I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17726

Does Your Doctor Matter?

Rita Ginja
Julie Riise
Barton Willage
Alexander Willén

FEBRUARY 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17726

Does Your Doctor Matter?

Rita Ginja
University of Bergen, Institute for Fiscal Studies and IZA

Julie Riise
University of Bergen and IZA

Barton Willage
University of Colorado – Denver, University of Bergen and NBER

Alexander Willén
Norwegian School of Economics, CESifo, IZA and UCLS

FEBRUARY 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Does Your Doctor Matter?*

We estimate doctor value-added (VA) combining population-wide patient-doctor register data with exogenous variation in the assignment of patients to GPs. We find substantial variation in the quality of GPs as measured by patients' post-assignment mortality. Certain doctor characteristics and practice styles predict VA, but most of the variation in GP quality is driven by differences in GPs' unobserved ability. VA variation across GPs primarily reflects differences in GPs' ability to engage in prevention and assign the right procedure to the patient. Finally, patients are unable to identify who the high-quality doctors are, and patient-generated ratings are uncorrelated with GP VA.

Corresponding author:
Rita Ginja
Department of Economics
University of Bergen
Fosswinckelsgate 14
5007 Bergen
Norway
E-mail: rita.ginja@uib.no

* The authors would like to thank seminar, workshop, and conference participants at the Norwegian School of Economics, the University of New Mexico, the Norwegian National Health Economics Seminar Series (2022), the Annual Meeting of the Society of Labor Economists (2022), the Nordic Summer Institute in Labor Economics (2022), the Essen Health Conference (2022), the Solstrand Workshop on Health, Human Capital and Social Insurance (2022), the FAIR Midway Conference in Tromsø (2022), the Annual Norwegian Research Meeting for Economists (2022), and the Norwegian Health Economics Conference (2022). Willén gratefully acknowledges financial support from the Research Council of Norway through its Centers of Excellence Scheme, FAIR project no. 262675.

1 Introduction

The average OECD country spends more than 10 percent of GDP on healthcare services each year. The goal of these expenditures is to create conditions for good and equitable health among the entire population. Despite this objective, there is substantial variation in the quality of healthcare services and patient outcomes. How much of these disparities can be explained by variation in doctor quality?

Even though there is a large interest in the role of doctor quality in the production of health, lack of exogenous variation in doctor-patient assignment coupled with limited linked doctor-patient data has made it difficult to conduct a comprehensive analysis on doctor value-added. We overcome these challenges and provide new evidence on the role of physicians in health production. This paper not only examines how much variation there is in doctor quality within a country, but also breaks new ground in helping us better understand who the good physicians are, why certain physicians are better than others, how doctor quality impacts patient outcomes, and if patients are able to identify who the good physicians are.

To conduct our analysis, we rely on rich doctor-patient register data from Norway and exploit exogenous variation in GP assignment for Norwegians. GPs represent a patient's first point of contact with the healthcare system and are responsible for initial examination, treatment, diagnosis, medication prescription, referrals, and sick note validation.[1] They thus represent the gatekeepers to the country's health care system, and ex ante it is likely that quality variation across GPs have a substantial impact on the health of patients.

When GPs retire or move – or for some other reason outside the patients control become unavailable – the Norwegian Health Economics Administration randomly and immediately reassigns patients to new local GPs conditional on municipality and availability. We use these GP reassignments as a source of exogenous variation in doctor-patient assignment to identify the quality of doctors. This randomization is done through an automated computer system, ensuring that neither the GP nor any other person can affect who is or is not transferred, and to whom the transferees should be reallocated.

[1] GPs are the initial contact with the health system not only in Norway, but also in many other countries. Examples include Croatia, Czech Republic, Denmark, Finland, Iceland, Ireland, Italy, Lithuania, the Netherlands, Poland, Portugal, Romania, Slovenia, Spain, Sweden, and the UK (Boerma et al., 1997).

To perform our analysis, we leverage matched doctor-patient administrative data on all individuals in Norway who were subject to an exogenous GP reassignment between 2006 and 2015. We link these data to detailed information on GP practice behavior and demographic characteristics as well as patient demographics and long-run health and labor market outcomes. Our primary measure of doctor quality is the 2-year post-assignment mortality of patients. The decision to focus on a mortality-based VA measure means that we are invoking the relatively strong assumption that keeping patients alive represents an unambiguous and objective measure of doctor quality. However, healthcare quality and health provision are multifaceted, and we acknowledge that there may be alternative indicators that can be used to provide additional insight into the quality of GPs.[2] We strongly encourage future research to explore such alternative indices. We focus on patients who are 55 or older because the mortality rate of Norwegians below this age is negligible.[3]

To ensure that our results are not driven by a systematic correlation between the characteristics of patients and the exogenously-assigned GPs, we incorporate a full set of pre-reassignment GP fixed effects in all our regressions. The thought experiment underlying our estimation approach is, therefore, to compare the mortality of patients who originally had the same GP and lived in the same municipality but then were exogenously assigned to new – and different – GPs through a computerized randomization procedure due to factors outside their control.[4]

After having estimated the quality of GPs in Norway, explored the impact on patients, and investigated potential mechanisms that may explain the variation in GP VA that we observe, we ask if patients are able to identify the quality of GPs themselves. To address this question, we merge our administrative data on GPs with proprietary data on patient-generated ratings of GPs. By correlating our measure of GP VA with the patient-generated ratings of these GPs, we can examine how patients' perceptions of GPs correlate with GPs' actual quality. A priori, it is not clear what direction of an association we would expect. First, objective indicators of quality, such as those based on mortality, may not be readily available and salient to patients. Second, objective measures may not capture all aspects of care that patients care about.

[2] We provide some such alternative measures in the online appendix.

[3] See https://www.ssb.no/256918/age-specific-death-rates-for-males-and-females.

[4] We use all exogenous switches regardless of whether the patient stays with the new GP after the switch, and rely on an intent-to-treat analysis.

We present six results. First, we identify large differences in the quality of physicians across Norway, with a standard deviation of 0.157 for our GP VA measure based on the 2-year post assignment mortality rate. This suggests that a one standard deviation change in the quality of a patient's GP is associated with a 15.7-percentage point increase in the probability of being alive two years after GP assignment. In terms of the magnitude of our estimated effects, moving from a median-quality GP to a GP at the 75th percentile reduces the two-year mortality risk with 4.9 percentage points, and moving from the bottom quartile of the GP quality distribution to the top quartile generates an increase in predicted life expectancy by 3.5 months. Note that conditional on post-swap mortality, the average length of time to death post-swap is 5.7 years, such that the effects we identify represent a 5 percent increase in time to death relative to the mean.

Second, we show that specific observable doctor characteristics and practice styles can explain only a small amount of the overall variation in doctor quality. These results are similar to those on worker quality in other professions, where observable characteristics have little explanatory power.[5] This is a particularly interesting result, as we have access to more detailed information on workers' behaviors and work habits than most of the previous work on this topic. Importantly, our findings demonstrate that doctor quality cannot be identified through studying the underlying characteristics and practice styles of GPs. This emphasizes the independent value of the VA approach in identifying GP quality and improving health production.

Third, we show that patients assigned to a high-quality GP do not receive a different quantity of treatment. Instead, the VA variation across GPs is coming from differences in the actual interaction between the GP and the patient, and the GP's ability to assign the right treatment or referral to the patient. This conclusion is based on three results in the analysis: (1) the mortality effects we identify are driven by a reduction in the likelihood of dying from diseases that are treatable if detected early; (2) there is a reduction in urgent care visits as a consequence of being assigned a high-VA GP; and (3) high-VA GPs have fewer consultations with patients but the patients' health outcomes are better.

Fourth, to understand if doctor quality has an impact on individual outcomes beyond health

[5]E.g., Hanushek 1992; Rockoff 2004; Rivkin, Hanushek and Kain 2005; Hanushek et al. 2005; Kane and Staiger 2008.

and healthcare utilization, we merge the patient data to population-wide labor market registers. We find some suggestive evidence that GP quality has a positive impact on patients' labor market outcomes, but these effects are economically modest and most effects are not statistically significantly different from zero. The one exception relates to welfare program usage for which we see significant negative effects. This implies a non-negligible fiscal benefit to the state from improved GP quality.

Fifth, we show that patients do not select high-quality doctors when given the chance. We do this by exploiting the fact that patients are allowed to independently change their GP twice per year, and examining if GPs with a higher VA are more likely to experience an influx of endogenous patient changes. The lack of a strong correlation on these dimensions could be driven by two distinct factors: An inability among patients to identify high-quality GPs or a desire to choose GPs based on some other metric. In auxiliary analyses, we therefore exploit rich information on GP practice style and behavior. We show that patients who endogenously swap to a new GP do so to get more popular and experienced GPs who are female and hold a specialization. This could be explained by patients erroneously associating these GP characteristics with GP quality, or because patients independently value these GP characteristics.

Finally, we find no relationship between subjective patient-generated ratings of GPs and the objective GP VA measure we construct based on post-assignment mortality. This result is consistent with the notion that patients either are unable to identify high-quality GPs, or they value aspects of GPs other than their ability to save lives. Understanding the relationship between our mortality-based GP VA measure and the patient-generated ratings of these GPs is of great independent value. Specifically, recent years have seen a large increase in the availability of patient-generated online rating platforms. If these platforms are used by patients to make health care choices, it is important to know to what extent they are reflective of GP quality, what information they convey, and to what extent they facilitate improvements in health care production.

To quantify the magnitude of the GPs' impact on health production, we adopt Hanushek's (2009) proposal to replace the worst performing 5 percent of GPs with GPs of average quality. Using $35,000 as the predicted value of an additional life year in Norway (Elvik 2018), we

calculate that the social benefit of replacing the bottom five percent of GPs in the quality distribution with GPs of average quality is $35,500 per patient, $11.7 million per GP, or 1.2 billion in total. Using instead the US Food and Drug Administration (FDA) value of a statistical life year ($369,000), we calculate that the social benefit is $372,075 per patient, $123 million per GP, or $12.4 billion in total.[6] At the same time, our results suggest that the practice styles and behaviors of higher-quality GPs are associated with a lower per-patient cost and a reduced dependence on the Norwegian welfare system. This implies that substantial investments in GP quality might be a cost-effective way to reduce mortality and improve general health. However, this depends on the cost of raising a GP's VA; we see this as a valuable avenue for future research.

The key contribution of our paper is to exploit exogenous variation in doctor-patient matches and leverage the full power of the Norwegian register data to identify the VA of doctors, how much variation there is in doctor quality within a country, who the high-quality physicians are, why certain physicians are better than others, and if patients are able to identify the high-quality physicians. We contribute to the existing literature in three main ways.

First, our analysis advances the rapidly expanding literature on doctor effectiveness. Within this literature, most studies have focused on the effect of provider practice style, hospitals, and treatment intensity on patient's short-term outcomes (e.g., Schnell and Currie 2018; Currie at al. 2016; McClellan et al. 1994; Bartel et al. 2014; Doyle et al. 2010). Due to the lack of detailed data on doctor quality linked to exogenous patient-doctor matches, only a small number of studies have attempted to identify doctor quality effects (Fletcher et al. 2014; Stoye 2022; Currie and Zhang 2021). Most relevant to our study is perhaps Currie and Zhang (2021), who use administrative data from the U.S. Veterans Health Administration (VHA) to examine primary care provider (PCP) effectiveness along three dimensions of hospitalizations, study how it is correlated with PCP characteristics, and what the implications are for mortality. Similar to us, they find substantial variation in effectiveness, and that effective PCPs do more with less.

[6]Interestingly, the per patient social benefit is qualitatively very similar to the teacher VA effect benefit found in Chetty et al. (2014). However, the social benefits in our setting are calculated over two years and only for individuals aged 55 and above. Thus, true total societal benefit could be significantly larger.

While providing novel evidence on the quality of attending physicians, data limitations have led these studies to focus on particular hospitals, patients, and physician specialties.[7] In addition, they have not been able to trace the impact on patients long-run labor market and health outcomes, or to explore the relationship between patient perceptions of GP quality and objective measures of GP quality.[8] We, therefore, add to the doctor effectiveness literature by identifying the distribution of physician quality across a representative sample in an entire country and its impact on later-in-life outcomes of patients, providing detailed evidence on the potential mechanisms through which GP quality may operate, and offering evidence on the relationship between GP VA and patient-generated ratings; something that has not been done before.

Second, our study helps advance the literature on patient choice in the health care sector. Most of this literature has focused on patients' choice of hospitals, and finds that hospital quality is associated with increased patient demand (e.g., Cutler et al. 2004; Howard 2006; Ho 2006; Pope 2009). A limited set of studies have explored the relationship between various observable doctor attributes (e.g., distance, opening hours, age, gender, practice style, and ethnicity) and patient demand (e.g., McLean and Sutton 2005; Godager 2012; Dixon et al. 1997; Santos, Gravelle and Propper 2017). Finally, a small set of studies have investigated the effects of patient-generated ratings on health care choices (Luca and Vats 2013; Chan 2022; Bensnes and Huitfeldt 2021).

Our main contribution to this literature is to examine whether individuals are able to identify, and select, GPs that are of higher quality; something that has not yet been done. Our results reveal that patients do not select physicians based on their VA, and that the correlation between

[7]There is another set of important papers that have provided evidence on other dimensions of practice styles. For example, Fadlon and Van Parys (2020) estimate the effects of practice styles on health care utilization by first identifying Medicare patients exposed to primary care relocations or retirements ("exits"), and then estimating event studies to compare patients who choose PCPs with different practice style intensities. Zhang (2022) estimates the effect of primary care physician retirement on the health care utilization of the elderly population that is on Medicare.

[8]For example, there are three major differences between our paper and Currie and Zhang (2021). First, while Currie and Zhang (2021) build their effectiveness measure based on male veterans using the VHA, we are able to access a broader population and physicians who are subject to a more heterogeneous set of health issues. Second, we have the ability to expand the set of outcomes beyond just health-related effectiveness measures, and we are able to explore impacts on a range of welfare and labor market outcomes. Third, by combining our registry data with proprietary data from a patient-generated doctor rating website (*legelisten.no*), we are able to extend the literature and explore the relationship between GP VA and patient rankings of these GPs.

objective GP VA and subjective patient-generated GP ratings is very small. These findings have important implications for ongoing debates on the organization of health care services and the distribution of information to patients.

Finally, there is a rich literature on cause-specific mortality and morbidity, and an emerging literature on inequalities in health outcomes across individuals over the course of the lifecycle (e.g., Angerer, Waibel, and Stummer 2019; Finkelstein et al. 2021).[9] Our paper advances these literatures by demonstrating to what extent primary care physicians contribute to variation in mortality and morbidity across individuals. We show that there are considerable social gains associated with improving GP quality, and that investments in GP quality may represent a cost-effective way to reduce mortality and improve general health. In addition, we uncover substantial heterogeneity in doctor quality, alluding to a potentially important role of doctors in explaining inequalities in health outcomes across space.

2 Institutional Background

Overview of the healthcare system. The per capita spending on healthcare in Norway is the third highest in the world, and the country consistently scores in the top of global healthcare performance rankings (OECD 2019). The healthcare system is based on the principle of universal access, and enrollment is automatic for all residents. While healthcare is not entirely free, it is heavily subsidized and annual copayment is capped at approximately $280. Children under age 16 are exempted from copayments.

Similar to most European countries, the healthcare system in Norway is split into two levels, with local municipalities providing primary care services and larger health regions providing specialist care.[10] Entry into specialist care and hospital services can only be obtained through referrals from the GP in the primary care sector (except for emergencies).

The Norwegian Health Economics Administration (which is part of the Norwegian Directorate of Health) assigns every resident of Norway to a local GP. In most cases, patients interact with their assigned GP every time they use the healthcare system, and the average patient in our

[9] For example, Clemens et al. (2014), Abaluck et al. (2021), McClellan et al. (1994), Bartel et al. (2014), Currie at al. (2016), Geruso et al. (2021), and Finkelstein et al. (2016; 2021).

[10] During our analysis period, there were 422 municipalities and 4 health regions in Norway.

sample sees the doctor once every seven weeks, or 7.5 times per year. However, there are a few rare exceptions. For example, if the patient is brought into the emergency department, there is no interaction between the patient and the GP (though the GP will be informed and updated on the condition of the patient). Individuals can independently change the GP they have been assigned twice per year. As of 2015, there were approximately 5000 GPs in Norway, and each GP had on average 1100 patients. The average GP was 47 years old, and 60 percent were male.

Most GPs are self-employed, with around 5 percent of GPs being salaried municipality employees during most of our time period. Municipalities contract with the self-employed GPs to provide services to residents by assigning them a patient list. GPs receive earnings through three different channels: (1) capitation from the municipalities – approximately 30 percent, (2) fee-for-service from the health administration – almost 70 percent, and (3) out-of-pocket payments from patients. The average GP earns approximately $100,000 in annual pre-tax wage in Norway, and GPs are among the most well-paid professions in the country.

Exogenous patient-doctor matches. To identify the effect of GP quality on patient outcomes, we would ideally randomize individuals to doctors that differ in their ability to treat patients. This randomization would break the correlation between GP quality and unobserved determinants of individual outcomes, and comparing the effects of individuals who are randomly assigned to differentially "able" doctors would give a reduced form estimate of the effect of being exposed to a better or worse GP.

The objective of our research design is to mimic this hypothetical experiment. Our source of exogenous variation comes from a novel feature of the Norwegian health care law in which patients are conditionally randomly allocated to new GPs in the event their current GP closes down or significantly reduces their practice. As GPs vary in their ability to treat patients, this randomization process allows us to construct a measure of GP VA that is unrelated to any patient characteristics.

Approximately 26 percent of all swaps in the Norwegian health care system are caused by doctors reducing or terminating their patient list, something we can observe directly in the data (Appendix Table A.1).[11] The other two main sources of swaps are ordinary exchanges

[11]The other two main sources of swaps are ordinary exchanges (65 percent) and automatic allocation (7 per-

(65 percent) and automatic allocation (7 percent). Ordinary swaps refer to endogenous swaps initiated by patients (individuals are limited to perform two such swaps per calendar year), while automatic allocations refer to the assignment of patients to doctors the first time they enter the system (which would be at the time of birth for Norwegian-born individuals). We do not use these swaps to identify GP VA since they suffer from endogeneity issues. To causally identify GP quality, we restrict attention to GP reassignments that are caused by GP list reductions and terminations, and that therefore are outside the patient's and the new GP's control. In addition, we always include pre-reassignment GP fixed effects, such that our quality measures are obtained by comparing patients who initially had the same GP but then got exogenously allocated to new, and different, GPs.[12] An added benefit with focusing on GP reassignments is that when patients match to new physicians, there is a consequential reassessment of patients' medical needs (Simonsen et al. 2021). This is ideal for the purpose of our exercise, as it improves our ability to identify quality effects.

Even if random assignment is ensured in our setting, there are two types of selection that can theoretically occur and that are worth noting: (1) Doctors may choose which patients to remove from their list in the event of list reductions and (2) individuals may choose to endogenously reallocate to a new GP in anticipation of a list reduction/termination. (1) would violate Norwegian law, and neither of these sources of selection would invalidate a causal interpretation of our results. However, they would impact the generalizability of our findings. In Section 3, we show that selection from either of these sources is minimal.

The mean number of patients that are being sent (exogenously) from one specific GP to another specific GP during our analysis period is approximately 22, with around 4200 old GP – new GP combinations (Appendix Table A.2). This generates a total sample of more than 92000 individuals. The average number of patients that leave a specific GP in the event of a closure or reduction is 71, with a large standard deviation and a maximum number of about 1300. This reflects the fact that the exogenous swaps can come either from a list reduction or

cent). Ordinary swaps refer to endogenous swaps initiated by patients (individuals are limited to perform two such swaps per calendar year), while automatic allocations refer to the assignment of patients to doctors the first time they enter the system (which would be at the time of birth for Norwegian-born individuals). We do not use these swaps to identify GP VA since they suffer from endogeneity issues.

[12]Should a new GP take over the entire patient list of a retiring GP those patients will not contribute to our identification, as they will be fully absorbed by our pre-assignment GP fixed effects.

a complete list closure. Appendix Figure A.1 provides a map showing the distribution of GPs per municipality across the country and Appendix Figure A.2 documents the average number of GPs by municipality. The figures demonstrate that there is a relatively constant number of GPs across space and that reassignments occur across the entire country.

3 Data and Method

3.1 Data

Our analysis data come from administrative registers covering the universe of Norwegian residents. Below, we provide a discussion on the relevant variables and data registers used to construct our final analysis data set. Appendix B includes a detailed overview of the data registers we use. Note that there is some variation in the years for which the various registers are available, such that the sample sizes for different outcomes vary slightly based on which register is used.[13]

GP Characteristics and Health Service Use. The Norwegian GP register contains detailed information on all active GPs in the country for each year. We use data for the years 2006 through 2020. Using unique GP identifiers, we combine this data with information from the Control and Payment of Health Refunds Database (KUHR in Norwegian), which provides information on all visits to GPs and primary care emergency rooms. GPs are required to report all consultations and relevant International Classification of Primary Care codes (ICPC-2) to this claims database to receive reimbursements from the government. ICPC-2 codes convey information about the GPs' assessment of the patient's health problems.[14] Additionally, the database contains information on all reimbursable actions taken by the GP, such as whether the GP engaged in a minor surgical intervention, performed laboratory tests, or conducted an administrative task.

We link the above data to information on the use of specialist health care services from the

[13] See also Ginja et al., 2024.

[14] Specifically, each ICPC-2 code is composed of one letter indicating where the symptoms or diseases are located in the body, and two numbers indicating symptoms and diseases diagnosed by the GP.

Norwegian Patient Register. This register is available for the years 2008-2020 and contains information on all patients who have received treatment from specialist health services. All visits are registered with date of stay, length of stay, and one or more related diagnoses (ICD-10 classification). We use this data to study impacts on hospitalizations (inpatient admissions) and consultations at outpatient clinics.[15]

Appendix Table A.3 provides descriptive statistics of the GPs in our sample (Panel A) and the patients they serve (Panel B). In columns (1) and (2) we show the mean and standard deviation for the sample of GPs that we use to compute the VA, and in columns (3) and (4) we show the same statistics for a broader set of GPs involved in exogenous swaps (i.e., that replaced other GPs). These two samples differ insofar as only those GPs who had at least one patient over the age of 55 are included in our VA calculation sample, while we include all doctors who had at least one patient over the age of 25 in the broader comparison group. In columns (5) and (6), we show the same statistics for all GPs who had at least one patient over the age of 55; irrespective of that swap being endogenous or exogenous. On average, the GPs in our sample are 44 years old, 39% of them have a specialization in general medicine, and 63% are males. The average GP is responsible for 1,108 patients. Comparing columns (1)-(2) with columns (3)-(6) reveals that the sample used to estimate GP VA is similar to the full sample with the exception of age and specialization. Specifically, the GPs in our sample are slightly younger, and slightly less likely to hold a specialization, than the average GP in the country.

GP swaps. Crucial to the analysis is our ability to identify whether an individual changed GP during the year, and the reason for that change. Specifically, we are interested in GP swaps that are outside the patient's control, which generate plausibly exogenous variation in the patients' new GPs. This allows us to obtain a causal estimate of doctor quality that is not driven by selective sorting of patients to doctors. To this end, we focus on GP swaps that are caused by the doctor terminating, or significantly reducing, the patient list.

We can directly identify who initiated the swap (i.e., the patient or the GP) through the register data, which contains a code for the type of swap that took place. Approximately 26

[15] An inpatient admission includes both overnight stays and day treatments, such as less invasive surgical procedures.

percent of all swaps that occur in the Norwegian primary care system are due to GPs reducing their patient lists or exiting the market. These are the swaps that we use to identify GP quality effects, as these patients are randomly reallocated to a new GP through a computer system. Appendix Figure A.1 provides a visual illustration of where in Norway new GPs are located. The figure demonstrates that there are exogenous swaps and re-assignments to new GPs across the entire country, though the more populated regions in the south experience more GP entries in absolute terms.

GP quality. Our primary measure of doctor quality is the 2-year post-assignment mortality of patients aged 55 and above.[16] To this end, we link all patients to the Norwegian cause of death register, and calculate the share of patients who are still alive two years after having been exogenously assigned to the GP.

Our decision to construct a GP VA measure based on underlying patient mortality makes an explicit assumption about the objective function of physicians and the patient utility from increased life expectancy. Specifically, it assumes that the goal of physicians is to extend the life of patients, and that patients benefit from increased life expectancy. Alternative interpretations are possible. For example, if physicians vary in their views on intensive care, or the value of early detection versus intervention, then there may be different perspectives on whether extended survival at all costs and all treatment intensities is the right goal. In addition, our measure does not account for patient discomfort from invasive procedures.

An alternative to using mortality as the VA measure would be to focus on aspects of the GP such as hospitalization rates, prescriptions, tests performed, and visits scheduled.[17] However, each of these shorter-term practice variables are ambiguous in terms of the direction of the effect of GP quality (increased hospitalization and additional tests may be reflective of a high-quality doctor who is attentive and careful, but it could also be reflective of a low-quality GP who is unable to diagnose the patient early). We therefore prefer to use mortality as our base quality measure, and we assume that lower mortality and increased life expectancy is a net

[16] We impose the age restriction because the mortality rate of Norwegians below age 55 is negligible. In addition, mortality below this age is most often characterized by accidents and injuries that could likely not have been avoided through repeated interactions with GPs. Appendix Figure A.3 illustrates this in detail.

[17] In Appendix Table A.4, we provide some such alternative measures and examine how they correlated with our mortality-based measure.

positive.

Panel A of Appendix Table A.5 provides descriptive statistics on the mortality of patients within 2 years after having been assigned to their GP, categorized by broad cause of death. The table shows that, on average, approximately 3 percent of patients aged 55 and older die within two years of exogenous assignment, with an implied standard deviation of approximately 18 percent. Thus, there is a large spread in mortality rates across patients. Among those who pass away at age 55 and older, 36 percent do so due to cancer, 26 percent do so due to cardiovascular conditions, and 11 percent do so due to respiratory illness.

Patient outcomes. To examine the effect of GP quality on non-health related patient outcomes, we follow patients across the administrative registers and collect key demographic and socioeconomic information. These data are obtained from population-wide registers, and include information on the year and month of birth, sex, immigration status, municipality of residence in each year, and highest attained educational credential. We merge these data to population-wide labor market registers. These data allow us to investigate if doctor quality has an impact on labor market outcomes such as employment status, labor market income, disability benefits, sick leave usage, and welfare support. Information on earnings is from the tax registers and information on welfare benefits comes from the social insurance database.[18]

Panel B of Appendix Table A.5 provides summary statistics on the demographics of the patients that are used to calculate the GP VA. On average, individuals in our swap sample are 64 years old, 54 percent are male, 97 percent are Norwegian-born, 67 percent are married, and the average number of years of education is 13. Comparing our sample to all adult patients involved in an exogenous swap reveals that our sample is older, slightly less educated, and more likely to be born in Norway. This is a mechanical effect caused by our decision to focus on patients aged 55 and over.

Panel C of Appendix Table A.5 provides descriptive statistics on the healthcare utilization of patients in our sample in the year prior to the exogenous swap. On average, individuals visit GPs more than six times a year. Approximately half of these visits are short consultations, and

[18]Note that the Education and Social Security records are available until 2019, while the tax/income records are available until 2018. Remember also that the two main health registries are available from 2006 (KUHR) and 2008 (NPR). This means that for the different outcomes the sample sizes will vary slightly.

about 60 percent of patients have GP visits that result in at least one lab test (e.g., blood tests). In addition, there is a 15% likelihood of having a sick leave certification provided to you by the GP in a given year.

Panel D of Appendix Table A.5 provides descriptive statistics on the labor market situation of patients in our sample.[19] On average, individuals have an annual labor income of 353,684 NOK (about $35,000), and older individuals have an annual labor income of 279,832 NOK ($28,000). About 50 percent of the patients receive some welfare transfers from the government, and the average number of sick leave days in a given year – conditional on working – is 15.[20]

Patient-generated GP ratings. To investigate if patients are able to identify the quality of GPs, we merge our administrative data on GPs with proprietary data on patient-generated ratings of GPs. We do this by taking advantage of a privately run online review platform, legelisten.no, which was launched in 2012. The website allows patients to anonymously rate their overall satisfaction with their GP on a scale from 1 (worst) to 5 (best). This is the only online platform for GP ratings available in Norway and is frequently used by patients. For example, during our analysis period, the average GP received 13 ratings per year (Appendix Table A.6).

It is not random who decides to provide a GP rating on this website, and the patients responsible for the rankings may be very different from the patients that make up our objective GP VA measure.[21] Yet, this online platform is one of very few available sources of information patients have at their disposal when trying to assess the quality of a specific GP. It is therefore not the composition of individuals who rate the GPs, but the correlation between the patient-generated ratings and the objective GP VA measure, that is of interest in our setting. This correlation will help us understand what information the patient-generated ratings convey, whether they provide accurate information about doctor quality, and to what extent they facilitate improvements in health care production.

[19] Monetary values are deflated to 2015 using the CPI (https://www.ssb.no/en/statbank/list/kpi).

[20] Social transfers include any retirement pensions, disability income, employment assessment allowance, contractual private and public pension, unemployment benefits, sick pay, parental allowance, child benefits, housing benefit, social assistance, basic and auxiliary allowance, and cash support.

[21] The data from *legelisten.no* does not allow us to identify the individual patient who provided the specific evaluation.

3.2 Method

Measure of doctor quality. To identify the effect of GP quality on patient outcomes, we exploit the random allocations of patients to new GPs in the event of GP list terminations or reductions. As GPs vary in their ability to treat patients, this randomization process allows us to break the correlation between GP quality and unobserved determinants of individual outcomes and to construct a measure of GP VA that is unrelated to any patient characteristics. We estimate the following equation:

$$h_{ijkt} = \mu_j + \theta_t + \pi_k + \chi_a + \varrho_g + \varepsilon_{ijkt} \tag{1}$$

where h_{ijkt} is a binary measure representing the mortality of patient i measured two years after exogenous assignment to GP j from GP k at time t. θ_t are year of swap fixed effects, π_k are pre-swap GP fixed effects, χ_a are age-at-assignment fixed effects, ϱ_g are sex fixed effects, and μ_j represent the exogenously-assigned GP fixed effects which is the basis of our GP VA. The thought experiment underlying our estimation approach is, therefore, to consider two patients of the same gender and age who originally had the same GP and lived in the same catchment area but then were randomly allocated to new, and different, GPs.[22] We use all exogenous swaps regardless of whether the patient stays with the new GP after the swap or decide to endogenously move to another GP. Our results should therefore be interpreted through the lens of an intention-to-treat framework.[23]

One challenge associated with estimating μ_j is sampling error because each GP has a different number of patients for which we can calculate the VA. As such, there may exist non-negligible variation in the degree of certainty associated with the VA measure across GPs. To account for such measurement error, we construct a Bayesian empirical estimator by adjusting the estimated VA according to the following equation (Kane and Staiger 2008, Chetty et al. 2014):

[22] Additionally, if an individual experienced several different GPs during the analysis period, we focus on the VA of the first exogenous GP swap.

[23] Our main specification contains a relatively limited set of controls. Adding more controls improves the precision slightly but has no impact on our results. We discuss this in the next section.

$$BE_j = \lambda_j VA_j \tag{2}$$

where the "shrinkage" factor is $\lambda_j = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2/n_j}$ and n_j is the number of patients of GP j. The term σ_u^2 represents the between-GP variance in the given outcome and σ_ε^2 is the within-GP variance in the given outcome.[24] Thus, to correct for sampling error we use the fact that we observe the full load of patients for a GP.

Another challenge associated with our identification strategy is that we require physicians to be connected to each other through the patients that they treat. Specifically, following Abowd, Kramarz, and Margolis (1999), the pre-swap GP fixed effects and the exogenously-assigned GP fixed effects are only separately identified within connected sets of GPs. These GPs would be connected by patients from each pre-swap GP having different exogenously-assigned GPs (which is part of the research design) and by exogenously-assigned GPs who receive patients from multiple pre-swap GPs. To this end, we restrict our analyses to the largest connected group that includes 99% of the patients.[25]

Identifying assumptions. The key assumption underlying our empirical design is conditional random assignment. This means that the GP-initiated patient reallocations that we exploit for identification cannot cause a systematic sorting of patients to GPs of different quality. In other words, patients who experienced the same GP exit cannot be systematically sorted across new GPs based on patient characteristics that also are correlated with their outcomes.

In theory, conditional random assignment is ensured through the legal framework governing the Norwegian healthcare system, which dictates that the Norwegian Health Economics Administration must randomly reassign patients to new local GPs conditional on municipality and availability in the event of GP list terminations and reductions. To ensure that the

[24] In practice, the parameters σ_u^2 and σ_ε^2 are replaced by their empirical estimates from our sample.

[25] To define our connected sets of GPs, we use all patients involved in exogenous GP re-assignments, which results in 23 connected sets. We only include the largest connected set in our analysis (the other 22 connect sets all have less than 10 patients in them). As a robustness exercise, we also present an alternative way of defining connected sets, using only patients who are 55 years of age or older (and are involved in exogenous GP re-assignments). This approach to defining connected sets generates a total of 248 connected sets. We show in Appendix Table A.12 (Panel B – strict mobility group) that the distribution of GP VA is similar to our baseline distribution if we restrict our sample to connected sets with more than 10 patients in them based on this alternative approach as well.

data is consistent with conditional random assignment and the legal framework of Norway, we first examine the relationship between our mortality-based VA measure and a rich set of pre-assignment characteristics of patients that may also predict their mortality risk. If there is no systematic sorting of patients across GPs that differ in their ability to treat patients, this exercise should return very small and not statistically significant coefficients.

The results from this exercise are provided in Table 1. All coefficients are small, and no coefficient is statistically significant even at the 10 percent level. This provides strong support in favor of the conditional random assignment required for interpreting our results as causal.[26] We also emphasize that sex and age fixed effects are included as controls in order to improve precision of our estimates, but that those fixed effects are not necessary for achieving balance on covariates (Appendix Table A.7). In addition to this balancing exercise, we also plot a series of binscatters that relate the GP VA to a measure of predicted mortality using the pre-assignment variables shown in Table 1 (Chetty et al. 2014). These results are provided in Appendix Figures A.4 and A.5, and they visually demonstrate balance.

In addition to the extensive balance test, we borrow insight from Abaluck et al. (2021) to provide further support for a causal interpretation of our results. Abaluck et al. (2021) leverages variation in health insurance plan choice interacted with plan market exit to examine the causal relationship between observed plan mortality rates and mortality effects. However, because patients are free to choose any plan they desire after their old plan has been removed from the market, there is a nonrandom selection issue the authors need to address. The authors accomplish this by deriving a "fallback condition" governing the subsequent choices of consumers affected by plan exits. This condition requires that the consumers, after their plans have been terminated, make similar choices from the remaining plans as new consumers in the market. This would ensure that the relationship between observed mortality and a particular plan-level unobservable are similar across terminated and non-terminated plans. If conditional random assignment is satisfied in our setting, then the Abaluck et al. (2021) fallback condition test for estimating causal effects in the event of endogenous sorting should be unproblematic

[26]Results from performing the balance exercise using all swaps are shown in Appendix Table A.8. The results suggest that the VA measure is correlated with pre-assignment patient characteristics when we do not account for selective sorting of patients to GPs during endogenous swaps. This supports our decision to focus on exogenous swaps.

to implement. The results from this exercise are provided in Appendix Table A.9 and provide additional evidence in support of our empirical design.[27]

Even if random assignment is ensured in our setting, there are two types of selection that one may in theory worry about: (1) Doctors choosing which patients to remove from their list in the event of list reductions and (2) individuals choosing to endogenously reallocate to a new GP in anticipation of a list reduction/termination. (1) would violate the Norwegian health care law and should not be possible given the institutional setting, but it remains a theoretical possibility. Importantly, neither of these sources of selection would invalidate a causal interpretation of our results, but they would impact the generalizability of our findings. Specifically, (1) would imply that our estimates are only valid for patients that GPs decide to remove from their lists, and (2) would imply that our estimates are only valid for patients who remain with the GP until the reallocation event occurs.

To examine if GPs are selectively choosing which patients to remove in the event of a list reduction (selection concern 1), we compared the health status of individuals who leave the GP due to an exogenous list reduction with the patients who are not induced to leave the GP at the time of the list reduction event. Across the health outcomes we look at, the coefficients for the differences between these two groups are very small and none of them are statistically significant even at the ten percent level (Appendix Table A.11). We interpret the result from this exercise to imply that GPs are not selectively choosing which patients to remove in the event of list reductions. This is consistent with the legal framework governing list reductions in Norway and provides additional support for our identifying assumption.

To examine if patients endogenously switch GP in anticipation of list closures and reductions (selection concern 2), we use our baseline model to estimate the probability that a patient switches GP endogenously in the current year as a function of the original GP reducing or terminating their list next year. This exercise returns a small and not economically significant point estimate of -0.002, providing strong evidence against this concern and threat to external validity.

[27] We have also implemented the balance condition test in Abaluck et al. (2021) to ensure that GP characteristics cannot predict observable differences in patient baseline characteristics. The results from this exercise are shown in Appendix Table A.10, also providing strong evidence in support of our empirical design.

Estimating impact on patients. In the next step we leverage the GP VA estimates to examine the effect of GP quality on patient outcomes. We do this by estimating the following equation:

$$w_{ijkt} = \beta\psi_j + \theta_t + \pi_k + \chi_a + \varrho_g + \varepsilon_{ijkt}, \tag{3}$$

where w_{ijkt} is an outcome of patient i measured after the exogenous assignment to GP j from GP k at time t, and ψ_j is a continuous measure of GP quality obtained from estimating the coefficients on μ_j in Equation (1). Again, χ_a represents age-at-assignment fixed effects and ϱ_g represents sex fixed effects. The coefficient β in Equation (3) captures the effect of a one SD increase in GP VA on the outcomes we examine.

To avoid a mechanical relationship between our VA measure and the patient outcomes that we investigate, ψ_j is based on a leave-one-out measure in which we exclude individual i from the VA calculation when examining the impact of GP quality on individual i's outcomes. We standardize this measure and take its negative such that a higher value means that the GP is of higher quality (fewer assigned patients dying within two years after assignment).

4 Results

4.1 GP Quality

Figure 1 shows the distribution of GP quality based on the 2-year post-assignment mortality of patients, obtained by estimating Equation (1) and plotting the predicted VA for each GP. Figure 1 illustrates that there is considerable variation in the distribution of GP quality across the primary care sector, and that the shrinkage procedure has little effect on the distribution of the GP effects that we obtain. Interestingly, the difference between the original distribution and the shrinkage-corrected distribution in our setting is extremely similar to that in Abaluck et al. (2021). Specifically, Abaluck et al. (2021) finds an average effective shrinkage coefficient close to 1, with 90% of plans having a coefficient greater than 92%. In our setting, we find an average shrinkage coefficient of 94%, with 98% of the GPs having a coefficient greater than 90%.

Based on the raw data underlying Figure 1, the standard deviation of GP value-added using the 2-year post assignment mortality rate is 0.157. A one standard deviation change in the quality of a patient's GP, therefore, is associated with a 15.7-percentage point increase in the probability of being alive two years after GP assignment.[28] In other words, there is a large and meaningful variation in GP quality with important implications for patients' health.[29]

Appendix Table A.12 provides information on the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles, across a wide range of specifications. Moving from a median-quality GP to a GP at the 75th percentile reduces the two-year mortality risk with 4.9 percentage points (approximately a 0.6 percentage point reduction per visit if we make a strong assumptions about mortality effects being linear with number of visits). Moving from a GP in the median of the VA distribution to one in the bottom quarter of the distribution, on the other hand, is associated with a two-year mortality rate increase of 6 percentage points.

To examine the sensitivity of our results to various modelling decisions and specification choices, Panel A of Appendix Table A.12 provides the key moments of the VA distribution for our baseline model shown in Figure 1, but also for a range of alternative model specifications. Specifically, (A) is the baseline moments; (B) includes municipality fixed effects; (C) includes previous GP-by-year fixed effects; (D) controls for any pre-swap chronic condition that the patient may have; and (E) defines the connected set based on individuals 55 or older, and excludes sets with less than 10 patients. While the standard deviation of GP VA varies slightly across some of these alternative specifications, the results show a clear and consistent pattern across all rows.[30]

One concern with our empirical approach is there may be statistical noise in our VA measure. We implement three robustness tests to explore this issue. First, we have constructed a

[28] Results based on a more saturated model in which we include additional controls (indicator of high school diploma before re-assignment, indicator for being born in Norway, number of children, birth other of individual, residence in a large city, total income, indicator for receiving welfare benefits, and number of GP visits two years before re-assignment) are shown in Figure A.6. We get slightly more precision when including more controls, though this difference is not economically meaningful

[29] These results are robust to using a non-linear model. Specifically, the Pearson correlation coefficient between the percentile rank of GPs base on a logit regression and based on an OLS regression is more than 87 percent.

[30] In Appendix Figure A.8, we have reestimated the GP VA distribution using all swaps rather than just focusing on the exogenous swaps. The figure illustrates that the distribution based on endogenous and exogenous swaps looks similar to the baseline density, though it has a more condensed distribution with less variation across GPs. This is consistent with the idea that selective sorting of patients to GPs generates an attenuation bias and pushes the VA towards zero across most GPs.

scatterplot of the GP VA estimates against the number of patients received by the GP to ensure that GPs with few patients are not clustered in either tail of the VA distribution. Such clustering would be indicative of the VA estimates reflecting statistical noise. Second, we have estimated the GP VA using different restrictions on the minimum number of patients a GP must have. If there are considerable changes to the GP VA as we alter the restrictions on the number of patients a GP must have, it would suggest that there may be statistical noise in the estimates. Third, we have examined how the GP VA estimates correlate with VA estimates that use a shorter or longer time horizon for post-swap mortality. If the within-GP VA correlation across shorter and longer time horizons is low, that provides suggestive evidence of statistical noise driving the results. The results from these supplemental exercises are shown in Appendix Figure A.7, Panel B of Appendix Table A.11, Panel C of Appendix Table A.12, and Appendix Table A.13. These results suggest that there is no noticeable issue with statistical noise in the distribution of physician effects.

To study if the variation in GP quality differs across specific subgroups of physicians, Appendix Figure A.9 and Appendix Table A.14 provide detailed information on the variation in physician quality across male and female GPs, old and young GPs, GPs with a light and a heavy patient load, and specialists and non-specialists. Interestingly, the variation in physician quality across these groups is noticeably different. For example, while the average quality of female physicians is considerably higher than the average quality of male physicians, the standard deviation is considerably larger as well. Thus, female physicians are more likely to be of very high-quality compared to male physicians, but they are also considerably more likely to be of very low-quality. The same pattern applies to old and young doctors as well as specialists and non-specialists. These results suggest that certain types of doctors provide higher-quality care on average, but that they also are associated with increased risk in terms of being located in the tail of the quality distribution.

The results presented and discussed up to this point can be viewed as the extensive margin effect of GP quality with respect to mortality: assignment to a high-quality GP leads to a substantial reduction in a patient's post-assignment mortality risk. However, there may be an important intensive margin effect associated with GP quality as well. Specifically, conditional

on dying, do patients who are randomly assigned to higher quality GPs die later? To address this question, we leverage the exogenous variation in patient-GP assignment to look at the relationship between age of death and GP quality among those who pass away before the end of our analysis period after being assigned (exogenously) to a new GP.[31] Panel A of Table 2 demonstrates that exogenous assignment to a high-quality GP has a substantial impact on a patient's intensive margin survival length. Specifically, being assigned to a GP whose VA estimate is one SD higher in the VA distribution increases the age of death with 4.14 months.

To put our result in perspective, Finkelstein et al. (2021) find that moving from the twenty-fifth to the seventy-fifth percentile mortality location in the US increases life expectancy at age 65 by 7 months. Our results imply that a SD change in GP quality thus generates about half the health effect, in terms of mortality, of moving from the bottom quartile to the top quartile location in the US. Nevertheless, caution should be exercised when comparing our results to those in Finkelstein et al. (2021), a paper which focuses on a population that is relatively sicker than the one examined in our analysis (65 and over US Medicare patients) and which finds that climate, pollution, homicides, urbanicity, and health behaviors (obesity, smoking) are all strong place-based predictors of mortality (while we focus exclusively on GP quality). That a SD change in GP quality in our setting can generate about half the health effect, in terms of mortality, of moving from the bottom quartile to the top quartile location in the US we believe can be explained by two factors. First, GPs play a much greater role in the health care of individuals in Norway relative to the US.[32] Second, Norway is a small country with low levels of, and little cross-municipality variation in, pollution, climate, homicides and health behaviors. This is also very different compared to the US, where these factors are much more varied across different parts of the country. Thus, these factors should play a much smaller role in explaining place-based mortality variation in Norway than in the US. [33]

[31] We condition on dying at some point during our analysis period as age of death would be undefined otherwise.

[32] As noted in Section 2, GPs represent the first point of contact between patients and the healthcare system in Norway, and are responsible for initial examination, treatment, diagnosis, medication prescription, and sick note validation. Should the GP consider it necessary that the patient receives specialist care, the GP will refer the patient to a specialist. There is no other way for patients to gain access to specialist care (except through emergency admission in the case of life-threatening health issues). This stands in stark contrast to the health care system in the US, in which there is considerably more flexibility in terms of which care to use, when, and where. Thus, it should be expected that the effect of GPs are large in our setting.

[33] It should also be noted that our measure of age conditional on death is not exactly the same as life expectancy at 65 used in Finkelstein et al. (2021). Specifically, Finkelstein et al. (2021) follow Chetty et al. (2016) and adopt

An interesting question is whether the effects we identify differ depending on the age of the patient at the time of the exogenous reassignment. To this end, we have estimated our main specification stratified by the mean age among the patients in our sample at the time of the exogenous reassignment (age 62). This analysis is interesting for two reasons. First, mortality due to natural causes and non-cancer conditions become more common at older ages, suggesting that the marginal impact that the GP may have on patients is smaller. Second, the younger age range of the patients included in our sample, 55-62, is the age range at which mortality begins to increase in Norway (Appendix Figure A.3), and individuals may not have had time to internalize the increased mortality risk by pursuing additional health checkups and screenings. Thus, this is an age at which patients may be particularly susceptible to the quality of the GP.

The results from this analysis are shown in Appendix Table A.15, both with respect to age at death as well as with respect to malign cancers. The results reveal a considerably smaller effect among patients aged 62 and older with respect to age at death, and a slightly smaller effect among patients aged 62 and older with respect to malign cancers. However, we are unable to reject that the estimates are the same across the two subgroups.

4.2 Predictors of GP Quality

GP Demographics. Having identified substantial variation in GP quality across the primary care sector in Norway, we ask whether observable GP demographic characteristics can help predict the variation in GP quality. Column (1) of Table 3 shows the explanatory power obtained from estimating a regression with the predicted GP value-added measure on the left-hand side, and a battery of GP demographic characteristics on the right-hand side: age, gender, specialization status, group practice indicators (whether the GP is part of a group practice), list

a Gompertz specification in which the log of the mortality hazard rate $m_{ij}(a)$ that individual i would experience at age a if she lived in location j is linear in age. Translating their setup to ours, this would imply that the "old" and "new" GPs in our setting would affect the level of mortality multiplicatively, while we assume a linear relation. While we acknowledge that this is an assumption, an important difference of our setup to that in Finkelstein et al. (2016) is that we model mortality in a short interval after the GP reassignment. We also estimate effects separately for different windows of length after the GP reassignment, which provides a check on the robustness of our linearity assumption. We further replace the main estimating equation with a binary logit model of two-years mortality, and show that the Pearson correlation coefficient between the percentile rank of GPs base on a logit regression and based on an OLS regression is more than 87 percent.

length (number of patients that are assigned to the GP), and shared list status (whether the GP shares the patient list with another GP).

The results provide little support for the notion that a GP's demographic characteristics can identify their VA, with each model having a very modest R^2. Having said that, two characteristics are found to be statistically significantly associated with GP quality: age and list length.

First, a 10-year increase in GP age is associated with a reduction in GP VA by five percent of a standard deviation. The magnitude of this effect is relatively small, but the finding is consistent with the descriptive evidence in Tsugawa et al. (2017). While we are unable to disentangle the reasons underlying this age-quality gradient, we note that prior studies have speculated that a quality decline in doctor ability by age could be driven by three different mechanisms: (1) a difficulty associated with keeping up with scientific and technological advancement over time, (2) differences in how younger and older physicians were trained, and (3) older physicians having acquired more specialized knowledge in their relevant fields but having forgotten more general knowledge in other fields. Interestingly, this result stands in relatively stark contrast to the effect of age in other occupations (such as teaching), where there appears to be a positive age-quality gradient for the first few years after which it flattens out.

Second, having an additional 100 patients is associated with an increase in the GP VA by 2 percent of a SD.[34] While speculative, we believe that this is reflective of high-quality GPs having a larger number of patients. We explore this in more detail below.

Taken together, the results in column (1) of Table 3 provide relatively little support for the notion that a GP's underlying demographic characteristics can help identify their VA. Despite statistically significant associations between certain demographic variables and GP VA, the R^2 remains less than 0.01 across all of these specifications. Almost all of the variation in GP quality is, therefore, coming from sources other than these demographics.

GP Behavior and Practice Style. In addition to observable demographic GP characteristics, we also have information on GP behaviors and practice styles. To this end, column (2) of Table 3 reports results from examining the relationship between GP consultation behavior and GP

[34]Longer lists are associated with urban settings, so this larger VA might also be capturing the doctor's learning and effectiveness due to his/her proximity to local health infrastructure such as university hospitals.

VA.

Overall, we find little evidence that GP behavior and practice style can be used to predict GP VA. Specifically, none of the variables display predictive power over GP VA in terms of statistical significance, and the magnitude of the point estimates are relatively small. We conclude that, similar to the GP demographic characteristics, it is not possible to infer GP VA based on the GPs observable practice behaviors. Specifically, accounting for all demographic information of the GPs, as well as all the observed practice styles and behaviors, explains less than 1 percent of the variation in GP VA (column (3)).

4.3 Effects on Patients

The results in Sections 4.1 and 4.2 demonstrate that there is substantial variation in physician quality in Norway, suggesting an important role for targeted policy interventions based on a GPs VA. To understand the policy implications of our findings in greater detail, we take a step back and ask how important GP quality is for patients' future outcomes.

Access to care and health care utilization. Tables 4 and 5 explore extensive and intensive margin effects of GP VA on GP treatment (Table 4) and hospitalization (Table 5). Each of the coefficients displayed in these tables show the effect of a one standard deviation change in GP VA on the outcome depicted at the top of the column. Note that the impact of GP VA on mortality identified above automatically leads to a change in the composition of patients that are used to identify these effects across the GP VA distribution. To reduce the impact of these compositional effects on our estimates, we examine medical treatment and hospitalization outcomes in the year immediately following the swap, when most patients in our sample are still alive.

The results in Table 4 indicate that patients visit their GPs on average seven times a year (or once every seven weeks). Patients assigned to a high-quality GP do receive a slightly smaller number of consultations per year (0.01), suggesting that high-VA GPs may be more efficient during the consultations and better at treating patients such that fewer follow-up consultations are necessary. That high-quality GPs may be more effective, or more able to immediately

assign the right procedure to the patient, is further supported by the finding that the cost (reimbursements) associated with high-quality GPs is lower than the cost associated with low-quality GPs. Reimbursements in this setting are a function of the tasks performed by the GP during the patient visit, as well as the time that the GP spends with the patient.

With respect to hospitalization effects, Table 5 provides clear evidence of a reduction in urgent ("acute") care as a consequence of being assigned a high-quality GP. Specifically, a 1 SD increase in GP VA is associated with a 2.4 percentage point reduction in seeking urgent care in the year after the first consultation with the new GP. At the same time, we find no reduction in overall hospitalization rates. This result is consistent with the notion that high quality GPs may be effective at early detection and in identifying the right type of treatment required by the patient, preventing the need for urgent care and emergency treatment. The result helps corroborate the discussions above, in which we suggest that variation across GPs in the primary care sector is coming from differences in the actual interaction between the GP and the patient during the patient visit, and the GPs ability to perform the right task for the specific patient.

The results in Tables 4 and 5 suggest that GP VA may not be driven by differences in the quantity of care, but rather by the quality, or effectiveness, of the care that GPs provide. An additional way of exploring this hypothesis is to examine what causes of death are driving the mortality reductions associated with high quality GPs. Results from such an exercise should show which causes of death that can benefit from early detection and treatment are driving the GP VA mortality effects. To this end, Panel B of Table 2 provides results from estimating equation (1) with cause of death as the dependent variable. The results from this exercise are consistent with the above hypothesis, demonstrating that the mortality effect induced by GP VA loads almost exclusively on cancer detection.

That higher quality GPs have a substantial impact on cancer mortality but no effect on cardiovascular mortality is an interesting finding, but perhaps not surprising. First, there are standardized procedures for testing general cardiovascular conditions and diseases at each GP appointment (e.g., blood pressure and heart rate), and these conditions are on average easier

to detect.[35] These procedures are unlikely to vary substantially across physicians.[36] Second, individuals experiencing severe cardiovascular issues (e.g., heart attack, pulmonary embolism, heart failure, arrhythmia) would likely be seeking urgent care and/or be taken to the hospital immediately, such that the role of the primary care physician would be relatively minor. Finally, there are relatively well-developed treatments for a large number of cardiovascular diseases (e.g., statins) which could make the role of the GP less crucial for the ultimate outcome of the patient compared to if the patient had an underlying malignant cancer.[37]

To better understand the mechanisms behind the cancer reduction effect that we identify, we explore which primary care diagnoses are associated with a malign cancer diagnosis. The results from this exercise can shed light on the type of GP behavior and practice style that drive the early cancer detection effects that we find.

To perform this analysis, we rely on a machine learning lasso procedure. Specifically, we take a random 50 percent sample of individuals aged 55 and older that visit GPs and that are not involved in an exogeneous re-assignment. In other words, these individuals are not included in our core analysis sample. Then, for these individuals, we take all groups of symptoms and medical diagnoses from the primary care services in year t and relate them to an indicator for a malign cancer recorded at a specialist care unit in year $t + 1$ (controlling for age at visit and year of visit). Using the letter and first digit from the ICPC-2 classification system, the lasso procedure identifies the following diagnoses/symptoms as those with the greatest cancer predictions: X7, Y7, D7, T7, and A7.

After having obtained predictions from the machine learning lasso procedure, we turn to our analysis sample (which was not used in the Lasso procedure) and relate our VA measure to these ICPC-2 codes through Equation (3) in the paper. The results from this exercise are shown

[35] See, for example, https://edition.cnn.com/2019/09/03/health/leading-cause-of-death-cancer-heart-disease-study/index.html

[36] While the physicians' recommendations could vary depending on the cardiovascular test results, there are standardized thresholds at which physicians are encouraged to perform specific procedures and referrals. We therefore do not believe that variation on this dimension would be particularly large.

[37] To explore the role of cancer in greater detail, we examine the reason patients visit specialists during the year following an exogenous assignment to a GP in Table 6. There is a relatively sizable reduction on the intensive margin of cancer-related GP visits. We interpret these findings as further supporting the idea that GP quality is driven by early detection and the ability of GPs to provide and identify the right treatment for the patient. Specifically, the results are consistent with the idea that a good GP is able to identify and refer a patient to the relevant specialist early on, ensuring a less intensive treatment with a higher success rate.

in Appendix Table A.16. These results provide suggestive evidence of the mechanism through which our cancer detection effect operates. Specifically, a high-quality GP appears more capable of identifying issues related to digestive diseases (gastrointestinal, colon, pancreas, stomach, etc.); diseases that are often hard to identify, that are not covered by the national cancer prevention programs, and that requires detection skills that the average GP may not possess. While these results are only suggestive of a potential mechanism through which our early cancer detection effect operates, we believe that it is an important finding that helps interpret the core results in our paper.

Labor market and welfare effects. If GP quality impacts the health of patients, and if positive health effects translate into improved labor market outcomes, then GP quality also may impact labor market outcomes. To this end, we explore the impact of GP quality on a range of labor market variables related to employment, earnings, hours worked, and welfare usage. When interpreting these results, it is important to bear in mind that the individuals in our analysis sample are aged 55 and more. Thus, if it takes time for health improvements to generate positive labor market effects, or if the relationship is stronger in the beginning or at the peak of individuals' careers, we would not observe it in our analysis.

Table 7 provides results from estimation of Equation (3) using a range of labor market outcomes as dependent variables. In Panel (A), we present results for our full sample. In Panel (B) we present results for a subsample of individuals who are at the lower end of the age distribution in our sample and therefore more likely to have stronger attachments to the labor market. Looking across the table, we find little evidence that GP quality has an impact on patient labor market outcomes as it relates to employment, earnings, or hours worked. However, we do see a statistically significant decline in the probability of receiving welfare support from the government. We hypothesize that this is due to improved care and early detection, enabling the patient to respond quicker and therefore not suffer any longer-term health complications that would generate large increases in welfare expenditure.

To better understand the reduced welfare probability effect, we estimate binarized versions of the largest components of the social welfare system for the younger sample (for whom we find a reduction in welfare support): unemployment benefits, sick leave, disability insurance,

and retirement (early retirement as well as regular old-age pension payments). The results are shown in Appendix Table A.17 and reveal that the reduction in welfare support is not driven by sick leave, unemployment benefits, or disability insurance – the three largest components of the national social welfare system. However, we do find a substantial reduction in AFP pension take-up. AFP is an early retirement scheme for people who have reached the age of 62 and are employed by an employer covered by a collective bargaining agreement that recognizes the AFP scheme. Currently, the entire public sector, and approximately 50 percent of the private sector, are members of the AFP scheme. This result is consistent with the idea that assignment to a higher quality GP generates positive health effects that enable workers to remain in the workforce for longer. While the result is suggestive, it sheds light on potential downstream effects of GP quality.

Finally, it is important to acknowledge that the labor market outcomes can only be estimated based on the individuals who are still alive. To better account for this dimension of GP effects, we have conducted a bounding exercise in which we assign the deceased to the 25^{th}, 50^{th}, and 75^{th} percentile earnings and see if including them in the sample with those imputed earnings generate estimated effects that deviate from main estimate when they are not included. The results from this exercise are shown in Panel A of Appendix Table A.18 and suggest that the omission of diseased individuals from these regressions have a minimal impact on the findings we present.[38]

4.4 Patients' Ability to Identify High-quality GPs

In this section, we advance the literature on patient choice by providing the first estimates on individual's ability to identify, and select, high-quality GPs. We do this by exploiting the fact that individuals are allowed to independently change GP twice per year and examining if GPs with a higher VA are more likely to experience an influx of endogenous patient switchers.

Table 8 shows that high-quality GPs do not receive a disproportionate influx of new patients.

[38]To examine if the results from this bounding exercise extend to another outcome that we examine in the paper, we have performed similar exercises for welfare transfers. These results are presented in Panels B of Appendix Table A.18. They differ slightly from the income exercise since this variable is binary. As such, we assign deceased individuals the lowest (zero) and highest (one) possible value for transfers to derive the bounds. Our core welfare transfer estimates in Table 7 are closer to the upper bound estimate. However, the results suggest that the omission of diseased individuals from our analysis have a minimal impact on the findings we present.

The lack of a strong correlation on this dimension could be driven by two distinct factors: An inability among patients to identify high-quality GPs or a desire to choose GPs based on some other metric. In column (2), we therefore use information on GP practice style and behavior to explore other potential GP attributes that could drive patients' endogenous switching behavior. The results show that patients who endogenously switch do so to get more popular and experienced GPs who are female and do hold a specialization. This sorting behavior could be explained by patients associating these GP characteristics with GP quality, or because patients independently value these GP characteristics. While our current analysis does not allow us to separately identify these two channels, we see it as a great avenue for future research on the topic.

The findings in Table 8 have important implications for ongoing policy debates on the organization of health care services and the distribution of information to patients. We view these findings as paving the way for a new strand of research, exploring whether the lack of a quality-driven switching behavior is due to information asymmetry (i.e., patients are unable to identify good GPs) or because patients value other aspects of a GP more.

4.5 Patient-generated GP Ratings and GP VA

To better understand patients' ability to identify high-quality GPs as measured by their ability to keep patients alive, we merge our administrative data on GPs with proprietary data on patient-generated ratings of GPs as discussed in Section 3. We then correlate our mortality-based measure of GP VA with the patient-generated ratings (which range from 1 to 5) of these GPs.

It is not random who decides to provide a GP rating on this website, and the patients responsible for the rankings may be very different from the patients that make up our objective GP VA measure. Yet, this online platform is one of very few available sources of information that patients have at their disposal when trying to assess the ability of a specific GP, and it is therefore not the composition of individuals who rate the GPs, but the correlation between the patient-generated ratings and the objective GP VA measure, that is of interest for understanding what information the patient-generated ratings convey, whether they provide true information, and to what extent they facilitate improvements in health care production.

The results from this exercise are shown in Panel A of Table 9 and reveal a small and not statistically significant correlation between patient-generated ratings of GPs and the GP VA measure we construct. At the same time, the results in Panel B of Table 9 demonstrate that the patient-generated ratings are correlated with a number of GP characteristics, including age, specialization status, list length, and gender.

While the patient ratings are uncorrelated with GP VA (Table 9), we show in Table 10 that GPs do receive a disproportionate influx of new endogenous patient swaps if they have a higher subjective patient-generated rating, or higher proportion of top scores on the subjective patient ratings.[39] We further show that GPs receive a smaller influx of endogenous patient swaps if they have a higher proportion of bottom scores on the subjective patient ratings. The patient-generated ratings are publicly available and likely more salient than the objective GP VA measure, and it is therefore not surprising that highly rated GPs are more likely to be chosen by patients. However, given the lack of correlation between the patient-generated ratings and the objective measure of GP quality that we construct, this has important policy implications. Specifically, recent years have seen a large increase in the availability of patient-generated online rating platforms. If these platforms are used by patients to make health care choices, but the rankings on those platforms are not associated with GP quality, then they do not necessarily facilitate improvements in health care production.

5 Policy Analysis

To quantify the value of improving the quality of GPs in the country, we adopt Hanushek's (2009) proposal to replace the worst performing GPs (bottom 5 percent of the VA distribution) with GPs of average quality. We estimate that this would generate a 12.1-month increase in the life expectancy of the patients who would otherwise encounter the low-quality GPs. To translate this increase in life expectancy into a monetary amount, we need a monetary estimate of the value of an additional life year. The methodologies used to calculate such values differ greatly across countries and agencies, and this has an implication for the estimate of total

[39] Using alternative GP quality measures based on e.g. hospitalization rates and sick leave also does not show any strong correlation (Appendix Table A.19).

social benefit we eventually derive. For example, the Norwegian government, operating via the recommendation of the EU commission, estimates the value of a statistical life year (VLSY) to be between USD 35,000 and USD 50,000. The US FDA, on the other hand, estimates a statistical life year to be worth as much as USD 369,000. Using the Norwegian VLSY, we calculate that the social benefit of replacing the bottom five percent of GPs in the quality distribution with GPs of average quality is $35,500 per patient, $11.7 million per GP, or $1.2 billion in total.[40] Using the US FDA guidelines, we calculate that the social benefit is $372,075 per patient, $123 million per GP, or $12.4 billion in total.

At the same time as we are finding positive social benefits, our results suggest that the practice styles and behaviors of higher-quality GPs are associated with a lower per-patient cost and that high-quality GPs reduce the fiscal burden on the state by lowering welfare transfers to individuals. This implies that substantial investments in GP quality through training and retraining may represent cost-effective ways to reduce mortality and improve general health. However, this depends on the cost of raising a GP's VA; we see this as a valuable avenue for future research.

6 Discussion

We exploit rich population-wide register data coupled with exogenous assignment of patients to general practitioners to estimate doctor value-added and provide novel evidence on the distribution of physician quality across an entire country, who the high-quality doctors are, what potential mechanisms can help explain the quality variation across doctors, and whether patients are able to identify the good doctors. In addition, the paper produces one of the first causal estimates of the impact of GP quality on patients' long-term health and labor market

[40]For individuals who are exogenously swapped at age 55 or older, the effect we estimate is equivalent to 4.14 extra months of life. The average quality of GPs at the bottom 5 percent of the distribution is 2.94 SD lower than the median, such that the effect on extra months of life among GPs at this part of the distribution is 12.1. Using $35,000 as the predicted value of an additional life year in Norway, this implies that the social benefit associated with each individual affected by this policy is $35,500. Each GP has on average 1100 patients, and there are 5800 GPs in the country. However, our analysis is restricted to patients over the age of 55, and to GPs who receive exogenous swaps. About 30 percent of patients are above 55, and the number of GPs in our analysis is 2023. Thus, the social benefit of replacing the bottom five percent of GPs in the quality distribution with GPs of average quality in our analysis is $35,500 per patient (35,000 * [12.1/12]), $11.7 million per GP ([1100*0.3] * 35,500) or $1.2 billion in total ([2023*0.05] * [1100*0.3] * 35,500).

outcomes.

We show that there is substantial variation in the quality of physicians in the primary care sector. While we find evidence of certain observable GP demographic characteristics and practice styles being able to predict GP VA, a standard decomposition exercise reveals that most of the quality variation is driven by unobserved ability differences across doctors. In terms of mechanisms, we show that patients assigned to high-quality GPs do not receive a different quantity of treatment, and that the VA variation across GPs is primarily reflecting differences in the GP's ability to engage in early prevention and assign the right procedure to the patient. Finally, we show that patients are unable, or unwilling, to identify who the high-quality doctors are by focusing on the switching behavior of patients across GPs.

In terms of policy implications, knowing the impact of physicians on their patients is important not only for establishing effective doctor compensation schemes and retention policies, but also for understanding the relative benefit of focusing on – and investing in – physician quality relative to other inputs that enter the healthcare production function. Applying Hanushek's (2009) canonical proposal to replace the worst performing GPs with GPs of average quality, we find that this policy would generate a total social gain of $1.2 billion (if using the Norwegian government's VSLY) or $12.4 billion (if using the VSLY of the US FDA). At the same time, our results suggest that the observed practice styles and behaviors of higher-quality GPs are not associated with a higher per-patient cost and that there are fiscal benefits to the state from having high-quality GPs. This suggests that substantial investment in GP quality through training and retraining may represent cost-effective ways to reduce mortality and improve general health.

In terms of extrapolating our results to countries outside of Norway, it is worth noting that the quality of medical school in Norway is relatively constant across the country, mainly due to the highly restricted and small set of medical schools in the country coupled with a strict national set of program regulations. In addition, it is useful to bear in mind that Norway ranks as one of the most equal countries in the world and operates a universal welfare system in which individuals are guaranteed free access to health care, substantial social support, and certain minimum living conditions. Both these features of Norway stand in stark contrast to

many other countries, in which there is greater variation in the quality of medical schools as well as substantially larger differences in living conditions and socioeconomic situations across individuals. This suggests that the distribution of GP VA we identify may be interpreted as a lower bound for other countries.

Independent of the social benefit discussed above, knowing the extent of variation in doctor quality helps us better understand inequalities in health outcomes across individuals, especially if specific subgroups of individuals have access to differently qualified doctors. We therefore see our results as not only opening up a new avenue of research through which we can build a better understanding of the relative role of doctor quality in the production of health, but also as opening up a new avenue of research in which we can better understand the relationship between doctor quality – and access to differentially-qualified doctors – and social goals such as equality and efficiency.

Data Availability

Code replicating the tables and figures in this article can be found in Ginja et al. (2024) in the Harvard Dataverse, https://doi.org/10.7910/DVN/7YXGYB.

References

[1] Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc. 2021. "Mortality effects and choice across private health insurance plans." Quarterly Journal of Economics, 136 (3): 1557-1610.

[2] Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High wage workers and high wage firms." Econometrica, 67 (2): 251-333.

[3] Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc. 2011. "Do value-added estimates add value? Accounting for learning dynamics." American Economic Journal: Applied Economics, 3 (3): 29-54.

[4] Angerer, Silvia, Christian Waibel, and Harald Stummer. 2019. "Discrimination in health care: a field experiment on the impact of patients' socioeconomic status on access to care." American Journal of Health Economics, 5 (4): 407-427.

[5] Bartel, Ann P., Nancy D. Beaulieu, Ciaran S. Phibbs, and Patricia W. Stone. 2014. "Human capital and productivity in a team environment: evidence from the healthcare sector." American Economic Journal: Applied Economics, 6 (2): 231-59.

[6] Behncke, Stefanie, Markus Frölich, and Michael Lechner. 2010. "A caseworker like me–Does the similarity between the unemployed and their caseworkers increase job placements?" Economic Journal, 120 (549): 1430-1459.

[7] Bensnes, Simon, and Ingrid Huitfeldt. 2021. "Rumor has it: How do patients respond to patient-generated physician ratings?" Journal of Health Economics, 76.

[8] Boerma, Wienke G. W., J. van der Zee, and Douglas M. Fleming. 1997. "Service profiles of general practitioners in Europe. European GP Task Profile Study." British Journal of General Practice, 47 (421): 481-6.

[9] Chan, Alex. 2022. "Discrimination Against Doctors: A Field Experiment." Mimeo.

[10] Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." American Economic Review, 104 (9): 2633-79.

[11] Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. "The association between income and life expectancy in the United States, 2001-2014." JAMA, 315 (16): 1750-1766.

[12] Clemens, Jeffrey, and Joshua D. Gottlieb. 2014. "Do physicians' financial incentives affect medical treatment and patient health?" American Economic Review, 104 (4): 1320-49.

[13] Cutler, David M., Robert S. Huckman, and Mary Beth Landrum. 2004. "The role of information in medical markets: an analysis of publicly reported outcomes in cardiac surgery." American Economic Review, 94 (2): 342-346.

[14] Currie, Janet, W. Bentley MacLeod, and Jessica Van Parys. 2016. "Provider practice style and patient health outcomes: The case of heart attacks." Journal of Health Economics, 47: 64-80.

[15] Currie, Janet and Jonathan Zhang. 2021. "Doing More with Less: Predicting Primary Care Provider Effectiveness", The Review of Economics and Statistics 2023; doi: https://doi.org/10.1162/rest_a_01290

[16] Dixon, Paul, Hugh Gravelle, Roy Carr-Hill, and John Posnett. 1997. "Patient Movements and Patient Choice." Report for National Health Service Executive, December 1997. Available at: http://www.york.ac.uk/media/che/documents/papers/Report%201997%20Patient%20Movement.pdf

[17] Doyle Jr, Joseph J., Steven M. Ewer, and Todd H. Wagner. 2010. "Returns to physician human capital: Evidence from patients randomized to physician teams." Journal of Health Economics, 29 (6): 866-882.

[18] Elvik, Rune. 2018. "The value of life: The rise and fall of a scientific research programme." Cambridge Scholars Publishing.

[19] Fadlon, Itzik, Jessica Van Parys. 2020. "Primary care physician practice styles and patient care: Evidence from physician exits in Medicare." Journal of Health Economics, 71: 102304

[20] Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of geographic variation in health care: Evidence from patient migration." Quarterly Journal of Economics, 131 (4): 1681-1726.

[21] Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2021. "Place-based drivers of mortality: Evidence from migration." American Economic Review, 111 (8): 2697-2735.

[22] Fletcher, Jason M., Leora I. Horwitz, and Elizabeth Bradley. 2014. "Estimating the value added of attending physicians on patient outcomes." NBER Working Paper No. w20534.

[23] Geruso, M., Danil Agafiev Macambira, Anthony Lollo, Chima Ndumele, and Jacob Wallace. 2021. "The Private Provision of Public Services: Evidence from Random Assignment in Medicaid." Mimeo.

[24] Ginja, Rita, Julie Riise, Barton Willage, and Alexander Willén. 2024. "Replication Data for: 'Does your Doctor Matter'," Harvard Dataverse, https://doi.org/10.7910/DVN/7YXGYB.

[25] Godager, Geir. 2012. "Birds of a feather flock together: A study of doctor–patient matching." Journal of Health Economics, 31 (1): 296-305.

[26] Hanushek, Eric A.. 1992. "The Trade-off Between Child Quantity and Quality." Journal of Political Economy, 100 (1): 84-117.

[27] Hanushek, Eric A.. 2009. "Teacher Deselection." In Creating a New Teaching Profession, edited by Dan Goldhaber and Jane Hannaway, 165–80. Washington, DC: Urban Institute Press.

[28] Hanushek, Eric A., John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin. 2005. "The Market for Teacher Quality" NBER Working Paper No. w11154.

[29] Ho, Katherine. 2006. "The welfare effects of restricted hospital choice in the US medical care market." Journal of Applied Econometrics, 21 (7): 1039-1079.

[30] Howard, David H. 2006. "Quality and Consumer Choice in Healthcare: Evidence from Kidney Transplantation." Topics in Economic Analysis and Policy, 5 (1): 1349.

[31] Kane, Thomas J., and Douglas O. Staiger. 2008. "Estimating teacher impacts on student achievement: An experimental evaluation." NBER Working Paper No. w14607.

[32] Luca, Michael, and Sonal Vats. 2013. "Digitizing doctor demand: The impact of online reviews on doctor choice." Mimeo.

[33] McClellan, Mark, Barbara J. McNeil, and Joseph P. Newhouse. 1994. "Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables." JAMA, 272 (11): 859-866.

[34] McLean, Gary, and Matt Sutton. 2005. "Competition and patient choice." Mimeo.

[35] Pope, Devin G. 2009. "Reacting to rankings: evidence from 'America's Best Hospitals'." Journal of Health Economics, 28 (6): 1154-1165.

[36] Riise, Julie, Barton Willage, and Alexander Willén. 2022. “Can Female Doctors Cure the Gender STEMM Gap? Evidence from Randomly Assigned GPs.” Review of Economics and Statistics, 104 (4): 621-635.

[37] Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. ”Teachers, schools, and academic achievement.” Econometrica, 73 (2): 417-458.

[38] Rockoff, Jonah E. 2004. ”The impact of individual teachers on student achievement: Evidence from panel data.” American Economic Review, 94 (2): 247-252.

[39] Santos, Rita, Hugh Gravelle, and Carol Propper. 2017. ”Does quality affect patients’ choice of doctor? Evidence from England.” Economic Journal, 127 (600): 445-494.

[40] Schnell, Molly, and Janet Currie. 2018. ”Addressing the opioid epidemic: is there a role for physician education?” American Journal of Health Economics, 4 (3): 383-410.

[41] Simonsen, Marianne, Lars Skipper, Niels Skipper, and Peter Rønø Thingholm. 2021. ”Discontinuity in care: Practice closures among primary care providers and patient health care utilization.” Journal of Health Economics, 80: 102551.

[42] Stoye, George. 2022. “The Distribution of Physician Quality in England.” Mimeo.

[43] Tsugawa, Yusuke, Anupam B. Jena, Jose F. Figueroa, E. John Orav, Daniel M. Blumenthal, and Ashish K. Jha. 2017. ”Comparison of hospital mortality and readmission rates for Medicare patients treated by male vs female physicians.” JAMA Internal Medicine, 177 (2): 206-213.

[44] Zhang, Xuan. 2022. “The effects of physician retirement on patient outcomes: anticipation and disruption.” Journal of Public Economics, 207: 104603.

Tables and Figures

Table 1: Relation Between GPs' VA and Baseline Characteristics of Patients: Sample of Patients Involved in Exogenous GP swaps

	(1) College Degree	(2) Married	(3) Born in Norway	(4) Nb. Children	(5) Total Income	(6) Hours	(7) Any Welfare Benefits	(8) Sick Leave	(9) Nb. GP Visits	(10) Reimbur-sements	(11) Any Chronic Conditions	(12) Nb. Chronic Conditions	(13) Any Hospital.	(14) Nb. Hospital.	(15) Charlson Index
Panel A: Restricted Sample															
Standardized VA	-0.003	-0.007	-0.002	-0.043	10.158	0.414	-0.007	-1.027	-0.347	-15.080	-0.003	-0.112	0.017	0.044	-0.026
	(0.011)	(0.012)	(0.003)	(0.032)	(8.470)	(0.343)	(0.010)	(1.764)	(0.284)	(43.602)	(0.013)	(0.227)	(0.013)	(0.098)	(0.017)
Mean of Depend. Var.	.246	.663	.967	2.15	468	15.6	.579	14.5	6.73	809	.443	3.5	.407	1.6	.149
N	54179	54179	54179	54179	54179	54179	54179	54179	54179	54179	54179	54179	54179	54179	54179
Joint F-Test	1.120														
p-value	.333														
Panel B: Unrestricted Sample															
Standardized VA	0.003	-0.001	-0.002	-0.039	8.152	-0.001	0.001	-1.655	-0.183	3.961	0.000	-0.021			
	(0.010)	(0.011)	(0.003)	(0.027)	(7.082)	(0.306)	(0.009)	(1.467)	(0.236)	(35.639)	(0.011)	(0.191)			
Mean of Depend. Var.	.232	.664	.968	2.15	449	15.2	.575	14.7	6.49	751	.432	3.43			
N	71419	71419	71419	71419	71419	71419	71419	71419	71419	71419	71419	71419			
Joint F-Test	.649														
p-value	.801														

Note: The dependent variables are characteristics of patients measured two years prior to the exogenous assignment to a new GP. Controls included in model but excluded from table are fixed effects for gender, the year of swap, age at swap, and the previous GP. The measure of VA is re-scaled such that a higher value means that the GP is of higher quality (fewer assigned patients dying within two years after assignment). Standard errors in parentheses clustered at the new GP level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 2: Impact on Mortality 2 Years After Swap: Age of Death and Causes of Death

	(1)	(2)	(3)	(4)	(5)
	Panel A: Age of Death				
	Age at Death	Additional Controls	Baseline Model		
Standardized VA	0.345***	0.200*	0.200*		
	(0.113)	(0.112)	(0.111)		
Mean of Depend. Var.	76.8				
N	13695	9061	9061		
	Panel B: Cause of Death				
	All Causes Mortality	Cancer ICD10 C	Cardiovascular Conditions ICD10 I	Respiratory Conditions ICD10 J	All Other
Standardized VA	-0.014*	-0.011**	0.001	-0.003	-0.002
	(0.008)	(0.005)	(0.003)	(0.002)	(0.005)
Mean of Depend. Var.	.031	.011	.008	.003	.009
N	90412	90412	90412	90412	90412

Note: The table present OLS estimates where the dependent variable in Panel A is age at death, and in Panel B the probability of dying within the first two years after an exogenous swap. In columns (1) of Panel A and (1), (2), and (3) in Panel B, controls included in model but excluded from the table are fixed effects for gender, the year of swap, age at swap, and the previous GP. In column (2) of Panel A the following additional controls are included: indicator of high school diploma before re-assignment, indicator for being born in Norway, number of children, birth order of individual, residence in a large city, total income, indicator for receiving welfare benefits, and number of GP visits two years before exogenous swap. Column (3) in Panel A is similar to column (1), but estimated on the smaller sample for which we have all the control variables we use in column (2). The outcome variable is measured between 2007-2017. The measure of VA is re-scaled such that a higher value means that the GP is of higher quality (fewer assigned patients dying within two years after assignment). Standard errors in parentheses clustered at the new GP level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 3: Correlation Between VA Measure and GP Characteristics

	(1)	(2)	(3)
GP Characteristics			
GP Age	-0.005*		-0.004*
	(0.003)		(0.003)
GP is male	-0.016		-0.002
	(0.048)		(0.050)
GP specialist	0.022		0.019
	(0.059)		(0.060)
GP list length (in 1,000)	0.189***		0.174***
	(0.067)		(0.066)
GP has a shared list	-0.005		-0.012
	(0.076)		(0.078)
GP in group practice	0.077		0.061
	(0.117)		(0.115)
GP Practice Style			
GP consultations		0.343	0.335
		(0.359)	(0.365)
Procedures		-1.276	-1.368
		(2.393)	(2.362)
Advising		5.542	4.735
		(5.437)	(5.354)
Blood Test		-5.170	-4.695
		(6.700)	(6.699)
Lab tests		0.499	0.494
		(0.504)	(0.509)
Sick Leave		0.346	-0.075
		(0.668)	(0.685)
Referrals		-22.683	-21.757
		(13.784)	(13.376)
Annual Reimbursement cost (in 100NOK)		0.001	-0.003
		(0.251)	(0.259)
P-values for joint test:			
P-value "GP Characteristics"	.082		.214
P-value "GP Style"		.398	.605
R^2	.006	.005	.010
N	2023	2023	2023

Note: The table presents coefficients from OLS estimations. The dependent variable is the standardized mortality-based VA measure. The regressors are characteristics of GPs in column (1) and practice style indicators in column (2). All variables are combined in the same regression in column (3). "GP Consultations" is the average annual number of regular patient visits. "Procedures" refers to minor surgical procedures (e.g., biopsy, implantation of drug implants, suture of wounds). "Advising" is the average annual number of visits during which the doctor provides counselling (e.g., on how to take medication, or nutrition). "Blood tests" is the average annual number of blood tests required by the GP (per patient). "Lab tests" is the average annual number of tests that require collecting a sample for analysis (e.g., blood tests for hemoglobin and hematocrit count, white and red blood cell count, and simple urine examinations). "Annual reimbursement cost" is the average annual reimbursement per each patient of the GP. "Referrals" is the average annual number of referrals to specialists. "Sick Leave" is the average number of sick leave certifications prescribed. Standard errors in parentheses clustered at the new GP level. * p<0.1, ** p<0.05, *** p<0.01

Table 4: Probability and Number of GP Visits in the First Year After GP Swap

	Panel A: Extensive Margin					
	(1) Consultations	(2) Procedure	(3) Lab Test	(4) Sick Leave		
Standardized VA	-0.010*	0.010	0.000	-0.001		
	(0.005)	(0.008)	(0.009)	(0.006)		
Mean of Depend. Var.	.856	.188	.680	.143		
N	90453	90453	90453	90453		
	Panel B: Intensive Margin					
	(1) Consultations	(2) Procedure	(3) Lab Test	(4) Sick Leave	(5) Reimbursements	(6) Reimbursements
					Log	Reimburs.>Median
Standardized VA	-0.290	0.002	-0.009	-0.027	-0.094*	-0.019**
	(0.207)	(0.023)	(0.074)	(0.031)	(0.049)	(0.010)
Mean of Depend. Var.	7.590	.312	2.390	.507	5.240	.500
N	90453	90453	90453	90453	90453	90453

Note: The table presents the estimates of β in model (3). The dependent variables are different measures of GP services. Column (1) includes the probability (Panel A) and number (Panel B) of annual visits to the GP in the first year after the exogenous GP swap. In column (2), the dependent variable is an indicator for whether the patient visit includes minor surgical procedures such as treatment of epistaxis (bleeding from nose), biopsy, removal of foreign bodies from the eyes, implantation of drug implants, surgical removal of small tumors, warts, part of nail and suture of wounds. The dependent variable in column (3) is an indicator for whether laboratory tests are preformed during the visit (for example, blood testing of total cholesterol, analyses of creatinine, potassium, glycosylated hemoglobin for the determination of long-term blood sugar or rapid test for the detection of helicobacter pylori infection, CPR test, pregnancy test, test for bacterial antigen for streptococci and mononucleosis or glucose chemical analysis). In column (4), the dependent variable is an indicator for whether the patient visit includes prescription of sick leave. In column (5), the dependent variable is the (log of) total reimbursement value associated with the activities performed by the GP during visits. In column (6) the dependent variable takes the value of 1 if the total reimbursement value is above median. Controls included in model but excluded from table are fixed effects for gender, the year of swap, age at swap and the previous GP. The outcome variables are taken from the Control and Payment of Health Refunds Database (KUHR; 2006-2020). Standard errors in parentheses clustered at the new GP level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 5: Hospitalization During First Year After GP Swap

	Panel A: Extensive Margin				
	(1)	(2)	(3)	(4)	(5)
	Any Hospitalization	Acute	Non-Acute	Inpatient	Outpatient
Standardized VA	-0.004	-0.024*	0.009	0.010	0.013
	(0.010)	(0.014)	(0.010)	(0.013)	(0.010)
Mean of Depend. Var.	.451	.346	.914	.26	.878
N	79013	35636	35636	35636	35636
	Panel B: Intensive Margin				
	(1)	(2)	(3)	(4)	(5)
	Any Hospitalization	Acute	Non-Acute	Inpatient	Outpatient
Standardized VA	-0.203	-0.226	0.023	0.107*	-0.084
	(0.299)	(0.143)	(0.209)	(0.061)	(0.185)
Mean of Depend. Var.	4.38	.878	3.5	.476	3.03
N	35636	35636	35636	35636	35636

Note: The table presents estimates of β in model (3). The dependent variables are different measures of specialist health care services. In column (1), the dependent variable is the probability (Panel A) or number (Panel B) of a specialist visit. In columns (2) and (3), the dependent variables indicate whether the patient visit is an acute visit or non-acute visit, respectively. In columns (4) and (5), the dependent variables indicate whether the visit is an inpatient admission (this includes overnight stays or day treatments, such as less minor surgical procedures) or consultations at outpatient clinics, respectively. Controls included in the model but excluded from the table are fixed effects for gender, the year of swap, age at swap and the previous GP. The outcome variables are taken from the Norwegian Patient Register (2008-2020). Standard errors in parentheses clustered at the new GP level. * p<0.1, ** p<0.05, *** p<0.01

Table 6: Probability and Number of GP Visits in the First Year After GP Swap: Symptoms and Diagnoses.

	(1)	(2)	(3)	(4)	(5)
	Panel A: Extensive Margin				
	Cancer	Musculoskeletal ICPC-2 L	Digestive ICPC-2 D	Cardiovascular ICPC-2 K	All except Cancer
Standardized VA	-0.011*	-0.002	-0.003	0.005	-0.006
	(0.006)	(0.008)	(0.007)	(0.009)	(0.006)
Mean of Depend. Var.	.069	.357	.136	.341	.851
N	90453	90453	90453	90453	90453
	Panel B: Intensive Margin				
	Cancer	Musculoskeletal ICPC-2 L	Digestive ICPC-2 D	Cardiovascular ICPC-2 K	All except Cancer
Standardized VA	-0.188**	-0.055	-0.130**	0.050	-0.102
	(0.073)	(0.065)	(0.064)	(0.086)	(0.190)
Mean of Depend. Var.	.275	1.390	.388	1.580	7.310
N	90453	90453	90453	90453	90453

Note: The table presents estimates of β in model (3). The dependent variables are different measures of GP services. Column (1) includes the probability (Panel A) and number (Panel B) of annual visits to GP one year after the exogenous GP swap. Column (1) includes all cancer and neoplasm diagnoses (ie, benign or malign cancers, which correspond to ICPC-2 codes: A79, B72-75, D74-78, F74, H75, K72, L71, L97, N74-76, R84-86, R92, S77-80, T71-73, U75-79, W72, W73, X75-81, Y77-79). Column (2) includes all symptoms and diagnoses related to the Musculoskeletal System (ICPC-2 codes starting with letter L). Column (3) includes all symptoms and diagnoses related to the Digestive System (ICPC-2 codes starting with letter D). Column (4) includes all symptoms and diagnoses related to the Cardiovascular System (ICPC-2 codes starting with letter K). Column (5) includes all conditions except those in column (1), ie, benign or malign cancer. Controls included in the model but excluded from the table are fixed effects for gender, the year of swap, age at swap, and the previous GP. The outcome variables are taken from the Control and Payment of Health Refunds Database (KUHR; 2006-2020). Standard errors in parentheses clustered at the new GP level. * p<0.1, ** p<0.05, *** p<0.01

Table 7: Labor Market Outcomes and Welfare Use 5 Years After GP Swap

	(1) Work	(2) Labor Income	(3) Any Self Empl. Income	(4) Sick Leave Days	(5) Hours	(6) Any Transfer
			Panel A: 55 or older			
Standardized VA	0.010	1949.522	-0.005	0.878	0.387	-0.016**
	(0.009)	(5254.260)	(0.005)	(0.774)	(0.282)	(0.006)
Mean of Depend. Var.	.375	187212	.0789	8.25	10.5	.833
N	82917	82917	82917	82949	83031	83113
			Panel B: 65 or older			
Standardized VA	0.000	1364.388	-0.012	0.005	0.126	-0.000
	(0.006)	(2235.988)	(0.009)	(0.019)	(0.152)	(0.000)
Mean of Depend. Var.	.0662	22581	.0492	.0433	1.62	1
N	27825	27825	27825	27824	27839	27868
			Panel C: 55 to 64 years old			
Standardized VA	0.015	4788.474	-0.003	1.169	0.521	-0.025**
	(0.014)	(7916.261)	(0.006)	(1.204)	(0.424)	(0.010)
Mean of Depend. Var.	.531	270361	.094	12.4	15	.749
N	55092	55092	55092	55125	55192	55245

Note: The table presents the estimates of β in model (3). Controls included in the model but excluded from the table are fixed effects for gender, the year of swap, age at swap, and the previous GP. Outcomes are taken from the tax registers (earnings) and the social insurance database for the years of 2007-2020. Social transfers include any retirement pensions, disability income, employment assessment allowance, contractual private and public pension, unemployment benefits, sick pay, parental allowance, child benefits, housing benefit, social assistance, basic and auxiliary allowance, and cash support. Standard errors in parentheses clustered at the new GP level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 8: Endogenous Swap of GP (after Exogenous Swap) and Value Added, GP-Level

	(1)	(2)
Standardized VA	-0.00027	
	(0.00324)	
GP 40+		0.04385***
		(0.00750)
GP specialist		-0.01666**
		(0.00841)
GP list length		0.00421
		(0.00917)
GP male		0.01336**
		(0.00644)
Sick leave		0.00063
		(0.00575)
Reimbursement		-0.00002***
		(0.00001)
N	2023	2023

Note: The table shows the correlation between the proportion of patients that a given GP loses via endogenous swap after an exogenous swap, and his/her VA (column (1)) and other GP characteristics (column 2). Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 9: Patients' Ratings and GP VA

	Panel A: Value Added				
	(1) # of Ratings	(2) Mean Rate	(3) Share Max. Rating	(4) Share Min. Rating	(5) No Ratings
Standardized VA	-0.0148	-0.0010	-0.0020	-0.0011	-0.0202**
	(0.2675)	(0.0202)	(0.0061)	(0.0045)	(0.0088)
Mean of Depend. Var.	13.2	4.11	.684	.132	.136
N	1747	1747	1747	1747	2023
	Panel B: Characteristics of GP				
	(1) # of Ratings	(2) Mean Rate	(3) Share Max. Rating	(4) Share Min. Rating	(5) No Ratings
GP 40+	-1.1068*	-0.2800***	-0.0864***	0.0543***	0.1163***
	(0.6230)	(0.0470)	(0.0149)	(0.0103)	(0.0191)
GP specialist	0.4907	0.1514***	0.0313**	-0.0397***	-0.0449**
	(0.6383)	(0.0479)	(0.0156)	(0.0105)	(0.0202)
GP list length	15.1460***	-0.2826***	-0.0794***	0.0485***	-0.1971***
	(0.9607)	(0.0546)	(0.0177)	(0.0118)	(0.0235)
GP male	-0.5715	0.0856**	0.0123	-0.0247***	0.0444***
	(0.5043)	(0.0385)	(0.0124)	(0.0083)	(0.0155)
Sick leave	-0.7442***	0.0223	-0.0045	-0.0098*	-0.0417***
	(0.2526)	(0.0281)	(0.0126)	(0.0053)	(0.0129)
Reimbursement	-0.0004	0.0000	0.0000**	0.0000	-0.0000*
	(0.0004)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
N	1747	1747	1747	1747	2023

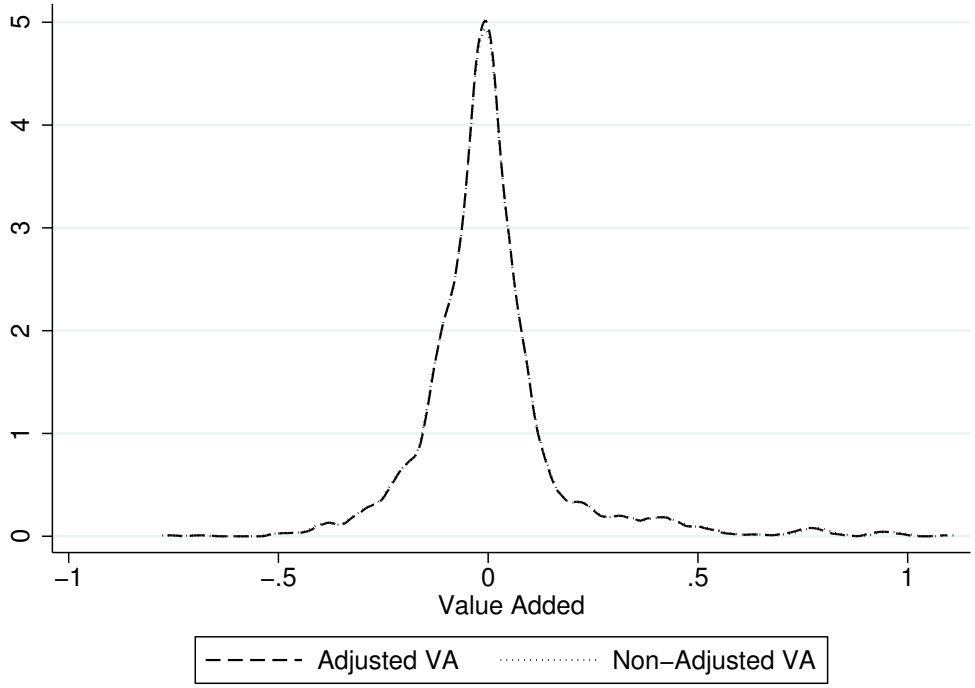
Note: The table shows the correlation between the standardized mortality-based VA measure (Panel A) and GP characteristics (Panel B), and patients' rating of the GP. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 10: Patients' Ratings and Endogenous Swaps

	(1) # of Ratings	(2) Mean Rate	(3) Share Max. Rating	(4) Share Min. Rating	(5) No Ratings
Endogenous Swap	-0.0022***	0.0209***	0.0395*	-0.1092***	0.0584***
	(0.0005)	(0.0069)	(0.0220)	(0.0317)	(0.0158)
N	1747	1747	1747	1747	2023

Note: The table shows the correlation between the share of new patients on a GP's list that are endogenous swaps and patients' rating of the GP. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Figure 1: Distribution of Value Added Measure, Mortality 2 Years After GP Swap



Note: The figure shows the distribution of the non-Bayes and the Bayes adjusted mortality-based standardized VA measure as described in the text.

Does Your Doctor Matter?

By Rita Ginja, Julie Riise, Barton Willage, and Alexander Willén

ONLINE APPENDIX, Not For Publication

Table A.1: Identification of Motives for GP Re-assignment

Motive	Frequency	Percent
Automatic allocation	1,385,521	6.62
Children reunited with parents	873	0.00
Resident returns to previous GP	1,195	0.01
Correction	28,648	0.14
Doctor has reduced or stopped practice	5,412,853	25.88
Manual assignment	1,189	0.01
Ordinary exchange	13,598,091	65.00
Allocated from waiting list	490,304	2.34
Total	20,918,674	100

Note: This table provides information on the causes (and their frequency) for re-assignments of GPs in Norway between 2001 and 2020.

Table A.2: Mobility Between Previous and Exogenously Assigned GP

	(1) Mean	(2) SD	(3) N
Panel A: Patients per swap (previous-GP by new-GP level)	21.91	50.62	4,129
Panel B: New GP per previous GP (previous-GP level)	3.25	2.50	1,272
Panel C: Patients per previous GP (previous-GP level)	71.11	83.89	1,272

Note: The table shows information about mobility between previous and exogenously-assigned GPs for individuals aged 55 or old at the time of re-assignment. Panel A shows the mean and standard deviation for the number of patients that are transferred from each previous GP to each new GP. Panel B shows the mean and standard deviation of the number of GPs that receive patients from each previous GP. Panel C shows the mean and standard deviation of the number of patients that exogenously swap from a given previous GP to a new GP.

Table A.3: Characteristics of GPs Involved in Swaps

Sample of Patients	(1)	(2)	(3)	(4)	(5)	(6)
	Exogenous Swaps				Any Swap	
	55 and Older		25 and Older		55 and Older	
	Mean	SD	Mean	SD	Mean	SD
Panel A: Characteristics of GP						
GP age	43.47	10.72	45.89	11.17	46.39	10.92
GP specialist	0.37	0.47	0.45	0.48	0.47	0.46
GP list length	1095.17	356.53	1140.10	369.39	1147.75	363.09
GP male	0.62	0.49	0.62	0.49	0.63	0.48
GP shared list	0.09	0.28	0.08	0.27	0.08	0.27
GP group practice	0.93	0.24	0.92	0.26	0.92	0.27
GP max list length	1168.85	354.67	1213.69	364.15	1222.83	359.30
Panel B: Characteristics of Patients Served						
Average Annual Number of GP consultations	6.42	5.19	5.58	6.40	6.24	3.88
Proportion of Patients with at least one lab test annually	0.58	0.34	0.49	0.34	0.57	0.25
Proportion of Patients with at least one surgical procedure in primary care	0.14	0.19	0.11	0.19	0.14	0.14
Proportion of Patients with at least one sick leave annually	0.16	0.20	0.24	0.27	0.17	0.16
N	2023		4834		5328	

Note: The table provides descriptive statistics of a series of selected characteristics for the GPs included in the main analysis. There is one observation per GP.

Table A.4: Correlations Between Different GP VA Measures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Hospitalizations					Sick Leave	Labor
	Total	Non acute	Acute	Inpatient	Outpatient	Days	Income 55
Mortality-based	0.002	0.016	-0.068*	0.043	-0.053*	-0.171***	0.086***
	(0.030)	(0.028)	(0.039)	(0.026)	(0.031)	(0.038)	(0.021)
N	1977	1977	1977	1977	1977	1999	1999

Note: The table shows correlations between the main VA measure based on mortality and alternative VA measures based on other outcomes as defined in the text. The measure of mortality-based VA is re-scaled such that a higher value means that the GP is of higher quality (fewer assigned patients dying within two years after assignment). All measures are calculated two years after the exogenous re-assignment. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.5: Descriptive Statistics for Patients

Sample of Patients	(1)	(2)	(3)	(4)	(5)	(6)
	Exogenous Swaps				Any Swap	
	55 and Older		25 and Older		55 and Older	
	Mean	SD	Mean	SD	Mean	SD
Panel A: Mortality						
Mortality 2 years after swap	0.03	0.18	0.01	0.11	0.04	0.19
Causes (Conditional on death)						
Cardiovascular conds	0.26	0.44	0.26	0.44	0.26	0.44
Cancer	0.36	0.48	0.36	0.48	0.34	0.48
External conds	0.04	0.19	0.04	0.19	0.04	0.19
Respiratory conds	0.11	0.31	0.11	0.31	0.11	0.32
Panel B: Demographics						
Age at swap	63.79	7.60	47.22	13.79	63.53	7.47
Male	0.55	0.50	0.53	0.50	0.53	0.50
Born in Norway	0.97	0.17	0.94	0.24	0.97	0.18
Years of education	13.15	2.47	13.65	2.45	13.12	2.47
Married	0.67	0.47	0.49	0.50	0.64	0.48
Panel C: Health and Health Care						
GP consultations	6.52	7.76	5.30	7.45	6.40	7.84
Lab tests	0.63	0.48	0.53	0.50	0.62	0.49
Surgery	0.16	0.36	0.12	0.33	0.15	0.36
Sick leave	0.17	0.38	0.24	0.43	0.17	0.38
Reimbursement	0.81	0.39	0.74	0.44	0.79	0.41
Hospitalization	0.41	0.49	0.34	0.47	0.41	0.49
Panel D: Labor Market						
Labor income	279832.28	410143.33	353684.65	388736.22	278313.38	404378.27
Total income	436034.79	680456.19	445458.19	621741.47	441847.84	761416.16
Any social transfer	0.57	0.49	0.53	0.50	0.58	0.49
Sick leave days	14.65	60.18	16.06	59.67	15.19	61.39
N	92,205		299,322		166,823	

Note: The table provides descriptive statistics of a series of selected characteristics of patients included in our analysis. One observation per patient is used. In Panel B, the education indicator is measured two years before the exogenous GP re-assignment. In Panel C and D the variables for health and health care, and the labor market outcomes, are measured two years before the exogenous GP swap. Monetary values are deflated to 2015 using the CPI.

Table A.6: Descriptive Statistics for Ratings

	All GPs Rated		GPs in Main Estimation Sample	
	Mean	Proportion w/ Top Rating	Mean	Proportion w/ Top Rating
Number of ratings	11.439	10.606	13.458	11.163
Rating	4.150	0.702	4.110	0.684
N	6,119			1,797

Note: The table provides descriptive statistics of the subjective GP ratings obtained from legelisten.no. The table includes both the mean and the proportion with top ratings. One observation per GP is used.

Table A.7: Relationship Between GP VA and Baseline Characteristics of Patients: Additional Sensitivity Analysis For the Sample of Patients Involved in Exogenous GP Swaps (not controlling for gender and age of patient at the time of swap).

	(1) Age	(2) Male	(3) College Degree	(4) Married	(5) Born in Norway	(6) Nb Children
Standardized VA	0.033	0.006	0.002	-0.000	-0.002	-0.034
	(0.181)	(0.011)	(0.010)	(0.011)	(0.003)	(0.026)
Mean of Depend. Var.	63.9	.539	.232	.664	.968	2.15
N	71419	71419	71419	71419	71419	71419
	(7) Total Income	(8) Hours	(9) Any Welfare Benefits	(10) Sick Leave	(11) Nb. GP Visits	(12) Reimbursements
Standardized VA	7.863	-0.186	0.008	-1.891	-0.169	7.503
	(7.831)	(0.374)	(0.011)	(1.490)	(0.236)	(35.683)
Mean of Depend. Var.	449	15.2	.575	14.7	6.49	751
N	71419	71419	71419	71419	71419	71419
	(13) Any Chronic Conditions	(14) Nb Chronic Conditions	(15) Any Hospitalization	(16) Nb. Hospitalizations	(17) Charlson Index	
Standardized VA	0.003	0.019	0.017	0.043	-0.025	
	(0.011)	(0.195)	(0.013)	(0.098)	(0.018)	
Mean of Depend. Var.	.432	3.43	.407	1.6	.149	
N	71419	71419	54179	54179	54179	

Note: The dependent variables are characteristics of patients measured two years prior to the exogenous assignment to a new GP. Controls included in the model but excluded from the table are fixed effects for the year of swap and for the previous GP. Standard errors in parentheses (SE) clustered at the new GP level. * p<0.1, ** p<0.05, *** p<0.01

Table A.8: Relationship Between GPs' VA and Baseline Characteristics of Patients: All Swaps.

	(1) College Degree	(2) Married	(3) Born Norway	(4) Nb. Children	(5) Total Income	(6) Any Welfare Benefits	(7) Sick Leave Days
Standardized VA	0.005***	0.010***	-0.001	0.015***	12082.918***	-0.007***	0.213
	(0.002)	(0.002)	(0.001)	(0.005)	(2564.647)	(0.002)	(0.359)
Mean of Depend. Var.	.242	.638	.965	2.12	441848	.58	22.4
N	165830	166437	166823	166823	166627	166627	112558
	(8) Nb. GP Visits	(9) Any Chronic Conditions	(10) Nb. Chronic Conditions	(11) Reimbursements	(12) Any Hospitalization	(13) Nb. Hospitalizations	
Standardized VA	-0.179***	-0.004	-0.121***	-25.102***	-0.005*	-0.014	
	(0.042)	(0.002)	(0.036)	(7.566)	(0.003)	(0.031)	
Mean of Depend. Var.	6.4	.419	3.3	754	.405	1.63	
N	122210	122210	122210	122210	89436	89436	

Note: The dependent variables are characteristics of patients measured two years prior to the assignment to the new GP. Controls included in model but excluded from table are fixed effects for gender, the year of swap, age at swap, and the previous GP. Standard errors in parentheses clustered at the new GP level. * p<0.1, ** p<0.05, *** p<0.01

Table A.9: Fallback Condition from Abaluck et al. (2021)

	(1)
Standardized VA	-0.0005
	(0.0003)
N	90453

Note: This table presents estimates for the Fallback Condition of Abaluck et al. (2021). The dependent variable is the prediction of the mortality residual based on GP characteristics. We construct the residual by regressing observed mortality on a set of observable doctor characteristics (square of GP's age, gender, an indicator for GP specialization, square of actual list length and maximum length, an indicator for whether the GP shares the list with other GPs and an indicator for whether the GP works in a shared office practice). Patient-level controls included in the model are fixed effects for gender, the year of swap, age at swap, and the previous GP. Standard errors in parentheses clustered at the new GP level. * p<0.1, ** p<0.05, *** p<0.01

Table A.10: Relationship Between GPs' Predicted Mortality and Baseline Characteristics of Patients: Sample of Patients Involved in Exogenous GPs Swap.

	(1) College Degree	(2) Married	(3) Born in Norway	(4) Nb of Children	(5) Total Income	(6) Hours	(7) Any Welfare Benefits	(8) Sick Leave Days
Linear Prediction	-1.757*	-1.629	0.029	-0.768	-615.450	-36.019	-0.183	-64.806
	(0.953)	(1.081)	(0.346)	(2.708)	(801.630)	(31.788)	(0.888)	(116.152)
N	89841	89841	89841	89841	89841	89841	89841	89841
	(9) Nb. GP Visits	(10) Reimbursements	(11) Any Chronic Conditions	(12) Nb. Chronic Conditions	(13) Any Hospitalization	(14) Nb. Hospitalizations	(15) Charlson Index	
Linear Prediction	-26.190	-2583.683	-0.356	2.787	-1.528	-14.822	0.915	
	(24.114)	(3843.261)	(1.244)	(19.115)	(1.220)	(11.229)	(2.047)	
N	71419	71419	71419	71419	54179	54179	54179	

Note: This table presents estimates for the balance condition proposed by Abaluck et al. (2021). We test whether GP characteristics predict observable differences in patient baseline characteristics (measured two years before the swap). To do so, we obtain the predicted two-years patient mortality from the characteristics of the newly assigned GP at the year of swap (gender, quadratic on GPs age, quadratic on the GPs actual and maximum patient list, and indicators for whether the GP is specialized in general medicine, whether he or she practice in a joint office and have a shared patient list). We then estimate models of patients' characteristics on this prediction, controlling for fixed effects for gender, the year of swap, age at swap, and the previous GP. The estimates are based on the sample of patients involved in exogenous GP swaps. Standard errors in parentheses clustered at the new GP level. * p<0.1, ** p<0.05, *** p<0.01

Table A.11: External Validity Exercise

	(1) Nb. GP Visits	(2) Any Chronic Condition	(3) Annual Reimbursements	(4) Any Hospitalization
1[Exogenous Re-assignment]	0.031	0.003	-1.570	0.001
	(0.091)	(0.005)	(12.616)	(0.003)
Mean of Depend. Var.	7.52	.493	929	.688
N	354246	354246	354246	274709

Note: The dependent is taken in the year the GP experiences between 10 to 1000 exogenous removals from her/his list. Controls included in model but excluded from table are fixed effects for gender, the year and age of patient. Standard errors in parentheses clustered at the new GP level. * p<0.1, ** p<0.05, *** p<0.01

Table A.12: Key Moments of GP Value-Added: Sensitivity Analyses

	(1) N	(2) P10	(3) P25	(4) P50	(5) P75	(6) P90	(7) Mean	(8) SD
	Panel A: 2-Year Mortality, 55 : Different Models							
A: Baseline	2023	-0.150	-0.062	-0.004	0.045	0.116	0.000	0.157
B: Add muni FE	2023	-0.153	-0.068	-0.005	0.060	0.130	0.000	0.160
C: Pre-swap GP-year FE	2023	-0.144	-0.054	0.003	0.055	0.120	0.000	0.142
D: Pre-swap chronic conds.	2023	-0.156	-0.054	0.005	0.052	0.126	-0.000	0.162
E: Strict mobility group	1986	-0.154	-0.063	-0.004	0.045	0.116	0.000	0.159
	Panel B: 2-years mortality for 55 : Minimum Number of Incoming Patients							
A: 25 patients	700	-0.127	-0.042	-0.004	0.041	0.126	0.000	0.201
B: 50 patients	561	-0.111	-0.037	0.002	0.026	0.064	0.000	0.155
C: 100 patients	353	-0.122	-0.050	0.004	0.066	0.092	0.000	0.128
	Panel C: Mortality, 55 : Different Time Spans							
A: 1-year mortality, 55	2023	-0.115	-0.045	-0.006	0.026	0.073	0.000	0.131
B: 3-year mortality, 55	2023	-0.171	-0.082	-0.006	0.064	0.164	0.000	0.178
C: 4-year mortality, 55	2023	-0.191	-0.090	-0.008	0.069	0.191	0.000	0.186
D: 5-year mortality, 55	2023	-0.219	-0.104	0.003	0.081	0.203	0.000	0.201
	Panel D: Mortality for 55 : All swaps							
2-year mortality, 55	5328	-0.055	-0.035	-0.011	0.023	0.067	0.000	0.062
	Panel E: 2-Year Mortality, 65							
2-year mortality, 65+	1269	-0.178	-0.076	-0.003	0.063	0.131	0.001	0.182

Note: Each row reports various moments for the GP VA measure using different specifications or samples. The numbers reported are the estimates for the GP fixed effects, referring to the estimates of μ_j in equation (1). Row A of Panel A shows the distribution from the baseline model. In row B of Panel A, we include municipality fixed effects. In Row C of Panel A, we include previous GP-by-year fixed effects. In row D of Panel A, we include controls for any pre-swap chronic condition that the patient may have, such as cancer, chronic conditions of the digestive system (ulcerative colitis, cholecystitis or cholelithiasis), chronic cardiovascular conditions (ischaemic heart disease, acute myocardial infarction, heart failure, atrial fibrillation, paroxysmal tachycardia, cardiac arrhythmia, heart/arterial murmur, pulmonary heart disease, heart valve disease, elevated blood pressure, hypertension, postural hypotension, transient cerebral ischaemia, stroke/cerebrovascular accident, cerebrovascular disease, atherosclerosis, pulmonary embolism, phlebitis/thrombophlebitis, varicose veins of leg or hemorrhoids) and endocrine, metabolic or nutritional chronic conditions (goitre, obesity/overweight, hyperthyroidism, hypothyroidism, hypoglycaemia, insulin or non-insulin dependent diabetes, vitamin deficiency, gout, lipid disorder)). In Row E of Panel A, we run a model based on equation (1) using a stricter sample definition where we exclude connected sets with fewer than 10 patients from the sample. In Panel B, we present the distribution of VA restricting the sample to patients allocated to GPs with different minimum number of incoming patients. In Panel C, we present the distribution of VA based on mortality 1, 3, 4 or 5 years after the exogenous re-assignment. In Panel D, we present the distribution of VA based on mortality 2 years after any type of re-assignment. In Panel E, we use mortality among individuals 65 or older at time of the exogenous swap to compute VA.

Table A.13: Correlations Between Standardized VA Measures based on Mortality Within Different Time Horizons

	(1) t 1	(2) t 3	(3) t 4	(4) t 5
Standardized VA t 2	0.7566***	0.7588***	0.6858***	0.6327***
	(0.0145)	(0.0145)	(0.0162)	(0.0172)
N	2023	2023	2023	2023

Note: The table shows correlations between the standardized mortality-based VA measure within different time horizons after an exogenous swap. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.14: Value Added Based on 2-Year Mortality: Heterogeneity by GP Characteristics

	(1) N	(2) Mean	(3) SD
Panel A: Practice Type			
GP not group practice	111	-0.014	0.128
GP group office	1889	0.003	0.166
Panel B: List Type			
GP not shared list	1843	0.002	0.191
GP shared list	169	-0.004	0.071
Panel C: Gender			
GP is female	738	-0.009	0.266
GP is male	1212	0.005	0.218
Panel D: Specialization			
GP does not have specialization in general medicine	1261	0.001	0.199
GP has specialization in general medicine	717	0.013	0.372
Panel E: List length			
GP has short list ($<$ 1100 patients)	1071	0.001	0.257
GP has long list ($\geq$ 1100 patients)	918	0.003	0.173
Panel F: Age			
GP is less than 40 years old	921	-0.011	0.212
GP is 40 or older	1044	-0.008	0.234

Note: This table provides information on the heterogeneity of GP VA by GP characteristics. Each row reports the mean and the standard deviation for the GP VA for the indicated group.

Table A.15: Impact on 2-Year Mortality After Swap: Age of Death and Causes of Death (Heterogeneity by Age at Exogenous Re-assigment)

	(1) Age at Death	(2) All Causes Mortality	(3) Malign Cancer ICD10 C	(4) Cardiovascular Conditions ICD10 I	(5) Respiratory Conditions ICD10 J	(6) All the Rest
			Panel A: 62 or older at swap			
Standardized VA	0.213	0.001	-0.009	0.003	-0.001	0.008
	(0.165)	(0.013)	(0.008)	(0.006)	(0.004)	(0.007)
Mean of Depend. Var.	80	.050	.016	.014	.006	.014
N	10874	47222	47222	47222	47222	47222
			Panel B: 55-61 at swap			
Standardized VA	0.693***	-0.027***	-0.012*	-0.001	-0.003	-0.010
	(0.148)	(0.010)	(0.007)	(0.003)	(0.003)	(0.006)
Mean of Depend. Var.	64.3	.011	.005	.002	.001	.003
N	2821	43190	43190	43190	43190	43190

Note: The table present OLS estimates where the dependent variable in Panel A is age at death, and in Panel B the probability of dying within the first two years after an exogenous swap. In columns (1) of Panel A and (1), (2), and (3) in Panel B, controls included in model but excluded from the table are fixed effects for gender, the year of swap, age at swap, and the previous GP. In column (2) of Panel A the following additional controls are included: indicator of high school diploma before re-assignment, indicator for being born in Norway, number of children, birth order of individual, residence in a large city, total income, indicator for receiving welfare benefits, and number of GP visits two years before exogenous swap. Column (3) in Panel A is similar to column (1), but estimated on the smaller sample for which we have all the control variables we use in column (2). The outcome variable is measured between 2007-2017. The measure of VA is re-scaled such that a higher value means that the GP is of higher quality (fewer assigned patients dying within two years after assignment). Standard errors in parentheses clustered at the new GP level. * p<0.1, ** p<0.05, *** p<0.01

Table A.16: Predictors of Malign Cancer Diagnoses.

ICPC2 Code Groups	(1) D7	(2) T7	(3) A7	(4) X7 Y7
Standardized VA	-0.128**	0.000	0.002	-0.011
	(0.052)	(0.001)	(0.007)	(0.034)
Mean of Depend. Var.	0.075	0.005	0.035	0.114
N	78474	78474	78474	78474

Note: Note: The dependent variables are characteristics of patients measured two years prior to the exogenous assignment to a new GP. Controls included in model but excluded from table are fixed effects for gender, the year of swap, age at swap, and the previous GP. The measure of VA is re-scaled such that a higher value means that the GP is of higher quality (fewer assigned patients dying within two years after assignment). The conditions included as outcomes are different types of infections, neoplasms or injuries. Column (1) includes an indicator for "ICPC2 D7 Conditions" which include digestive system infections, neoplasms or injuries, such as gastrointestinal infections, mumps, viral hepatitis, benign/uncertain/malignant neoplasms of stomach, colon/rectum or pancreas and a foreign body in the digestive system. Column (2) includes an indicator for "ICPC2 D7 Conditions" that include. endocrine/metabolic and nutritional conditions such as endocrine infections and benign /malignant neoplasm thyroid. Column (3) includes an indicator for "ICPC2 A7 Conditions" that are a set of general and unspecified conditions such as tuberculosis, measles, chickenpox, malaria, rubella, infectious mononucleosis, and other viral diseases or infection and nonspecific malign neoplasm. Column (4) includes an indicator for "ICPC2 X7 and Y7 Conditions" that include, respectively, female and male genital related conditions such as syphilis, gonorrhoea, genital candidiasis on females, genital trichomoniasis on females, pelvic inflammatory disease, malignant neoplasm cervix, fibromyoma uterus, benign/malign neoplasm, male genital herpes male, prostatitis/seminal vesiculitis, epididymitis, balanitis (on males), condylomata acuminata (on males), neoplasm prostate or other male genital. Standard errors in parentheses clustered at the new GP level. * p<0.1, ** p<0.05, *** p<0.01. See here the full list of ICPC-2 codes in English (accessed on February 22. 2024): `https://www.icgp.ie/go/in_the_practice/it_in_the_practice/gpit/publications_reports/5939A766-D385-4191-A7A946A8153C37E4.html`.

Table A.17: Welfare Use 5 Years After GP Swap: Sample of Individuals Re-assigned to a new GP between 55 and 64 years old.

	(1) Any Transfer	(2) Old-Age Pensions	(3) DI	(4) AFP Pensions	(5) UI	(6) Sick Leave Income	(7) Any Other
Standardized VA	-0.025**	-0.003	-0.010	-0.048***	-0.001	0.013	-0.002
	(0.010)	(0.011)	(0.013)	(0.011)	(0.002)	(0.010)	(0.002)
Mean of Depend. Var.	.749	.628	.240	.529	.019	.177	.009
N	55245	55245	55092	55092	55092	55092	55092

Note: The table presents the estimates of β in model (3). Controls included in the model but excluded from the table are fixed effects for gender, the year of swap, age at swap, and the previous GP. Outcomes are taken from the social insurance database for the years of 2007-2020 and are indicators that take value one if the individual receives a given type transfers and zero if does not receive. Social transfers include old retirement pensions (column 2), disability income (column 3), contractual private and public early-retirement (AFP) pensions (column 4), unemployment benefits (column 5), sick leave pay (column 6) or any other (such as parental allowance, child benefits, housing benefit, social assistance, basic and auxiliary allowance, and cash support; column 7). Standard errors in parentheses clustered at the new GP level. * p<0.1, ** p<0.05, *** p<0.01

Table A.18: Labor Market Outcomes and Welfare Use 5 Years After GP Swap: Bounds

	(1)	(2)	(3)
	Panel A: Labor Income		
	Percentile 25	Percentile 50	Percentile 75
Standardized VA	4417.756	4384.258	-2032.235
	(4112.605)	(4109.948)	(4375.132)
Mean of Depend. Var.	174978	175093	196997
N	88714	88714	88714
	Panel B: Any Transfers		
	Lower Bound	Upper Bound	
Standardized VA	0.005	-0.014***	
	(0.009)	(0.005)	
Mean of Depend. Var.	.779	.844	
N	88910	88910	

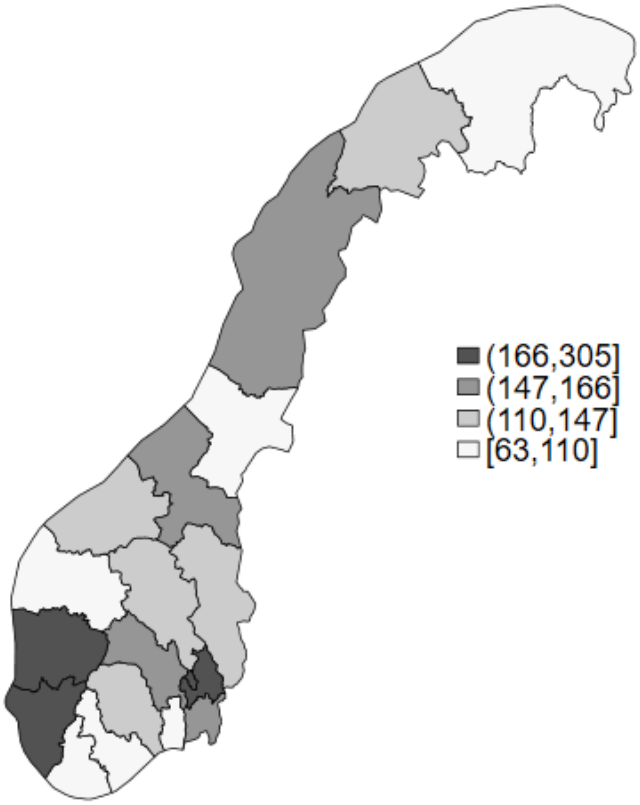
Note: The table presents the estimates of β in model (3). Controls included in the model but excluded from the table are fixed effects for gender, the year of swap, age at swap, and the previous GP. In Panel A, we conduct a bounding exercise in which we assign the deceased the 25th, 50th, and 75th percentile earnings. In Panel B, we assign deceased individuals the lowest (zero) and highest (one) possible value for transfers to derive the bounds. Standard errors in parentheses clustered at the new GP level. * p<0.1, ** p<0.05, *** p<0.01

Table A.19: Patient Ratings and Different Measures of GP VA

	(1)	(2)	(3)	(4)	(5)	(6)
			Hospitalizations			Sick Leave days
	Total	Non acute	Acute	Inpatient	Outpatient	for 55+
Mean rating	-0.0430	-0.0313	-0.0708*	-0.0190	-0.0349	-0.0369
	(0.0358)	(0.0351)	(0.0401)	(0.0344)	(0.0402)	(0.0361)
Observations	1716	1716	1716	1716	1716	1728
R^2	0.001	0.001	0.002	0.000	0.000	0.001

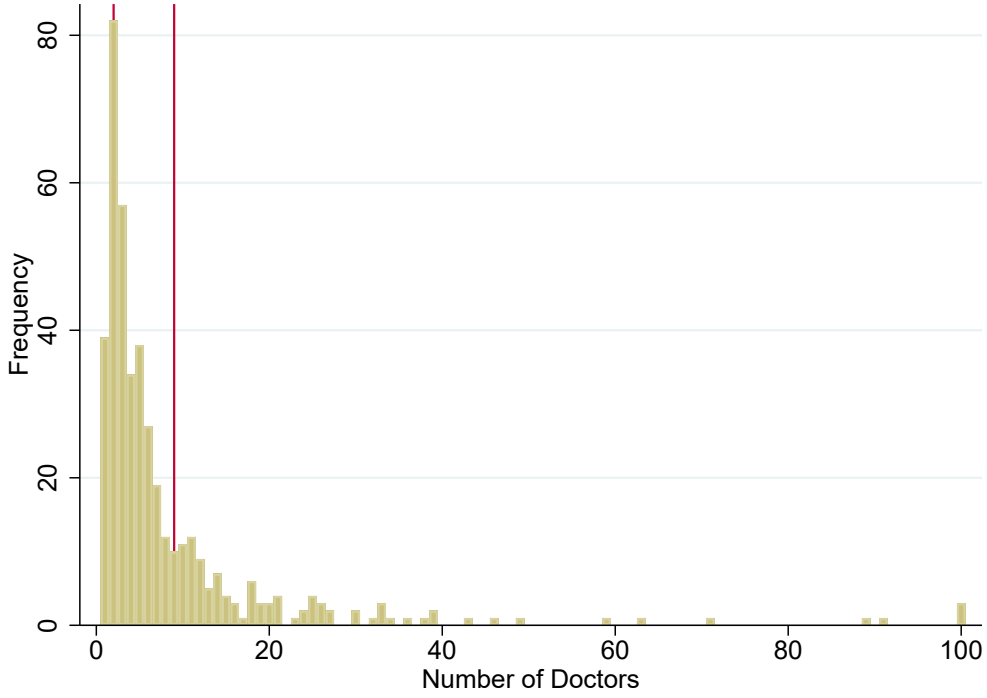
Note: This tables shows correlations between different VA measures and patient ratings. In column (1) the VA measure is based on the total number of annual hospital visits two years after an exogenous swap, while in columns (2)-(5), we narrow to the number of specific types of hospital visits. In column (6) the VA measure is based on the total number of sick days two years after an exogenous swap. Robust standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Figure A.1: Regional Variation in the Number of New GPs



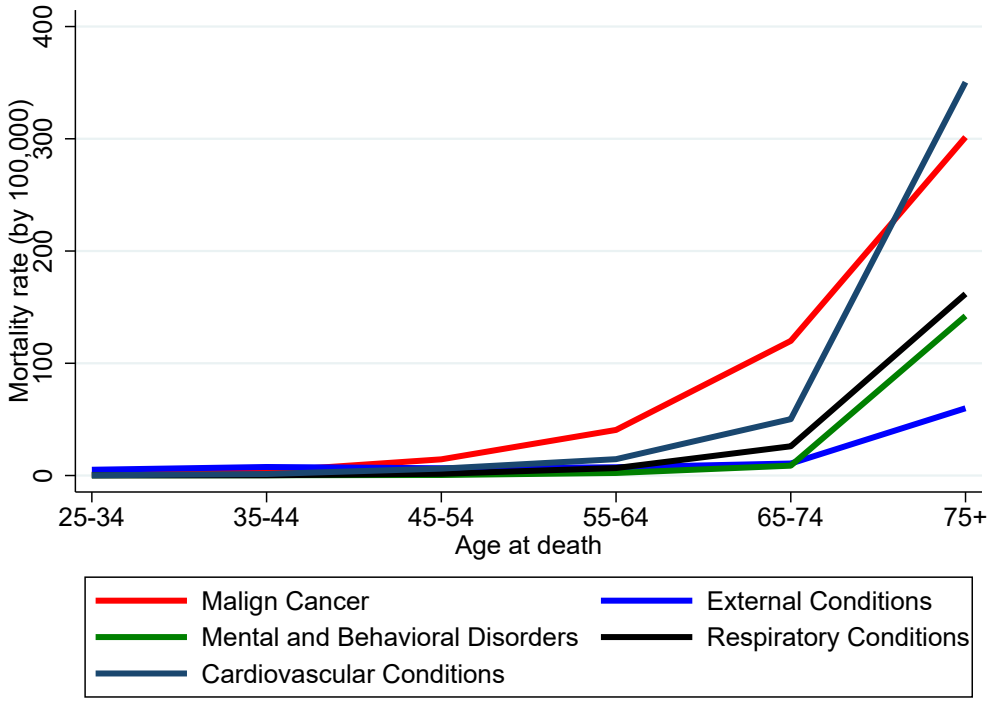
Note: This figure presents the average number of new GPs receiving patients between 2006 and 2015 across the 19 counties of Norway.

Figure A.2: Number of GPs per Municipality



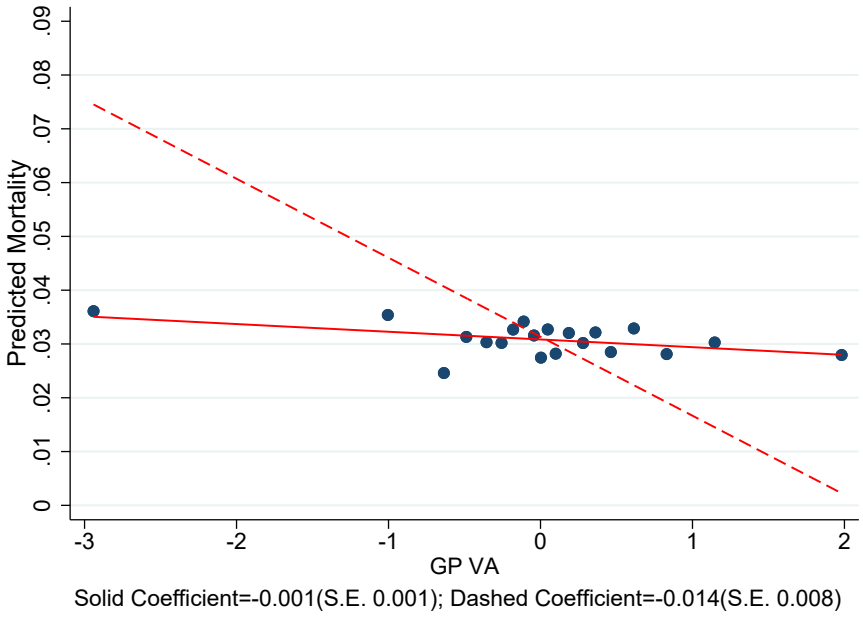
Note: The figure includes the distribution of the number of GPs working in each municipality. The vertical red lines indicate the 25th and 75th percentiles. The X-axis is right censored at 100, so that all values above 100 are assigned 100.

Figure A.3: Mortality Rate Due to Specific Conditions



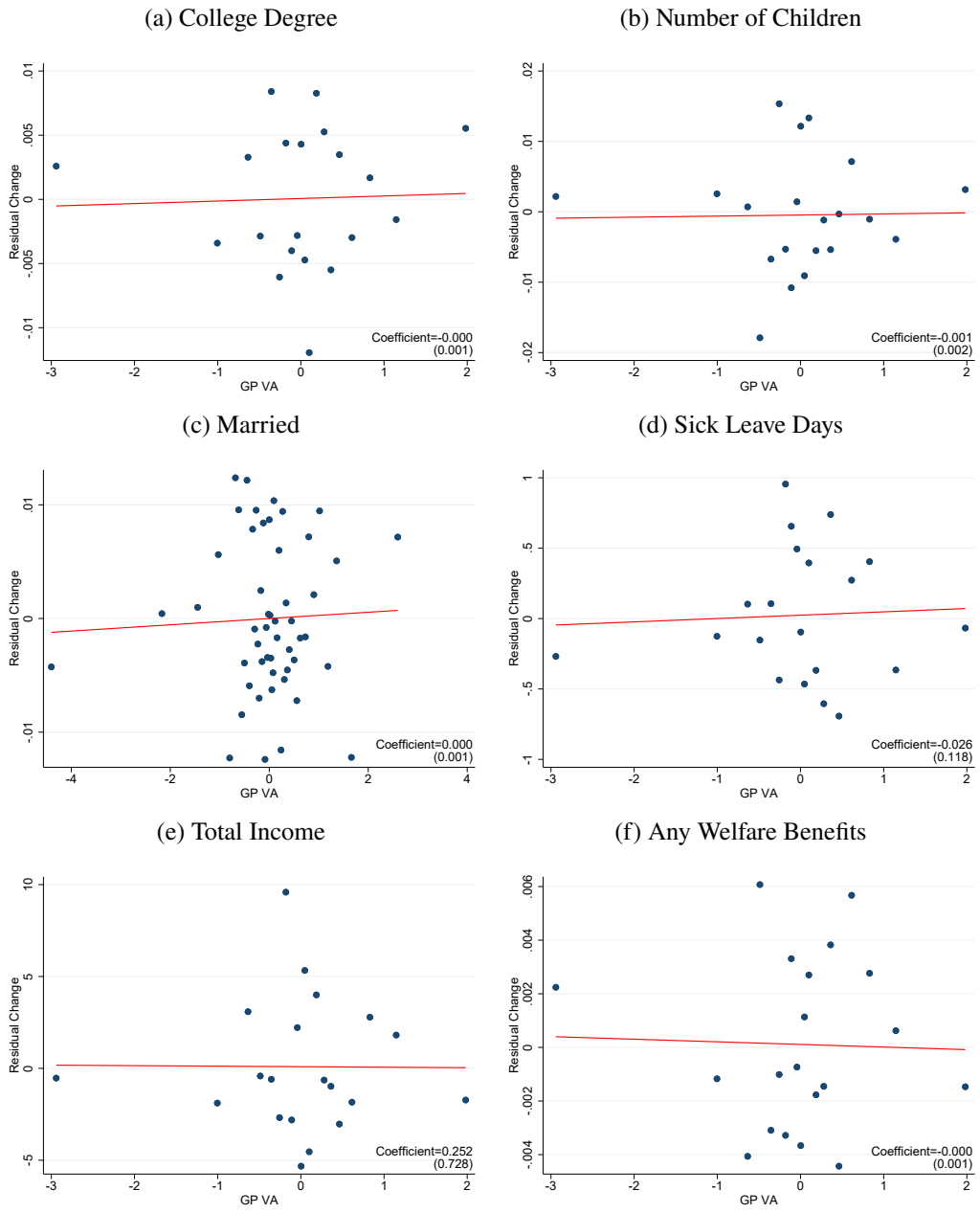
Note: This figure presents the age-adjusted mortality rate for the main causes of death for different age groups in 2020.

Figure A.4: Relationship between VA and Predicted Mortality



Note: To construct the binned scatter plot, we first predict the y-axis variable from to baseline characteristics listed in Panel B of Table 1 (education, marital status, immigrant status, total income, hours worked, an indicator for any social transfers, sick leave days, number of GP visits, annual primary care reimbursements, an indicator for any chronic conditions, number of chronic conditions) and gender, fixed effects for the year of swap, age at swap and for the previous GP. We then divide the VA estimates into 20 equal-sized groups (vingtiles) and plot the means of this predictions within each bin against the mean value of each bin. The solid red line shows the best linear fit. The coefficients show the estimated slope of the best-fit line, with standard errors clustered at the pre-swap GP level reported in parentheses. The dashed line is the predicted value for mortality on our measure of VA and gender, fixed effects for the year of swap, age at swap and for the previous GP.

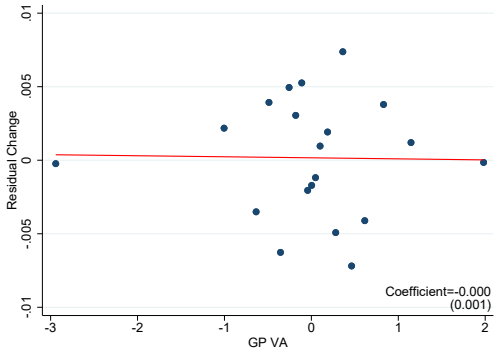
Figure A.5: Relation between VA and Pre-Swap Characteristics of Patients.



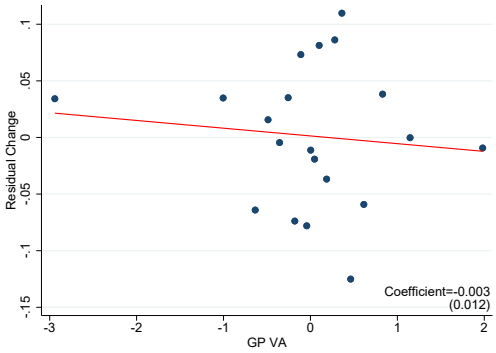
Note: To construct this binned scatter plot, we first residualize the variable with respect to a baseline control vector (gender, fixed effects for the year of swap, age at swap and for the previous GP.). We then divide the VA estimates into 20 equal-sized groups (vingtiles) and plot the means of the y-variable residuals within each bin against the mean value of each bin. The solid red line shows the best linear fit. The coefficients show the estimated slope of the best-fit line, with standard errors clustered at the pre-swap GP level reported in parentheses.

Figure A.5: Relation between of VA and Pre-Swap Characteristics of Patients (Cont).

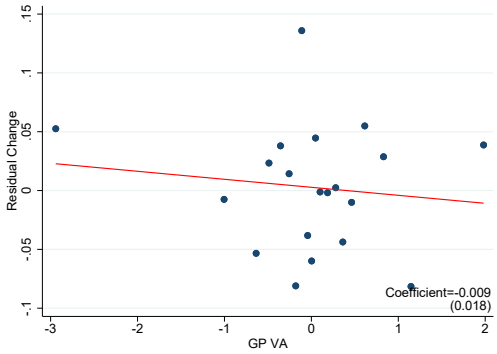
(g) Any Chronic Condition



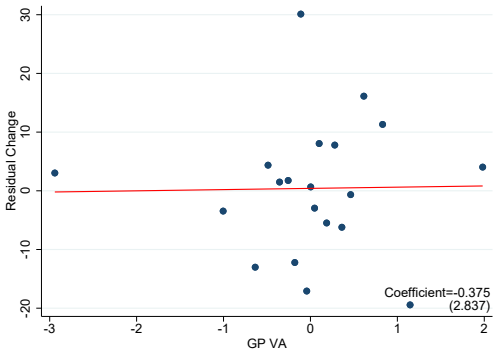
(h) Number of Chronic Conditions



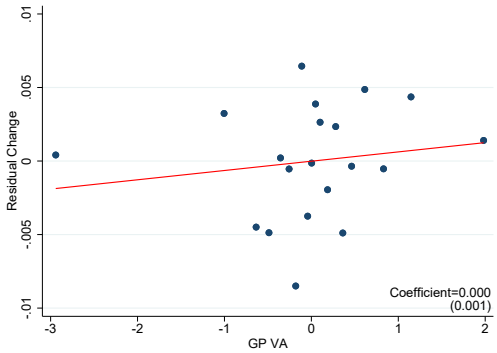
(i) Number of GP Visits



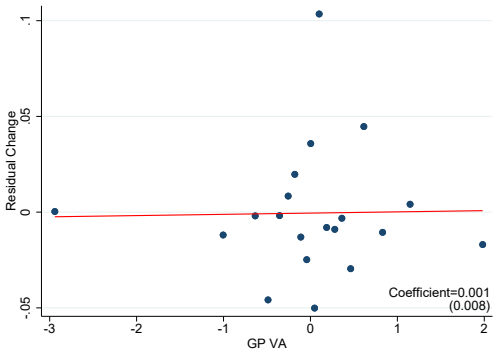
(j) Reimbursements



(k) Any Hospitalization

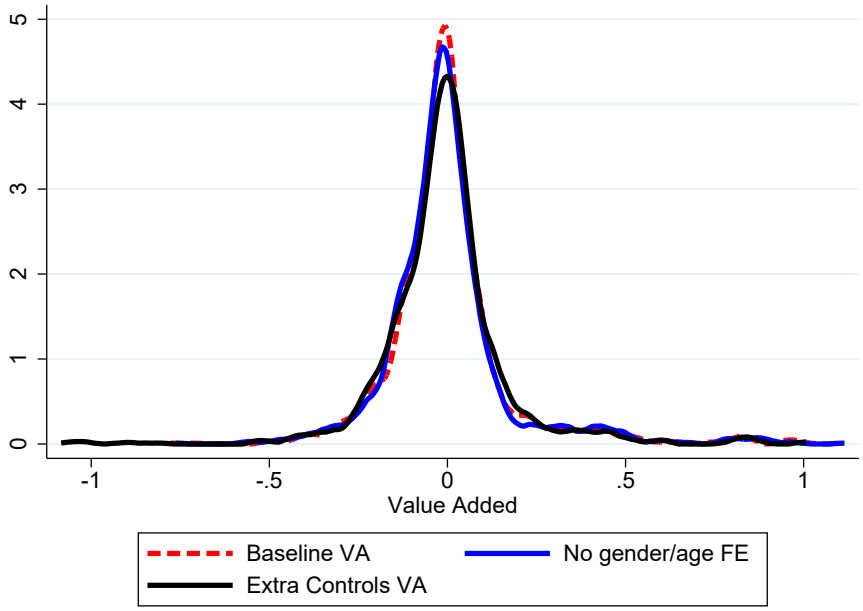


(l) Any Hospitalization



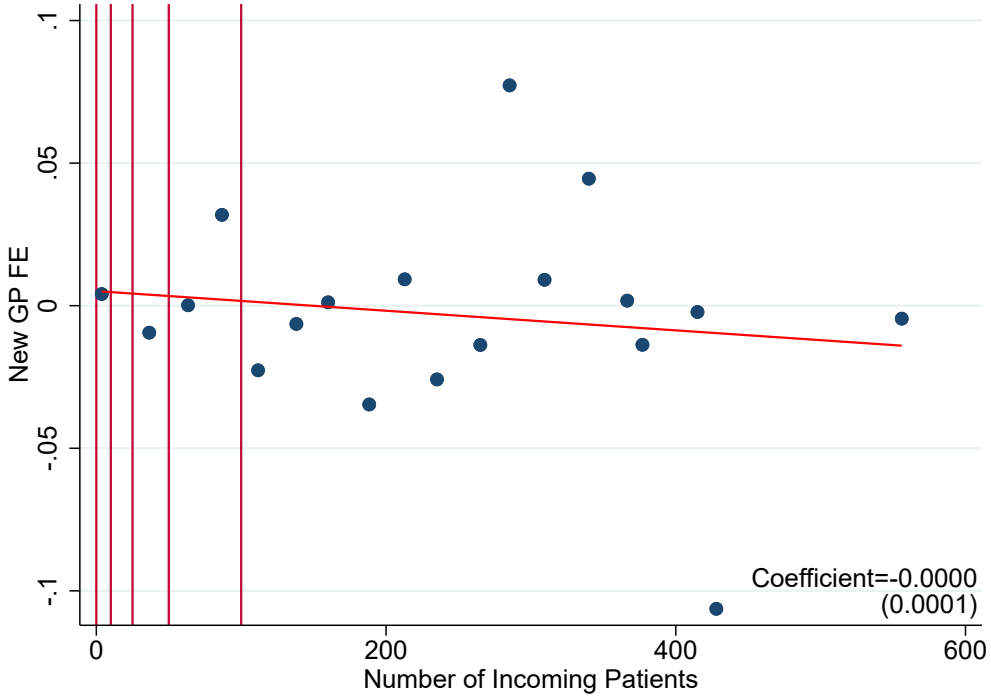
Note: To construct this binned scatter plot, we first residualize the variable with respect to a baseline control vector (gender, fixed effects for the year of swap, age at swap and for the previous GP.). We then divide the VA estimates into 20 equal-sized groups (vingtiles) and plot the means of the y-variable residuals within each bin against the mean value of each bin. The solid red line shows the best linear fit. The coefficients show the estimated slope of the best-fit line, with standard errors clustered at the pre-swap GP level reported in parentheses.

Figure A.6: GP Value Added Distributions: Heterogeneity by Model Specification



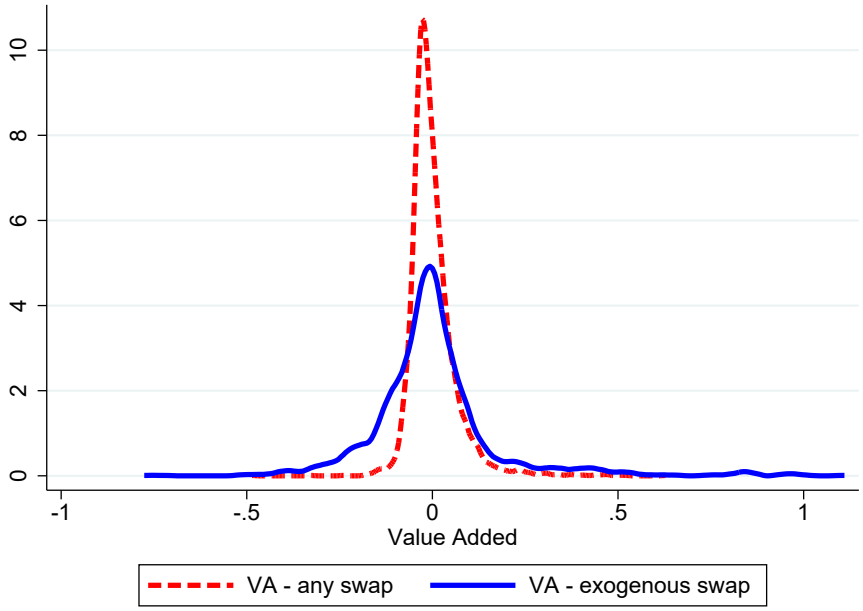
Note: The figure shows the distribution of the standardized VA measures based on different model specifications.

Figure A.7: GP Fixed Effects by Number of Incoming Patients.



Note: The figure plots the mean GP fixed effect by the number of incoming patients that underlie that specific GP VA estimate. Intervals are equally spaced and include groups of 25 patients. The vertical red lines are added to facilitate a visual inspection of the figure, and indicates that GPs with 0, 10, 25, 50, or 100, incoming patients were used in the construction of the VA.

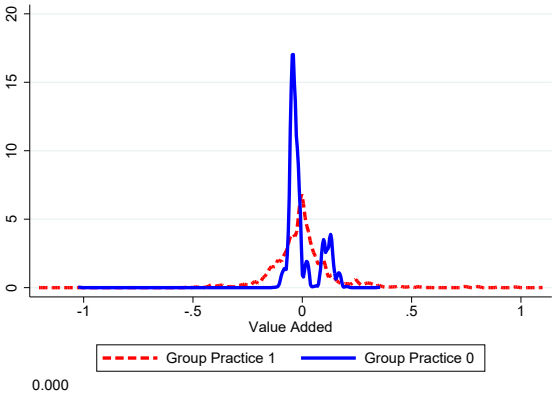
Figure A.8: Comparison of Standardized VA Distribution: All Swaps vs. Exogenous Swaps Only



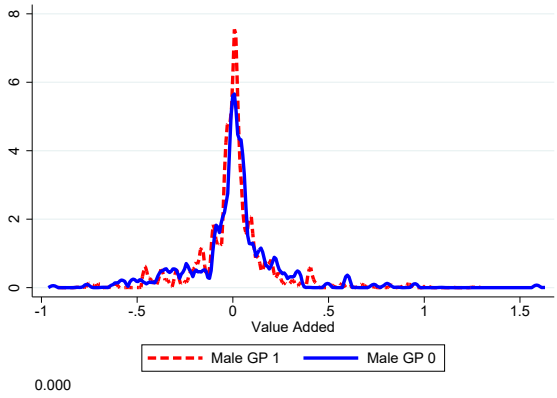
Note: The figure shows the distribution of the mortality-based standardized VA measure based on exogenous GP swaps only (solid blue line), and the distribution of the mortality-based standardized VA measure based on all GP swaps (dotted red line).

Figure A.9: GP Value Added Distributions: Heterogeneity by Doctor Characteristics

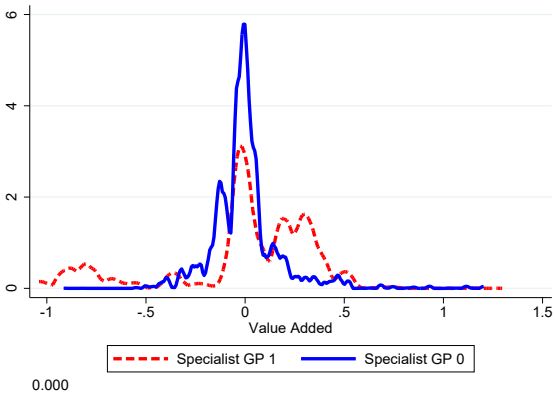
(a) Group Practice



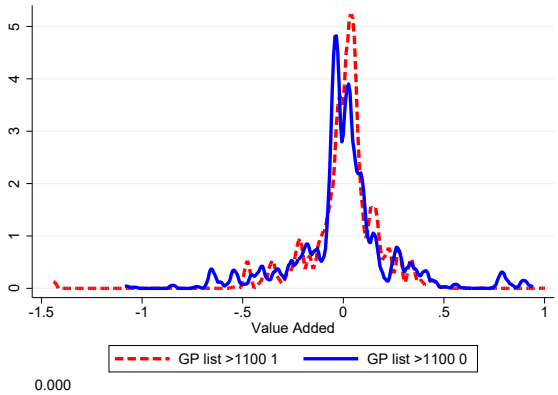
(b) Gender



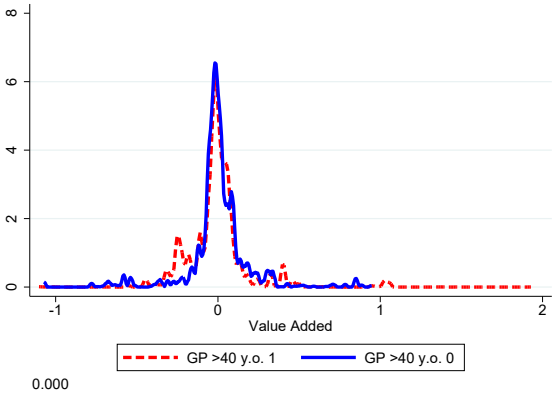
(c) Specialist



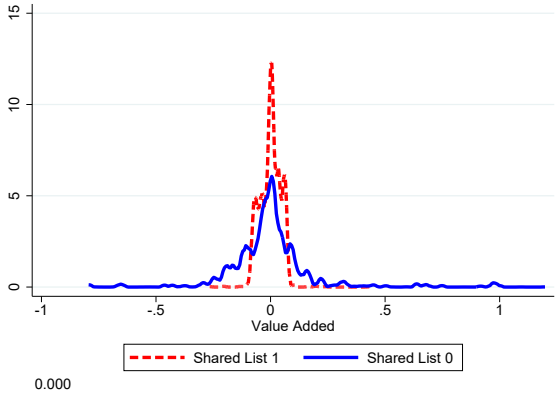
(d) GP List Length: 1100 or <1100



(e) GP Age: 40 or <40



(f) GP Shared List or Not Shared List



Note: The figure shows the distribution of the mortality-based standardized VA measure, separately for different GP characteristics. The number at the bottom left of each panel is the p-value for the Kolmogorov test for the difference in distributions.

Table B.1: Overview of Main Registers and Control Variables

Data source	Main description	Data used in analysis	Years available
Control and Payment of Health Reimbursement (KUHR) Database	Individual primary care visits (to GPs or emergency rooms)	Dates (year) of visits, with related diagnoses and symptoms (ICPC2), reimbursements and procedures	2006-2020
Norwegian Patient Registry (NPR)	Individual inpatient and outpatient visits in specialist care.	Entry & discharge dates (year), diagnoses (ICD10), planned and urgent admissions	2008-2020
GP-Patient Records	Information about each GP, incl. the in- and out-flow of patients on their list	Current list length, wanted list length, gender, age, specialisation, new patients on the list, reason for receiving new patient	2001-2020
Legelisten.no	Patients' subjective ratings of GPs	Ratings of GPs along several dimensions. We use the general score.	2012-2020
Mortality Records	Individual death event	Date (year) & cause of death (ICD10)	1996-2020
Education Records	Individual highest achieved education	Highest achieved education before GP-reassignment	1970-2019
Tax Authority Records	Annual information for eligible individuals (those formally employed).	Labor earnings and income from other sources	1993-2018
Social Security Records	Annual information for eligible individuals (those formally employed who falls ill)	Sick Leave, disability, unemployment, and social insurance	1992-2019
Population Records	Annual demographic information about all individuals	Marital status, municipality of residence, gender, age, nationality	1967-2020

Note: This table provides background information about the registers from which several of our key variables are derived.