



# Regional variation in health care utilization and mortality<sup>☆</sup>

Anna Godøy<sup>a,b,c</sup>, Ingrid Huitfeldt<sup>c,d,\*</sup>

<sup>a</sup> Institute for Research on Labor and Employment, University of California, Berkeley, United States

<sup>b</sup> Institute for Social Research, Norway

<sup>c</sup> Statistics Norway, Norway

<sup>d</sup> Frisch Centre, Norway



## ARTICLE INFO

### Article history:

Received 17 April 2019

Received in revised form 25 October 2019

Accepted 27 October 2019

Available online 19 February 2020

### JEL classification:

H51

I1

I11

I13

### Keywords:

Health care supply

Health care demand

Health care spending

Regional variation

Health outcomes

## ABSTRACT

Geographic variation in health care utilization has raised concerns of possible inefficiencies in health care supply, as differences are often not reflected in health outcomes. Using comprehensive Norwegian microdata, we exploit cross-region migration to analyze regional variation in health care utilization. Our results indicate that place factors account for half of the difference in utilization between high and low utilization regions, while the rest reflects patient demand. We further document heterogeneous impacts of place across socioeconomic groups. Place factors account for 75% of the regional utilization difference for high school dropouts, and 40% for high school graduates; for patients with a college degree, the impact of place is negligible. We find no statistically significant association between the estimated place effects and overall mortality. However, we document a negative association between place effects and utilization-intensive causes of death such as cancer, suggesting high-supply regions may achieve modestly improved health outcomes.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Geographic variation in health care utilization has raised concerns of possible inefficiencies in the supply of health care. In particular, we may be concerned that some regions are spending too much on health care, 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 health care utilization driven by place-specific factors, as opposed to variation in underlying patient health? Second, is higher regional supply of health care associated with better health outcomes?

We argue that both questions are central to policymakers seeking to understand regional variation in health care utilization. In principle, regional variation in health care

<sup>☆</sup> 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 the editor, two anonymous referees, Anna Aizer, Amy Finkelstein, Simen Gaure, Sverre A.C. 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 ESOP lunch seminar at the University of Oslo, 2016 are gratefully acknowledged.

\* Corresponding author at: Statistics Norway, Akersveien 26, 0177 Oslo, Norway.

E-mail addresses: [anna.godoy@berkeley.edu](mailto:anna.godoy@berkeley.edu) (A. Godøy), [ingrid.huitfeldt@ssb.no](mailto:ingrid.huitfeldt@ssb.no) (I. Huitfeldt).

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 contrary, 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 health care 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 health care utilization (Finkelstein et al., 2016; Song et al., 2010; Baicker et al., 2004; Fisher et al., 2009, 2003a,b). Finkelstein et al. (2016) find 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 health care spending is primarily driven by the supply side (see, e.g. 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 health care system, where hospitals are similar in terms of payment schemes and physician incentives, and patients face no to negligible copayments. 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.<sup>1</sup>

The present paper is also related to a large literature studying the link between education and health care. There is a well-documented socioeconomic gradient in health outcomes (see, e.g. Cutler et al., 2008, for a review). Education is associated with better self-reported health, lower risk of being diagnosed with several conditions and lower mortality rates. Evidence suggests there might also be a socioeconomic gradient in health care utilization. Papers from European countries and the U.S. find that high income groups are more likely to access specialist health services compared to lower income groups who are, if anything, more likely to use general practitioner care (Van Doorslaer et al., 2000; d'Uva and Jones, 2009). Similar patterns are found in Norway: Vikum et al. (2012) find that high income and more educated patients are more likely to see a med-

ical specialist or receive outpatient treatment at hospitals, but no relationship is found for general practitioner visits (or inpatient hospital care). Moreover, Fiva et al. (2014) document that highly educated individuals utilize centralized specialized treatment to a greater extent than do less educated patients. These findings are consistent with a pattern where the local availability of hospital services are less binding for more educated patients, leading us to expect a smaller impact of place for this group compared to less educated patients.

Identifying and estimating hospital region effects in the presence of patient heterogeneity is complicated by the fact that patient demand for health care 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 health care 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 patterns are as good as random with respect to health. Our model thus mirrors a difference in differences 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 health care utilization around the time of migration.

By observing patterns of individual utilization when patients move between regions, the two-way fixed effects model is able to credibly identify the relative impacts of each region on healthcare utilization. However, the estimated region fixed effects are not by themselves sufficient to draw conclusions on policy recommendations. First, while we use the terms supply and demand factors throughout the paper, we acknowledge that the research design of this paper is not ideal for distinguishing between the two. Under the assumptions of our model, the two-way fixed effects model allows us to identify an aggregate place effect. This aggregate comprises a number of factors, including hospital practice styles, physician practice styles, peer effects and geographic characteristics of the region. Second, unless these fixed effects are anchored to resulting health outcomes, we cannot know if regions with high fixed effects have an inefficiently high supply of healthcare,

<sup>1</sup> Data contain all public hospitals as well as private providers contracting with the health authorities. Very few health care institutions operate as for-profit institutions without any contract with public health authorities.

or whether it is the low utilization regions that provide too few services.

However, while the two-way fixed effects model is well suited to study utilization, the model may be less well suited to study these resulting health outcomes. One reason is that a number of potentially observable health outcomes, including mortality, by definition are once in a lifetime events. These outcomes are not possible to model directly in the two-way fixed effects model. Moreover, while healthcare utilization patterns can change very quickly, resulting health outcomes may be thought of as a slower process, where the quality and quantity of care affect outcomes with significant lags. The two-way fixed effects model identifies short run effects from the within person variation, precluding the study of such delayed impacts.

In the second part of the paper, we address these shortcomings by 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, 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 health care, 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 health care, 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 health care 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 health care shifts mortality rates, we might expect more significant correlations between region effects and mortality for conditions where patients tend to use more hospital services in the time leading up to death, such as cancer. Meanwhile, these associations should be weaker for causes associated with lower average utilization rates, like deaths from external causes.

Our linking of estimated hospital region effects to mortality also relates more generally to the literature that links school or teacher value added estimates to long run effects (see, e.g. Chetty et al., 2014a; Rothstein, 2010). As in this literature, a causal interpretation of the fixed effects on long run outcomes requires strong assumptions on the selection on observables. In particular, unobserved determinants of mortality, such as unobserved health, must be unrelated to

the estimated place fixed effects conditional on the observable characteristics. To be clear, we are not claiming to estimate true causal effects of spending, rather, the models should be seen as predictive.<sup>2</sup>

Our results show that place factors account for roughly half of the gap in average utilization between 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 richer model specifications. Disaggregating results by educational attainment, extended models document that place-specific factors are more important to people with low education compared to people with higher education. Estimated event study models indicate that place factors account for approximately 75% of the difference in average utilization between high and low utilization regions for high school dropouts, compared to 50% for patients with a high school diploma. For patients with a college degree, the event study models fail to detect a clear shift in utilization at the time of move, suggesting negligible impacts of place for this group.

The heterogeneous effects across education groups are not likely to be due to variation in patients' direct costs of treatment, as copayments are small and do not vary across regions. Existing evidence indicates that more educated patients may be better equipped to search out information on risks and benefits of different treatment options. Thus, it may appear that even in countries with low financial barriers in accessing specialist care, more educated patients may maintain an advantage in accessing specialist health care.

These findings are also related to the broader literature on the effects of place, in particular Chyn (2018)'s study of neighborhood effects, which documents differential impacts of place by background characteristics. Following the demolition of public housing in Chicago, affected children were forced to move out of disadvantaged neighborhoods, leading to significantly improved outcomes later in life: Chyn's subsample analysis treatment effects are larger for children from families where no adults are working, as well as for children from housing projects with the highest poverty rates. Our finding that place has a greater impact on utilization rates of less educated adults suggests that this pattern could hold more generally, even in a very different context.

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.

<sup>2</sup> Our approach estimating impacts by cause of death can only be interpreted causally under a narrow set of assumptions, including the strong assumption that cause of death (but not death alone) should be uncorrelated with other place characteristics.

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

Health care expenses in Norway are mainly subsidized by national insurance schemes. Hospital services are rationed by wait time, aiming at prioritizing patients based on their medical needs for health care. Some services, such as outpatient admissions and visits to GPs are subject to small copayment 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 health care expenditures exceed about 2100NOK (~260USD) all further expenses within that calendar year are reimbursed.

Health trusts, which are independent administrative enterprises comprising several institutions, have the responsibility to deliver hospital care services to inhabitants residing in defined catchment areas. The health trusts are governed as a single administrative unit with a centralized management group, i.e. they have a CEO and a board of directors, and are themselves owned by four state-owned regional health authorities who have the overarching responsibility for providing specialist health care.<sup>3</sup> Patients who are referred to specialist health care are free to choose treatment at any hospital, but in practice, very few end up at a hospital in another health region.<sup>4</sup>

In this paper, we focus on care delivered by hospitals, which does not include the primary care sector or specialists operating outside of the hospitals. There are significant institutional differences in the provision of primary and specialist care. Hospitals are funded by the regional health authorities, following guidelines set by the national government. In particular, the activity-based part of the hospital reimbursement, which is our source for calculating individual utilization, contains no geographic components. Reimbursements are made based on diagnosis-specific prices which reflect the average cost of treating any patient with that specific diagnosis. This means that any diagnosis will trigger the same hospital reimbursement regardless of the location of the patient or the hospital, and regardless of the actual costs incurred in treating the patient.

The management of primary care services is much less centralized. Primary care physicians typically operate in private practice with reimbursements from the government; they will have contracts with the municipalities but not with the regional health authorities. In addition,

patients are free to choose their own primary care provider at any time (up to two times per year); this includes people who switch primary doctors following a move between regions, but it also includes patients who switch doctor for other reasons. On the one hand, this yields a greater variation in primary care providers between patients, on the other hand, we worry that the choice of primary doctor is likely to be endogenous to health, making the identifying assumptions less likely to hold for these changes. Meanwhile, variation in the provision of primary care is a potential driver of utilization differences in specialist care. These patterns will be discussed in more detail in Section 4.5. To summarize, the Norwegian hospital system is characterized by universal coverage, low copayments, 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 may 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), 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, date of referral, diagnosis groups and diagnosis cost weight. Each patient discharged at a somatic hospital is assigned a diagnosis group that uniquely determines the reimbursement rate.

Health care 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.

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. In addition, we exclude people who move between hos-

<sup>3</sup> There are four regional health authorities, and 31 health trusts per 1.1.2018 (24 health trusts per 1.1.2012). Reimbursement from the state to the regional health authorities entails a combination of fixed budget and activity-based financing.

<sup>4</sup> 90% of elective surgeries are performed within the patients' own region, and 22% chooses a hospital at another health trust, but still within the health region (Huitfeldt, 2016).



**Table 1**  
Descriptive statistics of estimation sample.

	Stayers	Movers
Female	0.49	0.46
Norwegian-born	0.86	0.74
Enrolled in education	0.08	0.15
Educational attainment		
High school dropout	0.42	0.32
High school graduate	0.27	0.23
College	0.30	0.45
Age first observed		
30–44	0.44	0.69
45–59	0.33	0.22
60–75	0.24	0.09
First observed residence		
North	0.10	0.10
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.

pital referral regions 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.<sup>5</sup> 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% in the stayer sample. The average educational attainment level is higher for movers – 45% have a college degree – compared to 30% of the stayers. See online Appendix Tables A2 and A3 for more descriptive statistics by educational group.

On average, movers are younger than stayers; the 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 com-

pared 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 online Appendix Fig. A1 we show the full distribution of utilization in logs and levels.

Note that the observed difference between stayers and movers does not in itself pose a threat to the internal validity of the two-way fixed effects model, as our model is identified exclusively from within individual variation in outcomes. The differences may, however, be informative about the external validity of our findings, and if one wishes to extrapolate the results to the full population. We discuss this further in Section 4.4.

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

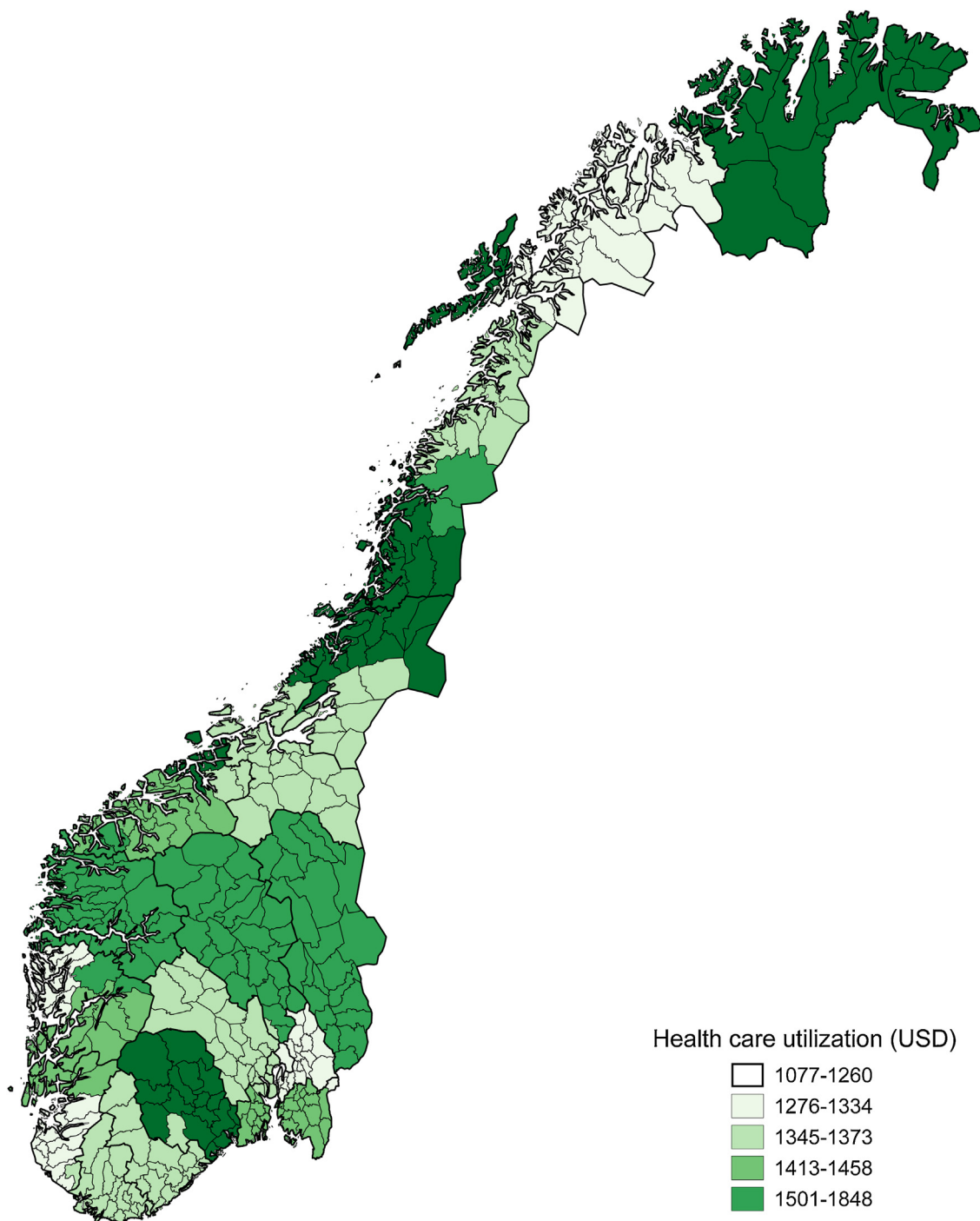
As discussed above, patients may seek medical care outside their own region of residence. In our sample, we calculate average utilization rates for the HRRs based solely on patients' residence region, regardless of where care was actually provided. About one fifth of total expenditures occur outside of a patient's HRR of residence.<sup>6</sup>

Fig. 1 shows the distribution of yearly average patient utilization across HRRs.<sup>7</sup> The average HRR has an average utilization of 1412USD per patient per year (standard devi-

<sup>5</sup> We additionally exclude individuals who move in the first or last year of our sample, as these do not provide any useful variation.

<sup>6</sup> This is a substantial share. Our regression model will identify and estimate a set of aggregate place effects, these estimated effects could reflect a number of factors including variation in how easy it is to access care in other regions, e.g. through variation in travel times or GPs culture of referring patients to out-of-region providers of specialist care. However, our analysis fails to find any systematic relationship between the share of utilization that occurs outside patients' region of residence and the estimated place effects. That is, it does not appear to be the case that places with more out-of-region utilization have systematically higher estimated place effects. See online Appendix Fig. A3.

<sup>7</sup> Fig. 1 indicates that there may be an outlier hospital region. This region is small, and accounts for only 0.45% of the sample, hence, the estimated

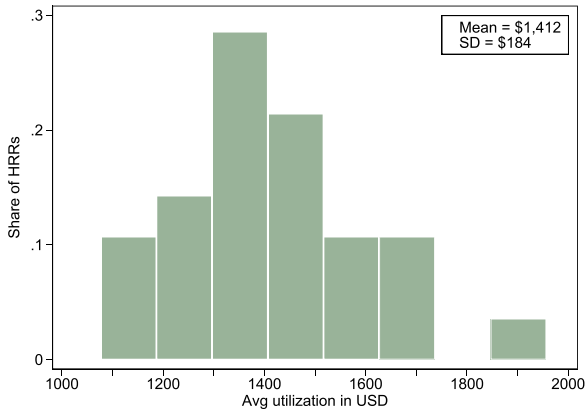


**Fig. 2.** Map of Norway. Utilization (in USD) by hospital referral region. Figure shows the geographic distribution of yearly average utilization per patient in the 28 hospital regions, divided into quintiles. Thick solid lines mark HRR borders; thin solid lines mark municipality borders. (For interpretation of the references to color in this figure citation, the reader is referred to the web version of this article.)

ation 184USD). In online Appendix Fig. 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 Fig. 2, where colors illustrate quintiles of health care utilization.

place share is virtually unchanged (47.6% compared to baseline 49%) when excluding this region.



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

### 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.<sup>8</sup>

#### 3.1. Fixed effects models

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

$$y_{it} = \alpha_i + \gamma_{j(i,t)} + X_{it}\lambda + \varepsilon_{it}, \tag{1}$$

where  $i = 1, \dots, N$ ,  $t \in \{n_{i1}, \dots, n_{iT}\}$ , and the function  $j(i, t)$  indicates the hospital region  $j$  of individual  $i$  in year  $t$ , where  $j = 1, \dots, J$ . There are  $T_i$  observations per individual and  $N^* = \sum_i T_i$  total observations.<sup>9</sup> The component  $\alpha_i$  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 calen-

dar year and age (in 5-year bins) only.<sup>10</sup> We explore richer versions of Eq. (1) in the robustness section.

Identification of individual and hospital region fixed effects hinges on the presence of movers, i.e. it requires the possibility to observe the same individual at different hospital regions (at different points in time).

As discussed in, e.g. [Bonhomme et al. \(2017\)](#), [Lamadon et al. \(2017\)](#) and [Finkelstein et al. \(2016\)](#), causal interpretation of the parameters in Eq. (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  $\alpha_i$  or  $\gamma_j$ . 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.

To study the sources of utilization differences across hospital regions, we use the estimated patient and hospital region fixed effects in a decomposition exercise. Precisely 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 (see, e.g. [Finkelstein et al., 2016](#); [Combes et al., 2008](#)).<sup>11</sup>

As a starting point for the decomposition exercise, we use the estimates from Eq. (1), and average over hospital referral regions:

$$\bar{y}_j = \bar{\hat{c}}_j + \hat{\gamma}_j, \tag{2}$$

where  $\bar{y}_j$  is the average utilization at hospital region  $j$ ,  $\bar{\hat{c}}_j$  are the estimated hospital region effects, and we label  $\bar{\hat{c}}_j$  as an average patient compound effect, comprising fixed effects for patient, age and year. Hospital referral regions are then split into two groups depending on the average patient utilization  $\bar{y}_j$  at the hospitals; average utilization is above median at hospitals in group A and below median

<sup>8</sup> Note that, the aim of this section 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. We return to the potential implications of this variation in Section 5.

<sup>9</sup> In estimation of model (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.

<sup>10</sup> Note that, as the individual fixed effects absorb the cohort effect, age and year are perfectly collinear. In [Table 3](#) 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 Eq. (1), but the robustness section shows that inclusion of such relative year dummies does not affect our results.

<sup>11</sup> See also online Appendix C for a variance decomposition exercise.

in group *B*. We next calculate the difference between the average hospital region effect estimates in the two groups, and finally divide by the difference in average utilization. This renders a hospital region share  $\frac{\hat{\gamma}_A - \hat{\gamma}_B}{\bar{y}_A - \bar{y}_B}$ . We similarly

construct the patient compound share as  $\frac{\hat{c}_A - \hat{c}_B}{\bar{y}_A - \bar{y}_B}$ .

The hospital region share tells us how much of the difference in utilization between high and low utilization regions can be explained by characteristics of the hospital region net of patient characteristics, while the patient compound share tells us how much of the difference in utilization between high and low utilization regions can be attributed to patient characteristics alone.

### 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  $\varepsilon_{it}$  in Eq. (1) consists of three separate random effects: a match component  $\eta_{ij(i,t)}$ , a unit root component  $v_{it}$ , and a transitory error  $\omega_{it}$ :

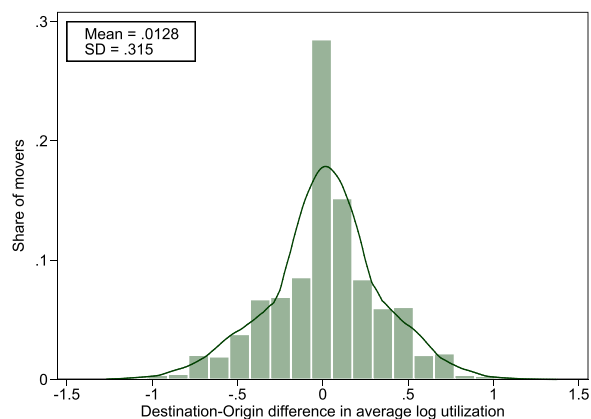
$$\varepsilon_{it} = \eta_{ij(i,t)} + v_{it} + \omega_{it}. \quad (3)$$

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  $v_{it}$  captures potential drift in the individual's utilization over time, such as health deterioration. The transitory component  $\omega_{it}$  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  $v_{it}$  and  $\omega_{it}$  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 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  $v_{it}$  will be positively correlated with the change in the hospital region effects. In other words, Eq. (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  $\omega_{it}$  may be associated with systematic moves between higher and lower utilization regions. For example, if individuals who experience a negative health shock are more likely to move to higher utilization regions, estimated hospital region effects might be amplified.



**Fig. 3.** Distribution of destination-origin difference in log utilization ( $\delta_i$ ). Figure plots the distribution of  $\delta_i$ , i.e. the difference in average log utilization in the destination and origin regions. Sample is all movers ( $N = 707,464$  person-years).

### 3.3. Event study framework

We now discuss whether the identifying assumptions are likely to hold in our data. Although we can never entirely rule out that changes in unobserved health predicts mobility patterns, hence violating the identifying assumptions, we can look for patterns in our data to mitigate potential concerns. We therefore introduce an event study framework tracking individuals' utilization before and after they move. The 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 of 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 health care 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.  $\delta_i$  can be interpreted as a scaling factor, capturing the direction and magnitude of *i*'s move. Fig. 3 shows the distribution of  $\delta_i$ . 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  $\delta_i$  is interacted with a set of dummy variables indicating event time  $k$ :

$$y_{it} = \alpha'_i + \sum_{k=-4}^4 \beta_k (\delta_i \times 1(t - t_i^* = k)) + X_{it}\lambda' + \varepsilon'_{it} \quad (4)$$

where  $t_i^*$  denotes the year of move of individual  $i$ . Here, as before,  $\alpha'_i$  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 calendar year dummies.<sup>12</sup>

The primary coefficients of interest are the  $\beta_k$ , capturing the effects of the event time coefficients multiplied by the scaling factor  $\delta_i$ . Our data allow estimation of  $\beta_k$  for  $k \in [-4, 4]$ . The coefficients  $\{\beta_{-4}, \dots, \beta_4\}$  are only identified relative to each other; we use the normalization that  $\beta_{-1} = 0$ .

In Online Appendix B we show that if the assumptions underlying the two-way fixed effects model hold, the coefficients  $\beta_k$  from Eq. (4) can be related to the parameters in Eq. (1) as follows:

$$\beta_k = \begin{cases} 0 & \text{if } k < 0 \\ \frac{\mathcal{Y}_{j''(i)} - \mathcal{Y}_{j'(i)}}{\bar{\mathcal{Y}}_{j''(i)} - \bar{\mathcal{Y}}_{j'(i)}} & \text{if } k > 0 \end{cases} \quad (5)$$

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.  $\beta_0 \in \left(0, \frac{\mathcal{Y}_{j''(i)} - \mathcal{Y}_{j'(i)}}{\bar{\mathcal{Y}}_{j''(i)} - \bar{\mathcal{Y}}_{j'(i)}}\right)$ . 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

Fig. 4 plots the estimated coefficients  $\beta_k$  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  $\beta_k$  before and after the move gives a direct indication of the presence of problematic drift. The figure shows no clear

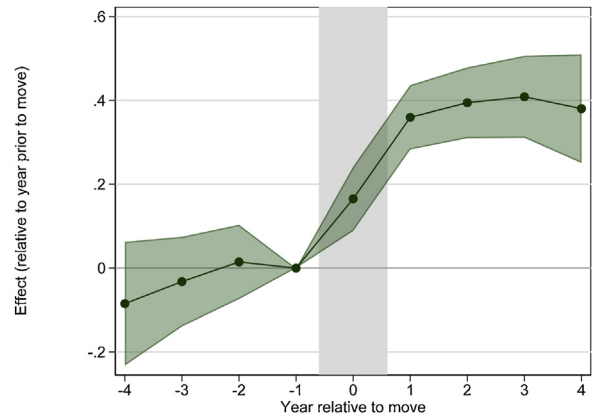


Fig. 4. Event study figure. Figure shows point estimates of  $\beta_k$  from Eq. (4).

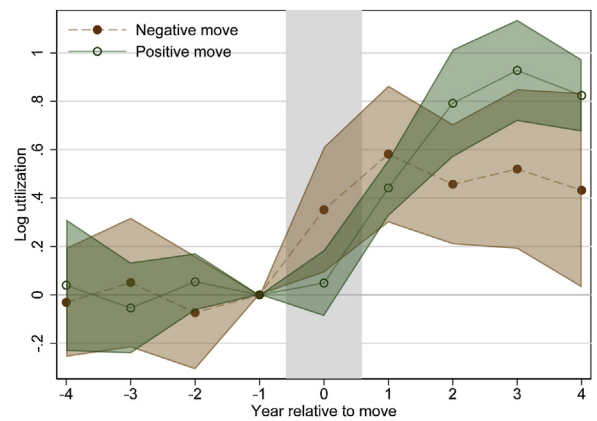


Fig. 5. Event study by direction of move. Figure shows point estimates of  $\beta_k$  from Eq. (4) when the event study is estimated separately by the direction of move. Solid dots/red area displays utilization for individuals moving from high- to low utilization regions ( $\delta_i < 0$ ), while hollow dots/green plots utilization for individuals moving from low- to high utilization regions ( $\delta_i > 0$ ). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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  $\omega_{it}$  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  $\beta_k$  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 today's patient preferences are 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 (Finkelstein et al., 2016). 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

<sup>12</sup> Note that the individual fixed effects  $\alpha'_i$  also absorb move year effects, as we restrict the model to individuals who move exactly once. As with the main model, the event study specification could additionally include event year fixed effects. This would control for differences in the propensity to move, but would not account for differences in the choice of destination. However, event study estimates are not sensitive to the inclusion of event year fixed effects.

utilization regions. Fig. 5 indicates that both the size of the jump and the post-trend are similar in the two cases.

Similarly, migration due to latent demand for health care represents a potential threat to our identification strategy. Consider a patient who experiences a negative health shock that requires a specific type of treatment that is not easily available in her region of origin. This could cause her to move to a high supply region motivated by a need for this service. If this health shock happens gradually over time, it would show up as pre-trends in the event-study figures. However, a move which is precipitated by a sudden shift in latent demand for health will bias our results, similarly to other unobserved concurrent health shocks. We believe this is unlikely; if a health shock induces individuals to move to high-utilization regions, one would expect to see elevated utilization in the first year after move for patients moving from low- to high-utilization regions - Fig. 5 displays no such patterns.

Fig. 5 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 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. In Fig. 5, individuals moving from low to high utilization regions seem to experience utilization changes that are relatively equal in magnitude (but of different sign) to individuals moving from high to low utilization regions: i.e. while the post move point estimates may appear to differ by the direction of move, the differences are non-significant in all years.<sup>13</sup> This provides suggestive evidence against the possibility of sorting on match effects.

To further assess the importance of match effects, we follow Card et al. (2013) and 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.<sup>14</sup>

The absence of match effects also provides justification for our log additive model. As a further justification for log additivity, we follow Finkelstein et al. (2016) and additionally plot the change upon move in individual log utilization against the average destination-origin difference in log utilization ( $\delta$ ). To this end, we divide the average difference between destination and origin into ventiles, and plot the average individual change in utilization upon move for each ventile. Online Appendix Fig. A4 shows that this relationship is linear, and symmetric above and below

zero, lending further support to our assumption of log additivity.<sup>15</sup>

As discussed in the previous section, if fluctuations in the transitory error  $\omega_{it}$  systematically affect mobility patterns through gradual health deterioration, we would expect to see an increasing trend in the estimated event time coefficients  $\beta_k$  for  $k < 0$ . The estimated coefficients plotted in Fig. 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. Fig. 3 also gives a first indication of the relative importance of hospital region effects. The estimated relative year coefficients  $\beta_k$  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 versa, that approximately  $1 - 0.4 = 0.6$  is the patient share. Next, we present results from the baseline two-way fixed effects model.

#### 4.2. Fixed effects estimates

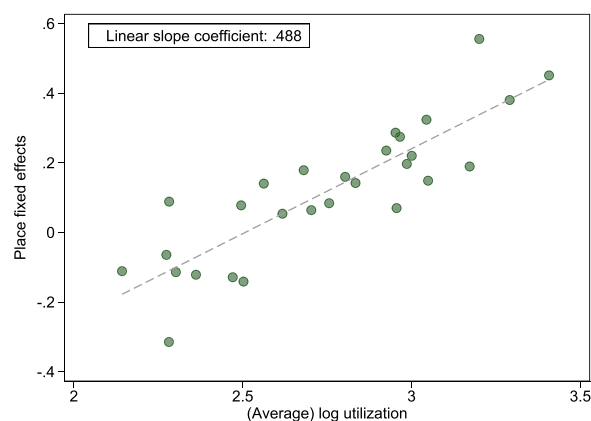
Estimation of Eq. (1) by ordinary least squares produces coefficient estimates  $\hat{\alpha}_i$ ,  $\hat{\gamma}_{j(i,t)}$ ,  $\hat{\lambda}$ , and  $\hat{\epsilon}_{it}$ . Figure 6 plots the estimated hospital region effects against average log utilization. The figure shows an upward sloping, fairly 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 decomposition exercise. Table 2 shows that place factors account for 39–59% of the difference in utilization between hospi-

<sup>13</sup> Recall that event time is scaled by both the magnitude and direction of move. Hence, utilization for individuals moving from high- to low utilization hospital regions displays a positive jump upon move although individuals decrease their utilization.

<sup>14</sup> Baseline model:  $R^2 = 0.4657$ ,  $Adj. R^2 = 0.3478$ . Saturated model  $R^2 = 0.4693$ ,  $Adj. R^2 = 0.3494$ .

<sup>15</sup> Note that log additivity does not 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 individ-



**Fig. 6.** Hospital region effects and (average) log utilization across hospital regions. Figure shows estimated hospital region effects and average patient utilization by hospital regions (from Eq. (1)). The figure shows estimated fixed effects for all 28 hospital regions, including the reference region. Estimates are scaled to the average of the estimated fixed effects.

**Table 2**

Additive decomposition of hospital level log utilization.

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

Notes: Additive decomposition of log utilization based on estimation of Eq. (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.

tal 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.

We have also used the estimated fixed effects to implement a standard variance decomposition exercise. Results, presented in online Appendix C, indicate substantial sorting of patients to regions: approximately 34% of the variance in average log utilization can be attributed to sorting, with approximately 32% and 34% explained by place and patient specific factors respectively.<sup>16</sup>

#### 4.3. Robustness tests

Our model is identified from individuals moving between hospital regions. To get a fuller understanding of the migration patterns we now turn to an analysis of

whether pre-move health predicts moving behavior. To this end, we have estimated models linking movers' pre-move health care utilization to the difference in average utilization between destination and origin ( $\delta$ ).

Online Appendix Fig. A2 shows a binned scatterplot of  $\delta$  and pre-move utilization: the figure does not indicate any systematic relationship between the two variables, moreover the slope coefficient is small and not statistically significantly different from zero.<sup>17</sup> To the extent that health care utilization proxies underlying health, this suggests health is not a main driver of mobility patterns.

For a fuller understanding of migration patterns, we have analyzed self-reported data on reasons for migration using the 2008 Survey of Relocation Motives. The survey is presented in more detail in Online Appendix D, together with additional results. Overall, the survey indicates that health concerns are a relatively uncommon reason for moving. About 8% of movers report health as a factor in the decision to move, while 3% report health concerns as their most important reasons for moving (2.5% excluding respondents older than 66).

It is worth noting that health-related migration is not in itself a threat to our identification strategy. We require only that changes in health be uncorrelated with movers' destination conditional on origin. To test this assumption, we have estimated a set of predictive models regressing (i) the estimated destination fixed effects and (ii) the difference between destination and origin place fixed effects, on an indicator variable equal to one for patients who moved due to health reasons, with and without origin fixed effects. The models, presented in online Appendix Table D2, find no indications that migrants who report moving for health reasons are any more likely to move to high utilization regions.

To further 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.<sup>18</sup> Model (1) in Table 3 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

uals, and that more weight is put on differences in the lowest part of the utilization distribution.

<sup>16</sup> Bias-corrected estimates (Dhaene and Jochmans, 2015) find the variance shares attributable to sorting, place and patients at 46%, 26% and 28% respectively.

<sup>17</sup> The regression of  $\delta$  on log pre move utilization gives a point estimate of  $-0.00077$  (0.00110).

<sup>18</sup> Fig. 1 indicated that there may be an outlier hospital region. This region is small, and accounts for only 0.45% of the sample, hence, the estimated place share is virtually unchanged (47.6% compared to baseline 49%) when excluding this region.

**Table 3**  
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 <sup>2</sup> + age <sup>3</sup> ) * 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 for alternative specifications and samples. See text for details. Standard errors are calculated using 500 bootstrap replications at the patient level.

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 fixed effects for relative year of move (i.e. event time). 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 row (7) through (10), while event study estimates for each subsample are shown in online Appendix Fig. A6. 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 online Appendix Fig. A6.

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.<sup>19</sup> This corresponds with the “medicare sample” used by Finkelstein et al. (2016). The event study plot (Online Appendix Fig. A5) 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 turn supports the exclusion of elderly patients from the main analysis. Estimating the two-way fixed effects model on this sample yields a place share of 0.67, compared with 0.49 for the baseline sample (see online Appendix Table A5). 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 online Appendix Fig. A5.

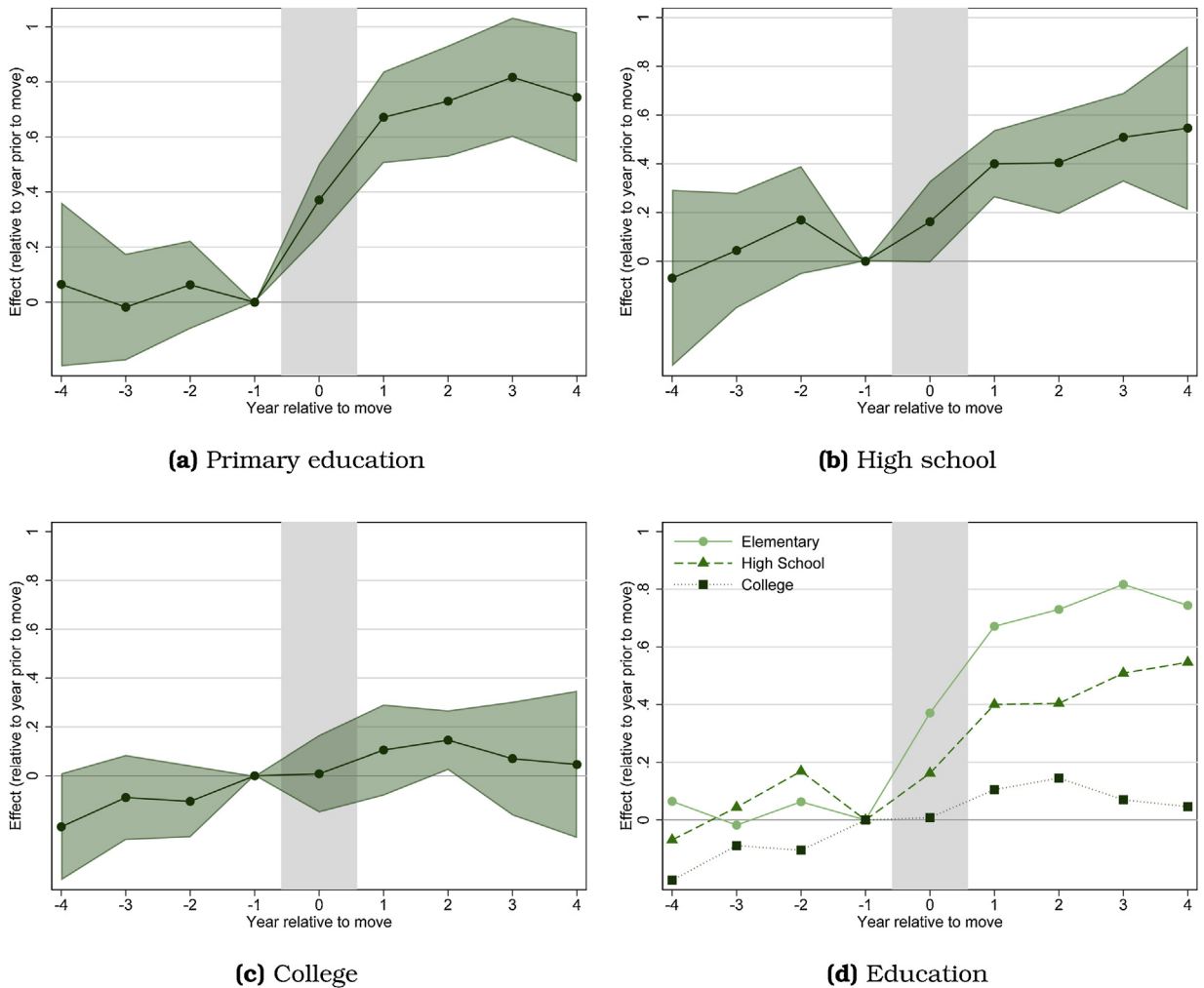
#### 4.4. Heterogeneity in the importance of place

In this section, additional models are estimated to see if the impact of place-specific factors varies with educational attainment. As discussed in the introduction, it is well established that education is a significant predictor of health outcomes and utilization patterns. Let  $educ_i$  denote a vector containing 3 indicator variables for education level: primary education (persons without a high school degree), high school graduates (who might have some post-secondary education, but not more than three years), and college (persons who have completed at least three years of post-secondary education).<sup>20</sup>

The event study model in Eq. (4) is then estimated separately for each of the three education groups. Estimated

<sup>19</sup> Recall that our baseline sample excludes individuals who are younger than 30 or older than 75.

<sup>20</sup> High school graduates include graduates from vocational tracks. College is equivalent to having at least a bachelor's degree.



**Fig. 7.** Event study model by education. Event study by educational group. Each figure shows estimates of log utilization by relative year of move. Education groups are defined as: primary education (persons without a high school degree), high school graduates (who might have some post-secondary education, but not more than a three year degree), and college (persons who have completed at least a three-year degree).

coefficients on the interacted event time dummies are plotted in Fig. 7. As before, the model predicts a discontinuous jump at the time of the move if place effects are significant; the size of the jump can be interpreted as an estimate of the relative importance of place factors in determining health care utilization.

Less educated adults are less geographically mobile than average; conditional on moving, this sample could be relatively more likely to move due to health concerns.<sup>21</sup> We may therefore worry that the assumption that mobility be conditionally random is less likely to hold for this group, potentially biasing the estimated place effects. In particular, if less educated movers are more likely to self-select to high utilization regions following negative health shocks, estimated models might overstate the impacts of place. If

this were the case, we would expect event-study models for less educated adults to exhibit evidence of increasing pre-trends in the years leading up to a move, as patient health starts deteriorating. However, the estimated event-study models of Fig. 7 give no indications that this is the case: estimated event study coefficients are close to zero for less educated patients.

The plots for individuals with primary education and high school graduates both exhibit a significant discontinuous jump at the time of move: For the primary education group, the jump is approximately 0.75 compared to 0.4 for the high school graduates. However, for college graduates, the jump is much smaller and difficult to distinguish from a linear trend: the estimated jump after the move is around 0.2. In other words, the effect of place factors on health care utilization appears to be declining with education.

From this analysis alone, it is not clear that the differential impacts of place illustrated in Fig. 7 reflect the impact of education per se. That is, these patterns could reflect variation in place impacts by correlates of educa-

<sup>21</sup> Our analysis of survey data (Online Appendix D) indicated a geographic gradient in health related migration. Conditional on moving, college educated adults are less likely to report that the move was due to health reasons than movers with only primary education.



tion, if highly educated movers tend to be systematically different from less educated movers along other dimensions, such as age and gender.<sup>22</sup> To address this, we have estimated additional models where the event time indicators are interacted with a set of dummy variables for age and gender as well as education. Results from this exercise, shown in online Appendix Fig. A7, show that place effects tend to be have larger impacts for older movers, with no differential impacts by gender. Meanwhile, the differential effect of place by education is qualitatively robust to adding these interactions, though point estimates are reduced somewhat: the model shows an estimated jump in event study coefficients of 0.5 for patients with only compulsory education, while there is no discernible impact for college-educated movers.

Estimates from studies like ours, which rely on movers, may not readily extrapolate to non-movers. For instance, we find that place matters more for those with low education; individuals with low education are also less likely to move. This suggests that place may matter more, on average, for non-movers. To assess this possibility, we have estimated the model reweighting the movers sample to be representative of the non-mover population in terms of education, age and gender.<sup>23</sup> Results are shown in online Appendix Table A4. In these models, place effects have a bigger relative impact (62% vs 49% in baseline), consistent with what we would expect given the higher residential mobility of college educated adults.

#### 4.5. Discussion

To study the drivers of regional variation, we have calculated correlations between the estimated 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, health expenses as share of local budget, and primary care visits.<sup>24</sup> Online Appendix Fig. A8(a) presents results from bivariate regressions, where the hospital region effects are regressed separately on each observable characteristics; while online Appendix Fig. A8(b) presents results from a multivariate regression where all observables are included in one regression. These correlations should be interpreted

with some caution, in particular, we have a limited sample size given that the sample only includes 28 regions. Nonetheless, we identify two suggestive patterns. First, rural regions tend to have larger estimated hospital region effects: we find statistically significant positive associated with travel time to nearest hospital and primary care physician (GP), and negative correlations with population size in both the bivariate and multivariate models.

Second, we find 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 health care.

If we could observe utilization outcomes of cross-region migrants who moved to the same GP but different HRRs, we could infer the relative importance of GPs by estimating three-way fixed effects specifications. However, in practice hospital referral regions are large enough that GPs are almost perfectly nested within regions if patients prefer GPs reasonably close to home. As a consequence, we cannot meaningfully estimate a model with both GP and region fixed effects.

To shed light on the relative importance of GPs, we have instead estimated a variation of our baseline model of individual utilization that substitutes the HRR fixed effects with GP fixed effects. This model includes individual and GP fixed effects (omitting the HRR fixed effects), as well as dummies for age (5-year groups) and calendar year, and is identified off of all individuals who change GPs over the sample period; this includes patients who did not move geographically during the sample period, and patients who moved within HRRs, in addition to patients who moved between HRRs. The details of this model are presented in online Appendix E.

By construction, the HRR-average GP fixed effects closely track the estimated HRR fixed effects from our baseline specification (see online Appendix Fig. E1). Online Appendix Figs. E2 and E3 illustrate the distribution of GP fixed effects within and between regions: these figures show substantial variation in the estimated GP fixed effects within HRRs.

If variation in utilization across regions was driven entirely by region-specific factors that were orthogonal to GP practice styles, such as hospital-specific supply factors, the estimated GP fixed effects should reflect only variation in HRR-specific factors, and we would not expect any within-region variation in these coefficients. The fact that we find substantial variation in within-HRR GP fixed effects thus gives some indications that GP practice styles play a role in determining specialist healthcare utilization.<sup>25</sup>

The proportionate impact of place is remarkably similar to the fraction found in [Finkelstein et al. \(2016\)](#). Compar-

<sup>22</sup> Online Appendix Table A3 shows this is indeed the case: movers with only primary education tend to be older than more educated movers.

<sup>23</sup> The reweighting approach can only account for differences between non-movers and movers in the distributions of observable characteristics. Movers may differ from non-movers in terms of unobservable characteristics such as underlying health status. If impacts of place vary with these unobserved characteristics, reweighting on observables will not fully ensure external validity.

<sup>24</sup> Note that the variables "health expenses as share of local budget" and "and primary care visits" relate to utilization in the primary care sector, which is different from the hospital utilization measure used to estimate the hospital region fixed effects. The local health budget only includes services organized by the municipalities, such as primary care and nursing homes. The local budget variables are retrieved from [Fiva and Natvik \(2017\)](#).

<sup>25</sup> Though there are several caveats: the estimated GP fixed effects will be estimated with relatively low precision compared to the HRR fixed effects of the baseline model, GPs patient groups may differ in their demographic characteristics in ways that are associated with differential impacts of place, and unbiased estimation of GP fixed effects requires stronger assumptions on conditionally random mobility. These factors are discussed in more detail in online Appendix E.

ing these figures is complicated by key differences in our underlying data: while Finkelstein et al analyze Medicare patients (age 65+), our paper analyzes effects on adults age 30–70, excluding elderly patients. Given our heterogeneity results which suggest that the impact of place on utilization varies with age, this age difference makes us cautious in reading too much into a direct comparison of the two figures. Nonetheless, it is perhaps surprising that our findings indicate that place factors account for about the same share of overall regional variation in healthcare utilization in Norway and the US.

This in turns suggests that drivers of place-based variation in utilization do not primarily reflect US-specific factors. In their correlational exercise, Finkelstein et al found that the share of doctors favoring aggressive treatment styles, sicker patients, and the share of for-profit hospitals all correlate positively with the estimated place effects. The first two factors – physician practice styles and endogenous supply – may also be present in the Norwegian setting. However, none of the hospitals in our data are for-profit entities, suggesting place variation exists even in the absence of for-profit providers.

On the other hand, the institutional context suggests that differences in patient demand for healthcare may be smaller in our setting. In particular, patients face low copays, implying barriers to access are lower across income groups. In other words, geographic variation in patient demand is unlikely to reflect variation in the ability to afford care, meaning we would expect patient factors to have a smaller impact on utilization if the underlying variation in health status is similar.

Our findings also contrast with existing studies of regional variation in healthcare utilization in western European countries. Using a similar approach, [Salm and Wübker \(2017\)](#) find that regional variation in ambulatory care utilization in Germany is almost entirely driven by variation in patient demand, while [Moura et al. \(2019\)](#) estimate that demand factors account for 70% of variation in healthcare utilization in the Netherlands. While Germany and the Netherlands are for the most part fairly densely populated, rural regions of Norway have comparatively low population density and long travel times to hospitals. The provision of healthcare in these areas may be more utilization intensive, e.g. it may be optimal from a medical perspective to keep recovering patients in the hospital for an additional day when travel times are longer. Our correlational analysis suggests this may be a factor: lower population density is associated with higher estimated place effects.

To summarize, institutional factors suggest, everything else equal, there is reason to expect less regional variation in healthcare utilization in universal healthcare systems, potentially reflecting smaller impacts of both place and patient factors in a centralized system with low cost-sharing. Indeed, Finkelstein et al find that the overall variation in healthcare utilization is substantially higher in the US, with a standard deviation of USD 779 compared to USD 184 in Norway. Nonetheless, the *relative* impact of place and patient factors may still be similar across systems.

## 5. Health outcomes – cause-specific mortality

In the results so far, we have seen that there is substantial variation across regions in health care 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 suggests some places provide inefficiently high or low levels of care. The optimal utilization level is, however, difficult to pinpoint. Utilization might be high due to overtreatment, 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 health care 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. gender 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:

$$d_{gj} = x_{gj}\beta^x + \hat{\gamma}_j\beta^\gamma + \bar{c}_j\beta^c + \varepsilon_{gj}, \quad (6)$$

where  $x_{gj}$  is a vector of gender and age dummies.  $\hat{\gamma}_j$  is the estimated  $\gamma$  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 Eq. (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, 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 Eq. (6) is  $\beta^\gamma$ . 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. In general, causal interpretation in this type of model requires strong assumptions on selection on observables ([Chetty et al., 2014a](#); [Rothstein, 2010](#)). Hospital region effects may be correlated with other place characteristics such as climate, pollution or economic opportunity, that affect mortality independent of health care spending (see also [Finkelstein et al., 2018](#)). Moreover, there could be non-

random sorting on health, e.g. if places with high health care spending conditional on patient demand also have healthier residents.<sup>26</sup> These concerns complicate the interpretation of the overall mortality model in Eq. (6).

Meanwhile, there is considerable variation between the different causes of death in how much specialist health care patients utilize in their last years of life. Total health care 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 health care supply has a negative impact on mortality, one would expect larger associations between estimated hospital region effects and mortality for relatively utilization intensive causes of death. On the other hand, a lack of association may indicate that the relationship between health care supply and mortality is driven by unobserved place heterogeneity. To study this hypothesis, we estimate the regression model in Eq. (6) separately for each cause of death  $m$ , and link the estimated  $\hat{\beta}^m$ 's to the average utilization intensity within each cause.<sup>27</sup> ICD10 codes are used to define grouped causes of death  $m = 1, \dots, M$ .

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.<sup>28</sup> For overall mortality, the hazard function at age  $\tau$  takes the following form:

$$r(\tau) = h(\tau) \exp(x_{it}\beta^x + \hat{\gamma}_j\beta^y + \bar{c}_j\beta^c), \quad (7)$$

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. In other words, we include all individuals of any age in the mortality analyses – though the hospital region fixed effects are estimated in a sample of individuals aged 30–65 only. 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 two-way fixed effects model of utilization rates, which requires only exogeneity with respect to changes in health status. Third, to the extent that health

care 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 patterns.

## 5.2. Mortality results

Table 4 presents selected estimates from linear models of death rates. The first column of Table 4 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  $H\hat{R}R$ . That is, places that have higher health care 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 4. To address issues of multiple hypothesis testing, we have also calculated two sets of adjusted  $p$ -values. First, we have implemented the standard Bonferroni correction, multiplying the Wild bootstrap  $p$ -values by 3 to account for the multiple testing of  $k = 3$  hypothesis. However, since the tests are likely not independent, a Bonferroni correction is likely to be overly conservative. We have also calculated adjusted  $p$ -values following the procedure of Romano and Wolf (2016). Results are robust to this correction, with the caveat that this calculation does not account for the low number of clusters. The model finds a significant and negative estimate of higher  $H\hat{R}R$  on deaths from cancer. A one standard deviation increase in the hospital region effect predicts 22 fewer cancer deaths per 100,000, or a 9.8% 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 4 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.<sup>29</sup>

Estimated duration models (online Appendix Table A6) yield qualitatively similar results: A one standard deviation increase in hospital region effects predicts 15% reduction in all-cause mortality. Competing risks models estimate a corresponding 15.5% 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.

In Eq. (6) the explanatory variables are estimates, and treating them as known will generally yield incorrect standard errors (Murphy and Topel, 1985). We therefore additionally show bootstrapped standard errors for effects of hospital regions.<sup>30</sup> Note that, as with the adjusted  $p$ -values, the bootstrapped standard error does not account for the low number of clusters. This means that although we are not simultaneously correcting for all potential sources of bias (due to estimated explanatory variables, few clus-

<sup>26</sup> Note that our empirical model ensures that variation in health care spending that results directly from variation in patient demand is purged from the estimated hospital region effects.

<sup>27</sup> Alternative measures such as the share admitted for at least one inpatient stay yield similar results.

<sup>28</sup> Duration models with a Gompertz baseline hazard provide qualitatively similar results.

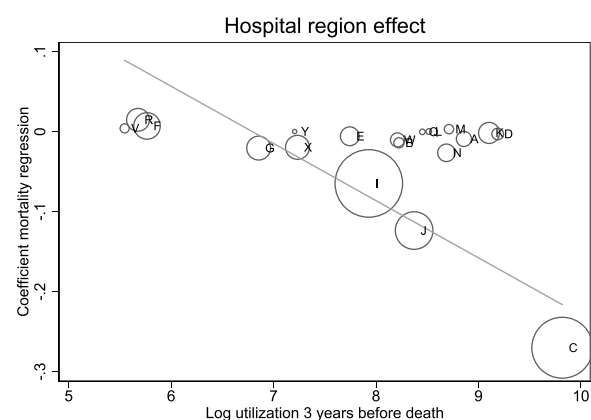
<sup>29</sup> The estimated coefficient is significant at the 5% level, however, this figure does not account for multiple hypothesis testing, meaning significance may be overstated.

<sup>30</sup> To this end, we use the set of bootstrapped HRR fixed effects from Eq. (1). For each of the 500 draws of HRR fixed effect, we estimate the effect of HRR on mortality, as written in Eq. (6). The bootstrapped standard error shown in Table 4 is the standard deviation of these 500 estimates.

**Table 4**  
Mortality rates.

	(1) All causes	(2) Cancer	(3) Heart	(4) External
<i>Model 1: HRR average utilization</i>				
Log utilization	-0.401	-0.282**	-0.0172	-0.00965
<i>p</i> (Wild cluster-bootstrap)	[0.464]	[0.00100]	[0.955]	[0.720]
<i>p</i> (Bonferroni)		0.003	1	1
<i>p</i> (Romano-Wolf)		0.0153	0.913	0.899
<i>Model 2: Estimated place (<math>\hat{H}RR</math>) and patient effects (<math>\bar{c}</math>)</i>				
$\hat{H}RR$	-0.365	-0.216*	-0.0565	-0.00875
Bootstrapped s.e.*	(0.264)	(0.113)	(0.078)	(0.019)
<i>p</i> (Wild cluster-bootstrap)	[0.237]	[0.0220]	[0.610]	[0.559]
<i>p</i> (Bonferroni)		0.0660	1	1
<i>p</i> (Romano-Wolf)		0.00400	0.741	0.741
$\bar{c}$	-0.0929	-0.180	0.0728	-0.000563
<i>p</i> (Wild cluster-bootstrap)	[0.839]	[0.0660]	[0.696]	[0.985]
<i>p</i> (Bonferroni)		0.198	1	1
<i>p</i> (Romano-Wolf)		0.117	0.732	0.981
$\bar{y}$	8.480	2.199	2.700	0.517
<i>N</i>	5920	5920	5920	5920
Sum of weights	29,193,796	29,193,796	29,193,796	29,193,796

Notes: Dependent variable is the death rate per 1000 inhabitants over the 2008–2013 period. Observations weighted by the population in each demographic  $\times$  HRR cell. Regressions include controls for log cell population, gender and 1-year age. *p* values in brackets are based on the wild bootstrap using the empirical *t* distribution, clustered at the HRR level: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .



**Fig. 8.** Hospital region effects and mortality, by cause of death. Figure plots estimated  $\hat{\beta}^y$  and average log health care utilization in the three years before death, by ICD 10 chapter. Observations weighted by number of deaths in each category.

ters, and multiple testing), results are robust to each of the corrections.

To see if the pattern holds more generally we use the first letter of the ICD 10 code to classify all deaths occurring in the sample period. Eq. (6) is then estimated separately for each group. Fig. 8 plots the estimated  $\hat{\beta}^y$  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  $\hat{\beta}^y$  to fall with our measure of treatment intensity: higher hospital region effects should yield larger estimates for conditions that typically involve more medical interventions in the years leading up to death. Fig. 8 finds that this pattern seems to be present in our data: the model finds larger, negative  $\hat{\beta}^y$  for conditions with high utilization rates, while the estimated  $\hat{\beta}^y$  are around zero for conditions that are less utilization intensive.

To summarize, while health care 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 health care utilization around the time of death. The estimates tend to be somewhat imprecise, and should moreover be interpreted with some caution as they are estimated from only 28 hospital regions. Still, these results suggest that high place-specific utilization may translate to better health outcomes, meaning high utilization regions are not necessarily inefficient.

## 6. Conclusion

This paper analyzes regional variation in health care 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 health care broadly defined. Following Finkelstein et al. (2016), we use migration data to decompose regional variation in health care utilization, finding that on average 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. Extended models suggest that this figure masks substantial heterogeneity across educational attainments. Patients with the lowest level of education experience the largest impacts of place, while place effects for college educated patients is negligible.

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 negatively asso-



ciated with overall mortality. The absence of association between health care spending and overall mortality is consistent with the work by Deryugina and Molitor (2018). They find that mortality fell among individuals exposed to Hurricane Katrina due to migration to lower-mortality regions, but that migrants' mortality is unrelated to local Medicare spending. Meanwhile, our data seem to display 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 health care delivery compared to a more decentralized system like in the U.S. Our estimated share of health care 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% and 60% of total variation reflects supply differences. Their paper uses data on elderly patients, while we look at the full population, making direct comparison difficult.<sup>31</sup> 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.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jhealeco.2019.102254>.

## References

- Abowd, J.M., Creecy, R.H., Kramarz, F., 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., Kramarz, F., Margolis, D.N., 1999. High wage workers and high wage firms. *Econometrica* 67 (2), 251–333.
- Andrews, M.J., Gill, L., Schank, T., Upward, R., 2008. High wage workers and low wage firms: negative assortative matching or limited mobility bias? *J. R. Stat. Soc.: Ser. A Stat. Soc.* 171 (3), 673–697.

- Anthony, D.L., Herndon, M.B., Gallagher, P.M., Barnato, A.E., Bynum, J.P., Gottlieb, D.J., Fisher, E.S., Skinner, J.S., 2009. How much do patients' preferences contribute to resource use? *Health Aff.* 28 (3), 864–873.
- Baicker, K., Chandra, A., Skinner, J.S., Wennberg, J.E., 2004. Who you are and where you live: how race and geography affect the treatment of medicare beneficiaries. *Health Aff.* 23, VAR–33.
- Bonhomme, S., Lamadon, T., Manresa, E., 2017. *A Distributional Framework for Matched Employer Employee Data*. Technical Report.
- Card, D., Heining, J., Kline, P., 2013. Workplace heterogeneity and the rise of west German wage inequality. *Quart. J. Econ.*
- Chandra, A., Cutler, D., Song, Z., 2012. Who ordered that? The economics of treatment choices in medical care. *Handb. Health Econ.* 2, 397–432.
- Chandra, A., Staiger, D.O., 2017. *Identifying Sources of Inefficiency in Health Care*. Technical Report. National Bureau of Economic Research.
- Chetty, R., Friedman, J.N., Rockoff, J.E., 2014a. Measuring the impacts of teachers I: evaluating bias in teacher value-added estimates. *Am. Econ. Rev.* 104 (9), 2593–2632.
- Chetty, R., Friedman, J.N., Rockoff, J.E., 2014b. Measuring the impacts of teachers II: teacher value-added and student outcomes in adulthood. *Am. Econ. Rev.* 104 (9), 2633–2679.
- Chetty, R., Hendren, N., 2018a. The impacts of neighborhoods on intergenerational mobility I: childhood exposure effects. *Quart. J. Econ.* 133 (3), 1107–1162.
- Chetty, R., Hendren, N., 2018b. The impacts of neighborhoods on intergenerational mobility II: county-level estimates. *Quart. J. Econ.* 133 (3), 1163–1228.
- Chetty, R., Hendren, N., Katz, L.F., 2016. The effects of exposure to better neighborhoods on children: new evidence from the moving to opportunity experiment. *Am. Econ. Rev.* 106 (4), 855–902.
- Chyn, E., 2018. Moved to opportunity: the long-run effects of public housing demolition on children. *Am. Econ. Rev.* 108 (10), 3028–3056.
- Combes, P.-P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: sorting matters! *J. Urban Econ.* 63 (2), 723–742.
- Cutler, D., Skinner, J., Stern, A.D., Wennberg, D., 2018. Physician beliefs and patient preferences: a new look at regional variation in health care spending. *Am. Econ. J.: Econ. Policy*.
- Cutler, D.M., Lleras-Muney, A., Vogl, T., 2008. *Socioeconomic Status and Health: Dimensions and Mechanisms*. Technical Report. National Bureau of Economic Research.
- Deryugina, T., Molitor, D., 2018. *Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina*. Technical Report. National Bureau of Economic Research.
- Dhaene, G., Jochmans, K., 2015. Split-panel jackknife estimation of fixed-effect models. *Rev. Econ. Stud.* 82 (3), 991–1030.
- Doyle, J., Graves, J., Gruber, J., Kleiner, S., 2015. Measuring returns to hospital care: evidence from ambulance referral patterns. *J. Polit. Econ.* 123 (1), 170.
- Doyle Jr., J.J., 2011. Returns to local-area healthcare spending: evidence from health shocks to patients far from home. *Am. Econ. J. Appl. Econ.* 3 (3), 221.
- d'Uva, T.B., Jones, A.M., 2009. Health care utilisation in Europe: new evidence from the ECHP. *J. Health Econ.* 28 (2), 265–279.
- Fine, J.P., Gray, R.J., 1999. A proportional hazards model for the subdistribution of a competing risk. *J. Am. Stat. Assoc.* 94 (446), 496–509.
- Finkelstein, A., Gentzkow, M., Williams, H., 2016. Sources of geographic variation in health care: evidence from patient migration. *Quart. J. Econ.*
- Finkelstein, A., Gentzkow, M., Williams, H., 2018. *Place-Based Drivers of Mortality: Evidence From Migration*. Working Paper.
- Fisher, E.S., Bynum, J.P., Skinner, J.S., 2009. Slowing the growth of health care costs – lessons from regional variation. *N. Engl. J. Med.* 360 (9), 849–852.
- Fisher, E.S., Wennberg, D.E., Stukel, T.A., Gottlieb, D.J., Lucas, F.L., Pinder, E.L., 2003a. The implications of regional variations in medicare spending. Part 1: the content, quality, and accessibility of care. *Ann. Intern. Med.* 138 (4), 273–287.
- Fisher, E.S., Wennberg, D.E., Stukel, T.A., Gottlieb, D.J., Lucas, F.L., Pinder, E.L., 2003b. The implications of regional variations in medicare spending. Part 2: health outcomes and satisfaction with care. *Ann. Intern. Med.* 138 (4), 288–298.
- Fiva, J.H.A.H., Natvik, G.J., 2017. *Local Government Dataset*.
- Fiva, J.H., Hægeland, T., Rønning, M., Syse, A., 2014. Access to treatment and educational inequalities in cancer survival. *J. Health Econ.* 36, 98–111.
- Gibbons, S., Overman, H.G., Pelkonen, P., 2014. Area disparities in Britain: understanding the contribution of people vs. place through variance decompositions. *Oxf. Bull. Econ. Stat.* 76 (5), 745–763.

<sup>31</sup> 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. See Section 4.4.



- 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. *Rev. Econ. Stat.* 95 (4), 1096–1116.
- Joynt, K.E., Jha, A.K., 2012. The relationship between cost and quality: no free lunch. *JAMA* 307 (10), 1082–1083.
- Lamadon, T., Mogstad, M., Setzler, B., 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. *J. Labor Econ.* 33 (3), 751–788.
- Molitor, D., 2018. The evolution of physician practice styles: evidence from cardiologist migration. *Am. Econ. J.: Econ. Policy*.
- Moura, A., Salm, M., Douven, R., Remmerswaal, M., 2019. Causes of regional variation in Dutch healthcare expenditures: evidence from movers. *Health Econ.* 28 (9), 1088–1098.
- Murphy, K.M., Topel, R.H., 1985. Estimation and inference in two-step econometric models. *J. Business Econ. Stat.* 379, 370.
- Romano, J.P., Wolf, M., 2016. Efficient computation of adjusted  $p$ -values for resampling-based stepdown multiple testing. *Stat. Prob. Lett.* 113, 38–40.
- Rothstein, J., 2010. Teacher quality in educational production: tracking, decay, and student achievement. *Quart. J. Econ.* 125 (1), 175–214.
- Salm, M., Wübker, A., 2017. Causes of Regional Variation in Healthcare Utilization in Germany, Number 675. Ruhr Economic Papers.
- Skinner, J., 2011. Causes and consequences of regional variations. *Handb. Health Econ.* 2.
- Song, Y., Skinner, J., Bynum, J., Sutherland, J., Wennberg, J.E., Fisher, E.S., 2010. Regional variations in diagnostic practices. *N. Engl. J. Med.* 363 (1), 45–53.
- Van Doorslaer, E., Wagstaff, A., Van der Burg, H., Christiansen, T., De Graeve, D., Duchesne, I., Gerdtham, U.-G., Gerfin, M., Geurts, J., Gross, L., et al., 2000. Equity in the delivery of health care in Europe and the US. *J. Health Econ.* 19 (5), 553–583.
- Vikum, E., Krokstad, S., Westin, S., 2012. Socioeconomic inequalities in health care utilisation in Norway: the population-based hunt3 survey. *Int. J. Equity Health* 11 (1), 48.