WORKING PAPERS IN ECONOMICS

No. 02/22

RITA GINJA, JULIE RIISE, BARTON WILLAGE & ALEXANDER WILLÉN

Does Your Doctor Matter? Doctor Quality and Patient Outcomes

Department of Economics
UNIVERSITY OF BERGEN

Does Your Doctor Matter?
Doctor Quality and Patient Outcomes

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

June 2022

Abstract

We estimate doctor value-added and provide evidence on the distribution of physician quality in an entire country, combining rich population-wide register data with random assignment of patients to general practitioners (GPs). We show that there is substantial variation in the quality of physicians, as measured by patients' post-assignment mortality, in the primary care sector. Specifically, a one standard deviation increase in doctor quality is associated with a 12.2-percentage point decline in a patient's two-year mortality risk. While we find evidence of observable doctor characteristics and practice styles influencing a GP's value-added, a standard decomposition exercise reveals that most of the quality variation is driven by unobserved differences across doctors. Finally, we show that patients are unable to identify who the high-quality doctors are, and that patient-generated GP ratings are uncorrelated with GP value-added. Using a lower bound of the predicted value of an additional life year in Norway ($35,000), our results demonstrate that replacing the worst performing GPs (bottom 5 percent of the VA distribution) with GPs of average quality generates a social benefit of $27,417 per patient, $9.05 million per GP, or $934 million in total. At the same time, our results show that higher-quality GPs are associated with a lower per-patient cost.

JEL CODES: Value-added, health behaviors, mortality rate
KEYWORDS: H75, I11, I14, J18

* Ginja: University of Bergen, Institute for Fiscal Studies, and IZA (rita.ginja@uib.no). Riise: University of Bergen (Julie.riise@uib.no). Willage: University of Colorado – Denver (barton.willage@ucdenver.edu). Willén: Norwegian School of Economics, CESifo, and UCLS (alexander.willen@nhh.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), and the Essen Health 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

A key national policy objective is to create conditions for good and equitable health across the entire population. The average OECD country spends around 10 percent of GDP in pursuit of this goal every year. Despite this objective, there is substantial variation in the quality of healthcare services and patient outcomes. While some of these differences reflect differences in patient needs and preferences, others do not. Instead, they are due to variation in access to health care, variation in the quality of health care, and variation in medical practice style (OECD 2014). In this paper, we ask: how much of the observed disparities can be explained by variation in doctor quality? In spite of a large interest in understanding the importance of doctor quality in health production, a lack of exogenous variation in doctor-patient assignment coupled with limited data on doctor-patient interactions has prevented a comprehensive analysis on this topic.

The goal of this paper is to move beyond the existing healthcare literature and provide evidence on the role of physician quality in the production of health. To study this question, we rely on rich doctor-patient register data from Norway and exploit exogenous variation in general practitioner (GP) assignment for Norwegians. Our primary measure of doctor quality is the 2-year post-assignment mortality of patients (GP value added or GP VA), and our estimating equations are based on conventional value-added methods used to identify labor productivity in other sectors. Similar to most European countries, GPs in Norway represent the first point of contact with the country's healthcare system. They are responsible for initial examination, treatment, diagnosis, medication prescription, and sick note validation. When GPs retire or move – or for some other reason outside the patients control become unavailable – the Norwegian Health Economics Administration randomly reassigns patients to new local GPs conditional on municipality and availability. We use GP reassignments due to GP retirement or other causes outside the patient's control as a source of exogenous variation in doctor-patient assignment to identify the quality of doctors.[1]

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

[1] Individuals have the right to independently change the GP they have been assigned twice per year. However, using information on the exact cause of the GP swap, we ignore endogenous swaps initiated by the patients, and use only exogenous swaps caused by factors outside the control of the patients (GPs' moving, reducing their workload or retiring). See Section 2 for more details.

characteristics as well as patient demographics and long-run health and labor market outcomes. We focus on patients who are 55 or older as the mortality rate of Norwegians below this age is negligible.[2] To ensure that our results are not identified off of a systematic correlation between the characteristics of patients and the plausibly exogenously-assigned GPs, we incorporate a full set of previous GP fixed effects in all our regressions. The intuition underlying our estimation approach is thus to compare the mortality of patients who originally had the same GP, but who were exogenously assigned to new – and different – GPs due to factors outside their control.

After having estimated GP VA and explored the impact on patients as well as the potential mechanisms underlying the VA of GPs, 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. We do this by taking advantage of the 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 to 5. By correlating our objective measure of GP VA based on mortality with the patient-generated ratings of these GPs, we are able to investigate 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.

In terms of identification assumptions, our measure of GP quality is identified under the assumption that the exogenous GP reassignments– conditional on previous GP fixed effects – are uncorrelated with GP quality and patient health. In theory, the validity of this assumption follows from the fact that the Norwegian Health Economics Administration randomly reassigns patients to new local GPs in the event of GP retirement, moving, or downsizing. In practice, it is possible to obtain suggestive evidence on the validity of this assumption by examining if the reassignments are correlated with observed patient characteristics. Using a rich set of patient characteristics, we find strong evidence in support of this assumption.

We present six results. First, we reveal large differences in the quality of physicians across Norway. Specifically, there is considerable variation in the 2-year mortality rate across the doctor quality distribution, with a standard deviation of 0.122. This implies that a one

[2] For age-specific death rates in Norway, see https://www.ssb.no/256918/age-specific-death-rates-for-males-and-females

standard deviation change in doctor quality is associated with a 12.2-percentage point lower mortality rate two years after a patient has been assigned to the doctor. Interestingly, the magnitude of this effect is similar to the VA estimates found for teachers in the economics of education literature (e.g., Rockoff 2004; Rivkin, Hanushek and Kain 2005; Jacob and Lefgren 2008; Rothstein 2010; Chetty et al. 2014). Among those patients who die, we find that patients who were treated by GPs 1 SD higher up in the quality distribution lived, on average, 3.2 months longer. To put these results in perspective, it is worth noting that 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. Based on these numbers, our results imply that a 2.2 SD change in GP quality has the same health effect as moving from the bottom quartile to the top quartile location in the US in terms of mortality, highlighting the value of GP quality for understanding disparities in health outcomes and making health policy.

Second, we find evidence of specific observable doctor characteristics and practice styles influencing a GP's value-added. However, we also show that these variables only can explain a small amount of the overall variation in doctor quality. Specifically, a variance decomposition demonstrates that the residual variation is more than 90 percent in regressions that control both for GP demographic characteristics (gender, age (used as a proxy for experience), and specialization status) and GP practice style (consultation rate, laboratory diagnoses tests, minor surgical procedures, reimbursements, sick leave approvals, number of patients, and shared practice status). These results are similar to those on worker quality in other professions, where observable characteristics have little explanatory power (e.g., Hanushek 1992; Rockoff 2004; Rivkin, Hanushek and Kain 2005; Hanushek et al. 2005; Kane and Staiger 2008). 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, while we show that patients assigned to a high-quality GP do not receive a different *quantity* of treatment, we provide suggestive evidence that the VA 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 GP's ability to assign the right procedure to the specific patient. First, the mortality effects we identify are driven exclusively by a reduction in the likelihood of dying from treatable diseases. Second, there is a strong reduction in urgent care visits as a consequence of being assigned a high-VA GP. Third, there is an increase in the likelihood of being sent to a specialist for treatable diseases following a visit to a high-VA GP. These results are consistent with the notion that high quality GPs may be more effective at early detection and in identifying the right type of treatment required for the patient. This is further supported by the lower patient cost associated with high-quality GPs (which is 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).

Fourth, to understand if doctor quality has an impact on individual outcomes beyond health and healthcare utilization, we merge the patient data to rich population-wide labor market registers. These data allow us to examine if doctor quality has an impact on the career prospects of individuals, examining outcomes such as employment status, labor market earnings, total income, sick leave benefits, and welfare receipt. 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. We conclude that the benefits associated with raising GP quality mainly operate through improved patient health. However, we note that the lack of labor market effects could be driven by our need to focus on patients who are 55 or older, and encourage caution in extrapolating these results to the entire population.

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 GP twice per year, and examining (1) if individuals' post-assignment decisions to endogenously switch GP is driven by the VA of their exogenously-assigned GP, and (2) if GPs with a higher VA are more likely to experience an influx of endogenous patient switchers. 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 the rich information on GP practice style and behavior that we have access to. We show that patients who endogenously switch do so to get more experienced GPs who hold a specialization. This sorting behavior 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 that they value aspects of GPs other than their ability to save lives. Understanding the relationship between our measure of GP VA based on mortality 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 the extent that the lack of correlation between objective and subjective ratings extend to other fields outside the health care section, these results also have important policy implications for other occupations and domains in which user-ratings are used as a way to convey quality (e.g., teachers, social workers, professors, schools).

To quantify the magnitude of the GPs' impacts on health production, 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 replacing the worst 5 percent of GPs with GPs of average quality would generate a 9.4-month increase in their life expectancy. 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 $27,417 per patient, $9.05 million per GP, or $934 million in total. Thus, while the value associated with each individual patient is relatively modest, the aggregate effect is substantial. At the same time, our results demonstrate that the practice styles and behaviors of higher-quality GPs are associated with a lower per-patient cost. 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, we note that this conclusion depends on the cost of raising a GP's VA from the bottom to the middle of the distribution; we see this as an extremely valuable avenue for future research.

Our analysis advances the small but rapidly growing literature on doctor effectiveness. Within this literature, most studies have focused on exploring the effect of provider practice style, hospitals, and treatment intensity on patient's short-term outcomes, offering important insights on the role of social institutions in the production of healthcare quality (e.g., Currie and Schnell

2018; Currie at al. 2016; McClellan et al. 1994; Bartel et al. 2014; Doyle et al. 2010). However, 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 through the use of conventional VA measures (Fletcher et al. 2014; Stoye 2022).[3] While providing novel evidence on the quality of attending physicians, limitations have forced these studies to focus on particular hospitals and narrow specialties. We add to this literature by identifying the distribution of physician quality across an entire country and its impact on later-in-life outcomes of patients, using exogenous variation in patient-doctor matches. Knowing the impact of physicians on their patients is important for understanding the relative benefit of focusing on – and investing in – physician quality relative to other inputs that enter the health production function. It also serves an important role for establishing effective doctor compensation schemes and retention policies. We therefore see our results as 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.[4]

Our study also makes an important contribution to the literature on patient choice in the health care sector. Most of this literature has focused on patients' choice of hospitals rather than individual doctors, and finds that hospital quality is associated with increased patient demand (e.g., Cutler et al. 2004; Howard 2005; Ho 2006; Pope 2009). A limited set of studies have also 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 2015). Finally, a novel set of studies have investigated the effects of patient-generated ratings on health care choices (Luca and Vats 2013; Chan 2022; Bensnes and Huitdeldt 2021).[5]

[3] Specifically, Fletcher et al. (2014) estimates the value-added of physicians at one hospital during one year in the US. However, while providing novel evidence on the quality of attending physicians, the authors do not have access to random allocation of patients to doctors, cannot observe patient outcomes after the patients have left the hospital, and work with a limited sample. Stoye (2022) focuses on the quality of cardiologists based on their ability to treat patients with high-risk cardiovascular conditions (myocardial infarctions) in emergency room settings.

[4] Our paper can also be viewed as extending the large overall value-added literature to the healthcare sector, enabling us to understand if the worker productivity effects that have been identified in other sectors such as teaching (e.g., Hanushek and Rivkin 2010, Andrabi et al. 2011, Chetty et al. 2014, Mulhern 2021) and social services (e.g., Behncke et al. 2010, Carneiro et al. 2020, Cattan et al. 2021) extend to doctors. The healthcare sector represents one of the biggest ticket items in a country's budget, with the average OECD country spending around 10 percent of their GDP on healthcare services. Being able to extend the worker productivity literature to the healthcare sector therefore has important policy implications. In terms of societal impact, examining doctor quality also has added value as it is relevant for the entire population, not just for specific groups.

[5] That individuals are responsive to information about care quality has been documented in several other setting as well. For example, a rich set of studies demonstrates that the introduction of report cards for health insurance plans

We contribute to this literature by examining whether individuals are able to identify, and select, GPs that have a higher VA. While the user-generated quality measures that have been studied in the prior literature are more readily available to patients, they are correlated with other dimensions of the patient-doctor interaction, and it is unclear to what extent these indicators represent objective measures of GP quality. Our results reveal that patients do not select physicians based on their VA, and that the correlation between objective GP VA and subjective patient-generated GP ratings is very small. These findings have important implications for ongoing policy debates on the organization of health care services and the distribution of information to patients.

More broadly, there is a rich literature on cause-specific mortality and morbidity,[6] 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). Our paper contributes to these strands of research by providing an additional dimension to the literature, demonstrating to what extent primary care physicians – and variation in the quality of primary care physicians – contributes to mortality and morbidity. 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 individuals. Specifically, if there is significant variation in the quality of physicians even within a given location, and if different subgroups of individuals have access to differently-qualified doctors, then access to high quality doctors may be an important mechanism underlying health inequalities in society; even in egalitarian societies when there is universal healthcare provision.

In terms of contextualizing our results, it should be noted that our GP VA measure is based exclusively on patient mortality among individuals aged 55 and above. The key advantage of this approach is that it provides an unambiguous and objective measure of quality directly aligned with

---

and hospitals in the US led patients to choose high rated plans and providers (Scanlon *et al*., 2002; Wedig and Tai-Seale, 2002; Kolstad and Chernew, 2009).

[6] For example, Clemens et al. (2014) examine the impact of physicians' financial incentives on patient health, Abaluck et al. (2021) study the competition effect of private health plans on patient welfare, McClellan et al. (1994) explore the impact of hospital treatment intensity on myocardial infarction mortality, Bartel et al. (2014) look at productivity effects of teamwork in the hospital sector, Currie at al. (2016) explore provider practice style and its effect on patient outcomes, Geruso et al. (2021) exploit random assignment of Medicaid beneficiaries to managed care plans to study plan-specific effects on healthcare, and Finkelstein et al. (2016; 2021) explore how much supply factors can help explain variation in morbidity.

the goal of the healthcare system; to reduce mortality and extend the lives of patients. In addition, it allows us to directly compare our results with the effects in Finkelstein et al. (2016; 2021) as well as Chetty et al. (2016), who have both focused on mortality effects among the older population. However, healthcare quality and health provision are multifaceted, and the usefulness of our VA measure will ultimately depend on the objective function of patients and policymakers.[7] We thus see the paper as opening up a new avenue of research on the multidimensionality of doctor quality, and how this multidimensionality interacts with the societal health objectives.

The rest of this paper proceeds as follows: In Section 2, we provide information on the healthcare system in Norway. In Section 3, we introduce our data and empirical method. In Section 4, we present the main results from our analysis. In Section 5, we provide a back-of-the-envelope calculation of the overall social benefit associated with improved GP quality. In Section 6, we conclude.

2. Background

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. Specifically, a one-time consultation has a copayment of approximately $18, and the maximum that an individual can spend on healthcare in a year is approximately $280.[8] Children under the age of 16 are exempted from copayments.

Similar to most European countries, the healthcare system in Norway is divided into two levels, with local municipalities providing primary care services and larger health regions providing specialist care.[9] Entry into specialist care and hospital services can only be obtained through referrals from the GP in the primary care sector (except for emergencies). Norwegian GPs therefore represent the first point of contact between patients and the healthcare system, and are

[7] For example, if the goal of the patient is to get a GP that is willing to validate a sick leave note, our quality measure will be of little interest to that patient. In addition, if a policymaker is concerned with ensuring health access to the population, rather than to minimize mortality, our measure would also be of little interest. However, the goal of this paper is to obtain a measure of GP quality, and we leave it to future work to investigate these alternative societal goals.
[8] For reference, see https://www.commonwealthfund.org/international-health-policy-center/countries/norway. Note that the numbers provided in-text refer to the year 2015. These numbers have been relatively constant in the last 30 years (see https://data.oecd.org/healthres/health-spending.htm).
[9] There are currently 422 municipalities and 4 health regions in Norway.

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.[10]

Beginning in 2001, 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, though 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 their assigned GP (though the GP will be informed and updated on the condition of the patient). Prior to the introduction of the GP scheme in 2001, individuals were not assigned to a particular GP, and had to find a new GP every time they needed care (Riise, Willage, and Willén 2022). This proved to be a relatively cumbersome and time-consuming exercise, and the GP scheme was meant to eliminate this inefficiency while at the same time improving doctor-patient relations. 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 around 1100 patients. The average GP was 47 years old, and 60 percent were male.

Most GPs are self-employed, with less than 5 percent of GPs being salaried municipality employees. 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 are among the most well-paid professions in the country.[11]

When GPs retire, move, or for some other reason decide to terminate/reduce their current practice, patients on the GP's list are reassigned to new GPs in the municipality. As described in Riise, Willage, and Willén (2022), there are two important aspects of this process. First, in the event of list reductions, which patients to be removed from the list must be randomly determined. Second, in the event of reassignment, patients should be randomly assigned to new GPs in the

[10] See https://www.commonwealthfund.org/international-health-policy-center/countries/norway for a more detailed introduction to the Norwegian healthcare system.
[11] See https://www.ks.no/fagomrader/statistikk-og-analyse/lonnsstatistikk-for-ks-tariffomrade/lonn-og-inntekter-for-fastleger/ for more information on GP wages and earnings.

municipality conditional on availability. It is worth noting that in certain cases, entering GPs can take over the entire list from a retiring GP.

To causally identify GP quality, we exploit GP reassignments induced by GP retirement or other causes that are outside the patient's and the new GP's control as a source of exogenous variation in doctor-patient match. We do not use the initial assignments, nor any swaps initiated by the patients, due to endogeneity concerns. In addition, we always include previous GP fixed effects, such that our quality measures are identified off of patients who had the same GP but then those patients were exogenously allocated to new, and different, GPs. Thus, should a new GP take over the entire patient list of a retiring GP, those patients will *not* contribute to our identification.[12] The intuition underlying our estimation approach is thus to compare the outcomes of patients who originally had the same GP, but who got exogenously assigned to new – and different – GPs due to factors outside their control. In our sample, each previous GP transfers patients to on average 7 new GPs (SD 10.8); in each of these transfers, an average of 11.5 patients are moved from a previous to a new GP (Appendix Table A.1). 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.

3. Data and Method

3.1 Data

Our analysis data come from administrative registers covering the universe of Norwegian residents. A unique individual identifier enables us to trace individuals across the various registers and build a rich data set encompassing all GP-patient interactions and behaviors, as well as individual demographics, labor market information, and family characteristics.

GP Characteristics. 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*), 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

[12] The reason is that the previous GP fixed effects will completely consume any variation between these patients.

from the government. ICPC-2 codes convey information about the GPs' assessment of the patient's health problems and the type of care provided.[13] Additionally, the database contains information on 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. Taken together, these two data sets provide us with comprehensive information on the characteristics of all GPs in the country, their practice styles and behaviors, and which patients belong to which GPs.

We link the above data to information on the use of specialist health care services from the Norwegian Patient Registry. This registry contains information on all patients who have received treatment from specialist health services. We use this data to study impacts on hospitalizations (inpatient admissions) and consultations at outpatient clinics.[14]

Appendix Table A.2 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 is 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, and 63% are males. The average GP is responsible for 1108 patients. Comparing columns (1)-(2) with columns (3) through (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, for our study we are interested in GP swaps that are outside the patient's control, which generates plausibly exogenous variation in

[13] 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.
[14] An inpatient admission includes both overnight stays and day treatments, such as less invasive surgical procedures.

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. This information is provided directly in the Norwegian GP Register. In Section 3.2, we provide evidence consistent with the notion that patient characteristics are uncorrelated with the VA of the newly assigned GP.

GP quality. Our primary measure of doctor quality is the 2-year post-assignment mortality of patients aged 55 and above. 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. The two choice parameters that make up our quality measure (the decision to focus on mortality and the decision to focus on individuals above age 55) deserve some discussion.

First, with respect to our decision to focus on mortality, the key advantage of this approach is that it provides an unambiguous and objective measure of quality directly aligned with the goal of the healthcare system – to save lives. In addition, it allows us to directly compare our results with the mortality effects in Finkelstein et al. (2016; 2021) and Chetty et al. (2016), who have both focused on mortality effects among the older population. 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. We strongly encourage future research to explore such alternative indices.

Second, with respect to the age restriction, we impose this restriction because the mortality rate of Norwegians below this age is negligible. In addition, morality below this age is most often characterized by accidents and injuries that could not have been avoided through repeated interactions with GPs.[15] Appendix Figure A.1 illustrate these points in detail. While this choice does affect our ability to discuss the role of GP quality among younger individuals (unless we are willing to impose relatively strong assumptions), it does equip us with a unique opportunity to investigate – and provide novel information on – the role of GP quality in influencing the health and wellbeing of a large and relevant subsample of the population. In addition, it allows us to directly relate our results to both Finkelstein et al. (2016; 2021) as well as Chetty et al. (2016).

[15] For age-specific death rates in Norway, see https://www.ssb.no/256918/age-specific-death-rates-for-males-and-females

Panel A of Appendix Table A.3 provides descriptive statistics on the mortality of patients within 2 years after having been assigned to their GP, categorized by broad cause of death categories. 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 understand the composition of the patients that are used to calculate the GP VA, we follow patients across the administrative registers and collect key demographic background information. The demographic and socioeconomic data is 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. To understand if doctor quality has an impact on individual outcomes beyond health and healthcare utilization, we also merge the patient data to population-wide labor market registers. These data allow us to examine if doctor quality has an impact on the career prospects of individuals, examining outcomes such as employment status, labor market income, disability benefits, and welfare receipts. Information on earnings is from the tax registers and information on welfare benefits comes from the social insurance database.

Panel B of Appendix Table A.3 provides descriptive statistics on the demographics of the patients that are used to calculate the GP VA. On average, the individuals in our swap sample are 64 years old, 54 percent are male, 97 percent are Norwegian-born, about 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.3 provides descriptive statistics on the healthcare utilization of patients in our sample. On average, individuals visit GPs 5 times a year. Approximately half of these visits are short consultations lasting at most 15 minutes. Conditional on going to the GP, an average of 1.6 lab tests are conducted (e.g., blood tests). In addition, there is a 40% likelihood of having a sick leave certification provided to you by the GP in a given year.

Panel D of Appendix Table A.3 provides descriptive statistics on the labor market situation of patients in our sample.[16] On average, individuals have an annual labor income of 386,292 NOK (about $46,000), and older individuals have an annual labor income of 183,041 NOK ($27,000).

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

About 20 percent of the patients receive some welfare benefit, and the average number of sick leave days in a given year – conditional on working – is 18.[17]

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 to 5. 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 (see Appendix Table A.4).

It is important to note that there is an endogenous selection of patients who sort into providing a GP rating on this website, and that 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. It is therefore not about the composition of individuals who rate the GPs, but about 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. Appendix Table A.4 provide descriptive statistics on the ratings.

3.2 Method

Measure of doctor quality. To measure doctor quality, we exploit a unique aspect of the Norwegian health care system in which patients are randomly allocated to new GPs in the event their current GP closes down or significantly reduces their practice. This enables us to obtain measures of doctor quality that are not identified off of a systematic correlation between patient characteristics and GPs. We estimate the following equation:

$$h_{ijkt} = \mu_j + \theta_t + \pi_k + \varepsilon_{ijkt}, \qquad (1)$$

[17] Information on welfare dependency comes from the social insurance database, which we use to construct an indicator of whether an individual received any welfare benefits each year. Welfare benefits include social assistance, unemployment benefits, work assessment allowance, and disability insurance.

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, and μ_j represent the exogenously-assigned GP fixed effects.[18] Under the assumption of conditional random assignment of patients to doctors, $cov(\mu_j, \varepsilon_{it} | \pi_k, \theta_t) = 0$, the μ_j vector provides an unbiased estimate of the quality of each GP in our sample as measured by the 2-year post-assignment mortality of his/her patients.

The inclusion of pre-swap GP fixed effects in Equation (1) means that the GP VA is identified off of a set of patients who had the same previous GP but then were randomly allocated to new and different GPs – at the same time – due to factors orthogonal to their health characteristics and the newly-assigned GP's quality. Thus, should a new GP take over the entire list of a retiring GP, those patients would not contribute to our identification.

One challenge while 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 follow Kane and Staiger (2008) and Chetty, Friedman and Rockoff (2014). Specifically, we construct a Bayesian empirical estimator by adjusting the estimated VA according to the following equation:

$$BE_j = \lambda_j VA_j, \qquad (2)$$

where the "shrinkage" factor is $\lambda_j = \sigma_u^2 / (\sigma_u^2 + \sigma_\varepsilon^2 / n_j)$ 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. Thus, to correct for sampling error we use the fact that we observe the full load of patients for a GP.

Identifying assumptions. Our measure of GP quality is identified under the assumption that the plausibly exogenous GP reassignments we exploit – conditional on previous GP fixed effects – are uncorrelated with GP quality and patient health: $cov(\mu_j, \varepsilon_{it} | \pi_k, \theta_t) = 0$. In theory, the validity of this assumption follows directly from the fact that the Norwegian Health Economics Administration randomly reassigns patients to new local GPs conditional on

[18] We do not directly include municipality fixed effects as they are perfectly absorbed by the previous GP fixed effects.

municipality and availability in the event of GP retirement, moving, or downsizing. In practice, it is possible to obtain suggestive evidence on the validity of this assumption by examining if the VA of the newly-assigned GP is correlated with observed patient characteristics. To this end, we conduct an extensive balancing test in which we regress our estimated doctor quality measure on a rich set of observable patient characteristics determined prior to the swap: education, immigrant status, number of children, employment status, income, unemployment insurance benefits, urbanicity, birth order, and pre-swap GP visits. Results from this exercise are provided in Table 1. All coefficients are economically small and no coefficient is statistically significant even at the 10 percent level. This provides strong support in favor of our identifying assumption.

Estimating impact on patients. After obtaining estimates of GP quality and providing evidence in support of the required identification assumptions, we leverage these estimates to examine the effect of GP quality on patient outcomes. We consider the impact not only on the health and health utilization of the individual patient, but also on the long-run labor market performance of these individuals. Specifically, we estimate the following equation:

$$w_{ijkt} = \beta\psi_j + \theta_t + \pi_k + \varepsilon_{ijkt}, \qquad (3)$$

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). To avoid a mechanical relationship between our VA measure and the patient outcomes that we investigate, ψ_j is based on a leave-on-out method in which we exclude individual *i* from the VA calculation when examining the impact of GP quality on individual *i*'s outcomes. The coefficient β is the effect of GP quality on the outcomes we examine. All other variables are defined as above. Provided that there is no systematic allocation of patients to new GPs of different quality, something which we provide support for in Table 1, Equation (3) enable us to estimate the causal impact of physician quality on patient short-and long-term health and labor market outcomes.

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 there is no difference between the raw and the Bayes adjusted distributions of GP VA. This is perhaps unsurprising, as each GP in our sample has a large number of patients that contributes to the calculation of the VA. This serves to minimize any measurement error that the shrinkage approach is used to adjust for.

To facilitate the interpretation of the value-added measure displayed in Figure 1, row (A) of Appendix Table A.5 provides statistics on key moments of the GP VA distribution. Looking across the columns of row (1), we see that a GP in the bottom decile of the quality distribution has a 11 percentage points higher mortality rate than the median doctor, and that a GP in the top decile of the quality distribution has a 12 percentage points lower mortality rate than the median doctor. More specifically, the standard deviation of GP value-added based on the 2-year post assignment mortality rate is 0.122. This suggests that a one standard deviation change in the quality of a patient's GP is associated with a 12.2-percentage point increase in the probability of being alive two years after GP assignment.

When interpreting the results from our GP VA investigation, a helpful exercise is to re-estimate the GP VA based on all swaps, endogenous as well as exogenous. This provides us with important information on the likely bias in doctor quality estimates induced by endogenous patient-doctor sorting. The results from this exercise are provided in row (B) of Appendix Table A.5. As expected, the quality distribution is considerably more compressed when including the endogenous swaps, biasing the results towards finding less variation in GP quality than there actually is.

In terms of policy implications, the variation in GP VA is similar to the estimates found for teacher effects in previous work (e.g., Rockoff 2004; Rivkin, Hanushek and Kain 2005; Jacon and Lefgren 2008; Rothstein 2010; Chetty et al. 2014). In addition, the GP quality variation in this analysis is associated with greater differences in patient mortality than more traditional health interventions such as altering physicians' financial incentives (e.g., Clemens et al. 2014), and similar to the place effects on mortality discussed in Finkelstein et al. (2021). This implies that there may be substantial welfare gains associated with focusing more on GP quality when attempting to bolster health production and creating conditions for good and equitable health across the population. However, we note that this depends on the cost of raising a GP's VA from the bottom to the middle of the distribution; we see this as an extremely valuable area for future research.

To study if the variation in GP quality differs across specific subgroups of physicians, Appendix Figure A.2 and Appendix Table A.6 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 relative to their male counterpart. The same pattern, though slightly less extreme, also 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.[19]

The results displayed in Figure 1 and Appendix Table A.5 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 associate with GP quality as well. Specifically, conditional on passing away, do patients who are randomly assigned to higher quality GPs live longer? 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 within two years of being assigned (exogenously) to a new GP. The results from this exercise are provided in Panel A of Table 2, and demonstrate that exogenous assignment to a high-quality GP has a substantial impact on a patient's intensive margin survival probability as well. Specifically, being assigned to a GP whose VA estimate is 0.122 higher (one SD higher in the VA distribution) increases the age of death with almost 3.2 months. Thus, high-quality GPs do not only reduce the mortality rate of patients at the margin of dying, but they also extend the life of those patients who do pass away within two years post-assignment.

[19] On the bottom of left of each graph we include the p-value for the Kolmogorov-Smirnov test for the equality of the distributions for two groups. We reject the equality between of distributions for all of the groups studied.

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. To this end, columns (1) through (3) of Table 3 provide results on the explanatory power obtained from estimating a series of regressions 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, list length, and shared list status. This is the full set of demographic characteristics that we are allowed to link to GPs in our data.

The results in columns (1) through (3) of Table 3 provide relatively little support for the notion that a GP's underlying demographic characteristics can help identify their VA, with each model having a very modest R^2. Having said that, three characteristics are found to be statistically significantly associated with GP quality: age, specialization status, and list length.

First, a 10-year increase in GP age reduces the 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 reason 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 a specialization reduces the GP VA by 0.11 SD. When interpreting this finding, it is important to note that we examine GP quality in the primary care sector, where GPs encounter patients with a broad range of health issues. This result should therefore not be interpreted as showing that specialists are of a lower quality than non-specialists, only that they appear to be of lower quality when dealing with the broad spectrum of health issues that they are exposed to in the primary care sector.

Third, having an additional 100 patients is associated with an increase in the GP VA by 2 percent of a SD.[20] 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 columns (1) through (3) 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 the demographics explored in columns (1) through (3) in this table.

GP Behavior and Practice Style. The lack of a strong association between GP VA and observable GP characteristics demonstrates that these variables can explain only a small amount of the overall variation in doctor quality. However, in addition to observable demographics GP characteristics, we also have information on GP behaviors and practice styles. To this end, columns (4) and (5) of Table 3 report the results from examining the relationship between GP consultation behavior and our VA measure.

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 very difficult to infer GP VA based on their 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 2 percent of the variation in GP VA (column (5)).

Overall explanatory power. Having identified significant variation in GP VA across Norway, and demonstrated that the majority of this variation is driven by unobserved characteristics and behaviors, we ask how important GP VA is for explaining the overall mortality of individuals in our sample. This exercise is instructive for understanding the overall role of GPs in the production of health, and to what extent improvements in GP VA can help reduce overall mortality. To this end, we perform a variance decomposition exercise in which we examine the relative importance of GP VA, as well as a rich set of observable characteristics of the patients,

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

in explaining the variation in patient mortality. The results from this exercise are shown in Appendix Table A.7.

The results from this exercise demonstrate that GPs overall can explain about 12 percent of the mortality among individuals in our sample (Column 2). This is a substantial amount, equivalent to the explanatory power of mortality rates across neighborhoods in the US, which Finkelstein et al. (2021) find to be 15 percent. Thus, most of the quality differences across neighborhoods in terms of health could be driven by care quality (though we note that our study takes place in a different country and context). Another interesting result in Appendix Table A.7 is that the explanatory power of the GP VA is larger than the explanatory power of observable patient characteristics (comparing the change in R-squared going from column (2) to column (1) with the change in R-squared going from column (2) to column (3)). GP quality therefore appears at least as important for identifying mortality risk as does a patient's underlying characteristics.

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. Specifically, assignment to a high-quality GP does not only generate an improvement on the extensive margin of mortality (probability of dying), but also on the intensive margin (age of death conditional on dying). 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. Specifically, how does the quality of the GP impact the short-and long-term health of patients, and how does it impact other individual outcomes such as employment and earnings?

Access to care and health care utilization. In terms of the health care of patients, Section 4.2 shows that the GP practice behavior has a limited ability to explain variation in GP VA. However, assignment to a high-quality GP may impact the way in which the patient interacts with the health care system, and this could provide valuable information on the mechanisms through which GP quality impacts patients.

To this end, Tables 4 and 5 explore extensive and intensive margin effects of GP treatment (Table 4) and of hospitalization (Table 5) associated with assignment to a better-quality GP. 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 effect on our estimates, we examine treatment and hospitalization outcomes immediately following the swap, a timeframe in which the majority of patients in our sample are still alive. However, we still encourage caution when interpreting the results in these tables.

Abstracting away from the above issues, the results in Table 4 indicate that patients assigned to a high-quality GP do not receive an economically meaningfully different *quantity* of treatment. This suggests that the VA variation across GPs is coming from differences in quality, or effectiveness, of the GPs. This is an interesting finding which we explore in greater length below, and which paves the way for future research on the interactions of patients and doctors during health visits. That high-quality GPs may be more effective, or more able to immediately assign the right task to the right 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 care as a consequence of being assigned a high-quality GP. Specifically, a 1 standard deviation increase in GP VA is associated with a three 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. 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.

To explore these health care visit effects in greater detail, we examine the type of diagnoses issued by specialists during patient visits at the hospital in the year after the GP visit. The results from this exercise are shown in Table 6. The results illustrate that there is a decline in severe diagnoses (e.g., malign cancers) and an increase in less severe diagnoses (e.g., benign cancers), 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 right patient.

The results in Tables 4 through 6 suggest that GP VA is not 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 that are driving the mortality reductions associated with high quality GPs. Results from such an exercise should show that causes of death that can benefit from early detection and treatment are driving the GP VA mortality effects (such as treatable cancers or cardiovascular diseases). To explore this question in detail, 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 on cancer detection.

Labor market and welfare effects. To examine if the health impact that high-quality GPs have on patients translate to other types of outcomes, such as employment and labor earnings, Table 7 provides results from estimation of Equation (3) using a range of labor market outcomes as dependent variables. In Panel (A), we present result 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. Specifically, even if some of the point estimates are economically meaningful, only one of them is statistically significant even at the 10 percent level. We therefore conclude that the benefits associated with raising GP quality mainly operate through improved patient health. However, we also highlight that our sample consist of patients aged 55 and over, and that the labor market effects may be very different among young patients who have most of their labor market careers ahead of them.

4.4 Patients' Ability to Identify High-quality GPs

Previous literature on patient choice in the healthcare sector has primarily focused on patients' choice of hospitals (e.g., Cutler et al. 2004; Howard 2005; Ho 2006; Pope 2009). A limited set of studies have also 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 2015). Finally, a novel set of studies have examined the effects of patient-generated ratings on health care choices (Luca and Vats 2013; Chen 2019; Bensnes and Huitdeldt 2021).[21] However, very few

[21] That individuals are responsive to information about care quality has been documented in several other setting as well. For example, a rich set of studies demonstrates that the introduction of report cards for health insurance plans and hospitals in the US led patients to choose higher rated plans and providers (Scanlon *et al*., 2002; Wedig and Tai-Seale, 2002; Kolstad and Chernew, 2009).

studies have been able to explicitly relate objective measures of doctor quality to patient choices, most likely due to lack of data on doctor quality linked to patient decisions.

In this section, we advance this literature by examining whether individuals are able to identify, and select, GPs that have a higher VA relative to their current GP. We do this by exploiting the fact that individuals are allowed to independently change GP twice per year, and examining if (1) patients' post-assignment decisions to endogenously switch GP is driven by the VA of their exogenously-assigned GPs, and (2) if GPs with a higher VA are more likely to experience an influx of endogenous patient switchers. The results from this exercise are in Table 8 and 9. Table 8 shows the probability of endogenously switching GP (after an exogenous swap) as a function of the quality of the exogenously-assigned GP (analysis conducted on the patient level). Table 9 shows probability that a GP receives endogenous patient swappers as a function of his/her VA (analysis conducted on the GP level).

The results in column (1) of Table 8 illustrate that patients do not select physicians based on their VA, and Table 9 illustrate that high-quality GPs do not receive a disproportionate influx of new patients. 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 column (2), we therefore exploit rich 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 do not hold a specialization and who more frequently refer patients to specialized care. 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 Tables 8 and 9 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 than the GPs ability to keep them alive (e.g., willingness to approve sick leave notes, willingness to prescribe medicines, e.tc.)

4.5 Patient-generated GP Ratings and GP VA

To better understand patients' ability to identify high-quality doctors as measured by their ability to keep their patients alive, we merge our administrative data on GPs with proprietary data on patient-generated ratings of GPs. We do this by taking advantage of the privately-run online review platform, *legelisten.no*, which was launched in 2012. The website allows patients to anonymously rate their overall satisfactory with their GP on a scale from 1 to 5. We then correlate our objective measure of GP VA based on mortality with the patient-generated ratings of these GPs.

When interpreting these results, it is important to note that there is an endogenous selection of patients who sort into providing a GP rating on this website, and that 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 about the composition of individuals who rate the GPs, but about 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 10. The results reveal a very 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 10 demonstrate that the patient ratings are correlated with a number of GP characteristics, including age, specialization status, list length, and gender.

The results in Table 10 are consistent with our finding on individuals' post-assignment decisions to endogenously switch GP, and with the notion that patients either are unable to identify high-quality GPs, or that they value other aspects of GPs than their ability to save lives. Understanding the relationship between our objective measure of GP VA based on mortality and the patient-generated ratings of these GPs is of great independent value. Specifically, recent years have seen an increase in the availability of patient-generated rating platforms. If patients use these platforms to make health care choices, it is important to know to what extent they are reflective of objective GP quality, and to what extent they facilitate improvements in health choices.

While the patient ratings are uncorrelated with GP VA (Table 10), we show in Table 11 that GPs do receive a disproportionate influx of new endogenous patient swaps if they have a higher subjective patient-generated rating, or a higher proportion of top scores on the subjective patient ratings. The patient-generated ratings are publicly available and perhaps 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.[22] 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. These findings therefore have important implications for ongoing policy debates on the organization of health care services and the distribution of information to patients.

5. Policy Analysis

In this section, we use our estimates to predict the potential social benefit from selecting and retaining GPs based on VA. 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.

To perform this exercise, we note that substituting the worst performing GPs with GPs of median quality leads to a 9.4-month increase in patients' life expectancy. Using $35,000 as the statistical 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 therefore is $27,417 per patient, $9.05 million per GP, or $934 million in total.[23] Thus, while the value associated with each individual patient is relatively modest, the aggregate effect is substantial. At the same time, our results demonstrate that higher-quality GPs are associated with

[22] See https://www.legelisten.no/leger.

[23] For individuals who are exogenously swapped at age 55 or older, the effect we estimate is equivalent to 3.2 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 9.4. 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 $27,417. 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 2064. 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 $27,417 per patient (35,000*[9.4/12]), $9.05 million per GP ([1100*0.3]* 27,417) or $934 million in total ([2064*0.05] * [1100*0.3]* 27,417).

a lower per-patient cost. 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, we note that this conclusion depends on the cost of raising a GP's VA from the bottom to the middle of the distribution; we see this as an extremely valuable avenue for future research.[24]

6. Discussion

The average OECD country spends more than 10 percent of GDP on healthcare services each year. The goal 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?

In this paper, we exploit rich population-wide register data coupled with exogenous assignment of patients to general practitioners to estimate doctor value-added and provide the first evidence on the distribution of physician quality across an entire country. In addition, the paper produces the first causal estimates of the impact of GP quality on patients' long-term health and labor market outcomes.

We show that there is substantial variation in the quality of physicians in the primary care sector, and that a one standard deviation increase in doctor quality is associated with a 12.2-percentage point decline in a patient's two-year mortality risk. While we find evidence of certain observable GP demographic characteristics and practice styles being able to predict GP VA, we also note that a standard decomposition exercise reveals that most of the quality variation is driven by unobserved differences across doctors. 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)

[24] To put our social benefit number into perspective relative to prior work on worker value-added, we note that Chetty et al. (2014) perform the same exercise for teachers and find that replacing a teacher whose current VA is in the bottom 5 percent with an average teacher would increase the mean present value of students' lifetime income by $250,000 per classroom over a teacher's career.

canonical proposal to replace the worst performing GPs (bottom 5 percent of the VA distribution) with GPs of average quality, we find that this policy would generate a total social gain of $934 million. At the same time, our results suggest that the practice styles and behaviors of higher-quality GPs are not associated with a higher per-patient cost. This suggests that substantial investment in GP quality through training and retraining may represent cost-effective ways to reduce mortality and improve general health.

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 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.

References

Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc. "Mortality effects and choice across private health insurance plans." *Quarterly Journal of Economics* 136.3 (2021): 1557-1610.

Abowd, John M., Francis Kramarz, and David N. Margolis. "High wage workers and high wage firms." *Econometrica* 67.2 (1999): 251-333.

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc. "Do value-added estimates add value? Accounting for learning dynamics." *American Economic Journal: Applied Economics* 3.3 (2011): 29-54.

Angerer, Silvia, Christian Waibel, and Harald Stummer. "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 (2019): 407-427.

Bartel, Ann P., Nancy D. Beaulieu, Ciaran S. Phibbs, and Patricia W. Stone. "Human capital and productivity in a team environment: evidence from the healthcare sector." *American Economic Journal: Applied Economics* 6.2 (2014): 231-59.

Behncke, Stefanie, Markus Frölich, and Michael Lechner. "A caseworker like me–Does the similarity between the unemployed and their caseworkers increase job placements?" *Economic Journal* 120.549 (2010): 1430-1459.

Bensnes, Simon, and Ingrid Huitfeldt. "Rumor has it: How do patients respond to patient-generated physician ratings?" *Journal of Health Economics* 76 (2021): 102415.

Carneiro, Pedro, Emanuela Galasso, Rita Ginja, and Lucy Kraftman. "The Heterogeneous Quality in Delivery of Welfare: Evidence from Social Workers in the Chile Solidario Program." (2021) Mimeo.

Cattan, Sarah, Gabriella Conti, and C. Farquharson. "Workforce quality in early years intervention: Evidence from a large-scale home-visiting program." (2021) Mimeo.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American Economic Review* 104.9 (2014): 2633-79.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. "The association between income and life expectancy in the United States, 2001-2014." *JAMA* 315.16 (2016): 1750-1766.

Clemens, Jeffrey, and Joshua D. Gottlieb. "Do physicians' financial incentives affect medical treatment and patient health?" *American Economic Review* 104.4 (2014): 1320-49.

Cutler, David M., Robert S. Huckman, and Mary Beth Landrum. "The role of information in medical markets: an analysis of publicly reported outcomes in cardiac surgery." *American Economic Review* 94.2 (2004): 342-346.

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

Doyle Jr, Joseph J., Steven M. Ewer, and Todd H. Wagner. "Returns to physician human capital: Evidence from patients randomized to physician teams." *Journal of Health Economics* 29.6 (2010): 866-882.

Elvik, Rune. “The value of life: The rise and fall of a scientific research programme.” Cambridge Scholars Publishing, 2018.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. "Sources of geographic variation in health care: Evidence from patient migration." *Quarterly Journal of Economics* 131.4 (2016): 1681-1726.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. "Place-based drivers of mortality: Evidence from migration." *American Economic Review* 111.8 (2021): 2697-2735.

Fletcher, Jason M., Leora I. Horwitz, and Elizabeth Bradley. “Estimating the value added of attending physicians on patient outcomes.” No. w20534. National Bureau of Economic Research, 2014.

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

Godager, Geir. "Birds of a feather flock together: A study of doctor–patient matching." *Journal of Health Economics* 31.1 (2012): 296-305.

Hanushek, Eric A., and Steven G. Rivkin. "Teacher quality." *Handbook of the Economics of Education* 2 (2006): 1051-1078. Amsterdam: North-Holland.

Hanushek, Eric A., and Steven G. Rivkin. "The distribution of teacher quality and implications for policy." *Annual Review of Economics* 4.1 (2012): 131-157.

Ho, Katherine. "The welfare effects of restricted hospital choice in the US medical care market." *Journal of Applied Econometrics* 21.7 (2006): 1039-1079.

Howard, David H. "Quality and Consumer Choice in Healthcare: Evidence from Kidney Transplantation." *Topics in Economic Analysis and Policy* 5.1 (2006): 1349.

Jacob, Brian A., and Lars Lefgren. "Can principals identify effective teachers? Evidence on subjective performance evaluation in education." *Journal of Labor Economics* 26.1 (2008): 101-136.

Kane, Thomas J., and Douglas O. Staiger. “Estimating teacher impacts on student achievement: An experimental evaluation.” No. w14607. National Bureau of Economic Research, 2008.

Kolstad, Jonathan T., and Michael E. Chernew. "Quality and consumer decision making in the market for health insurance and health care services." *Medical Care Research and Review* 66.1 Supplement (2009): 28S-52S.

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

McClellan, Mark, Barbara J. McNeil, and Joseph P. Newhouse. "Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality?: Analysis using instrumental variables." *JAMA* 272.11 (1994): 859-866.

McLean, Gary, and Sutton, Matt. "Competition and patient choice." (2005) Mimeo.

Mulhern, Christine. "Beyond teachers: Estimating individual guidance counselors' effects on educational attainment." (2020) Mimeo.

Pope, Devin G. "Reacting to rankings: evidence from 'America's Best Hospitals'." *Journal of Health Economics* 28.6 (2009): 1154-1165.

Riise, Julie, Barton Willage, and Alexander Willén, "Can Female Doctors Cure the Gender STEMM Gap? Evidence from Randomly Assigned GPs." *Review of Economics and Statistics* (2022).

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. "Teachers, schools, and academic achievement." *Econometrica* 73.2 (2005): 417-458.

Rockoff, Jonah E. "The impact of individual teachers on student achievement: Evidence from panel data." *American Economic Review* 94.2 (2004): 247-252.

Rothstein, Jesse. "Teacher quality in educational production: Tracking, decay, and student achievement." *Quarterly Journal of Economics* 125.1 (2010): 175-214.

Santos, Rita, Hugh Gravelle, and Carol Propper. "Does quality affect patients' choice of doctor? Evidence from England." *Economic Journal* 127.600 (2017): 445-494.

Scanlon, Dennis P., Michael Chernew, Catherine McLaughlin, and Gary Solon. "The impact of health plan report cards on managed care enrollment." *Journal of Health Economics* 21.1 (2002): 19-41.

Schnell, Molly, and Janet Currie. "Addressing the opioid epidemic: is there a role for physician education?" *American Journal of Health Economics* 4.3 (2018): 383-410.

Simonsen, Marianne, Lars Skipper, Niels Skipper, and Peter Rønø Thingholm. "Discontinuity in care: Practice closures among primary care providers and patient health care utilization." *Journal of Health Economics* 80 (2021): 102551.

Stoye, George. "The Distribution of Physician Quality in England." (2022) Mimeo.

Tsugawa, Yusuke, Anupam B. Jena, Jose F. Figueroa, E. John Orav, Daniel M. Blumenthal, and Ashish K. Jha. "Comparison of hospital mortality and readmission rates for Medicare patients treated by male vs female physicians." *JAMA Internal Medicine* 177.2 (2017): 206-213.

Wedig, Gerard J., and Ming Tai-Seale. "The effect of report cards on consumer choice in the health insurance market." *Journal of Health Economics* 21.6 (2002): 1031-1048.

1 Tables and Figures

Table 1: Relation Between GPs' and Baseline Characteristics of Patients: Sample of Patients at GP Swap.

	(1) Less than HS	(2) Born in Norway	(3) Nb Kids	(4) Work	(5) Log Income	(6) Any UI Benefits	(7) Large City	(8) Birth Order	(9) Pre-visits
Std. VA	0.001	0.002	-0.011	-0.000	0.009	0.002	-0.001	0.005	0.003
	(0.006)	(0.002)	(0.024)	(0.006)	(0.015)	(0.002)	(0.004)	(0.013)	(0.005)
Mean of Depend. Var.	.274	.971	2.18	.572	12.7	.0224	.14	1.69	.612
N	68535	68690	68690	68690	68482	68660	68690	68690	67625

Note: The dependent variables are characteristics of patients measured in the year prior to the exogenous assignment to a new GP. Controls included in model but excluded from table are: gender, fixed effects for the year of swap, age at swap and for the previous GP. Large city is an indicator that takes value 1 if the individual resides in one of the largest five cities of Norway (Oslo, Bergen, Trondheim, Stavanger or Kristiansand), and 0 otherwise. Standard errors in parentheses (SE) clustered at new GP. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

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

	Panel A: Age of Death		
	Age at death		
Standardized VA	0.269***		
	(0.097)		
Mean of Depend. Var.	77		
N	10877		

	Panel B: Cause of Death		
	Cancer ICD10 C	Cardiovascular Conditions ICD10 I	Rest
Standardized VA	-0.008**	0.002	-0.002
	(0.004)	(0.003)	(0.003)
Mean of Depend. Var.	.011	.008	.012
N	68664	68664	68690

Note: The table present the OLS estimates where the dependent variable in column (1) the age of individuals at death. Controls included in model but excluded from table are: gender, fixed effects for the year of swap, age at swap and for the previous GP. The outcome variable is measured between 2007-2020. Standard errors in parentheses clustered at GP level. * p<0.1, ** p<0.05, *** p<0.01

Table 3: Correlation Between VA Measure and GP Characteristics

	(1)	(2)	(3)	(4)	(5)
GP Characteristics					
GP Age	-0.0054**				-0.0064**
	(0.0023)				(0.0029)
GP is male		-0.0050			-0.0028
		(0.0529)			(0.0538)
GP specialist			-0.1098**		-0.0347
			(0.0535)		(0.0630)
GP list length			0.0002***		0.0002***
			(0.0001)		(0.0001)
GP has a shared list			-0.0039		0.0196
			(0.0761)		(0.0772)
GP in group practice			0.0511		0.0336
			(0.1232)		(0.1261)
GP Practice Style					
GP consultations				-0.0277	-0.0338
				(0.0360)	(0.0363)
Simple appointments				-0.0191	-0.0169
				(0.0167)	(0.0166)
Lab tests				0.0242	0.0284
				(0.0331)	(0.0334)
Surgery				-0.0236	-0.0107
				(0.0957)	(0.0971)
Advising				0.0024	-0.0736
				(0.3579)	(0.3650)
Annual Reimbursement cost				0.0000	0.0000
				(0.0001)	(0.0001)
Referrals				0.0002	0.0002
				(0.0002)	(0.0002)
Sick Leave				-0.0531	-0.0540
				(0.0513)	(0.0512)
P-values for joint test:					
”GP Characteristics			.034		.027
”GP Style”				.709	.731
R^2	.0026	4.37e-06	.0056	.0051	.0128
N	2064	2064	2064	2064	2064

Note: The table presents the coefficients from OLS estimations. The dependent variable is the standardized VA and the regressors are characteristics of GPs. ”Consultations” is the average number of patient visits in which a medical assessment is performed and the may include simple diagnose tests (for example, blood testing of total cholesterol, analyses of creatinine, potassium, CPR test, pregnancy test). ”Simple appointments” are the average number of visits in person, via messaging services, in written form or by telephone that result in writing off prescription, sick leave certificate, requisition of X-ray or physiotherapist or referral to specialist due to a non-urgent illness (ie, without a medical assessment). ”Laboratory tests” are the average number of test that require collecting a sample for analysis (for example, blood checks for hemoglobin and hematocrit count, counting of white and red blood cells and simple urine examination and microscopy of urine). ”Procedures” are minor surgical procedures (for example, 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). ”Advising” is the average number of visits during which the doctor provides counselling (for example, on how to take medication, nutrition). ”Annual Reimbursement cost” is the average reimburse per each visit. ”Referrals” is the average number of referrals to specialist care (ie, specialist visits within four weeks of the GP visit). ”Sick Leave” is the average number sick leaves prescribed. Standard errors in parentheses (SE) clustered at new GP. * p<0.1, ** p<0.05, *** p<0.01

Table 4: Probability and Number of GP Visits 1 Year After GP Swap

	(1) Consultations	(2) Lab Test	(3) Procedure	(4) Sick Leave	(5) Reimbursement
	Panel A: Extensive Margin				
Standardized VA	-0.007*	-0.001	-0.000	0.001	
	(0.004)	(0.006)	(0.006)	(0.004)	
Mean of Depend. Var.	.859	.683	.190	.144	
N	68690	68690	68690	68690	
	Panel B: Intensive Margin				
Standardized VA	-0.274*	-0.061	-0.007	-0.010	-57.699**
	(0.146)	(0.059)	(0.016)	(0.023)	(28.671)
Mean of Depend. Var.	7.460	2.440	.319	.516	1057
N	68690	68690	68690	68690	68690

Note: The table presents the estimates for β in model (3). The dependent variables are different measures of use of GP services. Column (1) includes the probability (Panel A) and number (Panel B) of annual visits to GP two years after the GP exogenous swap. The dependent variable in column (2) is an indicator for wether 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 (3) the dependent variable is an indicator for whether the patient visit includes minor surgical procedures such as treatment of epistaxis (bleeding from nose), treatment and instruction treatment of urinary incontinence with electrical stimulation, 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. 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 total reimbursement value associated to the activities performed by the GP during annual visits. Controls included in model but excluded from table are: gender, fixed effects for the year of swap, age at swap and for the previous GP. The outcome variables are measured from the Control and Payment of Health Refunds Database (KUHR; 2006-2020). Standard errors in parentheses clustered at GP level. * p<0.1, ** p<0.05, *** p<0.01

Table 5: Specialist Visits During 1 Year After GP Swap

	(1) Any Hospitalization	(2) Urgent	(3) Elective	(4) Inpatient	(5) Outpatient
	Panel A: Extensive Margin				
Standardized VA	-0.006	-0.028***	0.008	0.011	0.008
	(0.007)	(0.010)	(0.007)	(0.010)	(0.007)
Mean of Depend. Var.	.457	.345	.915	.260	.879
N	60162	27498	27498	27498	27498
	Panel B: Intensive Margin				
Standardized VA	-0.315	-0.215**	-0.100	0.051	-0.151
	(0.224)	(0.107)	(0.156)	(0.045)	(0.139)
Mean of Depend. Var.	4.400	.880	3.520	.477	3.040
N	27498	27498	27498	27498	27498

Note: The table presents the estimates for β in model (3). The dependent variables are different measures of use 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 (ie, including overnight stays or day treatments, such as less minor surgical procedures) or consultations at outpatient clinics, respectively. Controls included in model but excluded from table are: gender, fixed effects for the year of swap, age at swap and for the previous GP. The outcome variables are measured from the Norwegian Patient Register (2008-2020). Standard errors in parentheses clustered at GP level. * p<0.1, ** p<0.05, *** p<0.01

Table 6: Specialist Visits During 1 Year After GP Swap: Type of diagnoses

	(1) Malign Cancer (ICD10 C)	(2) Benign Cancer (ICD10 D)	(3) Cardiovascular (ICD10 I)	(4) Other
		Panel A: Extensive Margin		
Standardized VA	-0.002	0.008	0.017**	-0.007
	(0.008)	(0.005)	(0.009)	(0.009)
Mean of Depend. Var.	.099	.070	.200	.282
N	27498	27498	27498	27498
		Panel B: Intensive Margin		
Standardized VA	-0.235*	0.048***	0.053*	-0.007
	(0.128)	(0.018)	(0.032)	(0.009)
Mean of Depend. Var.	.662	.123	.447	.282
N	27498	27498	27498	27498

Note: The table presents the estimates for β in model (3). The dependent variables are different measures of use of specialist health care services. Controls included in model but excluded from table are: gender, fixed effects for the year of swap, age at swap and for the previous GP. The outcome variables are measured from the Norwegian Patient Register (2008-2020). Standard errors in parentheses clustered at 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) Working	(2) Labor Income	(3) Total Income	(4) Welfare Benefits	(5) DI
			Panel A: Age 55+		
Standardized VA	0.0005	2633.73	6867.74	-0.0116*	-0.0091
	(0.0073)	(4479.12)	(5111.04)	(0.0066)	(0.0066)
Mean of Depend. Var.	.339	183432	443583	.201	.169
N	54753	54753	54753	54753	54753
			Panel B: 55-70 years old		
Standardized VA	0.0002	2776.18	6304.60	-0.0136*	-0.0108
	(0.0088)	(5373.05)	(6150.89)	(0.0080)	(0.0079)
Mean of Depend. Var.	.394	213202	470039	.234	.197
N	46915	46915	46915	46915	46915

Note: The table presents the estimates for β in model (3). Controls included in model but excluded from table are: gender, fixed effects for the year of swap, age at swap and for the previous GP. Outcome are measured from the tax registers (earnings) and the social insurance database (information on welfare benefits and unemployment) for the years of 2007-2020. Standard errors in parentheses clustered at GP level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 8: Endogenous Swap of GP after Exogenous Re-assignment and Value Added, Patient-Level

	(1)	(2)
	Any Endogenous Swap	
Standardized VA	0.00087	0.00023
	(0.00094)	(0.00094)
GP 40y.o. or older		0.01580***
		(0.00196)
GP has specialization		-0.02546***
		(0.00234)
GP list length		-0.04407***
		(0.00285)
GP is male		0.00795***
		(0.00171)
Sick Leave		0.01230***
		(0.00393)
Reimbursement		-0.00008***
		(0.00000)
N	348865	348865

Note: The table includes correlation between an indicator for whether a patient swapped away endogenously after an exogenous re-assignment to a GP for whom the VA in period 2006-2015 is computed. Standard errors in parentheses (SE) clustered at new GP. * p<0.1, ** p<0.05, *** p<0.01

Table 9: Endogenous Swap of GP after Exogenous Re-assignment and Value Added, GP-Level

	(1)	(2)
	Any Endogenous Swap	
Standardized VA	-0.00202	
	(0.00470)	
GP 40y.o. or older		0.04291***
		(0.01147)
GP has specialization		-0.00228
		(0.01153)
GP list length		-0.05419***
		(0.01302)
GP is male		0.01933**
		(0.00971)
Sick Leave		0.01327
		(0.01397)
Reimbursement		-0.00003**
		(0.00001)
Mean of Depend. Var.	.546	.546
N	2064	2064

Note: The tables includes the correlation between the proportion of new patients a given GP receives via endogenous swap after an exogenous re-assignment and his/her VA (obtained for the years 2006-2015). Standard errors in parentheses (SE) clustered at new GP. * p<0.1, ** p<0.05, *** p<0.01

Table 10: Patients Overall Rating and GP Value Added

	(1) Nb of rating	(2) Mean Rate	(3) Share Max. Rating	(4) No Rating
		Panel A: Value Added		
Standardized VA	-0.1194	-0.0265	-0.0092	-0.0243***
	(0.2830)	(0.0175)	(0.0058)	(0.0085)
Mean of Depend. Var.	13.5	4.1	.682	.129
N	1797	1797	1797	2064
		Panel B: Characteristics of GP		
GP 40y.o. or older	-1.4217**	-0.2747***	-0.0867***	0.1062***
	(0.6003)	(0.0466)	(0.0152)	(0.0196)
GP has specialization	0.8949	0.0920**	0.0173	-0.0532***
	(0.6091)	(0.0466)	(0.0153)	(0.0202)
GP list length	16.3635***	-0.2735***	-0.0770***	-0.2001***
	(0.9462)	(0.0539)	(0.0173)	(0.0231)
GP is male	-0.8292*	0.1043***	0.0197	0.0473***
	(0.4965)	(0.0381)	(0.0123)	(0.0155)
Sick Leave	0.5884	0.0465	0.0120	0.0077
	(0.5848)	(0.0533)	(0.0172)	(0.0221)
Reimbursement	-0.0009*	0.0001	0.0000*	-0.0001***
	(0.0005)	(0.0001)	(0.0000)	(0.0000)
N	1797	1797	1797	2064

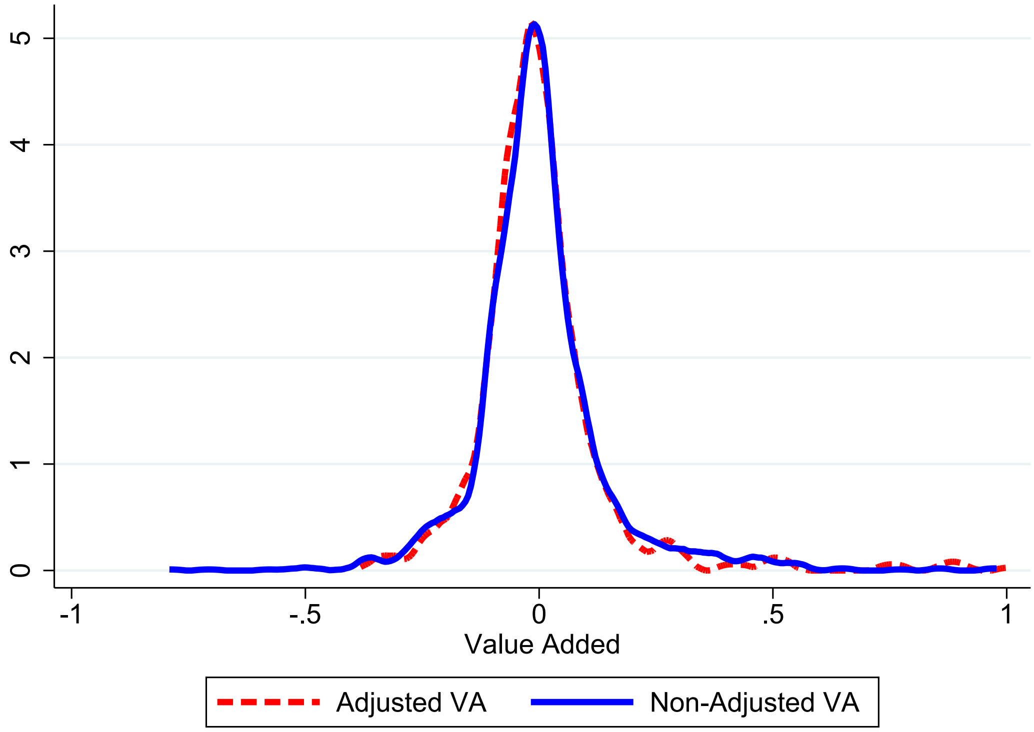
Note: Correlation between VA measure and Patient Ratings. Standard errors in parentheses (SE) clustered at new GP. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 11: Endogenous Swaps and Patient Ratings

	(1) Nb of rating	(2) Mean Rate	(3) Share Max. Rating	(4) No Rating
Endogenous Swap	-0.0025***	0.0249***	0.0610***	0.0923***
	(0.0004)	(0.0061)	(0.0193)	(0.0152)
Mean of Depend. Var.	13.5	4.1	.682	.129
N	1797	1797	1797	2064

Note: The dependent variable is share of new patients that are ”endogenous” switches. Standard errors in parentheses (SE) clustered at new GP. * p<0.1, ** p<0.05, *** p<0.01 (keep if VA not missing)

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



A Appendix Tables and Figures

Table A.1: Mobility Between Previous and Exogenously Assigned GP

	(1) Mean	(2) SD	(3) N
Number of new GPs per previous GP	7.243	10.806	4,528
Number of patients that move from each previous to new GP	11.528	73.414	32,795

Note: The table includes the mean and standard deviation for the number of new GPs that receive patients from each previous GP (top row) and the number of patients involved in an exogenous swap from a previous into new GP (bottom row).

Table A.2: Characteristics of GPs

	(1)	(2)	(3)	(4)	(5)	(6)
	Exogenous Swaps				Any Swap	
Sample of Patients	55 and Older		25 and Older		55 and Older	
	Mean	SD	Mean	SD	Mean	SD
Panel A: Characteristics of GP						
GP Age	44.11	10.78	46.35	11.15	45.11	10.21
GP has specialization	0.39	0.46	0.47	0.48	0.43	0.45
GP list length	1108.52	358.77	1142.35	367.21	1148.60	364.32
GP is male	0.63	0.48	0.62	0.49	0.64	0.48
GP has a shared list	0.08	0.27	0.08	0.27	0.08	0.27
GP in group practice	0.94	0.24	0.92	0.26	0.90	0.28
GP max list length	1180.27	356.89	1212.22	361.15	1234.30	365.65
Panel B: Characteristics of Patients served						
GP consultations	3.47	2.80	2.99	3.90	1.70	1.68
Lab tests	1.00	0.90	0.84	1.62	0.48	0.49
Procedure	0.12	0.20	0.11	1.05	0.06	0.09
Sick Leave	0.43	0.45	0.39	0.68	0.22	0.24
N	2064		5188		5768	

Note: The table includes selected characteristics for the GPs in the data used. There is one observation per GP.

Table A.3: Descriptive Statistics for Patients

	(1)	(2)	(3)	(4)
	55 and Older		25 and Older	
	Mean	SD	Mean	SD
Panel A: Mortality				
Mortality 2 Years After	0.03	0.18	0.01	0.11
Causes (Conditional on Death)				
Cardiovascular Conditions	0.26	0.44		
Cancer	0.36	0.48		
External Conditions	0.04	0.20		
Respiratory Conditions	0.11	0.31		
Panel B: Demographics				
Age at Swap	63.83	7.64	47.04	13.81
Male	0.54	0.50	0.52	0.50
Born in Norway	0.97	0.17	0.94	0.24
High School or Less	0.60	0.49	0.39	0.49
Years of Education	13.10	2.47	13.63	2.44
Married	0.67	0.47	0.49	0.50
Panel C: Health and Usage of Health Services				
GP Consultations	4.96	7.52	3.94	6.98
Surgery	0.19	0.83	0.14	0.72
Sick Leave	0.41	1.51	0.59	1.80
Blood Test	0.02	0.22	0.01	0.17
Annual Reimbursement cost	641.30	1305.03	507.74	1277.23
Any Hospitalization	0.46	0.50	0.38	0.49
Panel D: Labor Market				
Labor Income	183041.63	327043.92	386290.82	405718.02
Total Income	443268.79	483120.08	516005.12	536074.69
Disability income	0.17	0.38	0.12	0.32
Any Welfare Benefit	0.20	0.40	0.21	0.41
Sick Leave Days	18.10	66.41	18.62	63.92
N	70486		232884	

Note: The table includes the mean and standard deviation for selected characteristics of patients. One observation per patient is used. Monetary values deflated to 2015 using the CPI.

Table A.4: Descriptive Statistics for Ratings

	All GPs Rated		GPs used in our estimation sample	
	Mean	Proportion with top rating	Mean	Proportion with top rating
Number of Ratings	11.439	10.606	13.458	11.163
Rating	4.150	0.702	4.105	0.682
N	6,119			1,797

Note: The table includes the mean and proportion with top ratings for selected patient ratings. One observation per GP is used.

Table A.5: Key Moments of GP VA based on two-years mortality

	(1) N	(2) P10	(3) P25	(4) P50	(5) P75	(6) P90	(7) Mean	(8) SD
A: Exogenous swaps	2064	-0.124	-0.059	-0.004	0.053	0.117	-0.000	0.122
B: All swaps	5768	-0.038	-0.024	-0.011	0.010	0.041	-0.002	0.045

Note: Each row reports various moments for the GP effectiveness on different outcomes. The numbers reported are the estimates for the GP fixed effects.

Table A.6: Value Added Based on 2-Year Mortality: Heterogeneity

	(1) N	(2) Mean	(3) SD
GP not in group office	108	-0.011	0.110
GP in group office	1928	0.001	0.145
GP doesnt share list	1886	0.001	0.152
GP shares list	166	-0.005	0.065
GP is female	713	0.014	0.247
GP is male	1270	0.003	0.166
GP not specialist	1225	-0.002	0.179
GP specialist	796	0.001	0.204
GP has short list ($<$ 1000 patients)	1062	-0.002	0.300
GP has long list ($\geq$ 1000 patients)	970	0.005	0.188
GP is less than 40	865	-0.011	0.196
GP is 40+	1140	-0.005	0.199

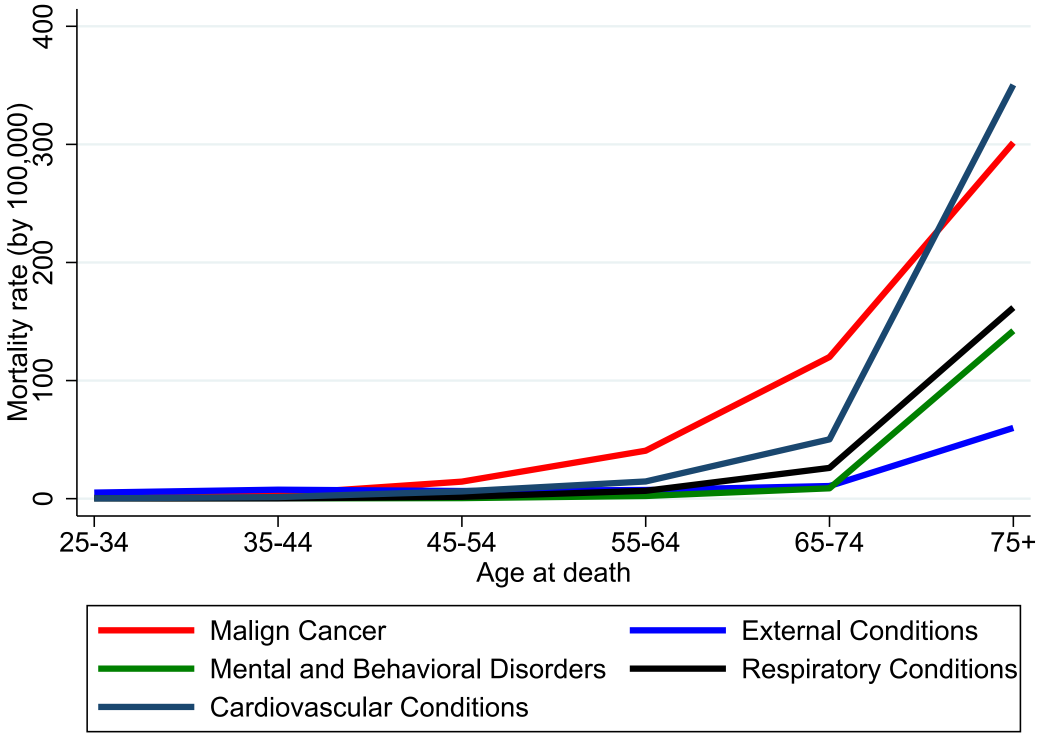
Note: Each row reports mean and standard deviation for the GP effectiveness for different groups of GPs.

Table A.7: Explanatory Power of Different Models

	(1)	(2)	(3)	(4)
R^2	0.105	0.120	0.130	0.114
Controls				
New GP FE		√	√	
Year of Swap	√	√	√	√
Previous GP FE	√	√	√	√
<u>Characteristics of Patient</u>				
Age at Swap FE			√	√
Gender			√	√
Education			√	√
Municip. of Residence at Swap FE			√	√
Married			√	√
Pre-swap Welfare Benefits			√	√
Pre-swap Income			√	√
Norwegian Born			√	√

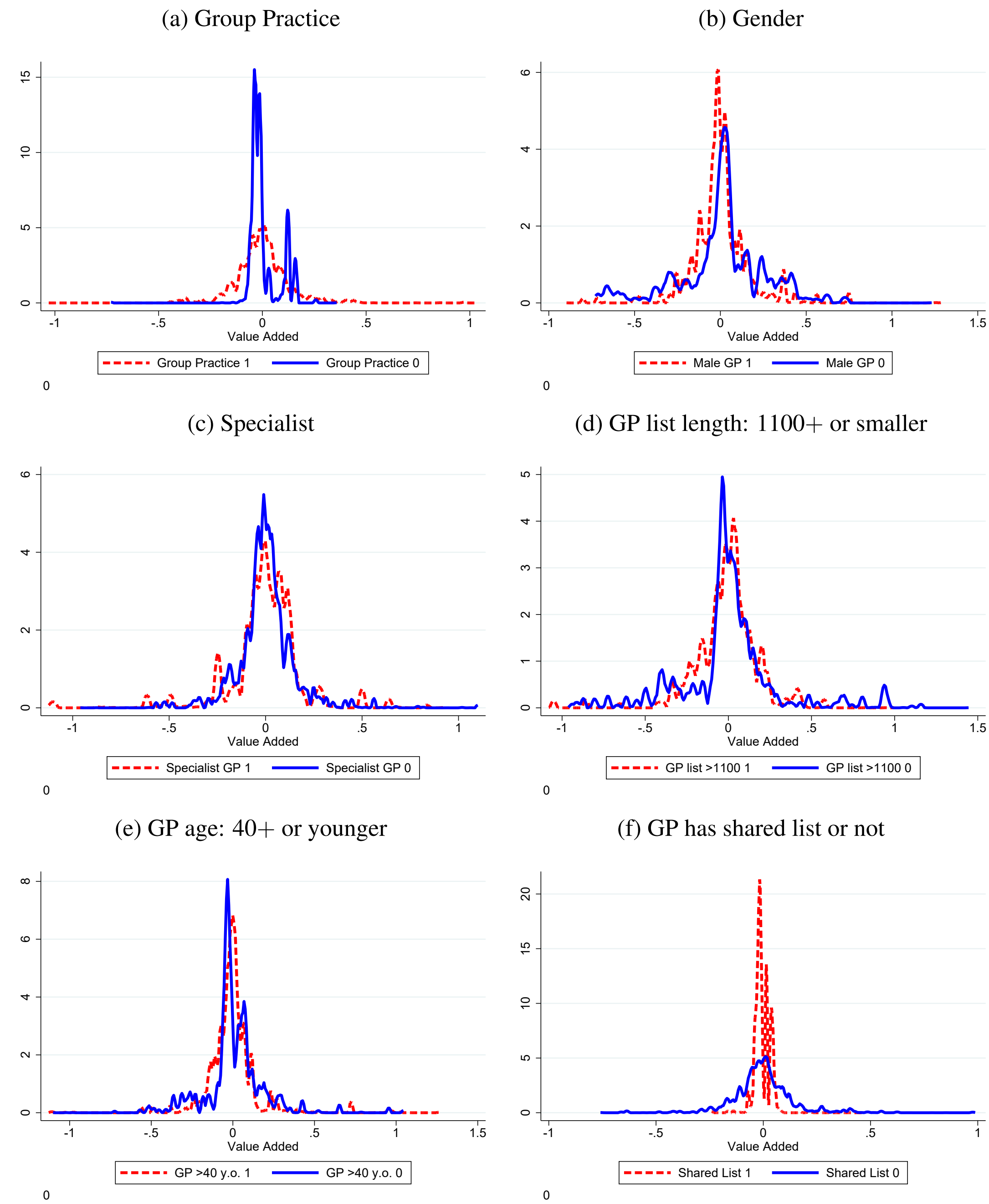
Note: This table presents the R^2 for different models, where the dependent variable is the an indicator that takes value 1 if the individual died within two years after GP swap and 0 otherwise, for individuals 55 years or older at the time of swap.

Figure A.1: Mortality Rate due to Specific Conditions



Note: This graphs presents the age-adjusted mortality rate for the main causes of deaths by age groups, in 2020.

Figure A.2: GP Value Added: Heterogeneity by Doctor Characteristics



Note: The number on the bottom left of each panel is the p-value for the Kolmogorov test for the difference in distributions.