

Haering, Alexander; Kaeding, Matthias; Werbeck, Anna

**Working Paper**

## Equal access to primary care: A reference for spatial allocation

Ruhr Economic Papers, No. 1089

**Provided in Cooperation with:**

RWI – Leibniz-Institut für Wirtschaftsforschung, Essen

*Suggested Citation:* Haering, Alexander; Kaeding, Matthias; Werbeck, Anna (2024) : Equal access to primary care: A reference for spatial allocation, Ruhr Economic Papers, No. 1089, ISBN 978-3-96973-266-3, RWI - Leibniz-Institut für Wirtschaftsforschung, Essen, <https://doi.org/10.4419/96973266>

This Version is available at:

<https://hdl.handle.net/10419/300565>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



# RUHR

ECONOMIC PAPERS

Alexander Haering  
Matthias Kaeding  
Anna Werbeck

## **Equal Access to Primary Care – A Reference for Spatial Allocation**

# Imprint

---

## Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung  
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics  
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences  
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics  
Universitätsstr. 12, 45117 Essen, Germany

## Editors

Prof. Dr. Thomas K. Bauer

RUB, Department of Economics, Empirical Economics  
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger

Technische Universität Dortmund, Department of Economic and Social Sciences  
Economics – Microeconomics  
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@tu-dortmund.de

Prof. Dr. Volker Clausen

University of Duisburg-Essen, Department of Economics  
International Economics  
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Ronald Bachmann, Prof. Dr. Almut Balleer, Prof. Dr. Manuel Frondel,  
Prof. Dr. Ansgar Wübker

RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

## Editorial Office

Sabine Weiler

RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

## Ruhr Economic Papers #1089

Responsible Editor: Ansgar Wübker

All rights reserved. Essen, Germany, 2024

ISSN 1864-4872 (online) – ISBN 978-3-96973-266-3

The working papers published in the series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

---

## Ruhr Economic Papers #1089

Alexander Haering, Matthias Kaeding, and Anna Werbeck

# Equal Access to Primary Care – A Reference for Spatial Allocation

## **Bibliografische Informationen der Deutschen Nationalbibliothek**

---

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;  
detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

<https://dx.doi.org/10.4419/96973266>

ISSN 1864-4872 (online)

ISBN 978-3-96973-266-3

---

Alexander Haering, Matthias Kaeding, and Anna Werbeck\*

# Equal Access to Primary Care – A Reference for Spatial Allocation

## Abstract

*We study the regional distribution of primary care physicians in Germany to learn about the extent and possible reasons of geographic maldistribution. For this aim, we apply a greedy capacitated algorithm on very fine spatial data. We compare this reference allocation of primary care physicians to the status quo. Our results suggest that Germany has a shortage of primary care physicians of 6% which is particularly evident in rural areas. Some municipality characteristics like purchasing power and number of schools can explain parts of the difference in the number of physicians between cities and rural areas. Large parts, however, remain unexplained.*

*JEL-Codes: I14, C61, R10*

*Keywords: Health and inequality; optimization techniques; regional economics*

*July 2024*

---

\* Alexander Haering, RWI and Hochschule Fresenius; Matthias Kaeding, RWI; Anna Werbeck, RWI and RGS Econ. – We thank Tim Brand, Stephan Hetzenecker, Lukas Hörnig, Maryna Ivets, Ingo Kolodziej, Philipp Markus, Simon Reif, Sandra Schaffner, Hendrik Schmitz, Jan Wickerath, Christiane Wuckel, Ansgar Wübker and Nicolas R. Ziebarth. Special thanks go to Astrid Wering. We also thank participants of the dggö health econometrics workshop, Lola HESG 2021 conference and the Ihea conference 2021. The authors have declared that no competing interests exist. – All correspondence to: Anna Werbeck, RWI, Hohenzollernstraße 1-3, 45128 Essen, Germany, e-mail: [Anna.Werbeck@rwi-essen.de](mailto:Anna.Werbeck@rwi-essen.de)

# 1 Introduction

Equal access to primary care is fundamental for a reliable health care system. Having access to a primary care physician (PCP) can have a huge impact on an individual's amount of received care, improve their health care experience and overall health and reduce follow-up visits (Levine et al., 2019; Riverin et al., 2018 and Macinko et al., 2007). In particular, a short distance to PCPs seems to make a difference to one's health (Kelly et al., 2016; Saijo et al., 2018). As such, it is beneficial for the individual. Good access to PCPs is also economically important. It can reduce overall health care costs and increase people's productivity and economic output (Clarke et al., 2021). Thus, in 2018, the member states of the World Health Organization (WHO) renewed their commitment to strengthen sustainable primary health care to ensure universal primary health care coverage (WHO, 2018). Yet, inequity in access remains a major problem in health care systems worldwide. Across industrialized countries, this inequity is often fueled by an uneven geographical distribution of PCPs (OECD, 2020).

The reason why geographical maldistribution drives inequities in health care access lies especially in the consequences of the demographic change. These have exacerbated the situation and are particularly evident in rural areas (Augère-Granier & McEldowney, 2020). The supply of primary care is decreasing as more and more PCPs reach their retirement age and there is no one to take over their practice. However, the demand for primary care is increasing in response to the growing number of older people with the highest needs for care (European Commission, 2020). In consequence, this high-need, least-mobile patients have to travel ever further distances to reach a PCP.

We address this problem of geographic maldistribution of PCPs, by answering two main research questions: "Where should PCPs be allocated to enhance equity in access to primary care?" and "Are there common characteristics of municipalities that can explain regional differences in PCP allocation?"

We employ a greedy capacitated algorithm on very fine spatial data to calculate a reference allocation and analyze the distribution of primary care physicians across different regions in Germany. We combine different data

sources on supply and demand for primary care. The resulting dataset contains information on the number and allocation of potential patients and primary care physicians, the demand for primary care visits, travel times between potential patients and primary care physicians and regional characteristics. Using these data we calculate the reference allocation and its difference from the current allocation. This difference serves as an indicator of maldistribution. Subsequently, we break down the results into endowment and coefficient effects by conducting a decomposition analysis of this indicator against municipality characteristics. As such, we identify potential factors influencing the misallocation. Through this approach, we investigate the nationwide allocation of PCPs to better understand the extent and common municipality characteristics of geographic maldistribution.

We focus on Germany because it is experiencing a comparably strong ageing of its society, which is especially pronounced in rural areas. The share of people aged 65 years and over across the country was one of the six highest in OECD countries in 2019 (OECD, 2021), with a share of 24% for rural areas, but only 19.6% in large cities (Destatis, 2020). This development is also evident for PCPs. Many PCPs are reaching retirement age and it is difficult to find replacements, especially in rural areas (Baum et al., 2022; KBV, 2022). This leaves the rural population with long distances to their PCPs.

Our results show a 6% shortage of PCPs in Germany in 2019. We also observe substantial geographic maldistribution of primary care physicians, which is particularly evident in rural areas, where there is a shortage of health care professionals. Our decomposition results show that municipalities with more educational institutions, more hospital beds and a higher trade tax multiplier tend to have more oversupply. Those with more purchasing power per capita and more particulate matter tend to have less oversupply. This further widens the gap in access between affluent and disadvantaged regions. However, most of this difference remains unexplained by the observed municipality characteristics.

There are many studies in health economics that investigate the importance of access to health care and explore existing inequity in access, its sources and its implications. In general, the literature shows that access to a PCP seems to increase quality and experience of received outpatient care (Levine



et al., 2019) and a timely visit to a PCP after hospitalization seems to reduce follow-up visits (Riverin et al., 2018). In addition, Kelly et al. (2016) and Saijo et al. (2018) show for UK and Japan that a short distance to physicians and especially to PCPs seems to make a difference to ones health. Finkelstein et al. (2016) find for the US that at least half of the regional differences in utilization are due to local supply factors. Salm and Wübker (2020) find that the reasons for regional differences in outpatient care in Germany vary by region, with demographics and other patient characteristics explaining most of it. This underlines the importance of an equal and timely access to care and approaches to reach it.

Existing approaches include physician planning mechanisms or monetary and non-monetary policy incentives. Ono et al. (2014) and Ozegowski and Sundmacher (2012) argue for different OECD countries and Germany respectively that the current planning mechanisms have shortcomings, such as relying on suboptimal past locations, and therefore induce rather than reduce inequities in access to care. We contribute to this debate by calculating a reference allocation of PCPs using techniques from operations research. Studies in operations research utilize algorithmic location-allocation planning (Daskin & Dean, 2005) which are applied to several types of health care facilities (see Rais & Viana, 2011, for a well-structured overview). We adjust it to the German setting and calculate a PCP allocation that ensures needs-based equity in access to care while abstaining from actual locations.

While monetary and non-monetary incentives seem to matter for location decisions (e.g. Günther et al., 2010, Baum et al., 2022), they are often only aimed at reducing disparities between rural and urban areas in general. Several studies (e.g. Newhouse, 1990) show that larger municipalities are indeed more likely to attract physicians than smaller ones. However, there is a growing literature that looks at more nuanced characteristics that correlate with the number of physicians. For Germany, Jürges (2007) and Sundmacher and Ozegowski (2016) find a positive influence of the population share of elderly on the number of physicians, Kuhn and Ochsen (2019) also find this result in urban areas but find a reversed correlation in rural areas. Bauer et al. (2016) use a principal component analysis and find that within metropolitan areas, central location, small householdsize and larger population density attracts more PCPs.

Closest to our approach, Vogt (2016) decomposes the variation in the density of primary care and specialist physicians in 2012 into the contribution of different municipality characteristics. She measures inequity using Gini coefficients based on physician density and finds that some factors such as a higher proportion of commuters and share of people with private health insurance can explain part of the variation. We contribute to these findings by suggesting an indicator of misallocation based on our reference allocation to measure inequity. Our reference allocation is needs-based, independent of historical location parameters and based on very fine 1km<sup>2</sup> grid cells. As such, we are not limited by current locations or variations within districts or municipalities. Based on this indicator of misallocation, our decomposition results suggest that other factors such as educational facilities and purchasing power contribute to explaining part of the variation.

With our reference allocation, we suggest an allocation approach that is independent of suboptimal historical practice locations. We identify regions where health care resources are disproportionately scarce and where interventions may be especially helpful in improving access to health care services and further explore potential municipality characteristics that correlate with over- or undersupply of PCPs.

## 2 Institutional Setting

In many countries, medical care is organized by the state and physicians are financed by taxes (e.g. Great Britain or Canada). This is not the case in Germany, where the health care system is self-organized and financed by the contributions of statutory and private health insured.

In the case of outpatient care, the German government has delegated part of the management of the health care system to insurance funds. The insurance funds then commission professional associations to provide the care. But the state reserves the right to withdraw this mandate at any moment<sup>1</sup>.

For physicians and psychotherapists who treat patients with statutory health insurance, the professional association responsible is the National Associa-

---

<sup>1</sup>The state remains ultimately responsible and can intervene in and override decisions made by the parties involved.

tion of Statutory Health Insurance Physicians (NASHIP), along with its 17 regional subsidiary associations (RASHIPs). Each RASHIP is responsible for one federal state, with the exception of North-Rhine Westphalia, which is split into two parts.

To regulate the reimbursements, a nationwide catalog of services is defined by the InBA (*Institut des Bewertungsausschusses*), while the base rate is set at the federal level. However, RASHIPs can add additional services and set their own base rate.

The mechanism implemented to plan access to outpatient care is called *Bedarfsplanung* (requirement planning) and is set at the federal level in Germany. This requirement planning is implemented by RASHIPs and health insurance funds. It builds upon historical health care supply and demand, adjusting the numbers for current number, age groups, gender, and a dummy for higher morbidity of residents. Using this information, they identify regions with a surplus or deficit of PCPs.

Regions with a health care supply rate of 110 % or above are classified as oversupplied. This implies that in those regions no further PCP gets a license to provide care for statutorily health insured. Yet, expiring licenses can still be passed on to successors. In highly oversupplied regions with a rate of 140 %, expiring licenses should not be passed on to potential successors. Regions with a health care supply rate of 75 % or less are classified as undersupplied. The RASHIPs are then requested to find a solution for this situation.

Such high levels of over- and undersupply are inefficient. Both economically and in terms of planning for efficient allocation, a range of 75 % to 140 % is too wide. It would be best to exactly meet the demand for PCPs. Therefore, we suggest a technical approach<sup>2</sup> that computes a nationwide reference allocation of PCPs on a small-scale that ensures equity and efficiency in access.

---

<sup>2</sup>The study by Klose and Rehbein (2017) is an example for the importance of a technical approach. The authors found a surplus of around 3,900 PCPs, but only compare and aggregate the planning regions, without considering variation within them.

### 3 Data

As supply and demand generally converge locally and are independent of regional borders, we need small-scale data to perform our analysis. By using grid-level data, we have a uniform classification of Germany in  $1km^2$  cells, which is not influenced by regional borders. On this grid-level, we combine different data sources with information on PCP supply and the demand for them. Our generated dataset gives us information on the number and allocation of potential patients and PCPs, the demand for PCP visits as well as driving times between potential patients and PCPs. With these data we calculate the reference allocation of PCPs as well as its difference to the status-quo.

On municipality level, we then add information on regional characteristics. We use population density and other regional variables to analyze their correlation with the calculated number of excess PCPs and decompose the endowment effect of these variables from the coefficient effect.

#### 3.1 Driving Times

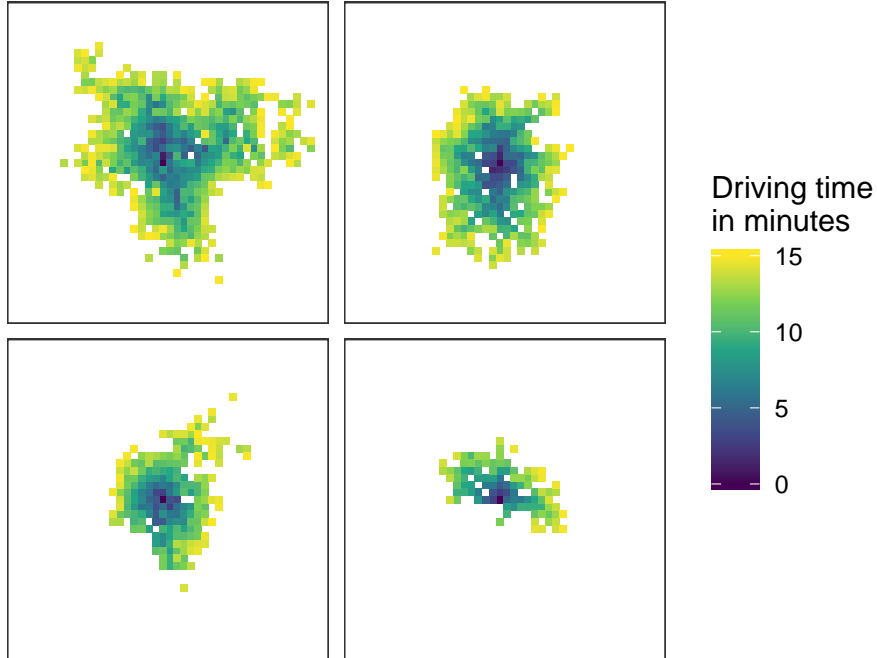
Our computed driving times are based on grid cells from the RWI-GEO-GRID (Breidenbach & Eilers, 2018). The RWI-GEO-GRID database contains socioeconomic data for Germany on 218,875 populated  $1km^2$  grid cells.<sup>3</sup> We determine driving time by car from each populated grid cell to all populated grid cells within  $40km$  radius using the Open Source Routing Machine program (Luxen & Vetter, 2011), based on OpenStreetMap (Haklay & Weber, 2008) road data, which contains nearly the complete German road network (Barrington-Leigh & Millard-Ball, 2017). A radius of  $40km$  ensures that we include all grid cells that can be reached within 15 minutes.

This procedure leaves us with around 600,000,000 driving times. Next, we discard driving times exceeding 15 minutes. We do so, to avoid unnecessary calculations and because BBSR (2005) demands that every person reaches one PCP within a maximum of 15 minutes. This results in around 170,000,000 driving times.

---

<sup>3</sup>The grid cells are aligned to the EU-wide INSPIRE Directive (Bartha & Kocsis, 2011).

Figure 1: Grids From Where the Center Grid Can Be Reached Within 15 Minutes



Note: The dark blue grid cell is at the center, driving times computed by using RWI-GEO-GRID (Breidenbach & Eilers, 2018) and Open Source Routing Machine program (Luxen & Vetter, 2011).

Figure 1 shows four exemplary sets from our data. In all four figures, the dark blue grid cell at the center depicts  $g_i$ , the potential location of a general practitioner. The differently colored cells in the surrounding show all grid cells for which  $g_i$  is reachable within 15 minutes. The cell colors represent the driving time to the center cell. We see that the shapes of the sets vary. This is driven by street infrastructure, classified by 15 speed types and geographical conditions.

### 3.2 Current PCP Allocation: the Status Quo

The RASHIPs contain information on all contract physicians in their respective region, also RASHIP and NASHIP provide aggregated numbers of

Table 1: Number of PCPs by Source

Data source	Number of PCPs
Only NASHIP	33,031
Only RASHIP	15,140
Both	6,835
Total	55,006

Note: State of data as of 2019, Number of PCPs shows the general practitioner entries in the respective search-engine. Only NASHIP/RASHIP lists PCPs entered in either of the two search-engines. Both lists the PCPs that have an entry in both search-engines, without duplicates.

PCPs on federal state levels. However, for our study we need the exact practice location. Therefore, we utilize data from the official on-line search engines of both organizations.<sup>4</sup> The search engines are accessible to the German population to search for doctors in all fields. Yet, physicians have to agree to be listed in the NASHIP and/or RASHIP online database. This potential self-selection into the search-engines may lead to two problems: Either, when focusing only on one database, we are left with an undercount as physicians that do not agree to be listed in the chosen search-engine are not part of our set; or, when using the two databases, we are left with an overcount when a physician is listed in both. To avoid these problems, we used both data sources, identified double entries, i.e., the intersection of both data sets, and dropped the duplicates. To verify our approach, we aggregated our PCP-count on federal state level and compared it to the official number provided by NASHIP (2018). Table 1 summarizes the number of PCPs obtained from the two sources. In total, there are 55,006 PCPs in the data set, representing the status-quo allocation.

### 3.3 PCP Visits

To calculate the adequate distribution of primary care physicians, we need to account for the expected number of cases (the demand for PCP visits) in each grid cell. There are two potential sources for information on the number of PCP visits. These are claims data and survey data. Although

<sup>4</sup>See appendix A for all 18 webpages.

Table 2: Summary Statistics Regression Tree

	Mean	SD
# of PCP visits p.a.	9.09	14.62
Age	48.66	17.62
Female	0.54	0.50

Source: SOEP wave 2016 (Goebel et al., 2019), Note: Summary Statics for a total of 24,418 observations, SD is the standard deviation.

claims data is not self reported and has a second advantage of additional patient characteristics, there is one major drawback for our setup: Follow up visits are not necessarily monitored as a new meeting, thus underestimating the demand for PCP visits. Therefore, we use data from SOEP (Goebel et al., 2019), which is the largest representative longitudinal panel of private households in Germany.

We utilize wave 2016. Besides age and gender<sup>5</sup>, it contains self-reported number of annual visits to the PCP<sup>6</sup>. Our data set contains 24,418 observations. Table 2 shows the mean and standard deviation for our variables of interest.

### 3.4 Characteristics of Municipalities

We use information on population density to assort all German municipalities<sup>7</sup> to cities, towns / suburbs and rural areas following the classification

<sup>5</sup>While additional variables might improve prediction, we can only use explanatory variables which are available in both the SOEP 2016 data and the RWI-GEO-GRID. However, we assume that age and gender are the most relevant variables and capture health demand fairly well.

<sup>6</sup>The precise question is: "Have you gone to a doctor in the last three months? If so, please state how many times.". The response is then aggregated at the annual level. Note that the question refers to any doctor visit (PCP and/or specialist). However, we follow the approach by Greiner et al. (2018) and use this number to approximate the quantity of PCP visits.

<sup>7</sup>Because the area of municipalities may vary in Germany over time, e.g. due to mergers, and our data sets are not from the exact date in 2019, there is some mismatch and we do not have information on all 11,125 municipalities.

Table 3: List of Municipality Characteristics

Characteristics / Indicators	Dimension
Number of Primary Care Physicians	Quantity
Number of schools per capita	Quantity
Number of universities per capita	Quantity
Number of hospital beds per capita	Quantity
Trade tax multiplier	Percent
Purchasing power per capita	100,000 Euro
Net commuters	1,000 Persons
Net migration	1,000 Persons
Particulate matter	$\mu$ g / cubic meter
Average noise	db
Average number of hot days	Quantity

Source: Federal Environment Agency (FEA), Federal statistics office (FSO) and German Weather Service (GWS)

of Dijkstra and Poelman (2014)<sup>8</sup> to examine whether a shortage of PCPs is driven by rural regions. Apart from population density, we suspect other regional factors to impact the number of excess PCPs in a municipality. Therefore, we include educational, economic and environmental variables as well as hospital beds in our analysis. We expect positive correlations with our educational and most of our economic variables as well as with hospital beds and negative correlations with our environmental variables. Table 3 lists all included characteristics and indicators.

## 4 Methodology

Our unique combination of data sets helps us to determine the amount and regional allocation of PCPs in Germany under the following constraints: All residents can reach at least one PCP within 15 minutes driving time and PCPs serve within their limited capacity of 13,000 annual cases. That is, we face a capacitated set cover problem. To determine the potential demand, we utilize the findings from our regression tree.

<sup>8</sup>They define a region as a city if at least 50% of the population is living in high-density grid cells, as towns and suburbs if less than 50% of the population is living in rural grid cells and less than 50% is living in a high-density grid cell, and as a rural area if more than 50% of the population is living in rural grid cells.



We solve this capacitated set cover problem by using a greedy approximation. This leaves us with a reference allocation of PCPs which we then compare to the status quo to get information on the extent of PCP shortage and maldistribution. In the following regression analyses, we use the municipality characteristics to investigate potential driving factors.

#### 4.1 Regression Tree

To calculate the expected number of cases (the potential demand) in each grid cell, we use a regression tree. We assume patient subgroups with an approximately constant number of PCP visits. Our regression tree approach is suitable for identifying these subgroups with age and gender as explanatory variables to forecast demand.

To construct our regression tree, we require a complexity parameter ( $cp$ ) that determines the number of branches and controls the tree size. The  $cp$  prunes off splits that are not worthwhile to keep in the tree. In general, any split that does not decrease the overall lack of fit by a factor of  $cp$  is not attempted. The complexity parameter  $cp$  is advisory. A value of  $cp = 1$  always results in a tree with no splits. For regression models the scaled  $cp$  has a clear interpretation: any split that does not increase the overall  $R^2$  of the model by at least  $cp$  is deemed not worth pursuing. In our case, we use the  $cp$  that minimizes the root mean squared error.

The  $cp$  is obtained through a cross-validation resampling method with 10 folds. A range of 0 to 0.01 is tested in steps of 0.00001 by a grid search. The optimal model is determined based on the lowest value of the error term. The process begins by fitting a fully-grown tree to the entire data set  $D$ , with  $T$  terminal nodes. Then, the tree is pruned as much as possible, based on the lowest misclassification loss. The following steps are performed:

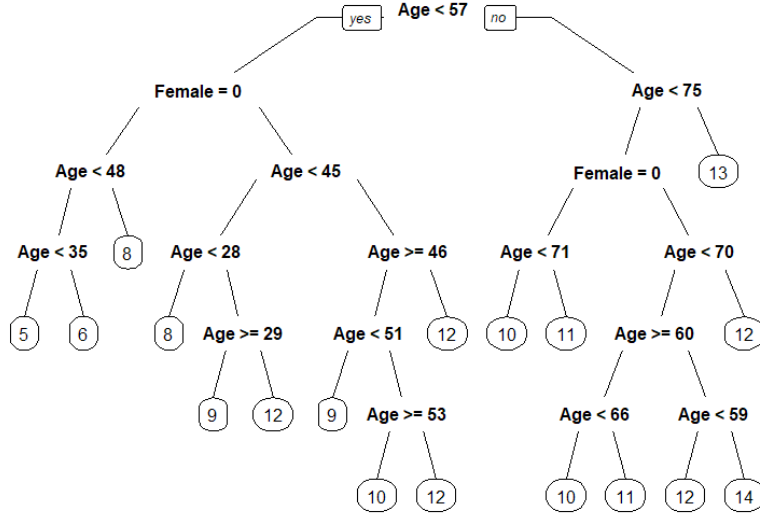
1. We split our data into  $n = 10$  randomly selected folds:  $F1$  to  $F10$ .
2. We use a 10-fold cross-validation and fit each sub-tree  $T_1 \dots T_m$  on each training fold  $D_s$ .
3. We compare the class predicted for the validation fold against the actual class to calculate the corresponding miss-classification loss (risk

value)  $R_m$  for each subtree. Then, we sum up the risk values for all subtrees and folds.

4. We select the complexity parameter  $cp$  with the lowest total risk value.

Figure 2 shows our final tree. The interpretation is straightforward, keeping in mind that the right branch means "no" and the left "yes". For example, a female subject ( $Female = 0$ , right branch:  $Female = 0$  is "not true")  $Age < 28$  is expected to visit the PCP's office 8 times a year.<sup>9</sup> We use our regression tree to predict the number of PCP visits for each 1km<sup>2</sup> grid cell using the demographic data provided by the RWI-GEO-GRID database. As an example, the city Essen has a population of some 0.6 million persons with an approximate demand of 5 million cases per year.

Figure 2: Regression Tree Number of Visits to the Doctor per Year



Note: Calculation based on SOEP wave 2016, endogenous variable is self-reported number of visits to the doctor per year.

<sup>9</sup>Interestingly: at  $Age > 75$  the number of expected visits to the PCP is independent of the gender with 13 visits a year.

## 4.2 Set Cover Problem

The set cover problem is a central problem in combinatorics with wide ranging applications, see, as examples, Vazirani (2001) and Korte and Vygen (2018). Initially, we ignore the limited capacity. We set up a general uncapacitated set cover problem, where one determines the smallest sub-collection from a collection of sets  $\mathcal{S} = \{S_1, S_2, \dots, S_l\}$  that covers a universe of elements  $\mathcal{U}$ , i.e. every element in  $\mathcal{U}$  appears in at least one of the subsets of  $\mathcal{S}$ . Here, the universe is  $\mathcal{U} = \{g_1, \dots, g_n\}$ , where  $g_i$  is the  $i$ th  $1\text{km}^2$  grid. We define  $S_i$  as

$$S_i := \{g_j \mid d(g_j, g_i) \leq 15\text{min}\},$$

where  $d(g_i, g_j) \geq 0$  is the asymmetric driving time from  $g_i$  to  $g_j$ . This implies that  $S_i$  consists of all grid cells from where grid  $g_i$  is reachable within 15 minutes as  $d(g_i, g_i) = 0$ ,  $S_i$  always includes  $g_i$ .

In the capacitated case, we restrict the demand to be covered by a single set to be below or equal to 13,000 patient-physician contacts, i.e., cases, per year, so that each set  $S_i$  consists of the grids containing the cases serviced by the  $i$ th PCP. This might require placing multiple PCPs in one grid to satisfy the demand.<sup>10</sup>

## 4.3 Greedy Approximation

Because the set of possible solutions grows exponentially with the number of sets in  $\mathcal{S}$ , the problem becomes too complex to be solved exactly. Therefore, we use the capacitated variant of the greedy approximation algorithm (see Chvatal, 1979 and Lovász, 1975). Feige (1998) shows that the basic greedy approximation gives reasonably good approximations to the true solution while scaling well to large data sets. Vazirani (2001, chap. 29) shows why the approximation is very hard to improve.

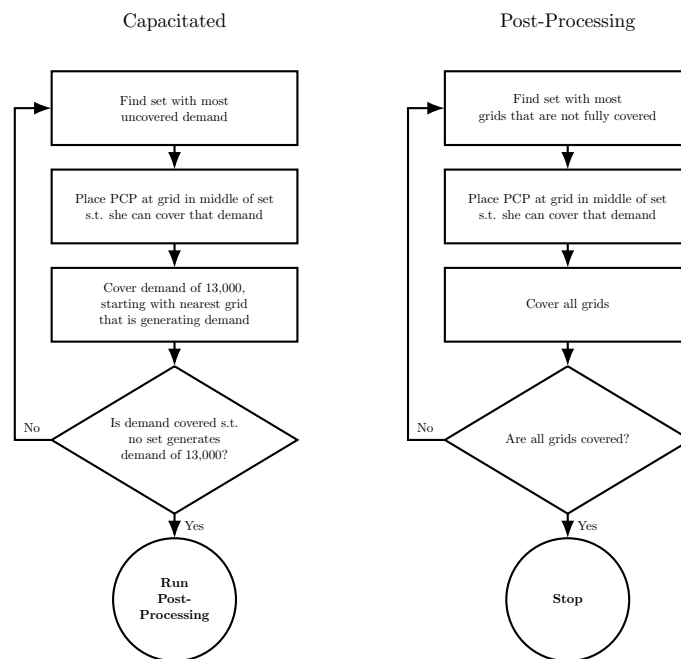
The basic idea of our greedy approximation is to iteratively distribute PCPs across the grid. Given that in Germany most PCPs serve in private practices with not more than a few primary care physicians working together, we introduce a penalty to avoid placing more than 10 PCPs per grid. We

---

<sup>10</sup>Because we operate within a resolution of  $1\text{km}^2$ , we do not differentiate between PCPs working in a joint practice and those who do not.

start from a green field without any PCP. In every iteration, we choose the set with the most uncovered demand, until every case is covered. Here, covering demand means associating cases with a PCP. Parka and Honga (2009) add a pre-processing step to the greedy approximation for the capacitated case, where one iteratively covers demand, here 13,000 cases, until no sets exceeding the capacity remain. Figure 3 gives a short illustration of the two algorithms.

Figure 3: Set Cover Algorithm Uncapacitated



Source: Own illustration

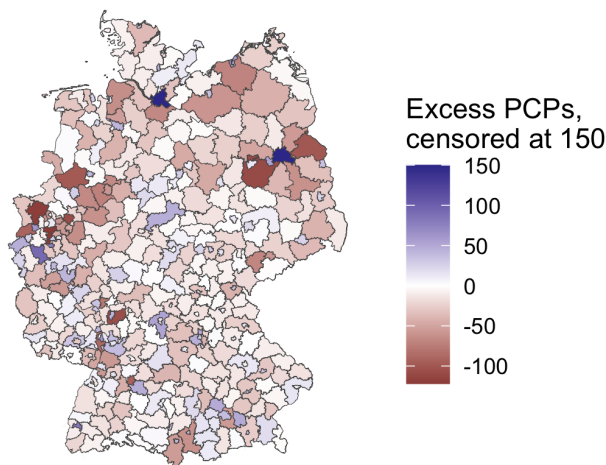
## 4.4 Regression Analyses

To examine whether an excess of PCPs is correlated with municipality characteristics, we run additional OLS regressions. We aggregate our grid data on municipality level, assigning each grid to the municipality with which it has the largest intersecting area. First, we define our dependent variable as

$$d_i := N_i^{observed} - N_i^{greedy}, \quad (1)$$

i.e., the difference between  $N_i^{observed}$ , the observed number of PCPs in the  $i$ th municipality, and  $N_i^{greedy}$ , the number of PCPs needed to achieve full support in the  $i$ th municipality. Therefore, we can interpret  $d_i$  as residual, measuring excess number of PCPs in a municipality. On these terms, an optimal allocation of PCPs would result in  $d_i$  equal to zero for all  $i$  and our coefficients should not pick up any correlation with the excess number of PCPs. Figure 4 shows a map of the dependent variable.

Figure 4: Excess PCPs in Germany by Municipality



Note: Excess PCPs range from -100 to +150. -100 corresponds to a shortage of PCPs of 100, +150 to an excess of PCPs of 150 in the specific municipality.

We use  $d_i$  from (1) and specify our regression as:

$$\begin{aligned}
 d_i = & \alpha_0 + \alpha_1 \text{commuter}_i + \alpha_2 \text{migration}_i + \alpha_3 \text{hospbed}_i \\
 & + \beta_1^\top \text{education}_i + \beta_2^\top \text{econpower}_i + \beta_3^\top \text{environment}_i \\
 & + \gamma^\top \text{state}_i + \epsilon_i.
 \end{aligned} \tag{2}$$

In equation (2),  $\alpha_0, \dots, \alpha_4$  are regression coefficients,  $\beta_1, \dots, \gamma$  are vectors of regression coefficients and  $\epsilon_i$  represents the error term. The variable  $\text{commuter}_i$  represents net commuter as the difference of in- and outflow of commuters in municipality  $i$ ,  $\text{migration}_i$  the equivalent net migration.  $\text{education}_i$  incorporates the variables schools and universities per capita,  $\text{econpower}_i$  the variables purchasing power per capita and trade tax multiplier.  $\text{environment}_i$  represents the variables number of hot days, particulate matter and average noise.  $\text{hospbed}_i$  is also standardized per capita.

We additionally add dummies for the federal states.<sup>11</sup> The fixed effects on federal state level account for heterogeneity in the institutional framework of Germany, for instance varying incentives for PCPs to practice in smaller towns.

After running standard OLS regressions, we use the Oaxaca-Blinder decomposition (Oaxaca, 1973 and Blinder, 1973) that compares the regressions that we ran for rural regions and for cities, where rural regions build our baseline.

$$\begin{aligned}
 \Delta d_i = & [E(X_r) - E(X_u)]' \gamma_r + E(X_r)' (\gamma_r - \gamma_u) \\
 & - [E(X_r) - E(X_u)]' (\gamma_r - \gamma_u)
 \end{aligned} \tag{3}$$

In equation (3) we follow the approach by Jann (2008). Here,  $\Delta d_i$  denotes the difference in excess PCPs between rural and urban areas. The  $X$ s include the municipality characteristics we use as our explanatory variables in equation (2). Subscript  $r$  represents rural regions, subscript  $u$  represents urban regions (cities). The first part of equation (3) is the endowment effect that explains the difference in excess PCPs due to a different level of municipality characteristics. The second part is the coefficient effect. It explains

---

<sup>11</sup>We add two dummies for North Rhine-Westphalia, because it is divided into two RASHIPs. In addition, we use a single dummy for all city-states, which only have one municipality each. Therefore, we add 15 federal state dummies to our regression analysis.

the difference in excess PCPs that evolves through a different impact of the municipality characteristics on PCP allocation. The third part is the interaction effect that explains the difference when endowment and coefficient effect occur together.

With the decomposition result, we separate endowment from coefficient effects and, thus, get information on how much the municipality characteristics influence the difference in the number of PCPs.

## 5 Results

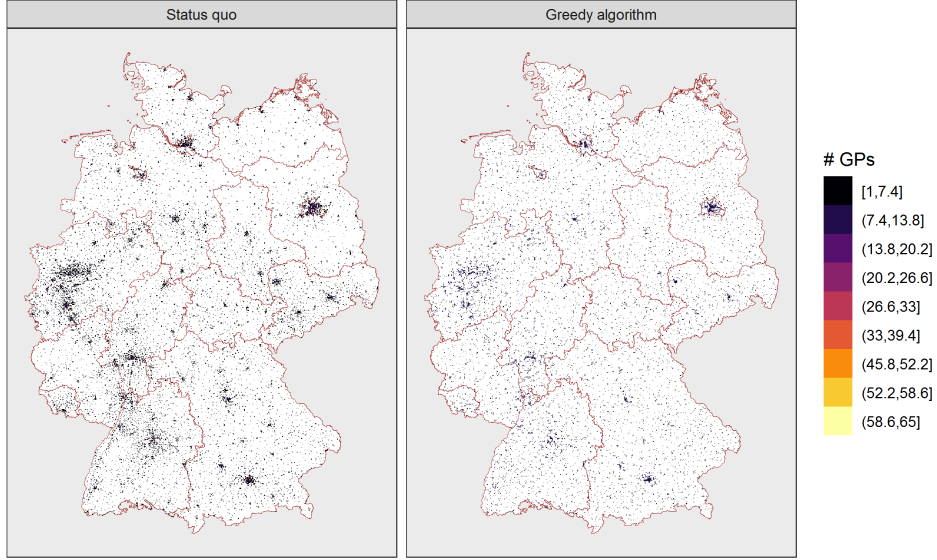
In our study, we investigate the overall shortage of primary care physicians, the undersupply of PCPs in rural regions and possible drivers of PCP allocation.

To inform our view on the general shortage of PCPs, Figure 5 shows the status-quo allocation compared to our counterfactual allocation on grid level. The left panel shows the distribution of primary care physicians in Germany as of 2019. The overall number of active primary care physicians was 55,006. As expected, densely populated areas like the city-states of Berlin and Hamburg as well as other large regions like Munich or the Ruhr area have a high density of PCPs. The right hand panel of Figure 5 shows our counterfactual distribution of PCPs in Germany. We determine the amount of PCPs needed in 2019 to be 58,144<sup>12</sup>. Remarkably, the calculated number of PCPs is fairly close to the actual number of active PCPs in 2019. We see a shortage of 3,138 PCPs ( $\approx 6\%$ ). This shortage can be interpreted as a lower bound. Note, that in our analyses, we assume that the PCPs work full time. However, since a certain proportion of PCPs also work part-time, the real shortage will be greater. As such, we take this difference as noteworthy. In our counterfactual, PCPs are more evenly spread out across Germany, resulting in complete coverage. The distributions differ mainly in densely populated areas such as Munich or the Ruhr area. Here, our counterfactual solution places fewer PCPs than there are in the status quo. Furthermore, more PCPs are located in the city-centers, which are easy to reach.

---

<sup>12</sup>We provide a sensitivity analysis in appendix B. We see that changing the maximum distance to PCPs and the capacity of the PCPs affects the number of PCPs needed as expected.

Figure 5: Distribution of Primary Care Physicians in Germany



Note: Comparison of PCPs allocation under status-quo and our counterfactual allocation in 2019 - grid level.

We now turn to the claim that Germany faces an especially severe under-supply in rural regions. We consider the excess number of PCPs in a municipality  $i$  calculated by  $N_i^{observed} - N_i^{greedy}$  (see equation (1)). Thereby, we take the allocation of PCP practices suggested by the greedy algorithm as the reference. Compared to this reference, we assess whether or not we face an over- or undersupply of PCPs in one region.

We report all values for the whole sample as well as for cities, towns / suburbs and rural areas. Table 4 summarizes the mean and standard deviation (SD) of our variables. The first row shows our dependent variable, the number of excess PCPs, i.e. the difference between the observed number of PCPs (second row) and the reference number of PCPs needed to achieve full support in a municipality (third row). The variable is 0 for an optimal allocation of PCPs in a municipality. Overall, we observe a shortage of  $-0.25$  PCPs per municipality. Yet, this is driven by rural areas, as in cities and towns / suburbs, we observe an overall surplus of PCPs.

In such a small scale analysis where population and their demand for health care should be the only influencing factors of PCP allocation, we should



observe an equilibrium and neither a surplus nor a deficit of PCPs. Therefore we use our educational, economic and environmental variables as well as the number of hospital beds to assess their correlation with excess PCPs. Our descriptive statistics show that cities in Germany are generally better equipped with educational facilities and hospitals than other regions. Trade tax multiplier and purchasing power are also highest here. In addition, more people move to the city or commute in to work. Environmental pollution is also highest in German cities. Towns and suburbs range in their characteristics between cities and rural regions.

Table 4: Descriptives: Municipality Characteristics

	All (N = 10,976)		Cities (N = 220)		Towns / Suburbs (N = 2,675)		Rural (N = 8,081)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Number of excess PCPs	-0.25	8.73	7.77	42.63	0.40	10.63	-0.69	3.86
Observed number of PCPs	5.01	35.51	102.38	227.89	8.49	8.58	1.21	2.13
Needed number of PCPs	5.26	32.29	94.61	203.25	8.09	12.05	1.90	4.17
Number of schools per capita	0.33	0.46	0.37	0.17	0.41	0.26	0.30	0.51
Number of universities per capita	0.00	0.02	0.03	0.06	0.00	0.03	0.00	0.02
Number of hospital beds per capita	1.68	10.02	7.73	6.56	4.42	16.82	0.61	6.08
Trade tax multiplier [%]	363.44	36.57	415.22	60.01	373.78	39.84	358.61	32.59
Purchasing power per capita [100,000 EUR]	0.22	0.06	0.25	0.05	0.24	0.05	0.22	0.06
Net commuters [1,000 persons]	0.03	5.95	16.60	37.35	-0.27	2.66	-0.33	0.56
Net migration [1,000 persons]	0.04	0.34	0.74	2.22	0.06	0.11	0.01	0.04
Particulate matter [ $\mu$ g / cubic meter]	10.08	1.03	11.08	1.09	10.25	1.00	10.00	1.01
Average noise [db]	4.35	10.87	22.47	15.19	9.21	14.41	2.25	8.03
Average number of hot days	19.30	7.41	22.11	7.83	21.97	7.97	18.34	6.95

Note: N is the number of observations, SD is standard deviation, the dimension is reported in square brackets, Sources: Federal Environment Agency (FEA), Federal statistics office (FSO) and German Weather Service (GWS), classification of cities, rural and suburban areas follows Dijkstra and Poelman (2014).

Table 5 summarizes our regression results for excess number of PCPs (see equation (2)). We first use the whole sample (All, column (1)), and then stratify our sample into Cities (2) and Rural areas (3). We compare cities with rural regions to gain most information on the influential municipality characteristics<sup>13</sup>.

When looking at the whole sample, we see a correlation of PCP shortage with all but one analysed regional variable. For the split samples we see a correlation for either cities or rural areas for all but two variables. This makes us confident that the chosen regional characteristics are meaningful for our analysis. For the whole sample, the number of education facilities, hospital beds, trade tax multiplier, net migration and hot days are positively correlated with the number of PCPs, noise and particulate matter as well as net commuters are negatively correlated.

For cities, we see a highly significant positive correlation of the number of PCPs and universities that is also large in size. We also find a highly significant positive correlation of the number of PCPs with net internal migration. We see a negative correlation only with net commuters. For rural areas, we see a positive correlation of the number of PCPs with schools, hospital beds, trade tax multiplier and net migration. The number of PCPs is negatively correlated with purchasing power and particulate matter in rural areas. To get an understanding, whether these differences between cities and rural areas are driven by the observed municipality characteristics or by other unobserved factors, we decompose our regression results.

---

<sup>13</sup>See appendix C for the results on Towns / Suburbs level.

Table 5: Regression: Excess Number of PCPs and Municipality Characteristics

	(1) All	(2) Cities	(3) Rural
Number of schools per capita	0.438** (0.160)	-4.496 (13.367)	0.203* (0.086)
Number of universities per capita	21.174*** (2.991)	122.664*** (31.143)	0.496 (2.120)
Number of hospital beds per capita	0.035*** (0.007)	0.081 (0.341)	0.014* (0.007)
Trade tax multiplier	0.011*** (0.003)	0.012 (0.053)	0.004* (0.002)
Purchasing power per capita	-3.941** (1.280)	11.557 (48.478)	-2.167** (0.750)
Net commuters	-0.152*** (0.014)	-0.261*** (0.063)	0.120 (0.086)
Net migration	12.868*** (0.292)	14.333*** (1.174)	2.533* (1.251)
Particulate matter	-0.604*** (0.121)	-2.328 (2.538)	-0.280*** (0.079)
Average noise	-0.017* (0.007)	-0.268 (0.145)	-0.008 (0.005)
Average number of hot days	0.024 (0.013)	0.353 (0.342)	0.009 (0.008)
N	10,976	220	8,081
R <sup>2</sup>	0.287	0.637	0.020
AIC	75,058.9	2,103.2	44,644.4

Note: We control for federal state fixed effects in all specifications, the three German city-states are represented by one fixed effect, N is the number of observations, standard errors reported in parentheses, \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ , classification of cities, rural / suburban areas follows Dijkstra and Poelman (2014). The coefficient estimates for the population variable in columns (2) and (3) are not comparable to the coefficient estimate in column (1) because the area type is mainly determined by population.

The descriptive results in table 4 show that cities have an average excess of PCPs and rural regions an average lack of PCPs. Hence, we observe a difference of 8.453 excess PCPs in cities compared to rural areas. The difference is highly significant. The decomposition results are summarized in Table 6. Our decomposition is formulated from the viewpoint of rural areas. The endowment effect shows the difference in the average number of PCPs in cities compared to rural areas if rural areas had the same level or amount of the municipality characteristics as cities. The coefficient effect reflects the difference in the number of PCPs if the influence of the characteristics in ru-

ral areas would be the same as in cities. The interaction effect quantifies the difference if endowment and coefficient effect occur simultaneously. The endowment and the coefficient effect show positive correlations, the interaction effect a negative correlation. All three overall effects are insignificant.

When looking at the contributions of individual variables, we see several significant effects. We observe significant positive endowment effects for schools, hospital beds and the trade tax multiplier and significant negative effects for purchasing power and particulate matter. For the coefficient effect, we observe significant positive effects for universities, net commuters and net migration and no significant negative effects. The same holds for the interaction effect.

Table 6: Oaxaca-Blinder Decomposition: Excess Number of PCPs and Municipality Characteristics in Rural and Urban Areas

Variable	Overall	Endowment effect	Coefficient effect	Interaction effect
Difference	8.453*** (2.874)			
Endowments	2.977 (3.436)			
Coefficients	6.165 (4.615)			
Interactions	-0.686 (6.408)			
Schools per cap.		0.012** (0.005)	-0.586 (3.265)	-0.117 (0.621)
Universities per cap.		0.012 (0.101)	0.096*** (0.035)	3.042*** (0.754)
Hospital beds per cap.		0.102** (0.051)	0.031 (0.166)	0.359 (1.908)
Trade tax multiplier		0.244** (0.099)	4.826 (23.974)	0.762 (3.814)
Purchasing power per cap.		-0.060** (0.025)	0.784 (7.943)	0.099 (0.991)
Net commuters		2.026 (2.750)	0.120* (0.066)	-6.227* (3.537)
Net migration		1.856 (1.431)	0.109*** (0.041)	9.710** (4.412)
Particulate matter		-0.301*** (0.098)	-22.368 (27.288)	-2.406 (2.964)
Average noise		-0.169 (0.123)	-0.510 (0.380)	-4.579 (3.435)
Avg. hot days		0.035 (0.032)	6.059 (5.282)	1.246 (1.101)
Base		-0.025 (0.018)	1.276 (0.777)	-0.321 (0.336)

Note: We control for federal state fixed effects in all specifications, therefore the coefficients in columns (2) to (4) do not sum up to the overall effects. The three German city-states are represented by one fixed effect, standard errors reported in parentheses, \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

## 6 Conclusion

The main objective of our study is to calculate a nationwide algorithmic reference allocation of PCPs and analyze the distribution of primary care physicians across different regions in Germany. With this approach, we

address the large discussion on geographical maldistribution of PCPs.

We apply a greedy capacitated algorithm on very fine spatial data that is independent of the current allocation and is not limited by variations within municipalities. We combine different data sources on supply and demand for primary care. The resulting dataset contains information on the number and allocation of potential patients and primary care physicians, the demand for primary care visits, driving times between potential patients and primary care physicians and regional characteristics. Using these data, we calculate a reference allocation and its difference from the status-quo, which serves as our indicator of misallocation. We use this indicator to inform our view on overall and regional specific shortage of PCPs as well as common characteristics of underserved areas.

Our results show an overall shortage of around 6% in Germany. We further see a geographical maldistribution of primary care physicians, which is particularly evident in rural areas, where there is a shortage of health care professionals. In towns / suburbs and especially in cities we observe an average excess of PCPs. However, the situation is heterogenous within these regional types. There are also cities with a shortage of PCPs and rural areas with an excess of PCPs.

Decomposing our results into endowment and coefficient effects, we find significant positive endowment effects for the number of schools, hospital beds per capita and the trade tax multiplier. The positive effect of the number of schools per capita could be explained by the importance of reconciliation of work and family (Küpper and Mettenberger, 2018 and Günther et al., 2010) and the positive effect of hospital beds per capita by physicians who worked or were trained in a hospital and then settled in the same area as a PCP. The positive effect of the trade tax multiplier may be driven by more attractive locations with a thriving economy that can afford a high trade tax multiplier. Still, there are regions with a high social expenditure that use high trade tax multipliers to offset their negative financing balance (Hüther et al., 2019). Purchasing power per capita and particulate matter show negative effects. Both seem to be more prevalent in rural areas. However, most of the difference in the number of excess PCPs between urban and rural areas remains unexplained by the analyzed municipality characteristics.

The significant coefficient effect of universities per capita is also noteworthy. There are practically no universities in rural areas, which is probably why we do not find a significant endowment effect. Nevertheless, our OLS results show a large positive correlation between universities and the number of excess PCPs for cities. This suggests that universities might be a more important pull factor for physicians than revealed by the endowment effect.

It's important to be aware of some limitations when interpreting our findings. First, for calculating the demand for PCP visits, we rely solely on demographic factors. We do not consider factors such as morbidity due to potential reverse causality. Second, we utilize survey data on physician visits. Thus, we might overestimate the demand. Also, survey responses may be subject to recall bias. Another option to get information on patient visits is to use administrative data. Yet, in administrative data follow up visits are not necessarily monitored as a separate visit. This might lead to an underestimation of the demand. Third, due to the cross-sectional design of the study we cannot interpret our results as causal.

In this paper, we suggest a reference allocation of PCPs independent of current practice locations. We identify regions where health care resources are disproportionately scarce and where interventions may be especially helpful in improving access to health care services and further explore potential municipality characteristics that correlate with over- or undersupply of PCPs. Our calculated reference allocation can therefore help to refine primary care location planning.

## References

- Augère-Granier, M.-L., & McEldowney, J. (2020). *Ältere Menschen in ländlichen Gebieten der Europäischen Union*. Retrieved April 10, 2024, from [https://www.europarl.europa.eu/RegData/etudes/IDAN/2020/659403/EPRS\\_IDA\(2020\)659403\\_DE.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2020/659403/EPRS_IDA(2020)659403_DE.pdf)
- Barrington-Leigh, C., & Millard-Ball, A. (2017). The world's user-generated road map is more than 80% complete. *PloS one*, *12*(8), 1–20. <https://doi.org/10.1371/journal.pone.0180698>
- Bartha, G., & Kocsis, S. (2011). Standardization of geographic data: The European inspire directive. *European Journal of Geography*, *2*, 79–89.
- Bauer, J., Brueggmann, D., Ohlendorf, D., & Groneberg, D. (2016). General practitioners in German metropolitan areas – distribution patterns and their relationship with area level measures of the socioeconomic status. *BMC Health Services Research*, *16*(672). <https://doi.org/10.1186/s12913-016-1921-5>
- Baum, C., Haering, A., Waehlert, L., & Werbeck, A. (2022). Hausärzteschaft und die Standortwahl– Ergebnisse einer Befragung. *Gesundheitsökonomie & Qualitätsmanagement*, *27*(3), 131–137. <https://doi.org/10.1055/a-1683-4298>
- BBSR. (2005). *Bundesamt für Bauwesen und Raumordnung, Raumordnungsbericht 2005*. Selbstverlag des Bundesamtes für Bauwesen und Raumordnung.
- Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 436–455.
- Breidenbach, P., & Eilers, L. (2018). RWI-GEO-GRID: Socio-economic data on grid level. *Jahrbücher für Nationalökonomie und Statistik*, *238*(6), 609–616. <https://doi.org/10.1515/jbnst-2017-0171>
- Chvatal, V. (1979). A Greedy Heuristic for the Set-Covering Problem. *Mathematics of Operations Research*, *4*(3), 233–235. <https://doi.org/10.1287/moor.4.3.233>
- Clarke, L., Anderson, M., Anderson, R., Klausen, M., Forman, R., Kerns, J., Rabe, A., Kristensen, S., Theodorakis, P., Valderas, J., Kluge, H., & Mossialos, E. (2021). Economic Aspects of Delivering Primary Care Services: An Evidence Synthesis to Inform Policy and Research



- Priorities. *The Milbank Quarterly*, 99(4), 974–1023. <https://doi.org/10.1111/1468-0009.12536>
- Daskin, M., & Dean, L. (2005). Location of health care facilities. In M. L. Brandeau, S. Francois, & P. William P (Eds.), *Operations research and health care* (pp. 43–76). Springer. <https://doi.org/10.1007/b106574>
- Destatis. (2020). *Ältere Menschen in Stadt und Land*. Retrieved April 10, 2024, from <https://www.destatis.de/DE/Themen/Querschnitt/Demografischer-Wandel/Aeltere-Menschen/stadt-land.html>
- Dijkstra, L., & Poelman, H. (2014). A harmonised definition of cities and rural areas: the new degree of urbanisation. *European Commission Working Paper*.
- European Commission. (2020). *European Commission Report on the Impact of Demographic Change*. Retrieved April 10, 2024, from <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0241&from=EN>
- Feige, U. (1998). A threshold of  $\ln n$  for approximating set cover. *Journal of the Association for Computing Machinery*, 45(4), 634–652. <https://doi.org/10.1145/285055.285059>
- Finkelstein, A., Gentzkow, M., & Williams, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. *Quarterly Journal of Economics*, 131(4), 1681–1726. <https://doi.org/10.1093/qje/qjw023>
- Goebel, J., Grabka, M., Liebig, S., Kroh, M., Richter, D., Schroeder, C., & Schupp, J. (2019). The German socio-economic panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik*, 239(2), 345–360. <https://doi.org/10.1515/jbnst-2018-0022>
- Greiner, G., Schwettmann, L., Goebel, J., & Maier, W. (2018). Primary care in Germany: access and utilisation—a cross-sectional study with data from the German Socio-Economic Panel (SOEP). *BMJ Open*, 8(10), e021036.
- Günther, O., Kürstein, B., Riedel-Heller, S., & König, H.-H. (2010). The role of monetary and nonmonetary incentives on the choice of practice establishment: a stated preference study of young physicians in Germany. *Health Services Research*, 45(1), 212–229.

- Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, 7(4), 12–18. <https://doi.org/10.1109/MPRV.2008.80>
- Hüther, M., Südekum, J., & Voigtländer, M. (2019). *Die Zukunft der Regionen in Deutschland. Zwischen Vielfalt und Gleichwertigkeit*. IW Köln.
- Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4), 453–479.
- Jürges, H. (2007). Health insurance status and physician-induced demand for medical services in Germany: new evidence from combined district and individual level data. *SOEP Paper*, (8).
- KBV. (2022). *Ärztmangel*. Retrieved April 10, 2024, from [https://www.kbv.de/html/themen\\_1076.php](https://www.kbv.de/html/themen_1076.php)
- Kelly, C., Hulme, C., Farragher, T., & Clarke, G. (2016). Are differences in travel time or distance to healthcare for adults in global north countries associated with an impact on health outcomes? A systematic review. *BMJ Open*, 6(11), e013059.
- Klose, J., & Rehbein, I. (2017). *Ärztatlas 2017: Daten zur Versorgungsdichte von Vertragsärzten*. WIdO, Wissenschaftliches Institut der AOK.
- Korte, B., & Vygen, J. (2018). *Combinatorial optimization* (Vol. 6). Springer. <https://doi.org/10.1007/978-3-662-56039-6>
- Kuhn, M., & Ochsén, C. (2019). Population Change and the Regional Distribution of Physicians. *The Journal of the Economics of Ageing*, 14C. <https://doi.org/10.1016/j.jeoa.2019.100197>
- Küpper, P., & Mettenberger, T. (2018). Professional and private location factors influencing general practitioners' settlement in rural regions. *Spatial Research and Planning*, 76, 229–245.
- Levine, D., Landon, B., & Linder, J. (2019). Quality and experience of outpatient care in the United States for adults with or without primary care. *JAMA Internal Medicine*, 179(3), 363–372.
- Lovász, L. (1975). On the ratio of optimal integral and fractional covers. *Discrete Mathematics*, 13(4), 383–390. [https://doi.org/10.1016/0012-365X\(75\)90058-8](https://doi.org/10.1016/0012-365X(75)90058-8)
- Luxen, D., & Vetter, C. (2011). Real-time routing with OpenStreetMap data. *Proceedings of the 19th ACM SIGSPATIAL International Con-*

- ference on Advances in Geographic Information Systems, 513–516. <https://doi.org/10.1145/2093973.2094062>
- Macinko, J., Starfield, B., & Shi, L. (2007). Quantifying the health benefits of primary care physician supply in the United States. *International Journal of Health Services*, 37(1), 111–126.
- NASHIP. (2018). *Regionale Verteilung der Ärzte in der vertragsärztlichen Versorgung*. Retrieved April 10, 2024, from <https://gesundheitsdaten.kbv.de/cms/html/16402.php>
- NASHIP Germany. (2019). Retrieved April 10, 2024, from <https://arztsuche.kbv.de/>
- Newhouse, J. (1990). Geographic access to physician services. *Annual Review of Public Health*, 11, 207–230. <https://doi.org/10.1146/annurev.pu.11.050190.001231>
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 693–709.
- OECD. (2020). *Health at a Glance: Europe 2020*. Retrieved April 10, 2024, from <https://www.oecd-ilibrary.org/content/publication/82129230-en>
- OECD. (2021). *Health at a Glance 2021*. Retrieved April 10, 2024, from <https://www.oecd.org/health/health-at-a-glance/>
- Ono, T., Schoenstein, M., & Buchan, J. (2014). *Geographic Imbalances in Doctor Supply and Policy Responses* (Working Paper). OECD Health Working Paper No. 69.
- Ozegowski, S., & Sundmacher, L. (2012). Ensuring access to health care—Germany reforms supply structures to tackle inequalities. *Health Policy*, 106, 105–109. <https://doi.org/10.1016/j.healthpol.2012.04.002>
- Parka, M.-J., & Honga, S.-P. (2009). Approximation of the capacitated set cover [Manuscript submitted for publication].
- Rais, A., & Viana, A. (2011). Operations research in healthcare: a survey. *International Transactions in Operational Research*, 18(1), 1–31. <https://doi.org/10.1111/j.1475-3995.2010.00767.x>
- RASHIP Baden-Württemberg. (2019). Retrieved April 10, 2024, from <https://www.arztsuche-bw.de>
- RASHIP Bavaria. (2019). Retrieved April 10, 2024, from <https://dienste.kvb.de/arztsuche/app/einfacheSuche.htm>

- RASHIP Berlin. (2019). Retrieved April 10, 2024, from <https://www.kvberlin.de/60arztsuche/index.html>
- RASHIP Bremen. (2019). Retrieved April 10, 2024, from <https://www.kvhh.de/arztsuche/>
- RASHIP Hamburg. (2019). Retrieved April 10, 2024, from <https://www.kvhh.net/de/physicianfinder.html?name=&location=>
- RASHIP Hessen. (2019). Retrieved April 10, 2024, from <https://arztsuche.hessen.de/arztsuche>
- RASHIP Mecklenburg-Vorpommern. (2019). Retrieved April 10, 2024, from <https://www.kvmv.de/ases-kvmv/ases.jsf>
- RASHIP Nordrhein. (2019). Retrieved April 10, 2024, from <https://www.kvno.de/20patienten/10arztsuche/index.html>
- RASHIP Rhineland-Palatinate. (2019). Retrieved April 10, 2024, from <https://www.kv-rlp.de/patienten/patientenservice-116117/>
- RASHIP Saarland. (2019). Retrieved April 10, 2024, from <http://arztsuche.kvsaarland.de/>
- RASHIP Saxony. (2019). Retrieved July 2, 2019, from <https://asu.kvs-sachsen.de/arztsuche/pages/search.jsf>
- RASHIP Saxony-Anhalt. (2019). Retrieved April 10, 2024, from [https://www.kvsa.de/service/arzt\\_und\\_therapeutensuche\\_in\\_sachsen\\_anhalt.html](https://www.kvsa.de/service/arzt_und_therapeutensuche_in_sachsen_anhalt.html)
- RASHIP Schleswig-Holstein. (2019). Retrieved April 10, 2024, from <https://www.kvsh.de/patienten/arzt-und-psychotherapeutensuche>
- RASHIP Thuringia. (2019). Retrieved April 10, 2024, from <https://www.kv-thueringen.de/arztsuche/>
- RASHIP Westfalen-Lippe. (2019). Retrieved April 10, 2024, from <https://www.kvwl.de/arztsuche>
- Riverin, B., Strumpf, E., Naimi, A., & Li, P. (2018). Optimal timing of physician visits after hospital discharge to reduce readmission. *Health Services Research, 53*(6), 4682–4703.
- Saijo, Y., Yoshioka, E., Kawanishi, Y., Nakagi, Y., Hanley, S., & Yoshida, T. (2018). Relationships between road-distance to primary care facilities and ischemic heart disease and stroke mortality in Hokkaido, Japan: A Bayesian hierarchical approach to ecological count data. *Journal of General and Family Medicine, 19*(1), 4–8.

- Salm, M., & Wübker, A. (2020). Sources of regional variation in healthcare utilization in Germany. *Journal of Health Economics*, *69*, 102271–102286. <https://doi.org/10.1016/j.jhealeco.2019.102271>
- Sundmacher, L., & Ozegowski, S. (2016). Regional distribution of physicians: the role of comprehensive private health insurance in Germany. *European Journal of Health Economics*, *17*, 443–451.
- Vazirani, V. (2001). *Approximation Algorithms*. Springer.
- Vogt, V. (2016). The contribution of locational factors to regional variations in office-based physicians in Germany. *Health Policy*, *120*(2), 198–204.
- WHO. (2018). *Declaration of Astana: Global Conference on Primary Health Care*. Retrieved April 10, 2024, from <https://www.who.int/publications/i/item/WHO-HIS-SDS-2018.61>

## Appendix

### A Summary of websites NASHIP and RASHIP

Table A.1 give the website for each health insurance association.

Table A.1: Scraped Websites

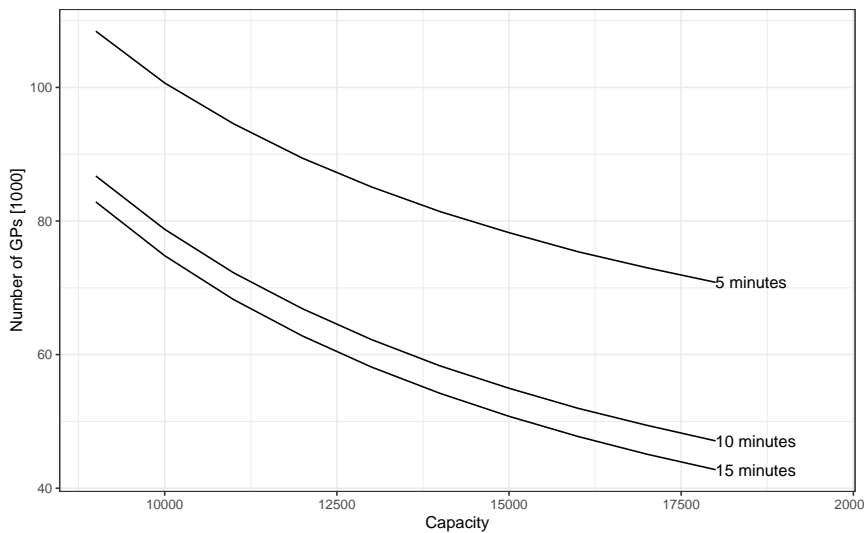
Association	URL
NASHIP Germany (2019)	<a href="https://arztsuche.kbv.de/">https://arztsuche.kbv.de/</a>
RASHIP Baden-Württemberg (2019)	<a href="https://www.kvbawue.de/">https://www.kvbawue.de/</a>
RASHIP Bavaria (2019)	<a href="https://www.kvb.de/">https://www.kvb.de/</a>
RASHIP Berlin (2019)	<a href="https://www.kvberlin.de/">https://www.kvberlin.de/</a>
RASHIP Berlin (2019)	<a href="https://www.kvbb.de/">https://www.kvbb.de/</a>
RASHIP Bremen (2019)	<a href="https://www.kvhb.de/">https://www.kvhb.de/</a>
RASHIP Hamburg (2019)	<a href="https://www.kvhh.net/">https://www.kvhh.net/</a>
RASHIP Hessen (2019)	<a href="https://www.kvhessen.de/">https://www.kvhessen.de/</a>
RASHIP Nordrhein (2019)	<a href="https://www.kvn.de/">https://www.kvn.de/</a>
RASHIP Mecklenburg-Vorpommern (2019)	<a href="https://www.kvmv.de/">https://www.kvmv.de/</a>
RASHIP Nordrhein (2019)	<a href="https://www.kvno.de/">https://www.kvno.de/</a>
RASHIP Rhineland-Palatinate (2019)	<a href="https://www.kv-rlp.de/">https://www.kv-rlp.de/</a>
RASHIP Saarland (2019)	<a href="https://www.kvsaarland.de/">https://www.kvsaarland.de/</a>
RASHIP Saxony (2019)	<a href="https://www.kvs-sachsen.de/">https://www.kvs-sachsen.de/</a>
RASHIP Saxony-Anhalt (2019)	<a href="https://www.kvsa.de/">https://www.kvsa.de/</a>
RASHIP Schleswig-Holstein (2019)	<a href="https://www.kvsh.de/">https://www.kvsh.de/</a>
RASHIP Thuringia (2019)	<a href="https://www.kv-thueringen.de/">https://www.kv-thueringen.de/</a>
RASHIP Westfalen-Lippe (2019)	<a href="https://www.kvwl.de/">https://www.kvwl.de/</a>

## B Sensitivity analysis

Our results may depend on the two input parameters, maximum distance to PCPs and capacity of PCPs.

We test the influence of the two constraints for the capacitated set cover algorithm, i.e., limited capacity of PCPs and driving time less than 15 minutes, on the required number of PCPs. To do so we vary each parameter, keeping the other parameter constant. We analyze the effect of PCP capacity and the threshold given the maximal driving time to the nearest PCP. Figure 6 shows the results. The effect of the capacity is approximately linear, while the effect of the threshold seems exponential.

Figure 6: Placebo Test: Input Parameters



Note: Sensitivity of required number of PCPs towards input parameters

## C Full regression results

Table C.1: Regression: Excess Number of PCPs and Municipality Characteristics

	(1)	(2)	(3)	(4)
	Pool	Cities	Towns Suburbs	Rural
Number of schools per capita	0.438** (0.160)	-4.496 (13.367)	2.238** (0.864)	0.203* (0.086)
Number of universities per capita	21.174*** (2.991)	122.664*** (31.143)	-0.272 (7.653)	0.496 (2.120)
Number of hospital beds per capita	0.035*** (0.007)	0.081 (0.341)	0.019 (0.012)	0.014* (0.007)
Tax rate	0.011*** (0.003)	0.012 (0.053)	0.028*** (0.008)	0.004* (0.002)
Purchasing power per capita	-3.941** (1.280)	11.557 (48.478)	-16.357*** (4.205)	-2.167** (0.750)
Net commuters	-0.152*** (0.014)	-0.261*** (0.063)	0.322*** (0.081)	0.120 (0.086)
Net migration	12.868*** (0.292)	14.333*** (1.174)	2.687 (1.886)	2.533* (1.251)
Particulate matter	-0.604*** (0.121)	-2.328 (2.538)	-0.822** (0.318)	-0.280*** (0.079)
Average noise	-0.017* (0.007)	-0.268 (0.145)	-0.014 (0.015)	-0.008 (0.005)
Average number of hot days	0.024 (0.013)	0.353 (0.342)	0.016 (0.035)	0.009 (0.008)
N	10,976	220	2,675	8,081
R <sup>2</sup>	0.287	0.637	0.057	0.020
AIC	75,058.9	2,103.2	20,128.7	44,644.4

Note: We control for federal state fixed effects in all specifications, the three German city-states are represented by one fixed effect, N is the number of observations, standard errors reported in parentheses, \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ , classification of cities, rural and suburban areas follows Dijkstra and Poelman (2014). The coefficient estimates for the population variable in columns (2-4) are not comparable to the coefficient estimate in column (1) because the area type is mainly determined by population.