to resource use? *Health Affairs 28*(3), 864–873.

Baicker, K. and A. Chandra (2004). The productivity of physician specialization: evidence from the medicare program. *American Economic Review 94*(2), 357–361.

Baicker, K., A. Chandra, J. S. Skinner, and J. E. Wennberg (2004). Who you are and where you live: how race and geography affect the treatment of medicare beneficiaries. *Health Affairs 23*, VAR–33.

Bonhomme, S., T. Lamadon, and E. Manresa (2017). A distributional framework for matched employer employee data. Technical report.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50*(2), 317–372.

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of west german wage inequality*. *The Quarterly Journal of Economics*.

Chandra, A., D. Cutler, and Z. Song (2012). Who ordered that? the economics of treatment choices in medical care. *Handbook of health economics 2*, 397–432.

Chandra, A. and D. O. Staiger (2017). Identifying sources of inefficiency in health care. Technical report, National Bureau of Economic Research.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *The American Economic Review 104*(9), 2593–2632.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014b). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *The American Economic Review 104*(9), 2633–2679.

Chetty, R. and N. Hendren (2018a). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics 133*(3), 1107–1162.

Chetty, R. and N. Hendren (2018b). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics 133*(3), 1163–1228.

Chetty, R., N. Hendren, and L. F. Katz (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review 106*(4), 855–902.

Combes, P.-P., G. Duranton, and L. Gobillon (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics 63*(2), 723–742.

Currie, J., W. B. MacLeod, and J. Van Parys (2016). Provider practice style and patient health outcomes: the case of heart attacks. *Journal of health economics 47*, 64–80.

Cutler, D., J. Skinner, A. D. Stern, and D. Wennberg (2018). Physician beliefs and patient preferences: A new look at regional variation in health care spending. *American Economic Journal: Economic Policy*.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies 82*(3), 991–1030.

Doyle, J., J. Graves, J. Gruber, and S. Kleiner (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *The journal of political economy 123*(1), 170.

Doyle Jr, J. J. (2011). Returns to local-area healthcare spending: evidence from health shocks to patients far from home. *American economic journal. Applied economics 3*(3), 221.

Fine, J. P. and R. J. Gray (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association 94*(446), 496–509.

Finkelstein, A., M. Gentzkow, and H. Williams (2018). Place-based drivers of mortality: Evidence from migration. *Working Paper*.

Finkelstein, A., M. Gentzkow, and H. Williams (forthcoming, 2016). Sources of geographic variation in health care: Evidence from patient migration. *Quarterly Journal of Economics*.

Fisher, E. S., J. P. Bynum, and J. S. Skinner (2009). Slowing the growth of health care costs - lessons from regional variation. *New England Journal of Medicine 360*(9), 849–852.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and E. L. Pinder (2003a). The implications of regional variations in medicare spending. part 1: the content, quality, and accessibility of care. *Annals of internal medicine 138*(4), 273–287.

Fisher, E. S., D. E. Wennberg, T. A. Stukel, D. J. Gottlieb, F. L. Lucas, and E. L. Pinder (2003b). The implications of regional variations in medicare spending. part 2: health outcomes and satisfaction with care. *Annals of internal medicine 138*(4), 288–298.

Fiva, J. H., A. H. and G. J. Natvik (2017). Local government dataset.

Gibbons, S., H. G. Overman, and P. Pelkonen (2014). Area disparities in britain: Understanding the contribution of people vs. place through variance decompositions. *Oxford Bulletin of Economics and Statistics 76*(5), 745–763.

Huitfeldt, I. (2016). Spending the night? effects of hospital reimbursement rates on medical treatment and health outcomes. *Working Paper*.

Jackson, C. K. (2013). Match quality, worker productivity, and worker mobility: Direct evidence from teachers. *Review of Economics and Statistics 95*(4), 1096–1116.

Joynt, K. E. and A. K. Jha (2012). The relationship between cost and quality: no free lunch. *JAMA 307*(10), 1082–1083.

Lamadon, T., M. Mogstad, and B. Setzler (2017). Imperfect competition and rent sharing in the u.s. labor market. Technical report, Working Paper.

Mansfield, R. K. (2015). Teacher quality and student inequality. *Journal of Labor Economics 33*(3), 751–788.

Molitor, D. (2018). The evolution of physician practice styles: Evidence from cardiologist migration. *American Economic Journal: Economic Policy*.

Roodman, D. (2015, Revised 18 Feb 2018). "boottest: Stata module to provide fast execution of the wild bootstrap with null imposed. Statistical Software Components S458121, Boston College Department of Economics.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics 125*(1), 175–214.

Skinner, J. (2011). Causes and consequences of regional variations. *Handbook of Health Economics 2*.

Song, Y., J. Skinner, J. Bynum, J. Sutherland, J. E. Wennberg, and E. S. Fisher (2010). Regional variations in diagnostic practices. *New England Journal of Medicine 363*(1), 45–53.

Appendix A

Table A1: Distribution of (level) utilization and log utilization

		Percentile				
	mean	50th	75th	90th	95th	99th
Log						
Stayers	2.36	0	5.64	7.99	8.65	9.95
Movers	2.21	0	5.40	7.76	8.45	9.55
Total	2.35	0	5.40	7.97	8.64	9.93
Levels						
Stayers	1184.6	0	280.5	2961.4	5736.6	20853.1
Movers	906.3	0	221.4	2334.9	4669.7	14014.1
Total	1173.5	0	221.4	2878.6	5676.6	20605.5

Notes: Distribution of utilization and log(utilization+1) in USD.

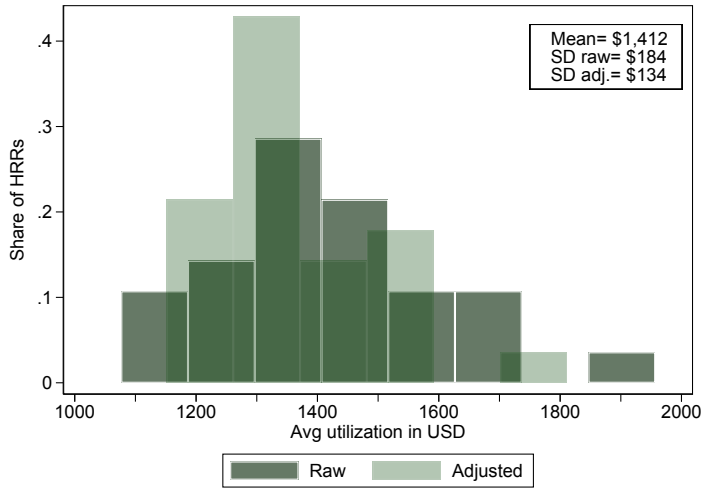


Figure A1: Distribution of utilization (in USD).
Figure shows the distribution of average utilization (in USD) per patient at the 28 hospital regions. Dark columns are based on raw data, lighter columns adjust for age, gender and education.

Appendix B

In Section 3.3, we presented an expression (Equation 3.6) linking the parameters of Equation (3.5) to model (3.1). In our case where people move only once this can be seen as follows. Observe that Equation (3.1) can be rewritten as

$$\begin{aligned} y_{it} &= \alpha_i + \gamma_{j'(i)} + I_{k(i,t)>0}(\gamma_{j''(i)} - \gamma_{j'(i)}) + X_{it}\lambda + \varepsilon_{it} \\ y_{it} &= \underbrace{\alpha_i + \gamma_{j'(i)}}_{\tilde{\alpha}_i} + I_{k(i,t)>0} \underbrace{\left(\frac{\gamma_{j''(i)} - \gamma_{j'(i)}}{\bar{y}_{j''(i)} - \bar{y}_{j'(i)}}\right)}_{\text{Hospital region share}} \underbrace{\left(\bar{y}_{j''(i)} - \bar{y}_{j'(i)}\right)}_{\delta_i} + X_{it}\lambda + \varepsilon_{it}. \end{aligned} \tag{B.1}$$

where I is an indicator function equal to one if relative year of moving, $k(i,t)$, is greater than zero (i.e. $k(i,t) = t - move\ year_i > 0$), and zero otherwise.[18] For non-movers, I is 0 for all t. $\gamma_{j'(i)}$ and $\gamma_{j''(i)}$ denote hospital region effects in origin and destination, respectively. Note that the term $\frac{\gamma_{j''(i)} - \gamma_{j'(i)}}{\bar{y}_{j''(i)} - \bar{y}_{j'(i)}}$ corresponds to the share of utilization difference between the origin and destination regions which can be explained by hospital region effects.

Using the definitions $\tilde{\alpha}_i = \alpha_i + \gamma_{j'(i)}$ and $\delta_i = \bar{y}_{j''(i)} - \bar{y}_{j'(i)}$ (defined in Section 3.3), Equation (B.1) can be rewritten as

$$y_{it} = \tilde{\alpha}_i + I_{k(i,t)>0} \left(\frac{\gamma_{j''(i)} - \gamma_{j'(i)}}{\bar{y}_{j''(i)} - \bar{y}_{j'(i)}}\right) \delta_i + X_{it}\lambda + \varepsilon_{it}. \tag{B.2}$$

Below, we reproduce the event-study equation from Section 3.3:

$$y_{it} = \alpha_i' + \beta_{k(i,t)}\delta_i + X_{it}\lambda' + \varepsilon_{it}'. \tag{3.5, revisited}$$

Equation (B.2) and equation (3.5) will be equivalent if the following holds true for β_k:

$$\beta_k = \begin{cases} 0 & \text{if } k < 0 \\ \frac{\gamma_{j''(i)} - \gamma_{j'(i)}}{\bar{y}_{j''(i)} - \bar{y}_{j'(i)}} & \text{if } k > 0 \end{cases}$$

[18] Note that since the variables are not measured in continuous time, the indicator function is misspecified in the year of move. In practice, we solve this by dropping the move year when estimating the two-ways fixed effects models.

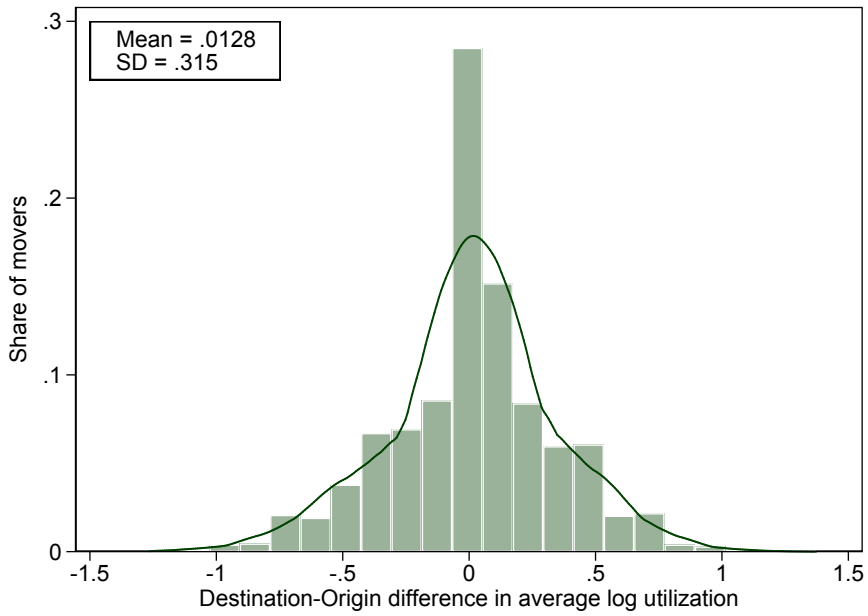
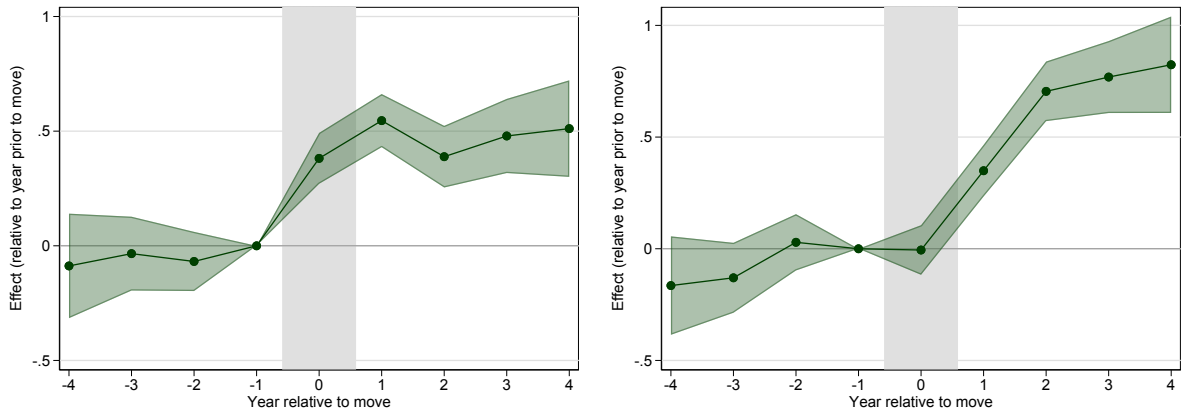


Figure B1: Distribution of destination-origin difference in log utilization (δ_i).
Figure plots the distribution of δ_i, i.e. the difference in average log utilization in the destination and origin regions. Sample is all movers (N=707,464 person-years)

Appendix C



(a) From high- to low utilization hospital regions **(b)** From low- to high utilization hospital regions

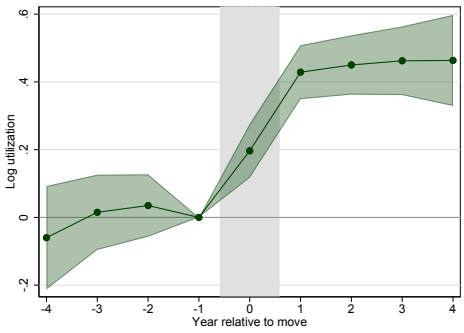
Figure C1: Event study by direction of move.
Figure shows point estimates of β_k from Equation (3.5) when the event study is estimated separately by the direction of move. Panel (a) displays utilization for individuals moving from high- to low utilization regions ($\delta_i < 0$), while panel (b) plots utilization for individuals moving from low- to high utilization regions ($\delta_i > 0$).

Appendix D

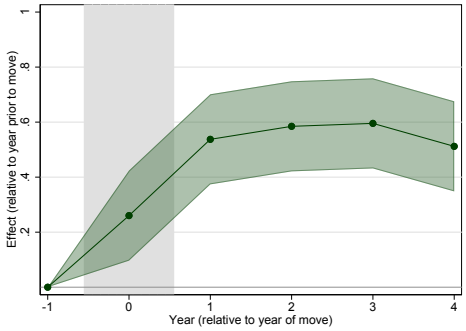
Table D1: Variance Decomposition of Hospital Level Log Utilization

	(1) Not bias corrected	(2) Bias corrected
Variance of log utilization	0.12	0.12
Variance of patient effects	0.041	0.034
Variance of year effects	0.0000012	0.0000012
Variance of hospital region effects	0.038	0.031
Covariance patient and year effects	-0.00017	-0.00017
Covariance patient and hospital region effects	0.020	0.027
Covariance year and patient effects	-0.00017	-0.00017
Covariance year and hospital region effects	-0.00016	-0.00016
Covariance hospital region and patient effects	0.020	0.027
Covariance hospital region and year effects	-0.00016	-0.00016
Correlation patient and year effects	-0.77	-0.82
Correlation patient and hospital region effects	0.50	0.73
Correlation year and patient effects	-0.77	-0.82
Correlation year and hospital region effects	-0.73	-0.81
Correlation hospital region and patient effects	0.50	0.73
Correlation hospital region and year effects	-0.73	-0.81

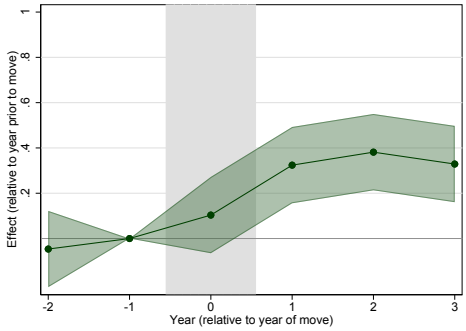
Notes: Table shows variance decomposition at the hospital region level. Parameters estimated in Equation (3.1) are averaged within hospital referral regions. Patient effects are compounds of id and age effects. Bias corrected variances of fixed effects are estimated using a split-sample approach (Dhaene and Jochmans, 2015).



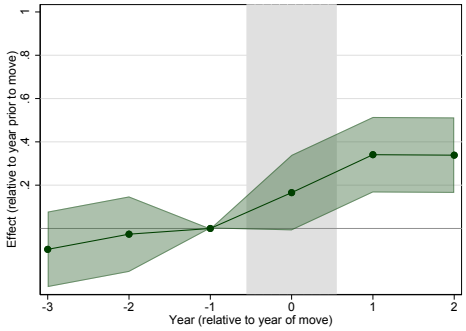
(a) Alternative market definition (19 health trusts)



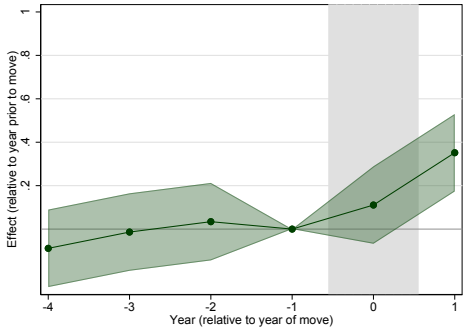
(b) Balanced population of movers in 2009



(c) Balanced population of movers in 2010



(d) Balanced population of movers in 2011



(e) Balanced population of movers in 2012

Figure D1: Event-study figures for alternative samples and hospital region definition. Panel (a) uses the 19 health trusts as hospital referral region. Panels (b)-(e) use balanced samples of movers depending on the year of move.

Appendix E

Table E1: Mortality: competing risks model

	(1) All causes	(2) Cancer	(3) Heart	(4) External
Model 1: HRR average utilization				
Log utilization	0.984	0.936**	1.123***	0.845*
	(-0.29)	(-2.20)	(2.94)	(-1.80)
Model 2: Estimated hospital region, patient effects				
$H\hat{R}R$	0.850**	0.845***	1.034	0.992
	(-2.54)	(-4.27)	(0.42)	(-0.06)
$\bar{c}$	1.118**	1.024	1.209***	0.737
	(2.10)	(0.71)	(3.61)	(-1.55)
N	4789765	4789765	4789765	4789765

Notes: Cox mortality estimates. Column (1) shows selected estimates from Cox models, columns (2) - (4) show selected estimates from competing risks regressions. Exponentiated coefficients; t statistics in parentheses. Regressions include controls for gender and 1-year age. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix F

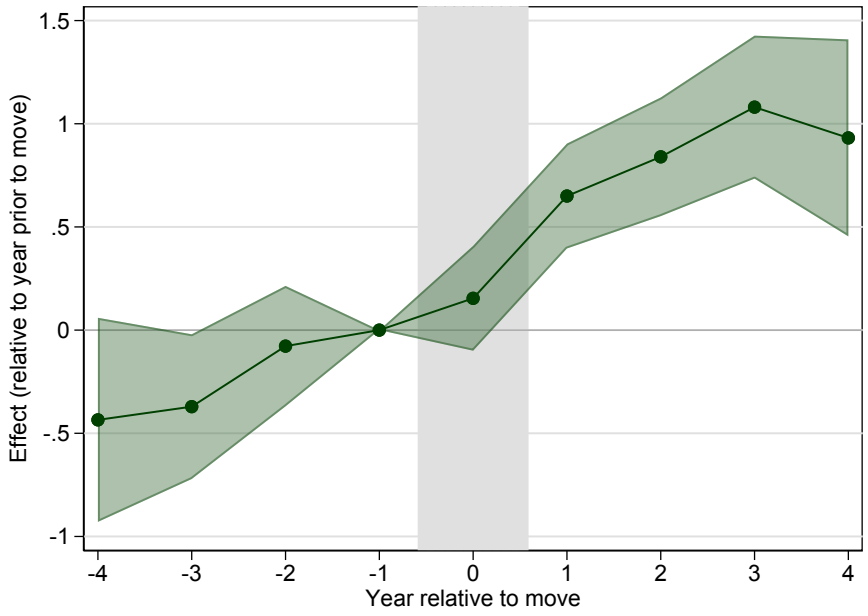


Figure F1: Event study on ages 65-99 ('Medicare population')

	N	Mean of y	Diff in y	Diff in hosp reg	Hosp reg share
Ages 65-99 ('Medicare sample')	4,694,649	3.94	0.57	0.38	0.67
Baseline sample	15,598,499	2.50	0.53	0.26	0.49

Statistics Norway

Postal address:
PO Box 2633 St. Hanshaugen
NO-0131 Oslo

Office address:
Akersveien 26, Oslo
Oterveien 23, Kongsvinger

E-mail: ssb@ssb.no
Internet: www.ssb.no
Telephone: + 47 62 88 50 00

ISSN: 1892-753X

Statistisk sentralbyrå
Statistics Norway

Design: Siri Boquist