Defying Distance?
The Provision of Services in the Digital Age

Link to Most Recent Version

Amanda Dahlstrand*

November 25, 2021

Abstract

Digital platforms are transforming services by making the physical distance between provider and user less relevant. I quantify the potential gains this flexibility offers in the context of digital primary care in Sweden, harnessing nationwide conditional random assignment between 200,000 patients and 150 doctors. I evaluate causal effects of matching patients of varying risks to doctors with different skills and assess counterfactual policies compared to random assignment. Matching patients at high risk of avoidable hospitalizations to doctors skilled at triaging reduces avoidable hospitalizations by 20% on aggregate – without affecting other adverse outcomes, such as counter-guideline antibiotics prescriptions. Conversely, matching the best triaging doctors to the richest patients leads to more avoidable hospitalizations, since the most vulnerable patients are often the poorest. Hence, remote matching can sever the link between local area income and service quality in favor of a needs-based assignment, improving the effectiveness and equity of service provision.

JEL codes: I14, J24, M54

*Department of Economics, LSE. Email: v.a.dahlstrand-rudin@lse.ac.uk. I acknowledge the collaboration with the digital care company and Nasim Bergman Farrokhnia, and grants from STICERD and CEP at the London School of Economics. I am grateful for advice and support from Nava Ashraf, Oriana Bandiera, Robin Burgess and Guy Michaels. I also thank Gharad Bryan, Gaia Dossi, Sacha Dray, Andreas Ek, Nicola Fontana, Anders Fridén, Rachael Meager, Peter Nilsson, Virginia Minni, Niclas Moneke, Petra Persson, Kilian Russ, Helena Svaleryd, Martina Zanella, Céline Zipfel and seminar and conference participants at LSE and MIT for helpful suggestions.

1 Introduction

A range of services is moving online – including healthcare, banking, and education[1]. In many countries, digitalization started before the pandemic, and has been accelerated by it. A direct implication is that geographical distance no longer by necessity factors into which service provider meets which user – these services are *defying distance*. This creates new opportunities to transform how services are provided by improving the matching between service providers and users to make better use of variation in provider skills.

This paper asks: to what extent can matching patients to online primary care physicians improve healthcare outcomes? In particular, the matching policy I consider is on doctor task-specific skill and patient outcome-specific estimated need or risk. I consider a setting in which the first doctor you see when contacting primary care can be based anywhere in the country, instead of a local physical primary care provider. This setting is ideal to study the potential effects of new technology to improve matching, as primary care is the front line of healthcare with the largest patient pool and the most heterogeneous patients and tasks. Hence, physician specialization and division of labor have the potential to increase output (Smith, 1776). I consider policies matching doctors with high ability in specific tasks, for instance risk prediction and triaging, to the patients most in need[2]. Such improvements to patient-doctor matching, independent of either party's location, can represent a cost-neutral policy in the digital age.

In order to overcome the endogenous selection between physical primary care providers and patients, which normally confounds causal effects of doctors on patients[3], I assemble a novel dataset of consultations, patients and doctors in digital primary care, available across an entire country – Sweden – in 2016-2018.[4] The proprietary data on over 200,000 patients and 150 doctors[5] comes from Europe's largest digital primary care provider. The key feature of the digital care data is that the allocation of doctors to patients is random, conditional on time and date.[6] This is a by-product of the first-come-first-served assignment procedure of

[1]Within education, this includes but is not limited to after-school tutoring, worker training programs and some university courses. Other services moving online are, e.g., therapy and counselling, exercise classes, real estate, financial advice and home improvement.

[2]Measured by their risk of each negative outcome, estimated from prior healthcare and demographic data.

[3]For instance, if in physical care some doctors meet patients with higher unobserved risk, selection bias would mean that those doctors appear worse.

[4]The fact that this is before the pandemic allows me to avoid any pandemic-related shocks to behavior, which particularly affected healthcare.

[5]The original dataset included approximately 380,000 patients. The sample restrictions are detailed in the Data section and in the Appendix.

[6]Conditional random assignment holds for 82% of the digital consultations, where patients have chosen to meet the first available doctor. Other options for patients include to book a specific time with a specified

patients to doctors, and neither party has the ability to intervene into this digital process.

To enable the analysis of healthcare outcomes in the physical care system, and to include patients' prior healthcare histories in physical care, this dataset is merged[7] on the individual patient level with physical healthcare data from the universal healthcare system. Here, patients are followed over six years, which allows me to measure patient risk in terms of past diagnoses and healthcare utilization history. Finally, the data is matched on the individual patient level with detailed socioeconomic and demographic variables from *Statistics Sweden*, to enable the addition of demographic variables to the patient risk estimation, and the study of redistributional effects of doctor reallocation across the income distribution.

In this paper, I compare counterfactual doctor skill-patient need matching policies to the most relevant other policies. These are, first, the status quo of digital time-conditional random matching between doctors and patients. Second, I simulate a second benchmark of positive assortative matching on patient income and doctor skill to approximate real-life existing healthcare inequalities in physical care. I provide evidence of such inequalities in physical primary care. Large location-based differences in healthcare outcomes persist within countries (see, e.g., Finkelstein, Gentzkow and Williams 2021) – even in countries with universal public health insurance such as Sweden (Chen, Persson and Polyakova 2021). I also study the *redistributional* effects of doctor skill-patient need matching policies along the patient income distribution. I provide evidence that changing doctor-patient matching can also allow us to address healthcare inequality by severing the link between the quality of local area service provision and patient income.

Estimating doctor ability in primary care has been a challenge, as important patient outcomes are often ambiguous, rare, or delayed.[8] Moreover, primary care physicians have multiple tasks, which opens the question of whether a single ability measure governs performance in all tasks, or whether doctors specialize. I address this by creating observable output measures of doctors in three key dimensions of a primary care physician's work: (1) identifying risky patients and triaging them to higher levels of care (2) providing guideline-consistent treatment for common conditions and (3) leaving the patient informed and sat-

doctor. I only use the conditionally randomly assigned consultations for analysis. Moreover, I restrict the data to one visit per patient, a patient's first visit in the service, to eliminate any concerns about endogeneity in any following consultations.

[7]Data merging and de-identification is done by *Statistics Sweden*.

[8]Mortality is the least ambiguous outcome, but the most rare and delayed as the conditions that people seek care for in primary care are often less serious. The main outcome I use (avoidable hospitalizations) can be seen as a proxy of mortality that is more commonly observed. Moreover, it is a preferable outcome to mortality as it is also more closely linked to the work of the primary care doctor, since this type of hospitalization is defined in the medical literature as preventable by primary care.

isfied so that they do not seek additional, costly, care more than necessary. I measure the outcomes in each task by *negative* patient outcomes: in the case of providing guideline-consistent treatment, I measure whether the patient has received a counter-guideline antibiotic. In the case of risk prediction, the negative outcome is an *avoidable hospitalization*, i.e. a hospital admission that could have been avoided with sufficient primary care. For the third outcome, I measure whether the patient has sought additional in-person primary care in the week following the digital care visit, for a subsample. For each of these outcomes, I estimate patient risk. To measure risk for avoidable hospitalizations, I generate a propensity score using pre-determined demographic and healthcare variables, such as age, a disease index of chronic diagnoses, and previous hospitalizations. These are variables available to the doctors in the patients' medical records, meaning that the risk prediction does not use additional data.

I implement an empirical method that allows for both measurement of doctor task-specific skill and estimation of doctor-patient match effects, where the latter uses the measures of doctor skill interacted with patient risk. This method avoids overfitting in two ways: first, it is based on a split-sample strategy, where I split the conditionally randomly assigned data into two: Sample 1 (a "hold-out sample") and Sample 2 (the "main sample").[9] Sample 1 is used to estimate physician effectiveness in each task with a value-added framework. Sample 2 is used to estimate the complementarities between different patient risk types and doctors of varying estimated ability in each outcome[10]. The second step that I take to reduce the noise in the doctor skill estimates is to shrink them using an empirical Bayes method.

In all outcomes, I find large and statistically significant differences across physicians in their task-specific effectiveness. However, the evidence is not consistent with a single latent ability variable governing all of the skills, meaning that doctors even within general practice have individual "specializations"[11]. These specializations are usually not taken into account in the organization of primary care, as a primary care doctor is expected to deal with all types of tasks.

The next step is to understand whether physician-patient matching matters for patient

[9] I verify the conditionally random assignment of patients to doctors in both Sample 1 and Sample 2.

[10] The doctor ability was estimated on different patients than those present in Sample 2

[11] This could be due to different innate ability, for instance some are better at speaking with patients and reassuring them (so that they do not seek additional care when not necessary), while others are better at being strict with antibiotics guidelines even if a patient argues that they want antibiotics. Or the absence of a positive correlation between skills could suggest that doctors reach different balances in some trade-off between the prioritization of different tasks of a general practitioner, perhaps due to different risk aversion, preferences or training.

outcomes in primary care. Indeed, the gains from matching are driven by another fact that I establish: that patients have predictable needs for different dimensions of doctor skills. In Sample 2 (the "main sample"), I estimate the effect of matching doctors with high skill in a task with patients who have a high estimated need for that task. One main result of this paper is that if we match a doctor who is among the top 10% at reducing avoidable hospitalizations, with a patient who is predicted to be among the top 1% risky for such adverse outcomes, we could reduce their number of such adverse outcomes by 90%. This is important, as avoidable hospitalizations are a sign of low quality primary care and are most common among low-income individuals. At the same time, patients who are not estimated as "risky" for this outcome have effects that are indistinguishable from zero from seeing a doctor among the top 10% at reducing avoidable hospitalizations. I will call this a complementarity between doctor and patient types.[12]

To increase the relevance of the causal treatment effects of some doctors on some patients, I assess the aggregate impacts of counterfactual policies of reallocations between doctors and patients, adapting a conceptual framework developed by Graham, Imbens and Ridder (2014). This framework enables us to answer a different question than the usual (what would the effect be of increasing a certain input?). In particular, as is especially relevant in healthcare, where the long education of doctors means inputs are difficult to increase, how can we reallocate existing inputs to get an output improvement? This conceptually simple framework relies on conditionally random matching to estimate an average match function (the average outcome when each doctor type meets each patient type), and then uses this function to evaluate effects of counterfactual reallocations. The reallocations chosen are based on an optimization problem, taking existing resources – doctor skills and work hours. The outcomes depend on the distribution and correlation of risks for each outcome in the patient population; the distribution and correlation of doctor skills; and the within-patient and within-doctor correlation of risk and skills across the different outcomes. The framework takes into account the externality on the patient *from* whom a high-skilled doctor is moved in a reallocation. Moreover, I add the consideration of possible correlations between doctor skills in different tasks, and study the effects on other outcomes than the main, in each reallocation.

A counterfactual simulation shows that we could reduce avoidable hospitalizations in the aggregate by 20% by matching doctors and patients, compared to random allocation. This reallocation does not negatively affect other main outcomes. The outcome is achieved

[12]This type of complementarity also exists for the other outcomes, even if the differences between patient groups are smaller.

by only reallocating of 2% of patients, since I show that I can accurately predict who the patients at risk for avoidable hospitalizations are using past healthcare data, and they are a small fraction of all patients. Moreover, the reallocation shifts this aspect of doctor skill (risk prediction and triaging) towards lower-income patients, who are the ones most in need.[13]

Matching is a resource-neutral policy that affects outcomes. However, its efficiency compared to resource-intence policy alternatives such as hiring and training, remains a priori ambiguous. To this end, I compare counterfactual doctor skill-patient risk matching policies to counterfactual physician hiring and selection policies, where doctors who have above median skill in three important tasks expand their hours of work at the expense of doctors with below median skill in these tasks. Even if these doctors expand their hours by as much as 70%, the gains are smaller[14] than from doctor-patient matching policies, and would moreover be more difficult to implement. Matching has larger effects as (1) patients in primary care have heterogeneous needs (and these needs can be identified with prior healthcare data) and (2) doctors have different skill sets that are important to some patients but not others.[15]

Matching of service providers to users is an under-utilized policy tool, which could be welfare improving at close to zero cost when distance is defied by digital services.[16] Algorithmic allocation means that machine prediction is used as a complement to human skill, as opposed to substitute[17]. The algorithm allocates patients to doctors, but the doctor makes the triage, diagnosis or treatment decision. This could make the policy less subject to "algorithm aversion" – that individuals trust recommendations from an algorithm less than from a human (Dietvorst et al. 2015, Yeomans et al. 2019). In fact, versions of matching are already being developed and used by digital platforms, including in digital primary care, without facing as much criticism as for instance artificial intelligence triaging. This paper establishes the potential impacts of such matching, and suggests new measures relevant for matching, such as doctor task-specific skill and patient risk.

[13]The estimated risk of having an avoidable hospitalization, as well as the number of prior avoidable hospitalizations, are concentrated in the lower end of the income distribution.

[14]No significant reduction in avoidable hospitalizations; 4% reduction in counter-guideline prescriptions.

[15]In the case of avoidable hospitalizations, it is also the case that the patients at risk are a very small subset of the total amount of patients. These patients are at risk for dangerous and costly complications, which is why focusing on them is important. The patients at risk for counter-guideline antibiotics are a much larger share of the total patient pool.

[16]The costs would be a small increase in waiting time for some patients, and the costs of importing data and developing the matching algorithm.

[17]If a substitute, the algorithm would make the medical decision. For a setting testing judges' predictions against algorithms, see Kleinberg et al. (2018).

The results on doctors' varying effects on heterogeneous patients could be generalizable also to physical care. The main reasons I focus on digital care are, first, that the policy of doctor-patient matching is feasible in digital care, due to the easing of shared location constraints, making some "nontradables" tradable (Muñoz 2021).[18] Second, digital services can be viewed as a "lab", which helps overcome endogeneity challenges endemic in physical primary care which have prevented the evaluation of causal effects of doctors. This is because, at least initially and for some parts of digital care, doctor-patient assignment has been conditionally random. In regular physical primary care, patient-doctor sorting confounds causal effects and all doctors do not meet all types of patients, meaning there is no common support for match effect estimators. The methods and conclusions of this study could speak also to other sectors, where the allocation of service providers, such as teachers, bank advisors, etc., to external clients could be key for effective production.

Digital provision services has become widespread in many sectors. This is the first paper to study nationwide digital service provision, and the first to show that digital services can defy inefficient and unequal matching due to distance and locational sorting. In addition, I bring a new source of conditionally random matching of service providers and external receivers to the literature. This complements the nascent empirical literature on reallocation and matching as mechanisms to improve outcomes instead of input augmentation (Aucejo et al. 2021, Bergeron et al. 2021, Fenizia 2020, Graham et al. 2021). These papers study teaching, tax collection and bureaucracies. I contribute by developing the ideas to a setting where there are lower obstacles and costs to matching on a large scale: digital service provision. Moreover, I add to this literature by studying matching in a medical setting, where the stakes are high and there is policy-relevant inequality in current resource allocation in many countries.

This paper also contributes to the literature on physician performance[19], by studying not only doctors' overall ability, but also specializations. Alsan et al. (2019) and Cabral and Dillender (2021) studied the effects of patient-doctor homophily on specific characteristics – gender and race, while the present paper is to the best of my knowledge the first to estimate causal effects of doctor skill on heterogeneous patients. Finally, I develop average reallocation effects to a new setting (Graham, Imbens and Ridder 2014, 2020), and implement them in a setting without pre-existing estimates of patient need and doctor skill.[20]

[18] Moreover, in digital care matching can be done by algorithms that quickly access patient and doctor data.

[19] See, e.g., Fadlon and van Parys 2020; Cutler et al. 2019; Abaluck et al. 2016; Doyle, Ewer and Wagner 2010; Grytten and Sørensen 2003.

[20] A more detailed literature overview is given in the appendix.

2 Institutional background

2.1 Digital and Physical Primary Care in Sweden

Sweden has a tax-financed universal public health insurance. Health expenditures accounted for 10.9% of GDP in 2016-2018.[21] Healthcare is provided by a mix of public (organized by 21 regions) and private providers. Only a small share of citizens - 6% in 2017 (Glenngard 2020) - have an additional private health insurance, mainly provided by employers. Private health insurance accounts for less than 1% of health expenditures (Glenngard 2020). Compared to other OECD countries, few people in Sweden (3.9%) skip a consultation due to cost (OECD 2017). Yet, patients complain of long waiting times for appointments in surveys, and the national goals of limiting waiting times are often unmet. In the few primary care outcomes that are measured and compared across countries, such as hospital admissions for asthma or chronic obstructive pulmonary disease, and congestive heart failure (related to avoidable hospitalizations), Sweden is above OECD average on one of the indicators and below on the other (OECD, 2017).

Primary care is the front line of healthcare, where the initial evaluation of a patient's condition, as well as cost-effective prevention takes place. In primary care in particular, patients are heterogeneous, as are the tasks facing primary care physicians/general practitioners (GPs), but the variation in doctor effectiveness with different patients is largely unknown. This is partly due to the endemic sorting between providers and patients in standard, physical primary care - sorting and selection has dominated this branch of healthcare more than others.[22] Yet, it is in primary care where patients and their health problems are most diverse, and the patient pool is largest, thus offering the largest possibilities for a better allocation.

Primary care physicians are institutionally positioned as a central gatekeeper in the access to healthcare. They are perhaps even more important in countries with universal health insurance, where access to specialists is more restricted, but they are central also in the US system (Fadlon and Van Parys 2020)[23].

Digital primary care, provided through smartphone video consultations, became widely available in Sweden in 2016. Digital primary care is not suitable for all conditions normally

[21] This a is slightly higher share than the OECD average, but lower than in the US.

[22] Previous research has exploited plausible randomization in, e.g., emergency hospital care to evaluate doctor effectiveness, but randomization has been harder to come by in primary care.

[23] Differences in how primary care works varies both within and across countries. For instance, referrals from the primary care provider to a specialist take place in 3% of consultations in our data. This is comparable to the lower end of GP referrals in the UK physical primary care setting, where in a meta-analysis, they range from 1.5% to 24.5% (O'Donnell 2000).

handled in primary care, since some conditions require physical examination or testing. However, many common conditions treated in primary care can be diagnosed and treated digitally. In Sweden, this is provided by private companies that are reimbursed by the regions, which are in turn responsible for the provision of healthcare from the universal public health insurance. Just as in physical primary care, which is provided by a mix of private (40%) and public providers (60%), doctors working in digital primary care are not paid fee for service but an hourly wage. The reimbursement level from the universal public health insurance for digital consultations has changed several times, while the fee faced by patients has remained at the level of fees for in-office primary care consultations during the study period 2016-2018. For children (under 18) and elderly (over 84 years old), the service is free from co-pay, just as in regular physical primary care.

How patients choose regular, physical primary care

Regular (physical) primary care is provided at GP clinics/primary care centers. Most patients are registered with one such clinic, but not registered with an individual doctor. Patients have the possibility to choose their GP clinic.[24] 92% of Swedish inhabitants live within 10 minutes' drive of the nearest physical primary care center, and 80% live within 5 minutes' drive of the second nearest (National Board of Health and Welfare, 2018). However, research indicates that a lower proportion (16% in 2011) of individuals with low education chose another center than their assigned default (compared to 29% among those with higher education) (Bendz 2011).[25] These results are in line with research showing that e.g. lower income students are less responsive to quality when choosing schools and need a larger quality increase to choose a school further away from them, than richer students (Bau 2021).

Physical care sorting: Lower-income areas have worse physical primary care scores

Aggregated public data indicate that patients across the country are less satisfied with their primary care in areas with lower income and higher share first-generation immigrants (Appendix Table 13).[26] As Appendix Table 13 does not directly connect a patient to a primary care center, I complement this with Table 1 below which includes the patients registered at the primary care clinics with non-missing data in one region, Skåne (there are around 150 primary care clinics in total in this region). In concordance with the country-

[24]In some regions, e.g. Stockholm, patients can remain unregistered with any GP clinic if they do not make an active choice, while in others, there is a default choice.

[25]This was early after the choice was introduced, and figures are likely to be different today.

[26]Table 13 covers most of Sweden, using a matching between municipality and 4-digit postcode-level observations, and the outcome variable is a patient-reported primary care clinic score from the national patient survey (NPE, 2019).

wide evidence, Table 1 also shows that patients have a less positive experience with primary care[27] in areas with a higher deprivation index[28]. Moreover, in more deprived areas, patients are also less satisfied with the information they receive in physical primary care. There is also a marginally significant negative relationship between deprivation and the share of patients who get to see a doctor instead of another profession (e.g., a nurse) when they visit primary care (Column 3). Column 4 measures one aspect of objective quality of care: whether patients diagnosed with diabetes also receive a lipid-lowering treatment. Here, there is no significant correlation with the deprivation index.

Table 1: Quality measures of physical primary care centers, patient-reported (1,2) and objective (3,4)

	(1) Positive experience	(2) Satisfied with information	(3) Met physician rather than other profession	(4) Recommended treatm. for diabetics
Deprivation index (win)	-10.60***	-6.26***	-0.02*	-0.14
	(2.15)	(2.022)	(0.01)	(3.17)
Constant	89.61***	80.26***	0.42***	63.39***
	(2.21)	(2.02)	(0.011)	(3.11)
N	120	120	149	115
R^2	0.17	0.07	0.02	0.00

Robust SEs in parantheses. Sample is primary care centers in Skåne. Source: Nationell Patientenkat and Region Skåne.

Sorting patterns into digital care

I assemble and analyze proprietary data from one digital primary care provider, which is the largest of providers in visit volume and which contributed with a majority of all such digital visits in the country during the study period. Patients sort freely into using the digital primary care service, and this is not the only option for primary care or digital primary care. When the service was started, advertisements were made on e.g. public transport, informing about the service and potential reasons to use it. To compare the sorting patterns into digital primary care to the sorting patterns into physical primary care, I study one Swedish region where I have the universe of physical primary care data.[29]

[27] The outcome variable in Columns 1 and 2 are from the National Patient Survey, *Nationell Patientenkät (NPE)*, 2019, and the variables in Columns 3 and 4 are from Region Skåne's publicly reported data.

[28] The deprivation index is used by the Region and is a weighted average of the variables (1) Born outside EU (2) Unemployed 16-64 year old (3) Single parent with child under 18 years old (4) low education 25-64 years old (5) over 65 years old and in a single household (6) Person over 1 years old who has moved into the area (7) Age below 5 years old.

[29] Primary care data is not collected by the national body (the National Board of Health and Welfare) which

This is Region Skåne, which is the southernmost region of 21 regions in Sweden, containing around 10% of the digital care users and the third largest city in the country.

Using the same index of low socioeconomic status among the patients registered at the clinic as above, I find that the deprivation index (also called the Care Need Index) is similar among digital users and non-users (Figure 13) (extensive margin). However, on the intensive margin (not comparing digital and physical anymore), individuals with higher deprivation index who use the digital service have more appointments in the digital service (Appendix Figure 11). This is corroborated when looking at individual income: lower-income users use the digital service more intensively (Appendix Figure 12). Figure 15 shows that digital care users are younger than non-users. There is a similar level of prior disease among digital users and non-users who are under 60 years old, measured by the sum of comorbidities from the Elixhauser index, a commonly used measure for summarizing disease burden (Elixhauser et al. 1998).[30] For users over the age of 60, non-users seem to have less prior disease.

In results available on request, I have compared the digital care users to the average Swedish citizen (the above is a comparison with the primary care users in one region). This shows that digital care users are more likely to live in cities than the average Swedish citizen. They are less likely to be a first generation immigrant, but more likely to be a second generation immigrant than the average Swedish citizen. In terms of income, adult patients have a somewhat higher median income than the average citizen.

Patients' use of other healthcare while using digital care
Patients take up the service freely, and are not obliged to change their relationship with their regular physical primary care clinic. Using data on physical primary care from Region Skåne, I find that around 4% of digital care users have a nurse contact in physical primary care the week after their digital care visit.[31]

2.1.1 The digital care provider

The healthcare provider contributing with proprietary, de-identified data for this study (in collaboration with Statistics Sweden) provides on-demand primary care via video consul-

contributes with the rest of the physical healthcare data to this study. To get access to physical primary care data in the entire country, separate applications and reviews have to be made to the 21 regions. I do not have data on individual socioeconomic variables of the patients in the region who do not use digital care, only their age.

[30] In this sorting analysis, the comorbidities are based only on data from primary care for both digital users and non-users, since I do not have data on other care for the digital non-users.

[31] This is consistent with evidence in Gabrielsson-Järhult et al. (2019), who find that 3.6% of digital care users in a different region (Jönköping) have a physical visit at a primary care centre within a week of using a digital care service.

tations with certified medical doctors. The physicians may have different specialties, but all are acting as general practitioners, and GP is the most common specialty. During the study period, the healthcare provider employed or contracted with around 500 doctors, while around 700 doctors had ever done a consultation in the service (some only had a test period and did not work further).

Patients access healthcare appointments by downloading the company's smartphone application and log in via Sweden's electronic identification system (Bank ID) which is used for all digital bank and governmental agency interaction. Adult patients access the system via their own Bank ID, while child patients need one of their parents or guardians to log in via the parent or guardian's Bank ID.

2.1.2 Randomization

A key feature for this study is that doctors and patients are as good as randomly assigned to each other, conditional on calendar date and time of day. This has not been the intended purpose of the service, but is a by-product of the aim to provide care as quickly as possible nationally. Doctors can choose their time shifts, and often choose them around 2-3 weeks ahead. When they are not busy with a patient or with follow-up work (such as writing prescriptions), or have a booked patient, they are in the roster of available doctors.[32] Patients who enter the system can choose between two tracks: meet the first available doctor ("drop in"), or meet a specific doctor at a specified time. Patients who choose the first track (82%) are effectively randomized to a doctor within this time period. One exception to this is that if there is a doctor in the roster of available doctors who has a pediatric specialty, then this doctor will be more likely to be matched to a child patient if such a patient is in the line. Therefore, I remove all pediatric specialists and the patients they are matched with (see further below in the definition of the analysis sample).

2.1.3 Doctors' incentives and work pattern

Doctors who work for the service almost invariably work part time from home and also work for other healthcare services, such as public or privately run hospitals or clinics. Doctors are recruited across the spectrum of experience, with the conditions that they (1) have a certification as MD (legitimerad läkare) in Sweden from the National Board of Health and

[32] Data from a later period may not be randomized to as large an extent since the healthcare provider after the study period started trying both matching on language and matching on geographical region (so that e.g. a patient based in the capital would meet a doctor in the capital, even if the meeting was digital, in order to enable a physical meeting as well if needed).

Welfare (Socialstyrelsen) which requires that they have finished the 18-21 months of intern period/residency (Allmäntjänstgöring, "AT") after medical school (2) that they have at least done 6 months of their intern period/residency (AT) in a Swedish GP clinic/primary care center *or* have at least 6 months of experience at a Swedish GP clinic after the intern period/residency (AT).

Table 2: Descriptive statistics of doctors included in the final sample.

	(1)				
	mean	sd	min	max	count
Specialist	0.31	0.47	0	1	143
In specialty training	0.36	0.48	0	1	143
MD + residency only	0.33	0.47	0	1	143
Speaks non EU15 language	0.36	0.48	0	1	143
GP specialist	0.40	0.49	0	1	143
Age	36.9	7.25	28	57	61
Female	0.43	0.50	0	1	61
Employed rather than contractor	0.38	0.49	0	1	52
Observations	143				

Doctors are paid per hour and there is no fee-for-service for the doctors, or bonus payments. Table 15 in the Appendix describes the characteristics of all doctors, which includes doctors who have worked very few consultations for the service during the study period. Table 2 shows the same characteristics for doctors who have worked at least 600 randomized consultations for the service and are included in the final sample.[33]

Around 50% of doctors are employed and the rest are hired as contractors, billing from their private company. Doctors can choose either of these methods when starting working for the digital care company. There are benefits to each option, with different tax liabilities, paperwork and pension contributions. The costs for the company are similar: around USD 70-95 per hour. Most doctors work part-time, and most also work in another type of healthcare provision, for instance in a public hospital or in a public physical primary care centre. Doctors are evaluated yearly on key performance indicators, and good performance can lead to a pay increase. The main performance indicators are patients per hour, fraction of patients who are helped, and patient satisfaction. Fernemark et al. (2020) studied the motivations and impressions of doctors working in digital care with e.g. the company studied here. They found that doctors perceive this type of work as highly autonomous, and choose

[33]These are 143 doctors. Data on the age and gender and employment status of these doctors are currently missing for a majority of doctors.

this partly because of the flexibility. They consider the stress level to be reasonably low, but want to complement this work with other types of work in order to continue developing their skills and abilities.

3 Data

Data	**Timing**	**N**	**D**
Digital care first visits	June 2016-Dec 2018	378,000	511
Hospital, acute & specialist	Jan 2013-Dec 2018	378,000	0
Prescriptions	Jan 2013-Dec 2018	378,000	0
Socioeconomics on adults	2013-2017	180,000	0
Demographics on patients	2013-2017	378,000	0
Primary care in 1 region	Jan 2013- Dec 2019	1.6mn	0

Table 3: Overview of data, timing and sample size.

3.1 Definition of analysis sample

The sample definition proceeds in three main steps. First, I start from the universe of patients who has had at least one digital consultation with the largest[34] provider of digital healthcare in Sweden, from the start of the service in mid-2016 to the end of 2018. I keep only the first visit for each patient, as these consultations are conditionally randomized, and I want to avoid any concern of endogeneity in following visits in terms of particular patients selecting in to a second visit. Hence, each patient has only one observation in digital care. I restrict the sample to "drop in" visits, that is visits where the patient had no way of specifying which doctor they want to meet, but rather meet the first available doctor. This is 82% of the first visit sample, and this is the sample where conditional randomization (conditional on time) holds. Moreover, I remove pediatricians and those children who are more likely to see a pediatrician (where randomization does not apply).

Second, I match this data to official registry data from Statistics Sweden on socioeconomic and demographic variables and data from the National Board of Health and Welfare (NBHW/ *Socialstyrelsen*) on diagnoses of chronic conditions from specialist, acute and inpatient care across the Swedish healthcare system in the three years preceding digital primary care, 2013-2015. In this physical healthcare dataset, there are many observations

[34]In terms of patient volumes in 2016-2020.

per patient. In addition, we match with data on physical primary care (2013-2019) from one Swedish region (Skåne), which matches for around 10% of the digital care sample.[35]

Third, for consistent definition of patient types according to their pre-digital physical healthcare utilization, I drop patients for whom I do not observe the full pre-period 2013-2016, i.e. patients who were born in or after 2013. Finally, I keep only doctors who have done >600 consultations and their patients, which leaves 210,171 patients (56% of original N) and 143 doctors (20% of original D). The reason is that many doctors were hired late in the sample period, since the service was expanding. These doctors have only done a few randomized consultations, many of them under 100. For more details on the sample definition, see the Data Appendix.

3.2 Measurement of outcomes

Estimating doctor performance in primary care has been a challenge, as important patient outcomes are often ambiguous, rare, or delayed. We might care most about mortality and quality of life, but also about costs to the rest of the healthcare system, where primary care physicians serve as gatekeepers. Mortality is the least ambiguous outcome, but the most rare and delayed as the conditions that people seek care for in primary care are often less serious. The main outcome I use (*avoidable hospitalizations*) can be seen as a proxy of mortality that is more commonly observed. Avoidable hospitalizations can even be seen as a preferable outcome to mortality as it is also more closely linked to the work of the primary care doctor, since this type of hospitalization is defined in the medical literature as preventable by primary care.

Moreover, primary care physicians have multiple tasks, which opens the question of whether a single ability measure governs performance in all tasks, or whether doctors specialize. I address this by creating observable output measures of doctors in three key dimensions of a primary care physician's work: (1) identifying risky patients and triaging them to higher levels of care (2) providing guideline-consistent treatment for common conditions and (3) leaving the patient informed and satisfied so that they do not seek additional, costly, care more than necessary. I measure the outcomes in each task by *negative* patient outcomes: in the case of providing guideline-consistent treatment, I measure whether the patient has received a counter-guideline antibiotic. In the case of risk prediction, the negative outcome is an avoidable hospitalization, i.e. a hospital admission that could have been

[35] Swedish physical primary care is devolved to 21 regions, which means that data from primary care is not included in the National Board of Health and Welfare data. Assembling primary care data in Sweden across all regions has eluded researchers, as policies, codings and applications vary across the country.

Negative outcome	**Frequency**	**Non-missing data**
Data on full sample		
Avoidable hospitalization 3 months	0.2%	100%
Counter-guideline prescription	2%	100%
Data on part of sample		
Contacted physical nurse week after	4%	11%

avoided with sufficient primary care. For the third outcome, I measure whether the patient has sought additional in-person primary care in the week following the digital care visit, for a subsample.

Rare vs. frequent

The main outcome in this paper is Avoidable hospitalizations, which is a rare event (0.2% of patients have an avoidable hospitalization in the 3 months following the digital consultation).[36] Yet, this is the most high stakes outcome of those which are measurable in the data and relatable to doctor inputs. The need to measure and understand rare and high-stakes events has been emphasized not least by the literature in financial economics (Dow and Bond 2021)[37] and the economics of disasters (e.g., Barro 2009)[38]. Another reason to focus on this outcome is that one of the main tasks of a primary care doctor is "looking for a needle in a haystack", i.e., sort the rare and seriously ill patients from the vast majority with minor complaints. Studies on healthcare in the United States often include mostly utilization and cost outcomes. A notable exception is Fadlon and Van Parys (2020) who include both outcomes such as avoidable hospitalizations and the doctor following guidelines, but not patient behavior outcomes or patient satisfaction.[39]

Avoidable hospitalizations (AH)

[36] That this adverse outcome is rare is, of course, a good outcome of primary care.

[37] This has also been at the forefront of public debate after the financial crisis and the pandemic. Dow and Bond cite Taleb (2007): "[w]hy do we keep focusing on the minutiae, not the possible significant large events, in spite of the obvious evidence of their huge influence?"

[38] Barro (2009) estimates the risk for disasters as 2% per year and shows that they have large welfare costs: society would be willing to reduce GDP by 20% each year to eliminate these rare adverse events. The rare events involved a contraction of GDP of 15-60%. An avoidable hospitalization involves not only the event per se, but can have large negative consequences as it is a negative health event that may lead to prolonged loss of productivity, and some risk of death.

[39] Only 0.24% of consultations in our sample end up in an avoidable hospitalization, as our sample is relatively young and seek for more minor conditions. In Fadlon and Van Parys (2020), 5% of the sample has an avoidable hospitalization each year. So in a sample like Fadlon and Van Parys', where the sample is over 65 years old and relatively sick.

The number of avoidable hospitalizations[40] is a widely accepted measure of healthcare performance, as the hospital admission could be avoided if primary care was timely and adequate. Bacterial pneumonia, urinary tract infection and congestive heart failure accounted for 77% of the AH costs in US (Rocha et al 2020). The Center for Medicare and Medicaid Services (CMS) states that high rates of avoidable hospitalizations could indicate that "beneficiaries are not receiving high-quality ambulatory care," and that low rates of ACSCs at the provider level "may signify that the provider is providing better primary care and coordinating effectively with providers in the continuum of care" (CMS, 2017). Avoidable hospitalizations are dangerous, both because of the inherent risks to a condition that has worsened unnecessarily, and because hospitalization in itself has risks such as hospital-acquired infections and risks from procedures carries out. It is estimated that 1.1 potential life year is lost from every AH (Rocha et al 2020). In both the United States and Sweden, AH decrease with income (McDermott and Jiang 2020), so reducing them could have an impact on health inequality.

Avoidable hospitalizations are also costly. In the US in 2017, 3.5 million adult AH (13% of hospitalizations) cost hospitals $33.7 billion (9% of costs for all adult non-childbirth hospital stays).[41] In Sweden, avoidable hospitalizations cost an estimated SEK 7.1 billion ($820 million) each year, and this represents 7% of all costs for inpatient curative and rehabilitative care.[42] The share of AH costs out of all national health expenditures is 1.3% in Sweden[43] (and also around 1% in the United States), and the share of these (purely hospital) costs out of GDP is 0.15% in Sweden.[44]

I create the variable measuring an avoidable hospitalization using the data from the National Board of Health and Welfare on all hospitalizations 2013-2018, where I code the hospitalization as an avoidable hospitalization if it has a diagnosis code (ICD 10) which is listed in Table A1 of Page et al. (2007). As a pre-digtial health risk factor, I use avoidable hospitalizations that took place within 3 years before the digital consultation. As an outcome of a digital consultation, I use avoidable hospitalizations that take place within 90 days of a digital consultation. Most of the avoidable hospitalizations within 90

[40] They are also called hospitalizations for ambulatory care sensitive conditions (ACSCs). This outcome is also used in Fadlon and Van Parys (2020) (using the 2017 CMS ACSCs).

[41] Jiang et al. (2009); McDermott and Jiang (2020).

[42] The number of hospital days for AH was around 1 million in Sweden in 2010 (Socialstyrelsen, 2011). The average cost per day in inpatient care is 7100 SEK (Socialstyrelsen, 2017). The exchange rate used as of 13 Sep 2021 is 8.64 SEK/USD. The costs for total inpatient and rehabilitative care are from Statistics Sweden Statistikdatabasen, 2021.

[43] Total expenditures were 528 billion SEK in 2018, from Statistics Sweden Statistikdatabasen, 2021

[44] GDP was 4 828 billion SEK in 2018, from Statistics Sweden Statistikdatabasen, 2021

days happen quite early after the digital consultation, and the mean is 33 days (see Figure 20 in the Appendix for the distribution of number of days). I conduct several checks to determine whether the avoidable hospitalization can actually be considered as preventable in the digital consultation.

First, the most common diagnosis groups[45] which are registered at the hospital as the primary diagnosis for the avoidable hospitalization within 3 months after the digital consultation are respiratory and genitourinary (connected to kidneys and e.g. complications of urinary tract infections), see Figure 17 in the Appendix. These are conditions which are commonly treated in digital care, for instance by prescribing antibiotics for urinary tract infections.[46] Second, patients who I have determined as risky for avoidable hospitalizations (see Sections 4.3 and 4.5) also are more likely to come to the digital service with symptoms that can later be related to avoidable hospitalizations: respiratory symptoms and urinary tract infection (see Figure 16 in the Appendix). Moreover, I compare the diagnosis group[47] set by the digital care doctor to the diagnosis group set as primary diagnosis by the hospital, and find that 33% concord in respiratory system, 20% concord in genitourinary system, and 27% concord in symptomatic diagnosis (these are the 3 most common groups for these avoidable hospitalizations).

Counter-guideline prescriptions (CGP)

Widespread non-adherence to medical guidelines contributes to hospitalizations, deaths, and spending (Neiman 2017). Such non-adherence has recently been studied with growing interest in economics, see e.g. Finkelstein et al. (2021) and Frakes et al. (2021). Non-adherence to guidelines on antibiotics prescriptions is particularly interesting since excessive antibiotics prescriptions lead to the negative externality of bacterial resistance.

Bacterial resistance means that the antibiotics that are usually effective in treating a bacterial infection will no longer work, which can lead to prolonged infection and mortality. The guidelines serve to limit the use of antibiotics to where the benefit outweighs the social cost of using them. Bacteria adapt under pressure[48], and if there is less prescription of antibiotics, it is possible to decrease the number of resistant bacterial infections (Bergman et al 2004). Sizing the costs of antibiotics resistance has proved challenging, but a very conservative estimate of the additional healthcare and prescription costs only (not counting

[45]This is a medical grouping of the ICD diagnosis codes into 23 categories related to the type of disease

[46]Figures 18 and 19 in the Appendix show that hospitalizations in general have a very different distribution of diagnosis groups.

[47]This is a medical grouping of the ICD diagnosis code that the doctor actually set into 23 categories related to the type of disease

[48]Penicillin was released in 1941, and a resistant germ to this antibiotic was identified in 1942. Another antibiotic, methicillin, was released in 1960 and the resistant germ was found in the same year.

lost productivity etc.) is SEK 160 million yearly in Sweden (Folkhälsomyndigheten 2013). This is likely under-estimating the total costs, due to the externality: resistance developed in Sweden also affects people in other countries.

I code non-adherence to 16 guidelines from Swedish strategic programme against antibiotic resistance on digital care (Strama 2017, 2019). All the guidelines are intended to limit the use of antibiotics or use a more narrow-spectrum antibiotic as a first line of response (which contributes less to resistance than a broad-spectrum antibiotic). Thus, to follow the guidelines, doctors sometimes need to say no to patients who think that they need antibiotics. To define the variable, I combine the incidence of prescription in the digital care data, conditional on the diagnosis (ICD) code, with data on the drug code from the NBHW's prescription register, which occurs once the patient has filled the presctiption.[49] The non-adherence measured in my sample is quite low by international standards. The Centers for Disease Control and Prevention (2019) estimate that 28% (47mn courses) of all antibiotics prescribed in doctors' offices and Emergency Departments in the United States are for infections that do not need antibiotics.

Contacted physical care within a week after the digital consultation
This is an outcome which is important for primary care costs and for patient satisfaction. If a patient contacts a physical primary care clinic in the week following the digital care consultation, this may indicate that they were not satisfied with the digital care consultation or the information given. This incurs additional costs to the universal health insurance in cases where the digital care consultation incurred a payment (which is not the case if the visit was deemed inappropriate for digital care by the doctor). I can measure this outcome in the region where I have primary care data, Region Scania, consisting of around 10% of the digital care sample.

4 Framework

This section lays out the econometric framework for estimating match functions between patients and doctors, and average reallocation effects from rearranging the doctor-patient matches. I follow Graham et al. (2014, 2020) in setting up the empirical framework, with some modifications and extensions. This framework takes into account the externality on the patient from whom the high skilled doctor is moved, and I add the consideration of the

[49] Thus, there will be a slight under-estimation of the counter-guideline prescriptions, since if patients do not fill the prescription the prescription, we will not know if it was an antibiotic. I do robustness analysis using only the incidence of prescription for diagnosis codes where no antibiotics should be prescribed in digital care, available upon request.

opportunity costs in terms of other outcomes when doctors are multitasking and skills are potentially correlated.

This study is complementary to the literature on mechanism design in matching markets (for an overview, see Roth 2012), where strategic incentives of agents are taken into account when studying matching problems. In this paper, I do not study strategic incentives of patients and doctors over whom they match with. There are two main reasons for this. First, in some settings (such as the new digital assignments in several markets), agents have little control over who they match with. Second, as Graham (2011) points out, the study of the effects of alternative assignments is the first step in a more complete policy formulation - before deciding if mechanism design of a decentralized system to implement a desired outcome is relevant, we need to know if there are large benefits to alternative allocations.

Consider D doctors and N patients. Doctors have observable characteristics W_j, which measure doctor effectiveness in different tasks, and patients have observable characteristics X_i which measure patient risk/need for different doctor inputs, and is predicted from patients' healthcare history. Patients also have unobserved attributes V_i and doctors have unobserved characteristics U_j. The potential healthcare output (healthcare outcome Y_{ij}) when patient i matches with doctor j:

$$Y_{ij} = g(W_j, U_j, X_i, V_i) \tag{1}$$

The research design is based on on random assignment (conditional on time[50]) of patients to doctors. Randomization of doctors to patients ensures that the joint density of patient observed characteristics X_i, unobserved characteristics V_i and doctor observed characteristics W_i and unobserved characteristics U_j can be factorized:

$$f_{X_i,V_i,W_i,U_i}(x,v,w,u) = f_{X_i,V_i}(x,v) f_{W_i,U_i}(w,u) \tag{2}$$

Under restriction (2) on the joint distribution of the characteristics of patients and doctors, the conditional mean of the outcome Y_i is:

$$\mathbb{E}\left[Y_{ij}|X_i = x, W_j = w\right] = \iint [g(x,w,v,u) f_{V_i|X_i}(v|x) f_{U_j|W_j}(u|w)] dvdu \equiv \beta(x,w) \tag{3}$$

The Average Match Function (AMF), $\beta(x,w)$, provides information on how match output

[50]The framework will omit the conditioning for simplicity, see Graham (2011, p. 989) for identification conditions under conditional random matching. The conditioning is on time of day (shift) and date of joining the queue for a consultation.

varies across different types of agent pairings, when both doctor and patient are random draws from their respective subpopulations x and w. Figure 1 shows an example of the AMF in our context.

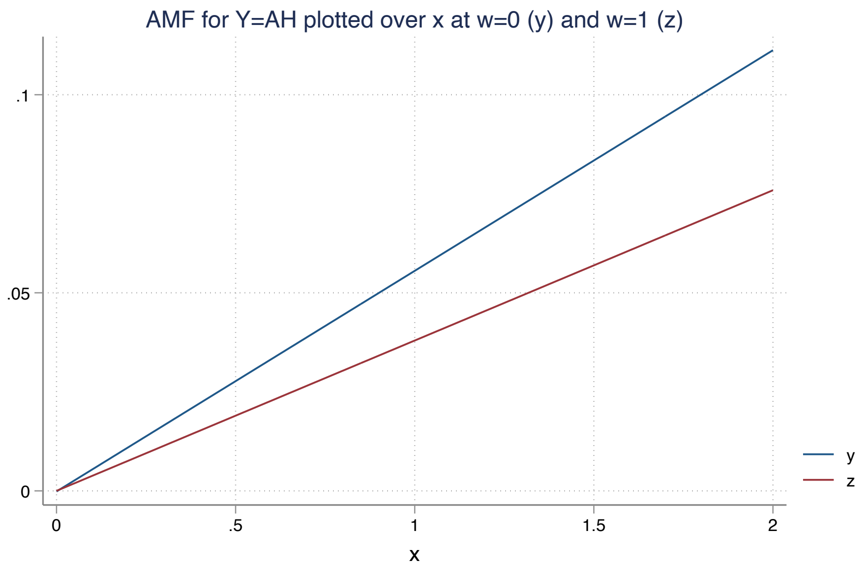


Figure 1: Illustration of the Average Match Function (AMF). The y-axis measures Avoidable Hopsitalizations (AH) and x-axis measures patient risk. W is doctor quality, where w=1 is 1 sd better than w=0, and w=0 measures the worst doctor at this outcome. The positive slopes of both graphs show that a risky patient has higher risk of an avoidable hospitalization, and the flatter slope of the z-graph (where w=1, i.e. a 1sd better doctor) shows the risk is reduced more for risky patients when they meet a better doctor at this task.

The Average Match Function is the main building block for conducting counterfactual analyses. Consider a counterfactual assignment of doctors to patients, i.e. a conditional distribution of doctor types $\tilde{W}_j$[51]:

$$\tilde{f}_{\tilde{W}_j|X_i}(w|x) \tag{4}$$

[51] $\tilde{W}_j$ has an equal marginal distribution to W_j (due to the feasibility condition) but the distribution conditional on patient attributes will differ.

which satisfies the feasibility condition:

$$\tilde{f}_{\tilde{W}_j|X_i}(w|x)f_{X_i}(x)dx = f(w) \tag{5}$$

for all $w \in W$.

The distribution of patients is kept fixed, i.e. $f_{X_i}(x)$ is left unmodified. Average healthcare outcomes under a counterfactual patient-doctor assignment equal:

$$\mathbb{E}\left[\tilde{Y}\right] = \int \left[\int \beta(x,w)\tilde{f}_{\tilde{W}_j|X_i}(w|x)dw\right] f_{X_i}(x)dx \tag{6}$$

where the average match function is used as a building block. The Average Reallocation Effect (ARE) from the reallocation $\tilde{f}$ is $\tilde{Y}$ relative to the average outcome under the status quo allocation, $\bar{Y}^{sq}$:

$$ARE(\tilde{f}) = \mathbb{E}\left[\tilde{Y}\right] - \bar{Y}^{sq} \tag{7}$$

Since everything to the right of the equality in equations (6) and (7) is identified, so is the Average Reallocation Effect (Graham et al. 2014, 2020). To calculate this, I first compute the expected outcome for each type of patient (e.g., $X_i = x$) given their new doctor assignment (e.g., to type $\tilde{W}_i = w$ - the inner integral in equation (6). I then average over the status quo distribution of X_i, which is left unchanged (the outer integral in equation (6)). This yields average patient outcomes under the new assignment of doctors to patients.

4.1 Problem: Reallocation of Fixed Healthcare Resources

The objective of this problem[52] is to improve healthcare outcomes, under the constraint that resources are fixed. Here, the fixed resources are the doctors, including their abilities and number of consultations. Hence, I assume that in the relatively short run I am considering, it is not possible to hire more doctors or increase their abilities. In an extension, I consider selective hiring policies where I extend the working hours of the doctors who have above median skill in several tasks. However, such policies yield smaller gains than matching policies, and are not directly related to the increased flexibility in matching that digital care enables.

[52]It can be interpreted as a problem of a social planner, or of a planner of healthcare provision, either in a healthcare system such as Medicare or a national healthcare system, or a planner in e.g. a Health Maintenance Organization.

I will make one main simplification: to focus on one outcome k at a time, e.g. reducing avoidable hospitalizations. This is reasonable since it is unclear how a planner should weight the different outcomes against each other. I will study what happens to other outcomes when I reallocate to improve one outcome, e.g. avoidable hospitalizations. In fact, it turns out that doctor skills are not positively correlated across outcomes, so there are no important trade-off between the different outcomes measured here. This further strengthens the case for simplifying the problem by focusing on one outcome at a time.

To be realistic, I assume that the planner does not observe U_j or V_i, hence I are restricted to consider only reallocations where unobserved traits are randomized. From now on, I let W_j and X_i be discretely-valued. This is motivated by the fact that I will reduce the dimensionality of doctor and patient types below in order to calculate a semi-parametric version of the average match function.

Suppose we know the AMF $\beta(w, x)$ $\forall(w, x) \in W \times X$ (up to sampling uncertainty), and the marginal distributions of doctor and patient characteristics: $\rho = (\rho_1, ..., \rho_D)'$ for $\rho_d = Pr(W_j = w_d)$ and $\lambda = (\lambda_1, ..., \lambda_P)'$ for $\lambda_p = Pr(X^i = x_p)$. The planner chooses the assignment function $\pi_{ij} = Pr(W = w_j, X = x_i)$ to minimize a negative healthcare outcome such as avoidable hospitalizations:

$$min_\pi Y^k(\pi) = \sum_{i=1}^{I} \sum_{j=1}^{J} \beta^k(x_i, w_j)\pi_{ij} \tag{8}$$

s.t. feasibility/status quo constraints:

$$\sum_{j \in J} N_p \pi_{ij} = N_x \qquad \forall x \in X \tag{9}$$

(each patient gets 1 doctor)

$$\sum_{x \in X} N^\pi(x, w) = N^\pi_{SQ}(w) \qquad \forall w \in W \tag{10}$$

(same workload as in Status Quo (SQ))

where N_p = total number of patients, N_x = number of patients of type x, $N^\pi(x, w)$ = number of patients of type x that doctor w meets in any assignment π, $N^\pi_{SQ}(w)$ = total number of patients that doctor w are assigned to in the status quo (SQ). This problem is similar to those found in Graham, Imbens and Ridder (2020) and Bergeron et al. (2021).

The difference between a candidate assignment and the completely random matching (i.e., the status quo situation where both observed and unobserved characteristics are randomized) is given by:

$$ARE = Y(\pi') - Y(\pi^{rdm}) = \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} (\pi'_{ij} - \rho_j \lambda_i)(\beta(w_J, x_I) - \beta(w_J, x_i) - [\beta(w_j, x_I) - \beta(w_j, x_i)]) \tag{11}$$

where the last term is a measure of the average local complementarity between W and X.

Outcome-maximising assignments will tend to be assortative in regions of complementarity $[\beta(w_J, x_I) + \beta(w_j, x_i)] - [\beta(w_J, x_i) + \beta(w_j, x_I)] > 0$ (Becker 1973, Graham 2011). I will show evidence of complementarities and evaluate ARE of assortative matchings. The average reallocation effect (ARE) takes into account the externality on the patient from whom the high skilled doctor is moved. For each counterfactual reallocation, I will not only compute the average reallocation effect for the main outcome which was intended to be improved with this reallocation, but also compute AREs for other outcomes. The latter will shed light on the opportunity costs of reallocation in terms of other outcomes when doctors are multitasking and skills are potentially correlated.

5 Empirical strategy

The empirical strategy is founded on two building blocks. The first is nationwide conditional random assignment between patients and doctors in digital primary care service. This generates both variation in patient types that each doctor meets - variation in geographic location, age, socioeconomic status, previous healthcare utilization, etc. Moreover, the random allocation allows for causal identification of doctor effects, in contrast to the usual patient-doctor sorting in primary care.

The second building block of the empirical strategy is a split-sample approach, in which I evaluate doctor effectiveness using a fixed effects method in a hold-out sample of randomized digital care (Sample 1). In Sample 2, I use the estimates of doctor skill to estimate causal match effects with patients. This creates the average match function: the expected adverse outcomes conditional on the doctor and patient types. It is also in Sample 2 that I estimate the effects of counterfactual assignments. The samples are completely disjoint and no patients exist in both samples. Both samples have conditional random assignment between doctors and patients.

The patient risk factors are estimated from a third separate sample, which consists of pre-digital (physical) healthcare data in 2013-2016 - the period preceding digital care. In this sample, I identify patient risk (X_i): using past healthcare records. I find logical ex ante patient characteristics which could indicate need for each outcome Y^k, and for the rare outcome avoidable hospitalizations, I predict the risk with a propensity score.

I choose to split the digital care sample in 40-60% shares for power reasons. Also, I want equal number of observations for doctors in Sample 1, thus using 600 visits for each doctor in the estimation of their effectiveness. I choose their *first* 600[53] randomized visits because that is how the procedure could be operationalized. It gives the employer ∼3 months of work by the doctor before they can evaluate the doctor and decide whom to match the doctor to.[54]

5.1 Validating the random assignment using pre-digital care administrative data

The identifying assumption is that within a time period (defined as a 3-hour shift, unique for each date), the allocation of doctors is orthogonal to any patient characteristics which affect the outcomes. To test this for observables, I regress doctor characteristics on patient characteristics when controlling for shift-by-date (randomization strata) fixed effects. Table 4 shows that characteristics are balanced.

5.2 Estimating doctor skill - in Sample 1

Primary care physician skill is challenging to evaluate for several reasons: (1) pervasive sorting between primary care physicians and patients, (2) a lack of linked patient-provider datasets followed over time (3) multitasking and the ambiguity of many measurable outcomes, (4) the delayed nature of the outcomes, and (5) the co-production of healthcare with the patient, where patient adherence, motivation and understanding is key. To overcome (1) and (2), I use the unique random patient-doctor allocation in a nationwide digital primary care service in Sweden in 2016-2018. I also match this with rich pre-digital care administrative data on both healthcare use and socioeconomics to validate the random assignment mechanism to doctors in digital care. For (3), I recognize that multitasking is at the core of possible specialization, and I will deal with this by defining several doctor tasks which

[53] Robustness to 500 and 600 available on request.

[54] The median number of randomized appointments/doctor/calendar day is 10, and I assume 60 working days in 3 months.

Table 4: OLS of doctor on patient characteristics for dropin first visit

	(1) Doc foreign	(2) Only MD	(3) Specialist	(4) GP specialist
Female patient	-0.001	0.003	-0.005**	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)
Patient age	-0.000*	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
1st gen immigrant	0.001	-0.001	0.002	0.002
	(0.004)	(0.004)	(0.004)	(0.004)
2nd gen immigrant	0.000	-0.002	0.009*	0.000
	(0.006)	(0.006)	(0.005)	(0.006)
Municip. density	-0.000	0.000	0.000**	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Sthlm county	-0.002	-0.003	-0.002	0.002
	(0.004)	(0.003)	(0.003)	(0.003)
Self-employed	0.004	-0.004	0.010**	-0.004
	(0.005)	(0.005)	(0.005)	(0.005)
Unemployed	0.009	-0.001	0.003	0.003
	(0.007)	(0.006)	(0.006)	(0.006)
University	0.003	0.002	0.001	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)
Yearly income SEK	0.000*	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Welfare	-0.003	-0.001	0.005	0.000
	(0.004)	(0.004)	(0.004)	(0.004)
Hypertension pre-2016	0.010	0.000	-0.021	0.003
	(0.016)	(0.015)	(0.014)	(0.015)
Asthma pre-2016	-0.006	0.004	0.011	0.002
	(0.010)	(0.010)	(0.009)	(0.010)
Diabetes pre-2016	0.018	-0.036	0.026	0.034
	(0.027)	(0.026)	(0.024)	(0.026)
Anxiety pre-2016	-0.010	-0.001	-0.012*	-0.003
	(0.007)	(0.007)	(0.006)	(0.007)
Depres. pre-2016	-0.006	-0.004	0.006	-0.011
	(0.009)	(0.009)	(0.008)	(0.009)
Constant	0.443***	0.369***	0.278***	0.354***
	(0.005)	(0.005)	(0.005)	(0.005)
N	130941	130941	130941	130941
R^2	0.000	0.000	0.000	0.000

Year-month-date-time shift fixed effects included stand in direct relation to measurable patient outcomes. I assemble data on all patients' prescriptions, acute and secondary healthcare utilization in the national healthcare system, and define several important dimensions of the doctor's outputs, some of which are standard (e.g., preventing avoidable hospitalizations) and some novel (e.g., motivating the patient to fill prescriptions).

To deal with the delayed nature of the outcomes, (4), I use a variety of outcomes, ranging from frequent and lower-stakes, to rare and high-stakes, but all of which are measurable within 3 months. Some of these are closely related by the medical literature to longer-term outcomes such as mortality. I address (5) by specifically studying the varying effectiveness of different doctors with heterogeneous patient types. The co-production with the patient is at the core of possible complementarities, and I use a set of outcome measures that are at varying proximity to the locus of control of the doctor.

In a sample consisting of doctors' first 600 randomized consultations (40% of the sample), I estimate the doctor effect for each task as the average of the effect across all the patients.

$$Y_{ij} = Z_i\Pi + \lambda_t + u_j + \epsilon_{ij} \tag{12}$$

where $\hat{u}_j = \hat{W}_j^{EB}$ is estimated as the Empirical Bayes shrunk random effect of doctor j.[55] This regression is estimated separately for all the outcomes k. λ_t capture date-shift fixed effects (randomization strata) and Z_i is a vector of patient characteristics.[56]

Given a large enough sample size (creating common support in patient types for all doctors) and random allocation, all doctors have a similar patient pool. $\hat{W}_j$ is unbiased due to random assignment and a common support assumption. The common support assumption is that all doctors meet all type of patients in the sample where I estimate the doctor effectiveness. This sample consists of each doctors' first 600 randomized consultations, and this ensures that >95% of doctors have met a patient with an avoidable hospitalization in the past 3 years. However, $Var(\hat{W}_j)$ is positively biased due to (1) regular sampling noise (2) additional noise from sampling noise in the patient sensitivity. I perform an Empirical Bayes shrinkage procedure for the doctor estimates, which results in a best linear predictor of the random doctor effect (Morris 1983). The noisy estimate of doctor quality from a value added regression is multiplied by a measure of its reliability, which in turn is the ration of signal variance to signal plus noise variance. Similar shrinkage is common in studies

[55] A Durbin Wu Hausman test between fixed and random effects does not reject random effects: $Prob > \chi^2 = 0.16$. Results with fixed effects instead of random are similar and are available upon request.

[56] In the case of Y_{ij} being the outcome variable avoidable hospitalizations, Z_{ik} includes the patients' past number of avoidable hospitalizations in the 3 years before digital care, as this is a rare outcome which is correlated over time within patient.

of teacher value-added (see e.g. Kane and Staiger 2008; Chetty et al. 2014; Hjort et al. 2021.). Table 20 in the Appendix shows this regression for the outcome negative number of avoidable hospitalization (so that the random effect is higher for a better doctor).

5.3 Defining patient types

We define patient types based on the corresponding traits which could logically complement the doctor observable tasks. These are listed in Table 5. For the rare outcome avoidable hospitalizations, I create a propensity score based on the lagged outcome variables from data before digital healthcare (2013-15):

$$P_i = X_i \Gamma + \upsilon_i \tag{13}$$

where P_i is the past number of avoidable hospitalizations and X_i are both demographic and healthcare-related variables. I generate a prediction $\hat{P}_i$ for each i, which is what I use as the patient risk variable in Sample 2 (a disjoint sample). Table 6 in the Appendix reports the regression used to create the propensity score for avoidable hospitalization risk.[57] Table 7 in the Appendix reports the regression used to create the propensity score for counter-guideline prescription risk.

Dimensionality reduction for types regarding avoidable hospitalizations
To make fewer assumptions about the nature of complementarities in the match function, I reduce dimensionality of the patient types to a binary variable measuring high and low risk, and use this variable in a semiparametric regression to estimate the average match function. Since around 1% of patients have an AH each year nationally, I characterize 1% of patients as risky ($X = 1$) based on the rank of the propensity score $\hat{P}_i$. To allow for a less time-restricted matching, I characterize 10% of doctors as high skilled in preventing avoidable hospitalizations ($W = 1$) based on the rank of $\hat{W}_{jk}^{EB}$.[58]

Figure 2 illustrates that the groups created based on the propensity score are closely related to the number of past avoidable hospitalizations of the patient. A patient in the risky group has had on average 0.35 AH in the past 3 years, while a patient classified as not risky has had on average 0.01 AH in the same period. Figure 3 shows how the risk groups (defined only based on past healthcare records and demographics) are highly predictive of

[57] This is done with a linear probability model, but robustness checks with ordered logit do not change the results.

[58] This will give a lower effect of the interaction effect than if I had also picked the top 1% of doctors in this skill, but since I do not want to make patients wait too long for the best doctor for them, I pick 10% so that there is a wider choice of good doctors in this skill in each time period.

future avoidable hospitalizations.

Quality measure for doctor	**Prior outcome for patient (2013-15)**
1. Preventing avoidable hospitalizations	Propensity score of nr. avoidable hospitalizations
2. No counter-guideline antibiotics prescription	Antibiotics consumption share of total prescriptions
3. Visit to physical nurse following week	Patient visiting physical nurse week before

Table 5: Overview of which patient prior variables are used to define target groups for different doctor quality measures.

5.4 Match effects: In Sample 2

By interacting doctor effectiveness with the relevant patient characteristic (X_i) in a second step, I estimate individual sensitivity to doctor input. Again, this is estimated in a different sample (Sample 2) from that where I estimated $\hat{W}_{jk}^{EB}$ (Sample 1). Sample 2 is each doctor's first visit randomized consultations *after* the 600th.

The estimation of the healthcare production function will be semi-parametric, to avoid making too restrictive assumptions on the form of complementarities. To do this, I reduce the dimensionality of doctor and patient characteristics, by making X and W binary. I characterize 1% of patients as risky ($X = 1$) based on the ranking of their propensity score $\hat{P}_i$ (around 1% of patients have an AH each year nationally). Moreover, I characterize 10% of doctors as high skilled in each outcome ($W = 1$) based on $\hat{W}_{jk}^{EB}$. Then, I estimate the effect of a top 10% doctor on top 1% risky patient:

$$Y_{ij} = \alpha + \beta_1 W_j + \beta_2 X_i + \beta_3 W_j X_i + \lambda_t + e_{ij} \tag{14}$$

where λ_t is date-time-shift (randomization strata) fixed effects. Standard errors are clustered on doctors. The main coefficient of interest is β_3. In addition, β_{2k} measures how different the patient group as I defined it is in the outcome variable on average.

5.5 Reallocation procedures and costs

Reallocation based on the semi-parametric AMF to reduce avoidable hospitalizations

The simplest reallocation procedure I carry out uses the semi-parametric match function with a binary definition of doctor and patient types. Here, I reallocate the top 10% doctors randomly to top 1% high-risk patients and let them swap doctors with some non-risky patients. Costs of reallocations are small in the digital setting compared to the physical setting where geographic distances play a big role. One cost in the digital setting could

Table 6: Nr. avoidable hosp. 3 years before consultation

	(1)
Disease index	0.0684***
	(0.0069)
Female	-0.0027*
	(0.0014)
Age	0.0001
	(0.0001)
2nd gen immigrant	0.0013
	(0.0021)
1st gen immigrant	0.0047*
	(0.0028)
Nr hosp 3 years before excl. AH	0.0049**
	(0.0020)
_cons	-0.0005
	(0.0018)
N	95816

This is used to create patient propensity scores for AH risk.
Robust SEs in parantheses.
Main sample: patients who had visits after doctors' 600th visit.
Patients born before 2013, to allow 3 years pre-data.
Disease index is sum of Elixhauser comorbidities.

Table 7: Nr. avoidable hosp. 3 years before consultation

	(1)
Disease index	-0.0482***
	(0.0008)
Female	-0.0252***
	(0.0013)
Age	-0.0026***
	(0.0000)
2nd gen immigrant	-0.0127***
	(0.0023)
1st gen immigrant	-0.0149***
	(0.0018)
_cons	0.2627***
	(0.0018)
N	116391

This is used to create patient propensity scores for CGP risk.
Robust SEs in parantheses.
Main sample: patients who had visits after doctors' 600th visit.
Patients born before 2013, to allow 3 years pre-data.
Disease index is sum of Elixhauser comorbidities.

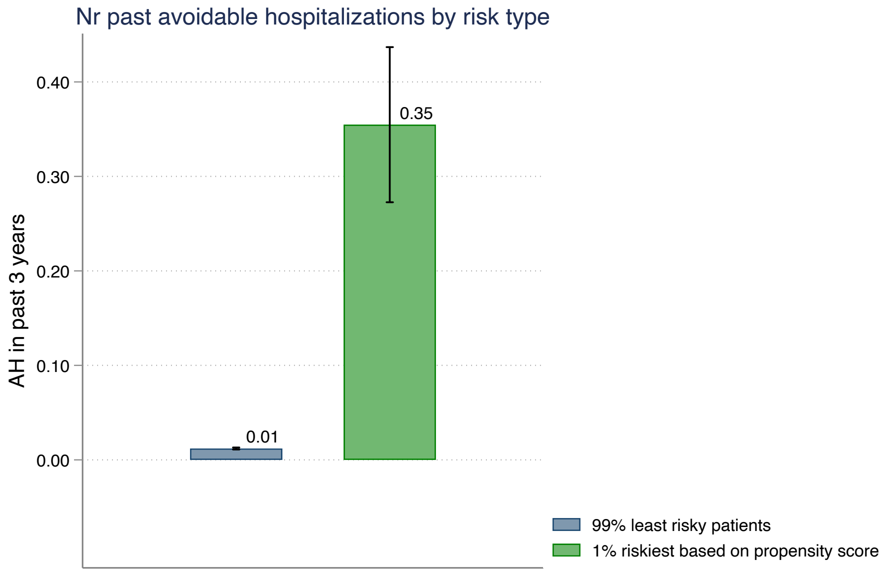


Figure 2: High and low risk patients: past 3 years' Avoidable Hospitalizations. 95% confidence intervals in black.

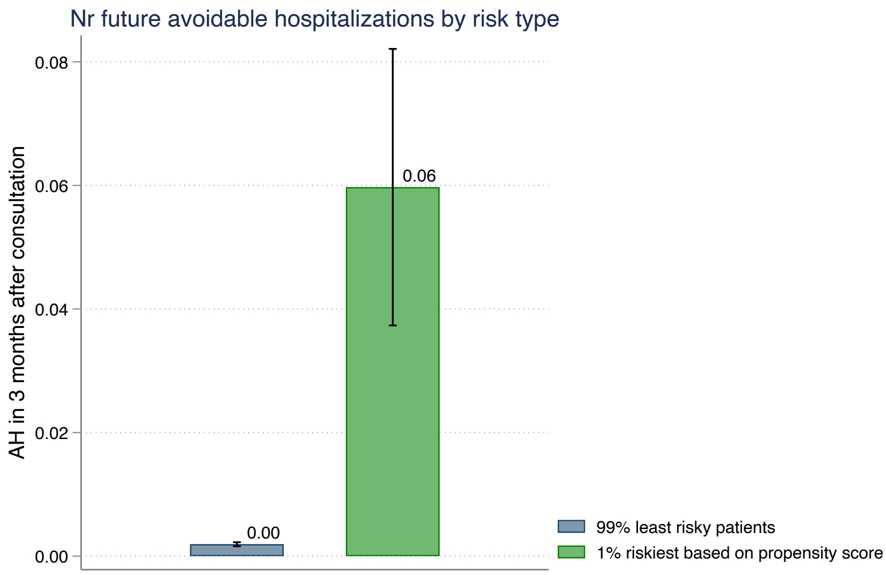


Figure 3: High and low risk patients: future 3 months' Avoidable Hospitalizations. 95% confidence intervals in black.

be longer waiting time for patients to get the reallocated doctor. These costs are small as we are only reallocating 2% of consultations (= the top 1% risky patients and the patients they swap doctor with) in the reallocation mentioned above. Moreover, among these 2%, 55% of patients can be reallocated to a doctor within the same time shift, meaning there is a negligible additional time cost for them. Hence, any additional waiting from the reallocation procedure would occur only for 0.9% of patients, and only half of them are high-risk patients.[59]

Reallocation based on the parametric AMF to reduce counter-guideline prescriptions
The reallocation procedure for the parametric version of the match function, where I use continuous measures of patient risk and doctor skill, is positive assortative matching (PAM): allocate the highest effectiveness doctors to the highest need/risk patients. This builds on the standard Becker (1973) results that the outcome-maximizing allocation when there are complementarities is assortative.

Informational requirements
The informational requirements to carry out these reallocations consists in having access to patients' past healthcare records. This can be compared to earlier research showing that electronic medical records reduce deaths by making information accessible (Miller and Tucker 2011). Specifically, for the avoidable hospitalizations reallocation, data is needed on the past three years' avoidable hospitalizations as well as the age and gender of the patient. Demographic data about patients is readily available to the healthcare provider, while data on past avoidable hospitalizations can be accessed in theory, if the electronic medical records are built to flag these events.

The data needed on patients for the reallocation reducing counter-guideline prescription is data on their past three years' antibiotics prescriptions as a share out of total prescriptions. This data also exists in patients prescription histories which is part of their electronic medical records. The data needed on doctors is data on their first 600 patients' outcomes and histories. In the case of counter-guideline prescriptions, the outcomes data already exists within the medical provider as the diagnosis and prescription drug are recorded and can be used to determine guideline adherence. For avoidable hospitalizations, three months' follow up hospitalization data is needed for the doctor's first 600 randomized patients, and this could be achieved by an integration of medical records where only patients who have avoidable hospitalizations are flagged and reported back to the digital healthcare

[59]Since I have defined high-skilled doctors as the top 10% in avoidable hospitalization reduction effectiveness, and depending on the year, day and time there are 10-30 doctors working at each shift, it is highly likely that if a high-risk patient cannot be allocated to a top 10% doctor in the same shift where she originally met a doctor, they would have to wait maximum 3 hours to the next shift.

provider. Such follow up data would be useful even in the absence of a reallocation objective. Currently, the ownership of the means of prediction remains with the governmental agencies that host patient data, as well as with the providers that produce the data.

6 Results

6.1 Reallocation results

The first part of the results covers counterfactual simulations: the Average Reallocation Effects (ARE). The following section presents results on what drives these effects in terms of causal match effects and stylized facts about skills. Finally, we study the healthcare production more in detail to make clear the mechanisms in terms of doctor actions.

The first set of Average Reallocation Effects are derived from the optimization problem in Section 4.1. This problem takes existing resources in terms of doctor skills and time worked as given, as it might be difficult if not impossible to increase all doctors' skills at several different tasks[60], and there are constraints to hiring new doctors. Moreover, retraining (and placing emphasis) on some skills may lead other skills to suffer in a multitasking setting. I evaluate the aggregate effects of reallocating doctors between patients with different needs, subject to the above-mentioned constraint. I consider reallocating doctors according to patients' risk for each outcome variable, as described above. I will focus here on reallocations to reduce the adverse outcome avoidable hospitalizations – other reallocations can be found in the appendix.

The first result (Figure 4) is that avoidable hospitalizations (AH) decrease by 20% when I have matched doctors and patients on doctor AH-prevention skill (skill in risk prediction/triaging) and patient AH risk as described in Section 5.5. At the same time, the aggregate number of counter-guideline prescriptions and double visits[61] do not change. Hence, the positive outcome (reducing AH) has been achieved without increasing other negative outcomes. There are also effects on healthcare inequality. Before reallocation, the probability of meeting a top 10% doctor in risk prediction/triaging was similar across patients' income distribution (Figure 5). After the reallocation, the chance of meeting a top 10% doctor in risk prediction/triaging increases for the bottom decile, without any large change for the rest of the income deciles (Figure 5). This is because the risk for avoidable hospitalizations is highest in the lowest income decile.

[60] Although I will also consider a selective hiring policy, which creates smaller benefits than matching.

[61] Double visit means that the patient contacted a nurse in physical care the week after the digital visit

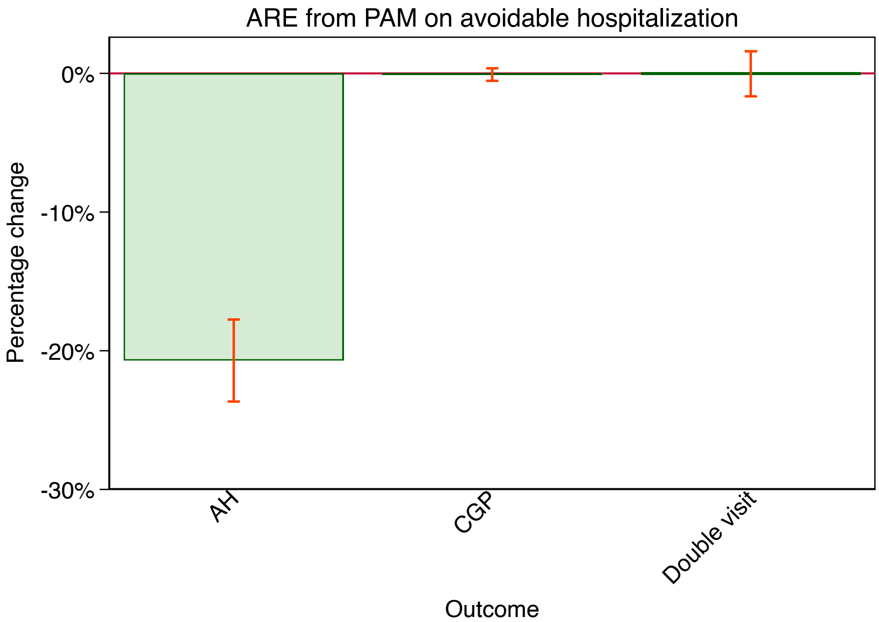


Figure 4: Average Reallocation Effects from the semi-parametric match function where a good doctor is defined as a top 10% in the AH outcome.

Figure 6 presents another way of understanding the income-health gradient aspect of doctor-patient matching. This figure shows Average Reallocation Effects from a reallocation where the highest-skilled doctors in reducing avoidable hospitalizations are matched with the highest-income patients (positive assortative matching (PAM) on doctor skill and patient income using the parametric match function[62]). This reallocation is compared to the random real-life digital assignment, and shows that avoidable hospitalizations would be around 5% higher if the highest income patients were matched with the highest skilled doctors in preventing avoidable hospitalizations[63].

These results can be interpreted in light of the results from the descriptive analysis earlier in this paper about a positive relationship between patient area-level income and perceived quality of local primary care, as well as results from other studies which indicate that higher-income patients get access to better doctors in physical care (e.g., Agency for Healthcare Research and Quality 2020). If this also applies to risk detection and triage skill for physical care doctors, Figure 6 suggests that avoidable hospitalizations after physical primary care could be lowered by up to 5% if patient-doctor matching changed to a random matching. Moreover, if we add together the results from Figures 4 and 6, they suggest that moving to an needs-based allocation on avoidable hospitalizations from a positive assortative matching on patient income and doctor skill could reduce the number of avoidable hospitalizations by 25%.

The gains from matching are much larger than the gains from a more selective doctor hiring policy, which I simulate by increasing the work hours of the doctors who are have above median skill ("good") in all three outcome measures (preventing avoidable hospitalizations, following antibiotics guidelines, and reducing double visits in digital and physical primary care in the same week). However, Figure 7 illustrates that even if we increased these doctors' work hours by as much as 70%, we still would see no significant improvement in aggregate avoidable hospitalizations, and only a 4% reduction in counter-guideline prescriptions (less than half of the reduction from matching doctors and patients to reduce counter-guideline prescriptions). Moreover, a 70% increase in these doctors' work hours would hard to achieve, even if digitalization can be expected to give room for some increase in hours.[64].

[62]The parametric match function means that I use each doctor's estimated effect from the hold-out sample, and each patient's estimated risk, instead of the binary dimensionality reduction of high/low types that I use in the semi-parametric match function

[63]The figure also shows that counter-guideline prescriptions would remain unchanged compared to the random allocation, which is expected given the zero correlation in those skills within doctors.

[64]For instance, if digital care saves commuting time for the doctors, we could imagine increasing the "good" doctors' working hours by 10-20%, but not by 70%. An average round-trip commute in Sweden is

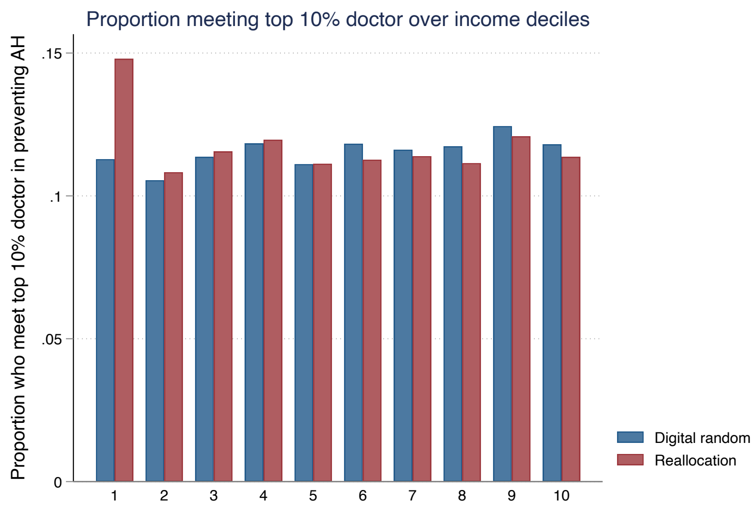


Figure 5: This figure shows what proportion of patients across income deciles who meet a doctor who is classified as top 10% in reducing avoidable hospitalizations, in the reallocation to reduce avoidable hospitalizations as well as in the random allocation. All income deciles have a slightly higher than 10% proportion of top doctors, which is because the top doctors work more consultations than other doctors. The patient income is the income of adult patients in 2017.

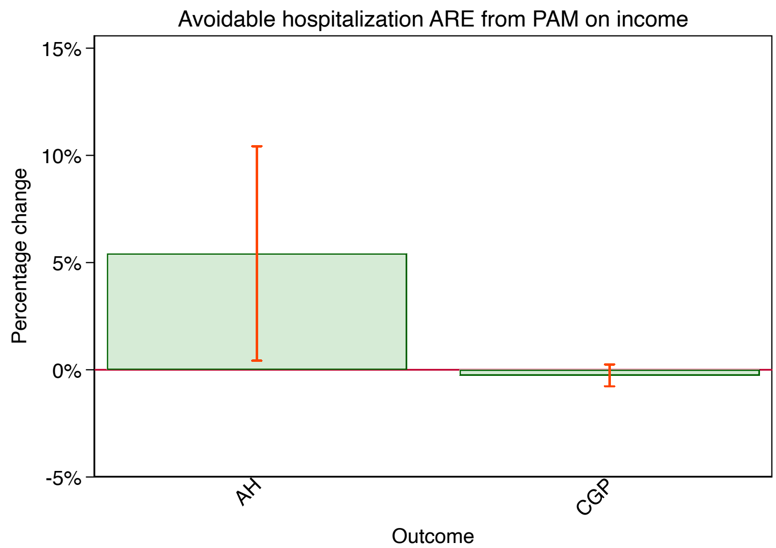


Figure 6: Average Reallocation Effects from Positive Assortative Matching (PAM) on doctor skill in reducing avoidable hospitalizations, and patient income, using the parametric match function.

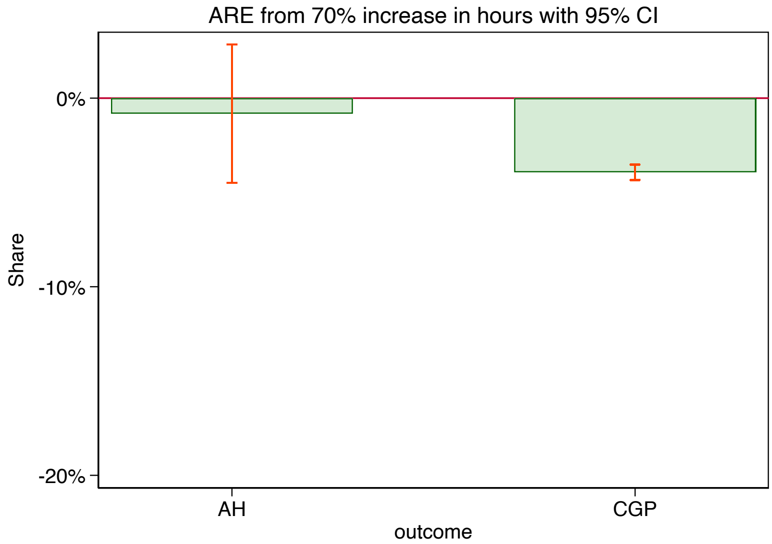


Figure 7: Average Reallocation Effects from an increase in work hours by 70% for the doctors who are above median in all three measures (preventing avoidable hospitalizations, double visits and counter-guideline prescriptions) without any matching to patients, using the parametric match function.

around 40 minutes and doctors work shorter shifts than 8 hours.

6.2 What drives the gains from matching?

6.2.1 Variation in doctor effectiveness within task

The first driver behind the gains from reallocation is that there is variation in doctor effectiveness in each task. Figure 8 shows that the share of a doctor's patients who end up having an avoidable hospitalization within 3 months after the consultation ranges from virtually 0% to 0.6% .

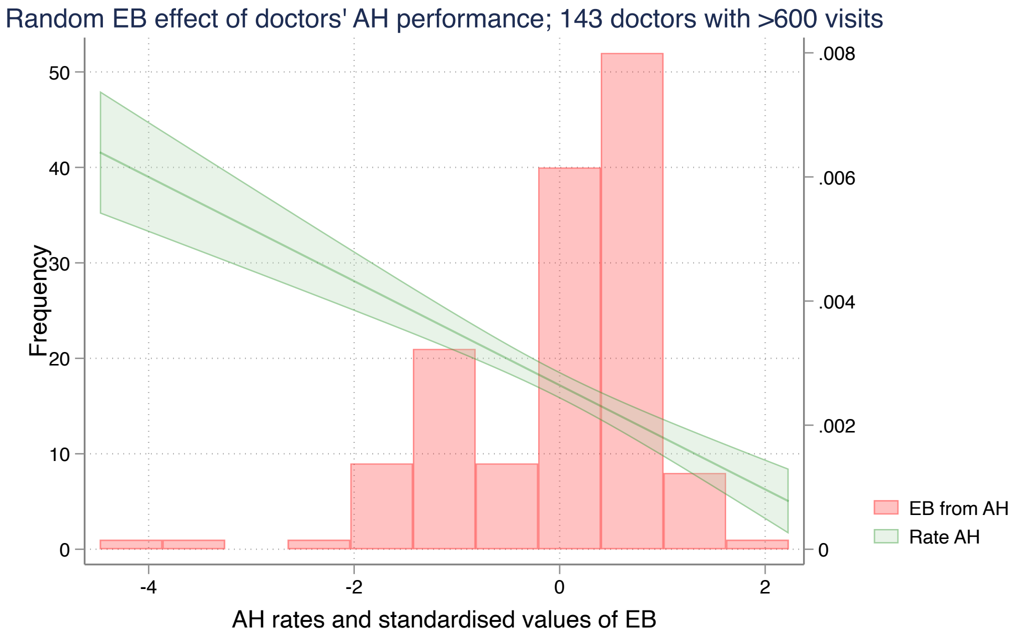


Figure 8: Histogram of the standardized (over doctors) Empirical Bayes (EB) estimates of doctor quality with the outcome *negative* avoidable hospitalizations (AH) in the 3 months after the visit (in red). Included are the 143 doctors with more than 600 randomized first visit consultations. Overlaid in green is the predicted rate of AH out of the doctor's total consultations, from a regression of the rate on the EB random effect, with a 95% confidence interval.

6.2.2 No positive correlation in effectiveness across tasks: specialization

If some doctors are better at all tasks, then reallocation would be more difficult as the planner would need to prioritize more between different patients who have needs for different doctor skills. However, Figure 9 and Table 8 show that there is no positive correlation

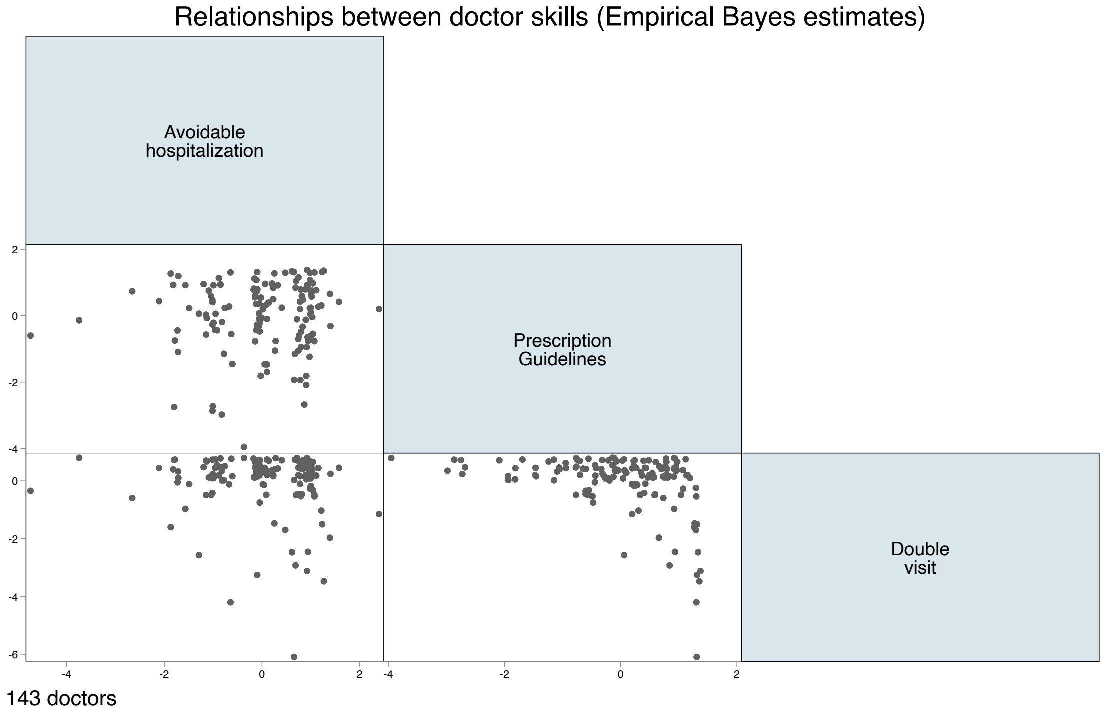


Figure 9: Scatterplots of different doctor skills.

between doctors' effectiveness in different tasks. In fact, that there is a negative within-doctor relationship between certain skills. For instance, a doctor who is better at following antibiotics guidelines is slightly worse at preventing double visits (when the patient seeks physical nurse care the week after the digital doctor appointment) (see also Appendix Figure 21). This can be conceptualized as specialization. It may also be related to patient behavior. Some patients may particularly want an antibiotic. If they do not get it from the digital doctor, because the doctor adheres to guidelines, then they might be more likely to call the nurse at the physical healthcare clinic the following week, to try to get antibiotics from there. But even in this case, it reflects a different balance struck by the doctor in the trade-off between following guidelines and satisfying the patient.

6.3 Match Effects

The final driver of the reallocation effects is evidence of strong complementarities or "match effects": causal treatment effects of matching doctors of higher effectiveness in outcome k

Doctor skill	AH	CGP
CGP	0.0655	
	(0.4368)	
Double visit	-0.0861	-0.3528
	(0.3068)	(0.0000)

Table 8: Spearman's rank-order correlation coefficient. In parantheses: **p-value** from test of H^0: the two effectiveness measures are independent. N= 143.

to patients with higher estimated need/risk in outcome k.[65] A doctor who is among the top 10% at reducing avoidable hospitalizations (AH) in the hold-out sample reduces AH by as much as 90% for the top 1% risky patients in the main sample, but has no effect on the rest of patients (Table 25). These complementarities in patient-doctor matching are illustrated graphically in Figure 10. Table 21 in the Appendix shows results from the parametric version of the match regression, and includes robustness checks.

Table 9: Number Avoidable Hospitalizations within 3 mo. after visit

	(1) Clustered SEs	(2) Bootstrapped SEs
Top 10% doctor X top 1% risky patient	-0.060***	-0.060***
	(0.014)	(0.016)
Top 10% doctor on AH	0.000	0.000
	(0.001)	(0.001)
Riskiest 1% patient in AH	0.067***	0.067***
	(0.012)	(0.014)
N	95816	95816
Mean	0.003	0.003
Mean_risky	0.062	0.062

All columns have date-time shift fixed effects. Sample is all doctors' randomized visits after the 600th' consultation

While the targeting of patients who are at risk for avoidable hospitalizations may be most important, there are also effects of matching patients on who have had a higher share

[65] This is not ex ante evident - it could have been that high risk patients are simply not possible to help from the bad outcome, and that it would be best to allocate the most effective doctors to patients who had less risk and were more amenable to change.

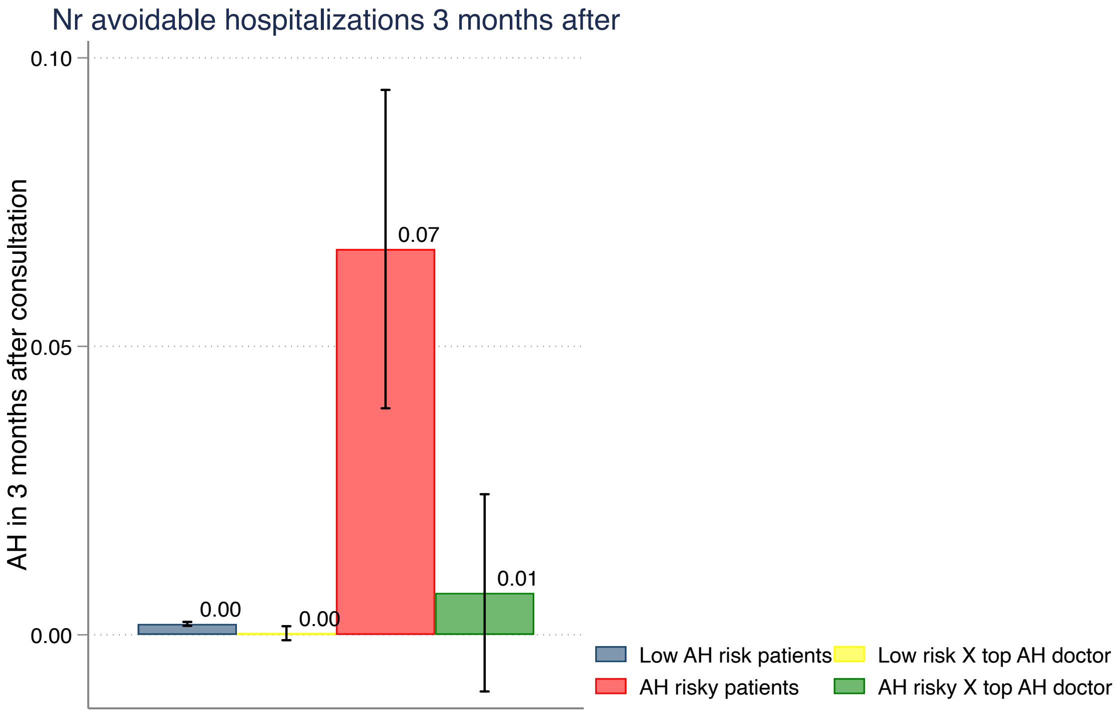


Figure 10: Bootstrap 95% confidence intervals in black.

Table 10: Definitive counter guideline prescription

	(1) OLS Simple	(2) Controls	(3) Bootstrap	(4) Time shift FE
Std doctor FE based on no CGP	-0.0028***	-0.0028***	-0.0028***	-0.0028***
	(0.0010)	(0.0010)	(0.0011)	(0.0008)
Share antib 3yrs b4	0.0489***	0.0529***	0.0529***	0.0526***
	(0.0031)	(0.0034)	(0.0033)	(0.0033)
Std doc FE X share antib 3yrs b4	-0.0131***	-0.0131***	-0.0131***	-0.0126***
	(0.0020)	(0.0020)	(0.0022)	(0.0020)
Controls	No	Yes	Yes	Yes
N	116396	116391	116391	116391
R^2	0.008	0.009	0.009	0.009
Mean	0.0172	0.0172	0.0172	0.0172

All columns: SEs in parantheses clustered on doctors. Col 3 has booststrapped SEs.

Sample is dropin first visits after doctor's 600th such visit. Only patients born before 2013.

Controls are Elixhauser sum of comorbidities, female, age, first- and second-generation immigrant.

of antibiotics prescriptions in the past to doctors better at following antibiotics guidelines.[66] A patient who had a 50% higher share of antibiotics out of their total pre-digital care prescriptions has 2.4%-2.6% higher risk of receiving a counter-guideline prescription, suggesting that the patient may want or need more antibiotics (Table 10).[67] Ex ante, it is not obvious that a doctor who has had a good track record in the holdout sample of not prescribing a counter-guideline prescription (CGP), would also be more restrictive with antibiotics when meeting a patient in the main sample who has a higher share of antibiotics in the past[68]. It could be the case that such a patient needs more antibiotics and any doctor would be willing to surpass the guidelines with such a patient.

However, it turns out that if a patient who has a 50% higher share of antibiotics out of their total pre-digital care prescriptions is matched with a doctor who is one standard

[66] In this regression, I have not reduced the dimensionality of doctor and patients types to binary for the semi-parametric specification. Instead, the regression specification has the continuous doctor skill and patient risk and their interaction

[67] Either that the patient is particularly fragile so that any doctor would prescribe a little more over cautiously for them - but I am controlling for age, gender and Elixhauser sum of comorbidities in Columns 2-4 which controls for their disease level. Otherwise it suggests that the patient is particularly keen on antibiotics, and potentially tries to pressure the doctor to get them.)

[68] We do not know if the antibiotics in physical healthcare in the patient's history were according to guidelines or not.

deviation better in the hold-out sample at following guidelines, their risk of getting a CGP is reduced by 24-27%.[69]

6.4 Mechanisms for preventing avoidable hospitalizations

To shed light on the mechanisms though which some doctors are particularly effective at preventing avoidable hospitalizations, I study the actions that the doctors take during the digital care consultation. Table 11 shows the most common outcomes (in terms of doctor actions) of a consultation, together capturing 98% of the consultations' outcomes. These outcomes are prescription, advice only, redirection, referral, and sick note. To redirect a patient means to tell them that their condition is not suitable for digital primary care, and that they should go to, e.g., a physical primary care center, possibly one with extended opening hours, or in some cases the Emergency Department.[70] The main take-away from Table 11 is that doctors who are among the top 10% at preventing avoidable hospitalizations (AH) are more likely to identify that the AH-risky patients need other care than digital and redirect them. At the same time, they are less likely than other doctors to only give advice to these patients. There are no significant differences in how the top 10% doctors at AH treat other patients than the risky – meaning that it is not the case that these doctors are simply more cautious and avoid false negatives at the expense of increasing false positives. False negatives in this case would be that patients who need additional checkups in physical care are not redirected, while false positives would be that patients who do not need additional checkups in physical care are redirected for these checkups, which would be costly.

These results indicate that the AH-skilled doctors are better at identifying the patients at high risk and determining that they (and not other patients) need more care (triaging), which can possibly prevent an avoidable hospitalization. Triaging is one of the key components of a primary care physician's job and can make the difference between appropriate, cost-effective care and poor outcomes at high cost (Vasilik 2021). Triaging is difficult and requires separating the few urgent patients from the many non-urgent patients. The medical literature indicates that while triage handbooks exist, they may be difficult to use in practice and there are no explicit guidelines at many primary care centers (Vasilik 2021). Hence,

[69] Table 22 in the appendix uses the number of antibiotics instead of the share for the patient risk variable, and the results are similar.

[70] A referral, on the other hand, means that the doctor writes a letter to a specialist clinic and the patient will in due course be called by the clinic. This can take weeks or months depending on the condition and wait list. In our data, the share of consultations ending in referrals from the primary care provider to a specialist clinic (3% of consultations) are comparable to the lower end of GP referrals in the UK physical setting (where in a meta-analysis, they range between 1.5% and 24.5% (O'Donnell 2000)).

the triaging process requires experience and knowledge within several fields of medicine (Göransson et al. 2021).

We have seen that the doctors who prevent more avoidable hospitalizations for risky patients do not do this at the expense of redirecting a higher share of non-risky patients for additional care. But are there other downsides to these doctors' work, potentially that they spend longer time with the patients, thus decreasing the time available for other patients? Column 1 of Table 12 shows that the consultation duration is no different when an AH-risky patient meets a top 10% doctor in preventing AH. Column 2 of Table 12 also shows that there are no significant differences in the administration time - the time that the doctor spends after the consultation on writing notes and prescriptions, etc.

A final question which bears on future possible strategic incentives and mechanism design, is whether patients recognize which doctors are most appropriate for their needs. Column 3 of Table 12 shows that patients in general are more satisfied with the top 10% doctors in AH prevention. However, patients who are at risk for avoidable hospitalizations are not differentially more satisfied with these doctors.

Table 11: Doctor actions during digital visit

	(1) Redirected	(2) Advice only	(3) Prescription	(4) Referral	(5) Sick note
Top 10% doctor on AH	0.00	-0.00	-0.00	0.01	0.00
	(0.02)	(0.02)	(0.02)	(0.01)	(0.00)
Riskiest 1% patient in AH	0.07***	0.00	-0.09***	-0.00	-0.01
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Top 10% doctorX 1% riskiest patient	0.11**	-0.09**	-0.03	0.00	-0.01
	(0.06)	(0.04)	(0.07)	(0.01)	(0.01)
N	91519	91519	91519	91519	91519
Mean	0.12	0.26	0.53	0.03	0.04

Date time shift FE included. SEs in parantheses clustered on doctors.

Table 12: Process outcomes during digital visit

	(1) Duration, mins	(2) Admin time, mins	(3) Score, 1-5
Top 10% doctor on AH	-0.0693	-0.2043	0.0812**
	(0.3180)	(0.9744)	(0.0385)
Riskiest 1% patient in AH	0.1956	0.1098	-0.1567***
	(0.1263)	(0.3186)	(0.0389)
Top 10% doctorX 1% riskiest patient	-0.0942	0.8547	-0.0452
	(0.1613)	(1.0113)	(0.1381)
N	93869	93868	70607
Mean	4.5226	11.7034	4.6331

Date time shift FE included. SEs in parantheses clustered on doctors.

In column 1 the outcome variable is patient-doctor consultation duration; in column 2 it is the doctor's administration time after the meeting, spent on e.g. issuing prescriptions and writing notes; and in column 3 it is the patient's satisfaction rating of the doctor, ranging between 1 and 5.

7 Conclusion

The digitalization of services, such as parts of healthcare and education, has many implications, three of which are especially important for the topic of this paper. First, new possibilities for matching service providers and users emerge as the number of potential providers that any given user could meet has increased. Video consultations mean that the constraint of doctors and patients sharing the same geographic location is less binding. Second, the digitalization of services results in detailed data about each agent's work and outcomes, which can be used to measure providers' task-specific skills and users' needs and used in algorithmic matching. Finally, this technology disrupts geographically related sorting patterns between service providers and receivers, which often resulted in inequality in service quality. Together, this creates new opportunities to transform the organization of service production by changing the matching between service providers and users to make better use of provider skills. This is a process that has already started with digital healthcare firms using algorithms to match users and patients, so we need to better understand this development.

Within healthcare, many countries face pressing cost challenges and human capital shortages. At the same time, digital technology and video consultations have grown fast in healthcare, just as in other sectors. I quantify the potential gains from a better allocation of scarce resources: doctor skills, by first using detailed data about both patients and doctors to classify types. Such a reallocation procedure is potentially cost-neutral, as opposed to training and hiring additional doctors, which is costly. Moreover, I show that the gains from a selective hiring policy are not nearly as large as those from a matching policy.

The matching gains are driven by another fact that I establish: that there is heterogeneity in skill among doctors in dimensions that vary in importance for heterogeneous patients, even within general practice which is studied in this paper. I have shown that physician effectiveness varies considerably in different tasks. The evidence is not consistent with a single latent ability variable governing doctor effectiveness on all the outcome measures, but rather with specialization. Moreover, doctors' effectiveness varies with different patients who have varying pre-existing risk relevant for the different doctor tasks. If we match a doctor who is among the top 10% at reducing the main outcome *avoidable hospitalizations*, with a patient who is predicted to be among the top 1% risky for such adverse outcomes, we could reduce their number of such adverse outcomes by 90%. However, we need to take these doctors from other patients who may themselves also have a small risk for the adverse outcome.

To understand the trade-off between the positive and the negative effects from this reallocation, I calculate the aggregate effects of reallocating doctors. Reallocating the best doctors at preventing avoidable hospitalizations (AH) to the patients at risk reduces AH by 20% without making other main outcomes worse, and by reallocating only 2% of patients. A back-of-the-envelope calculation shows that an AH reduction of 20% scaled up nationally could hypothetically save up to 2% of total hospital costs in Sweden (USD 160 million in Sweden)[71] and the US (USD 6.7 billion in the US for only adults in purely hospital costs), apart from lives saved.[72] Moreover, reallocating the best doctors at following antibiotics guidelines to patients who are intensive users of antibiotics reduces counter-guideline prescriptions by 10%, potentially contributing to the global battle against bacteria becoming resistant to antibiotics through externalities from over-prescription.

A main take-away is that in primary care, doctor heterogeneity in skill and patients' varying needs matter: there are gains to be made from a doctor-patient reallocation where provider specialized skills are put to better use. It is highly likely that this could also be the case in other service sectors, and when services move to digital, this is becoming a feasible and resource-neutral, low-cost way of increasing effectiveness of service provision. Moreover, this could have effects on inequality in access to high-quality services across the income distribution.

[71] Calculated from an estimate of the total costs of avoidable hospitalizations: 820 million USD per year in Sweden. The number of hospital days for AH was around 1 million in Sweden in 2010 (Socialstyrelsen, 2011, p.51). The average cost per day in inpatient care is 7100 SEK (Socialstyrelsen, 2017). The US figure on the total costs of avoidable hospitalizations is USD 33.7 for adults only (Jiang et al 2009; McDermott and Jiang 2020)

[72] In both countries, this saving represents around 0.03% of GDP.

8 References

- Abaluck, Jason, Agha, Leila, Kabrhel, Chris, Raja, Ali, and Venkatesh, Arjun. “The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care”. The American Economic Review 106.12 (2016): 3730-764.

- Abdulkadiroglu, Atila, Yeon-Koo Che, Parag A. Pathak, Alvin E. Roth, and Olivier Tercieux, “Efficiency, Justiffied Envy, and Incentives in Priority-Based Matching”. American Economic Review: Insights, Vol. 2, No. 4, (Pp. 425-42).

- Agency for Healthcare Research and Quality. 2020. “2019 National Healthcare Quality and Disparities Report”. Rockville, MD: Agency for Healthcare Research and Quality. AHRQ Pub. No. 20(21)-0045-EF. Link. Accessed 4 October 2021.

- Alsan, Marcella, Owen Garrick, and Grant Graziani. 2019. “Does Diversity Matter for Health? Experimental Evidence from Oakland”. American Economic Review 109.12: 4071-4111.

- Aucejo, Esteban M., Patrick Coate, Jane Cooley Fruehwirth, Sean Kelly and Zachary Mozenter. 2019. “Teacher Effectiveness and Classroom Composition”. CEP Discussion Paper No 1.

- Bau, Natalie. 2021. “Estimating An Equilibrium Model of Horizontal Competition in Education”. Unpublished. Accessed 14 September 2021. Link.

- Baumhäkel M, Müller U, Böhm M. 2009. “Influence of gender of physicians and patients on guideline-recommended treatment of chronic heart failure in a cross-sectional study”. Eur J Heart Fail, 11(3):299-303.

- Becker, Gary S. 1973. “A Theory of Marriage: Part I.” *Journal of Political Economy.* Volume 81 Number 4.

- Bendz, Anna (2011) “Vårdvalet i Västsverige” in Annika Bergström (ed.) *Västsvensk vardag.* Göteborgs universitet: SOM-institutet.

- Bergeron A, Bessone P, Kabeya JK, and Tourek G. “Quality and Optimal Matching of Bureaucrats: Evidence from Tax Collection in the DRC.” Unpublished. Link. Accessed August 21, 2021.

- Berthold HK, Gouni-Berthold I, Bestehorn KP, Böhm M, Krone W. 2008. “Physician gender is associated with the quality of type 2 diabetes care.” J Intern Med. 264(4):340-350.

- Bhattacharya, Debopam, and Pascaline Dupas. 2011. “Inferring welfare maximizing treatment assignment under budget constraints.” Journal of Econometrics 167 (2012) 168–196. Journal of the American Statistical Association 104 (486), 486–500.

- Bhattacharya, Debopam. 2009. “Inferring optimal peer allocation using experimental data”. Journal of the American Statistical Association 104 (486), 486–500.

- Bond, Philip and James Dow. 2021. “Failing to forecast rare events.” *Journal of Financial Economics*, 2021-06.

- Bonhomme, S., Lamadon, T. and Manresa, E. 2019. “A Distributional Framework for Matched Employer Employee Data.” Econometrica, 87: 699-739.

- Cabral, Marika and Marcus Dillender. 2021. “Gender Differences in Medical Evaluations: Evidence from Randomly Assigned Doctors.” Manuscript. Link. Accessed on August 3, 2021.

- Centers for Disease Control and Prevention. 2019. “Antibiotic Resistance Threats in the. United States.” Link. Accessed on September 1, 2021.

- Chen, Yiqun, Petra Persson and Maria Polyakova. 2021. “The Roots of Health Inequality and The Value of Intra-Family Expertise.” Mimeo. `https://web.stanford.edu/~perssonp/FamilyDoctors.pdf`

- Chen, Yiqun, Raj Chetty, Luke Chu, David Cutler, Lorena Di Bono, Andreas Haller, Katja Hofmann, Jonas Minet Kinge, Claus Thustrup Kreiner, Hsien-Ming Lien, Kevin Milligan, Benjamin Milner, Thomas Minten, Torben Heien Nielsen, Petra Persson, Maria Polyakova, Tammy Schirle, Benjamin Ly Serena, Johannes Spinnewijn, Stefan Staubli, Michael Stepner, Tzu-Ting Yang, Yuting Zhang and Josef Zweimüller. 2021. “Health Inequality Around the World: Comparing the Relationship Between Income and Life Expectancy in Ten High-Income Countries.” Unpublished.

- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. “Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates.” American Economic Review, 104 (9): 2593-2632.

- Condie, Scott, Lars Lefgren, David Sims. 2014. Teacher heterogeneity, value-added and education policy. Economics of Education Review, Volume 40, 76-92.
- Currie, Janet, and W Bentley MacLeod. 2017. "Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians." Journal of Labor Economics, 35(1): 18977.
- Cutler, David, Jonathan S. Skinner, Ariel Dora Stern, and David Wennberg. 2019. "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending." American Economic Journal: Economic Policy, 11(1): 192-221.
- Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey. 2015. "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err". Journal of Experimental Psychology: General, 144: 114–126.
- Elixhauser, A., Steiner, C., Harris, D. R., and Coffey, R. M. 1998. "Comorbidity measures for use with administrative data". Medical Care, 36(1), 8–27.
- Fenizia, Alessandra. 2020. "Managers and Productivity in the Public Sector." Manuscript. Link. Accessed 24 July 2021.
- Fernemark, Hanna, Janna Skagerström, Ida Seing, Carin Ericsson and Per Nilsen. 2020. "Digital consultations in Swedish primary health care: a qualitative study of physicians' job control, demand and support." BMC Family Practice, 21:241.
- Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2021. "Place-Based Drivers of Mortality: Evidence from Migration." American Economic Review, 111 (8): 2697-2735.
- Finkelstein, Amy, Petra Persson, Maria Polyakova and Jesse Shapiro. 2021. "A Taste of Their Own Medicine: Guideline Adherence and Access to Expertise." Unpublished. Link. Accessed 24 July 2021.
- Fitzmaurice, Garrett M., Nan M. Laird, and James H. Ware. 2004. "Applied Longitudinal Analysis." Hoboken, NJ: Wiley, ISBN: 0-471-21487-6. xix + 506 pp.
- Folkhälsomyndigheten. 2013. "Samhällsekonomiska konsekvenser av antibiotikaresistens." Link. Accessed on September 13, 2021.

- Frakes, Michael, Jonathan Gruber, and Anupam Jena. 2021. “Is great information good enough? Evidence from physicians as patients.” Journal of Health Economics, 75, 102406.

- Gabrielsson-Järhult, Felicia, Kristina Areskoug-Josefsson, and Peter Kammerlind. 2019. “Digitala vårdmöten med läkare: Rapport av kvantitativ och kvalitativ studie.” Manuscript.

- Gale, D., and Shapley, L. (1962). “College Admissions and the Stability of Marriage.” The American Mathematical Monthly, 69(1), 9-15.

- Glenngård, Anna H. 2020. “International Health Care System Profiles: Sweden.” In Tikkanen, Roosa, Robin Osborn, Elias Mossialos, Ana Djordjevic, and George Wharton, eds. 2020. *International Health Care System Profiles* Link.

- Goldfarb, Avi and Catherine Tucker. 2019. “Digital Economics.“ Journal of Economic Literature, 57(1): 3–43.

- Göransson M, Persson A-C, Abelsson A. 2020. “Triage in primary healthcare.“ Nordic Journal of Nursing Research: 40(4):213-220.

- Graham, Bryan S. 2011. “Chapter 19 - Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers.” in Benhabib, Jess, Alberto Bisin, Matthew O. Jackson, eds.: *Handbook of Social Economics* North-Holland, Volume 1, 965-1052.

- Graham, Bryan S., Guido W. Imbens and Geert Ridder. 2020. “Identification and Efficiency Bounds for the Average Match Function Under Conditionally Exogenous Matching.” Journal of Business and Economic Statistics, 38:2, 303-316.

- Graham, Bryan S., Geert Ridder, Petra Thiemann and Gema Zamarro. 2021. “Teacher-to-classroom assignment and student achievement.” Manuscript. Retrieved on 2 June, 2021. Link.

- Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch and Gabriel Zucman. 2021. “Tax Evasion at the Top of the Income Distribution: Theory and Evidence.” Mimeo. Link. Accessed 22 June 2021.

- Hill, Andrew, Daniel Jones and Lindsey Woodworth. 2018. “Physician-patient race-match reduces patient mortality.” Unpublished.

- Kasy, Maximilian. 2016. “Partial identification, distributional preferences, and the welfare ranking of policies.” The Review of Economics and Statistics, Vol.98 (1), p.111-131.

- Kasy, Maximilian and Alexander Teytelboym. 2021. “Learning by Matching.” Unpublished. Link. Accessed 21 Sep 2021.

- Kim C, McEwen LN, Gerzoff RB, et al. 2005. “Is physician gender associated with the quality of diabetes care?” Diabetes Care, 28(7): 1594-1598.

- Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, Sendhil Mullainathan. 2018. “Human Decisions and Machine Predictions.” *The Quarterly Journal of Economics*, 133(1): 237–293.

- Koopmans, T.C., Beckmann, M., 1957. “Assignment problems and the location of economic activities”. Econometrica 25 (1), 53–76.

- Lindenlaub, Ilse. 2017 “Sorting Multidimensional Types: Theory and Application.” The Review of Economic Studies, Volume 84, Issue 2, Pages 718–789.

- Lockwood, J. R. and Daniel F. McCaffrey 2009. “Exploring Student-Teacher Interactions in Longitudinal Achievement Data.” Education Finance and Policy, 4(4); 439-467.

- Manski, C., 2004. “Statistical treatment rules for heterogeneous populations”. Econometrica 72 (4), 1221–1246.

- Manski, CF. 2018. “Reasonable patient care under uncertainty. Health Economics.” 27: 1397– 1421. Link.

- Marx, Benjamin, Vincent Pons and Tavneet Suri. 2021. “Diversity and team performance in a Kenyan organization.” *Journal of Public Economics*, Volume 197, 104332.

- Miller, Amalia R. and Catherine E. Tucker. 2011. “Can Health Care Information Technology Save Babies?” *Journal of Political Economy*, 119(2): 289-324.

- Molitor, David. 2018. “The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration”. American Economic Journal: Economic Policy 2018, 10(1): 326–356. Link.

- Muñoz, Mathilde. 2021. Trading Non-Tradables: The Implications of Europe's Job Posting Policy. Unpublished.
- National Board of Health and Welfare / Socialstyrelsen. 2018. "Tillgänglighet i hälso- och sjukvården." 2018-2-16.
- OECD. 2017. "Health at a glance 2017: OECD indicators." Link.
- O'Donnell, Catherine A. 2000. "Variation in GP referral rates: what can we learn from the literature?" Family Practice, 17(6): 462–471.
- Page, Anthea, Sarah Ambrose, John Glover and Diana Hetzel. 2007. "Atlas of Avoidable Hospitalisations in Australia: ambulatory care-sensitive conditions." Adelaide: PHIDU, University of Adelaide.
- Papageorgiou, T. 2014. "Learning Your Comparative Advantages." Review of Economic Studies, 81, 1263–1295
- Reck, Daniel, Max Risch and Gabriel Zucman. 2021. "Response to a Comment by Auten and Splinter on 'Tax Evasion at the Top of the Income Distribution: Theory and Evidence'." Link. Accessed August 8, 2021.
- Rocha, J.V.M., Marques, A.P., Moita, B. and R. Santana. 2020. "Direct and lost productivity costs associated with avoidable hospital admissions." BMC Health Serv Res 20(210). Link.
- Roghman, K. J. and Zastowny, T. R. 1979. "Proximity as a factor in the selection of health care providers: emergency room visits compared to obstetric admissions and abortions." Social Science and Medicine 13, 61–9.
- Scholle, S. H., Roski, J., Dunn, D. L., Adams, J. L., Dugan, D. P., Pawlson, L. G., and Kerr, E. A. 2009. "Availability of data for measuring physician quality performance". The American journal of managed care, 15(1), 67–72.
- Shapley, L.S., Shubik, M. The assignment game I: The core. Int J Game Theory 1, 111–130 (1971). Link.
- Skinner, Lucy, Douglas Staiger, David Auerbach and Peter Buerhaus. 2019."Implications of an Aging Rural Physician Workforce." New England Journal of Medicine, 381(4):299-301.

- Smith, Adam. 1776. An Inquiry into the Nature and Causes of the Wealth of Nations. McMaster University Archive for the History of Economic Thought
- Socialstyrelsen och Sveriges Kommuner och Landsting. 2011. “Öppna jämförelser av hälso- och sjukvårdens kvalitet och effektivitet. Jämförelser mellan landsting.” Accessed 13 September 2021.
- Socialstyrelsen 2017. “Konsekvensutredning – förslag till föreskrift om belopp för vård av utskrivningsklara patienter.” Link. Accessed 13 September 2021.
- Statistics Sweden Statistikdatabasen, 2021. Link.
- Strama. 2019. “Rekommendationer för kvalitetsindikatorer vid digitala vårdmöten.” Link. Accessed 20 August 2021.
- Strama. 2019. “Rekommendationer för kvalitetsindikatorer vid digitala vårdmöten.” Link. Accessed 20 August 2021.
- Studnicki, J. 1975. “The minimization of travel efforts as a delineating influence for urban hospital service area”. International Journal of Health Service 5, 679–93.
- Theis, R. P., Stanford, J. C., Goodman, J. R., Duke, L. L., and Shenkman, E. A. 2017. “Defining 'quality' from the patient's perspective: findings from focus groups with Medicaid beneficiaries and implications for public reporting.” Health expectations : an international journal of public participation in health care and health policy, 20(3), 395–406.
- Tinbergen, Jan. 1962. “An Analysis of World Trade Flows” in Shaping the World Economy, edited by Jan Tinbergen. New York, NY: Twentieth Century Fund.
- Tsugawa, Yusuke, Joseph P Newhouse, Alan M Zaslavsky, Daniel M Blumenthal, Anupam B Jena. 2017a. “Physician age and outcomes in elderly patients in hospital in the US: observational study.” BMJ; 357.
- Tsugawa Y, Jena AB, Figueroa JF, Orav EJ, Blumenthal DM, Jha AK. 2017. “Comparison of Hospital Mortality and Readmission Rates for Medicare Patients Treated by Male vs Female Physicians.” JAMA Intern Med, 177(2):206–213.
- Tucker, Catherine E. and Shuyi Yu. 2019. “Does IT Lead to More Equal or More Unequal Treatment? An Empirical Study of the Effect of Smartphone Use on Social Inequality in Employee-Customer interactions”. Link. Accessed 2 August 2021.

- Vasilik, Ashley. 2021. “Triage Process in Primary Care Clinics”. DNP Scholarly Projects. 51. Link. Accessed 4 October 2021.
- Willis, R. J. and Rosen, S. 1979. “Education and Self-selection.” Journal of Political Economy, 87, S7–36.
- Yeomans, M, Shah, A, Mullainathan, S, Kleinberg, J. 2019. “Making sense of recommendations.” J Behav Dec Making, 32: 403–414.
- Zeltzer, Dan, Liran Einav, Joseph Rashba, and Ran D. Balicer. 2021. “The Impact of Increased Access to Telemedicine.” SIEPR Working Paper No. 21-038. Link. Accessed 20 August 2021.

A Appendix I: Additional Tables and Figures

Table 13: Physical Primary Care Clinic Scores, standardized

	(1)	(2)	(3)	(4)	(5)
Std(Foreign)	-0.24***		-0.25***	-0.24***	-0.24***
	(0.033)		(0.037)	(0.049)	(0.049)
Std(Income)		0.18***	0.14***	0.14***	0.15***
		(0.031)	(0.030)	(0.030)	(0.041)
Avg. age				-0.00	-0.00
				(0.032)	(0.032)
Gender				-1.55	-1.46
				(2.465)	(2.501)
Std(foreign)XStd(Income)					-0.02
					(0.028)
Region FE					
Robust SE					
N	1298	943	943	943	943

Table 14: Robust standard errors in parantheses. The unit of observation is a 4-digit post-code matched with municipality. Region fixed effects are included. Std(Foreign) measures the standardised share of foreign-born inhabitants in the area. Std(Income) measures the standardised mean income in the area. The outcome variable comes from the National Patient Survey (NPE).

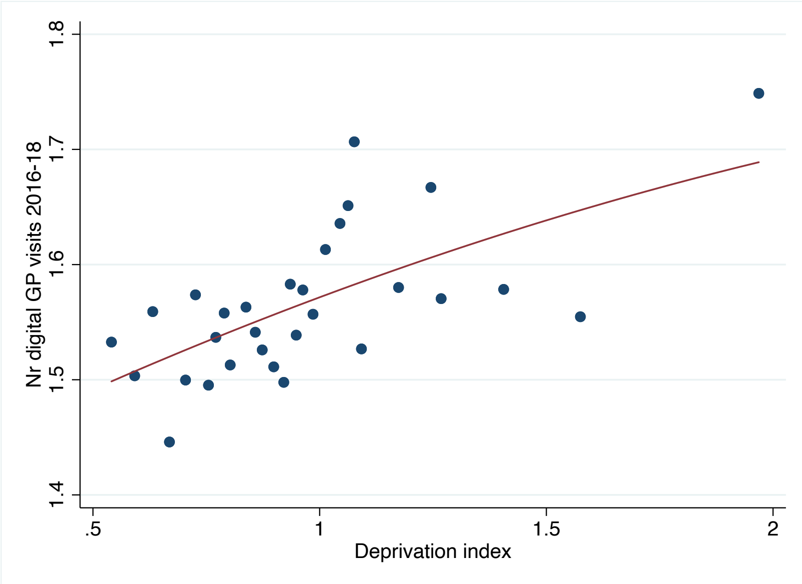


Figure 11: Binned scatterplot of number of digital GP visits in 2016-18 (individual level), controlling for age, against index of social deprivation (Care Need Index) of the individual's GP clinic in Scania.

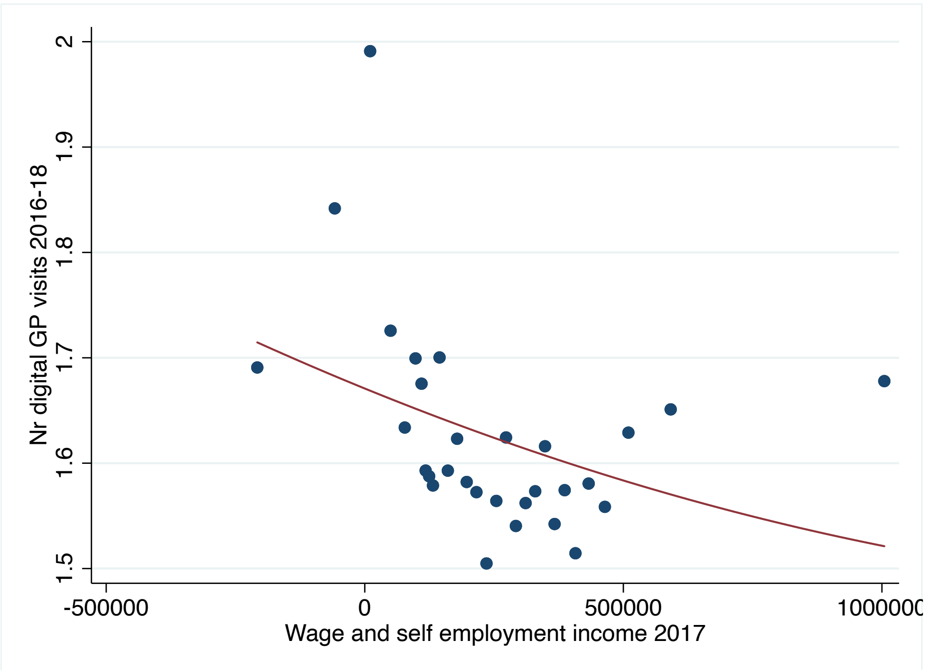


Figure 12: Binned scatterplot of number of digital GP visits (individual level) in 2016-18 against individual total income 2017, controlling for age.

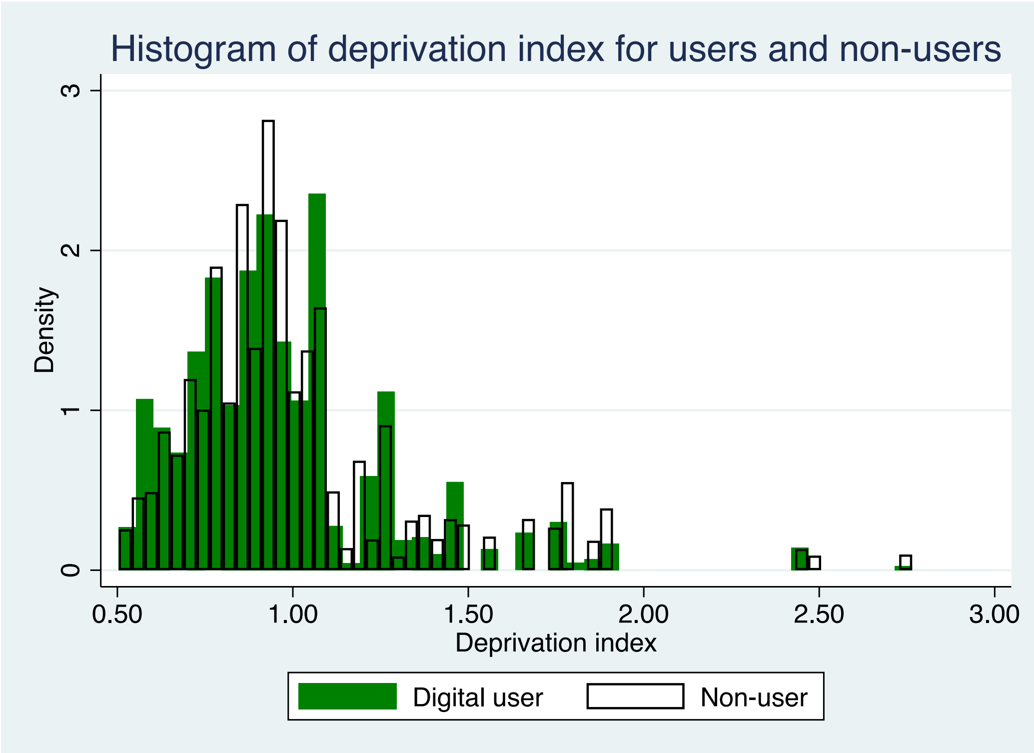


Figure 13: Data from Region Skåne. Deprivation index is the Region's *Care Need Index*. It is a weighted average of the variables (1) over 65 years old and in a single household (2) Born outside EU (3) Unemployed 16-64 year old (4) Single parent with child under 18 years old (5) Person over 1 years old who has moved into the area (6) low education 25-64 years old (7) Age below 5 years old.

Table 15: Descriptive statistics of all doctors

	(1)				
	mean	sd	min	max	count
Specialist	0.36	0.48	0	1	619
In specialty training	0.30	0.46	0	1	619
MD + residency only	0.33	0.47	0	1	619
Speaks non EU15 language	0.21	0.41	0	1	693
GP specialist	0.35	0.48	0	1	619
Age	41.2	10.9	25	80	262
Female	0.41	0.49	0	1	262
Employed rather than contractor	0.43	0.50	0	1	131
Observations	693				

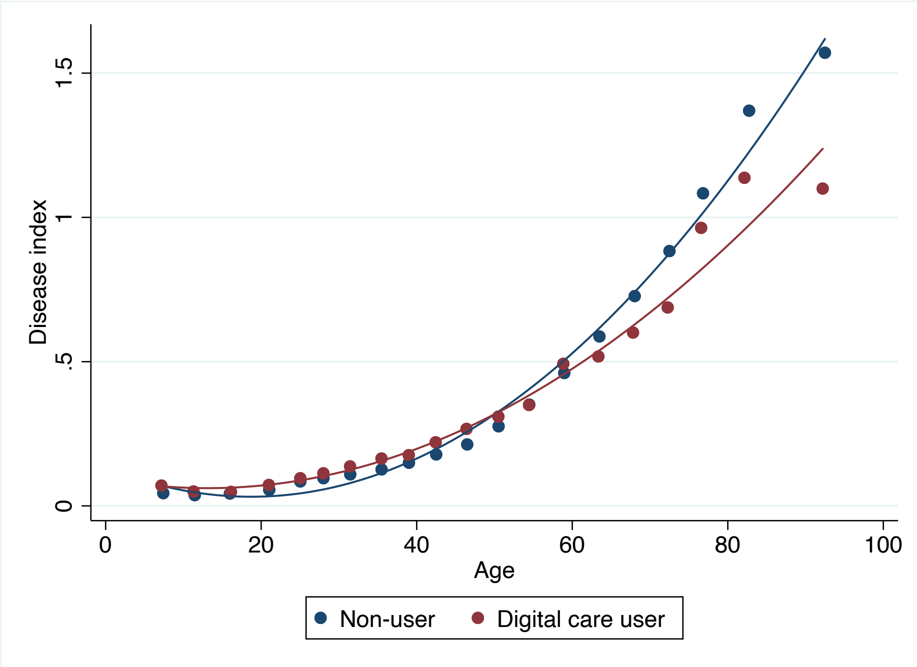


Figure 14: Elixhauser's comorbidity index using data from 2013-15 in Scania.

Variable	(1) Not included	(2) Included	(3) Difference
Total number of first and revisit consultations	323.9	3346.0	3022.2***
	(303.8)	(2584.7)	(138.9)
Seniority	1.1	1.0	-0.1
	(0.8)	(0.8)	(0.1)
Specialty	2.9	2.1	-0.8*
	(4.6)	(3.8)	(0.4)
Speaks non EU15 language	0.2	0.4	0.1***
	(0.4)	(0.5)	(0.0)
Average admin duration	13.4	11.3	-2.1***
	(4.3)	(2.3)	(0.4)
Average consultation duration	6.1	5.0	-1.2***
	(1.9)	(1.2)	(0.2)
Observations	357	143	500

Table 16: Comparison of doctors included in the final analysis and those who are not.

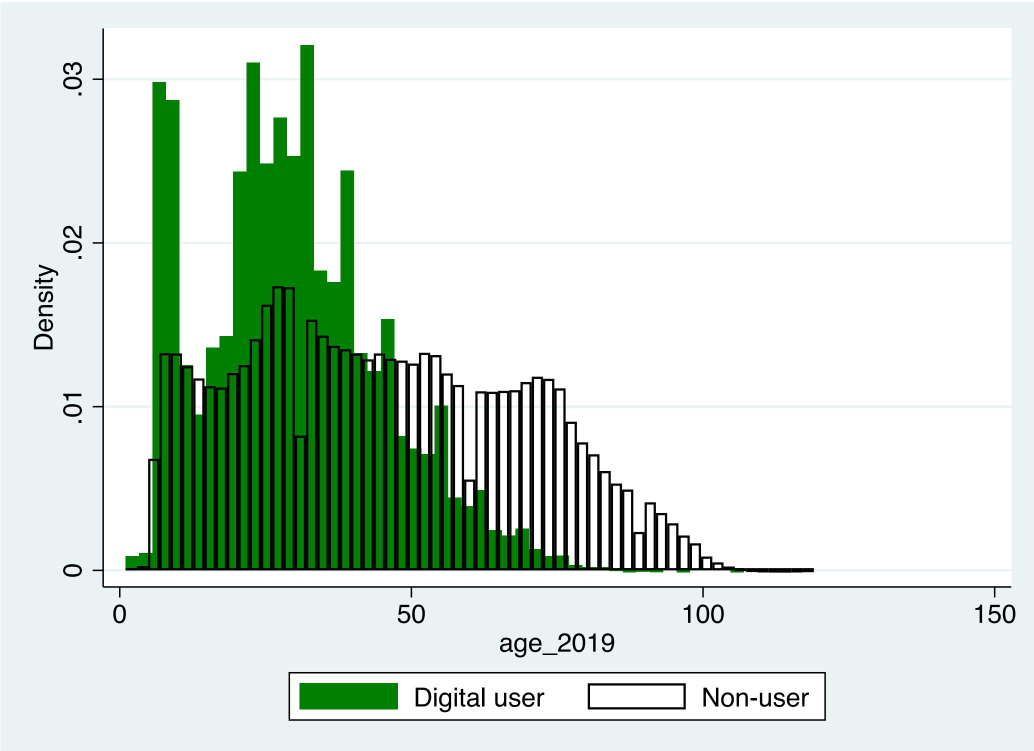


Figure 15: Data from Region Skåne.

Variable	(1) Not included	(2) Included	(3) Difference
Total number of first and revisit consultations	112.0	1188.2	1076.1***
	(489.3)	(1958.5)	(142.0)
Seniority	1.0	1.0	0.1
	(1.0)	(0.8)	(0.1)
Specialty	2.9	2.7	-0.2
	(4.4)	(4.4)	(0.5)
Speaks non EU15 language	0.0	0.3	0.2***
	(0.2)	(0.4)	(0.0)
Average admin duration	20.2	12.8	-7.4***
	(22.1)	(4.0)	(1.1)
Average consultation duration	4.9	5.8	0.9***
	(2.6)	(1.8)	(0.2)
Observations	195	500	780

Table 17: This table shows summary statistics of (Column 2:) the doctors who are (a) not pediatricians (who have a different assignment protocol to patients) (b) who have worked a sufficient number of consultations to merit inclusion in the sample of 500 doctors, compared to (Column 1:) the doctors who are either pediatricians or have worked very few consultations and are thus not included in any sample.

	(1)	(2)	(3)
	No AH	Past AH	Difference
below median income	0.452	0.619	0.167***
	(0.498)	(0.486)	(0.009)
adult without income	0.064	0.201	0.138***
	(0.244)	(0.401)	(0.005)
age	36.491	40.461	3.970***
	(12.456)	(16.043)	(0.227)
patient female	0.630	0.674	0.044***
	(0.483)	(0.469)	(0.009)
any benefit	0.134	0.305	0.172***
	(0.340)	(0.461)	(0.006)
disability insur	0.013	0.066	0.054***
	(0.111)	(0.249)	(0.002)
housing subsidy	0.039	0.065	0.026***
	(0.193)	(0.247)	(0.004)
employed	0.870	0.771	-0.100***
	(0.336)	(0.420)	(0.006)
self-employed	0.073	0.065	-0.008*
	(0.260)	(0.246)	(0.005)
unemployed (20-67)	0.047	0.127	0.080***
	(0.211)	(0.333)	(0.004)
minority	0.168	0.197	0.029***
	(0.374)	(0.398)	(0.007)
foreign born	0.108	0.135	0.026***
	(0.311)	(0.341)	(0.006)
born outside EU15,Scandi.	0.087	0.111	0.024***
	(0.282)	(0.315)	(0.005)
married	0.343	0.328	-0.015*
	(0.475)	(0.469)	(0.009)
inhabitants per km2 munic.	1,649.564	1,373.369	-276.195***
	(2,062.333)	(1,951.354)	(37.415)
Observations	157,475	3,115	160,590

Table 18: This table shows the difference in socioeconomic covariates (measured in 2017) for patients who have had no previous avoidable hospitalization (AH) in 2013-2016, compared with patients who have had at least one such hospitalization in the period before digital care. The socioeconomic variables do not exist for child patients.

	(1)	(2)	(3)
	No AH	Past AH	Difference
hypertension pre 2016	0.008	0.078	0.070***
	(0.089)	(0.269)	(0.002)
asthma pre 2016	0.018	0.056	0.038***
	(0.133)	(0.230)	(0.002)
diabetes pre 2016	0.002	0.083	0.081***
	(0.039)	(0.276)	(0.001)
depression pre 2016	0.025	0.057	0.032***
	(0.157)	(0.232)	(0.003)
anxiety pre 2016	0.040	0.095	0.056***
	(0.195)	(0.294)	(0.004)
hyperactivity pre 2016	0.019	0.041	0.022***
	(0.138)	(0.199)	(0.003)
had any visit pre 2016	0.733	0.937	0.205***
	(0.443)	(0.243)	(0.008)
nr acute visits pre 2016	0.069	0.263	0.193***
	(0.349)	(0.901)	(0.007)
never filled presc. 2013-16	0.096	0.016	-0.080***
	(0.295)	(0.127)	(0.005)
nr presc. filled201316	21.314	88.221	66.907***
	(54.234)	(189.340)	(1.083)
above median presc. 2013-16	0.499	0.834	0.336***
	(0.500)	(0.372)	(0.009)
Observations	157,475	3,115	160,590

Table 19: This table shows the difference in pre-digital care diagnosis and healthcare utilization covariates for patients who have had no previous avoidable hospitalization (AH) in 2013-2016, compared with patients who have had at least one such hospitalization in the period before digital care. The socioeconomic variables do not exist for child patients.

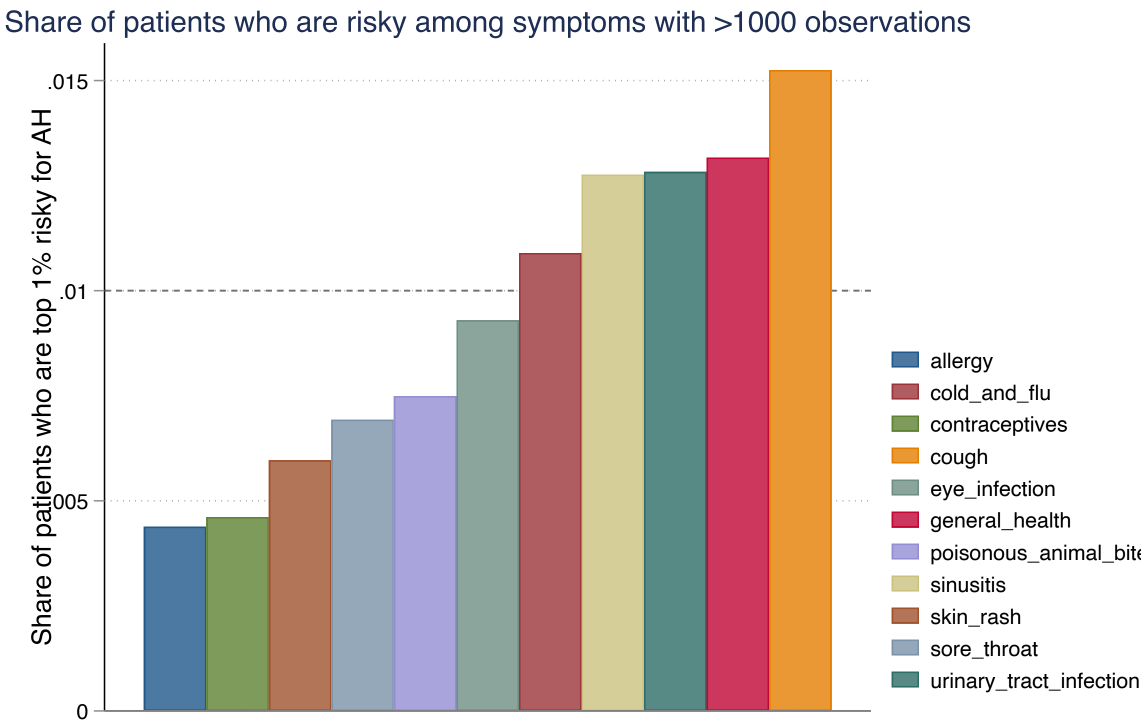


Figure 16: Patients are asked to fill in their main symptom/reason for seeking care just before the consultation. Exactly 1% of patients are risky for AH in the overall sample, so symptoms with over 1% risky patients are over- represented for risky patients and vice versa.

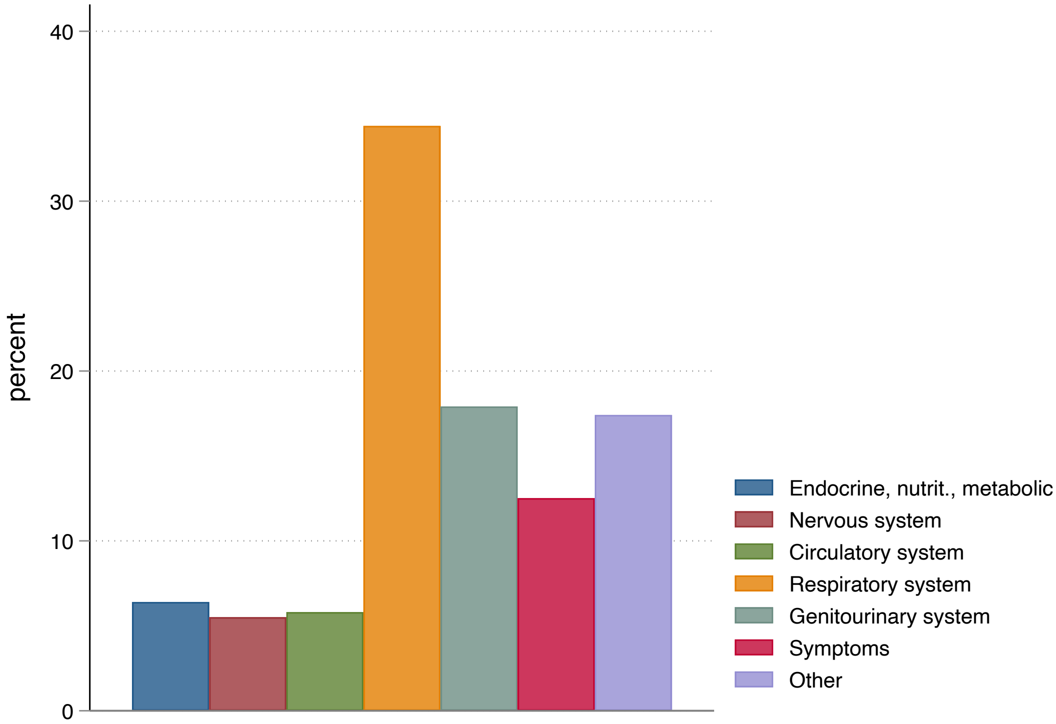


Figure 17: Grouping of the primary diagnosis code among the avoidable hospitalizations within 3 months after the digital visit.

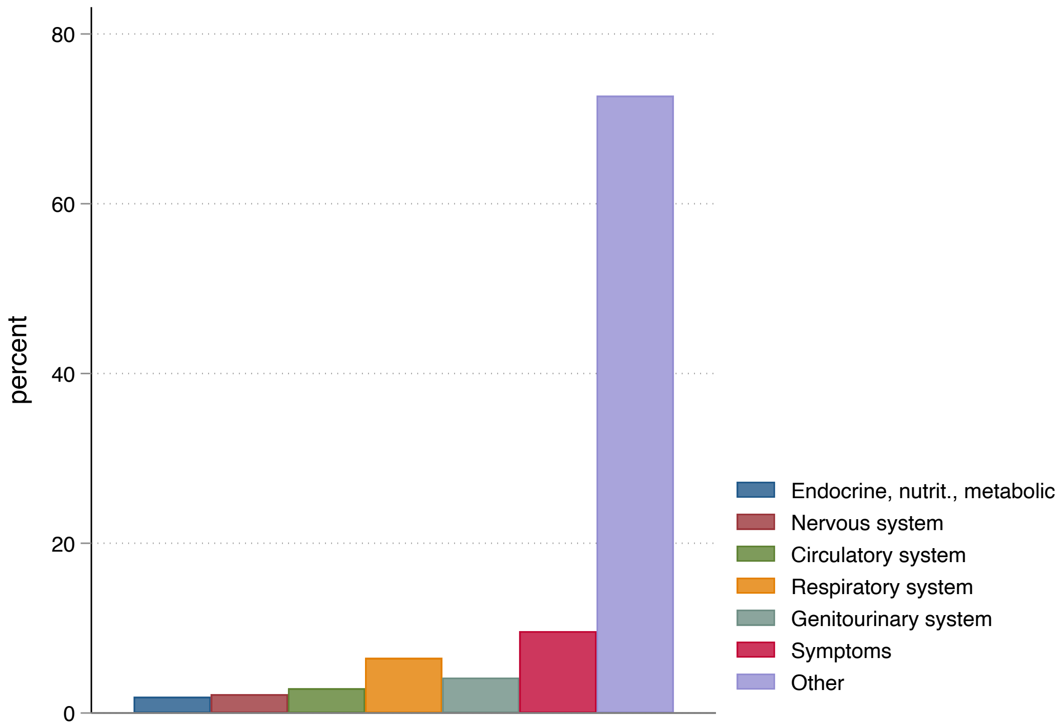


Figure 18: Grouping of the primary diagnosis code among all hospitalizations with the same groups as above.

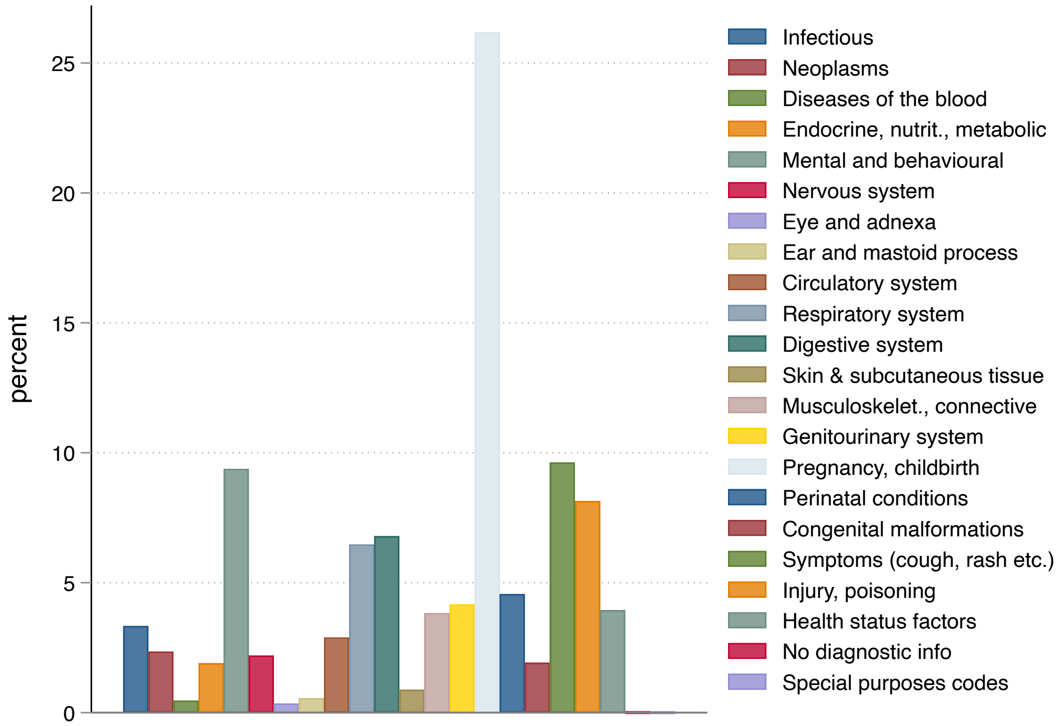


Figure 19: Grouping of the primary diagnosis code among all hospitalizations with all diagnosis groups.

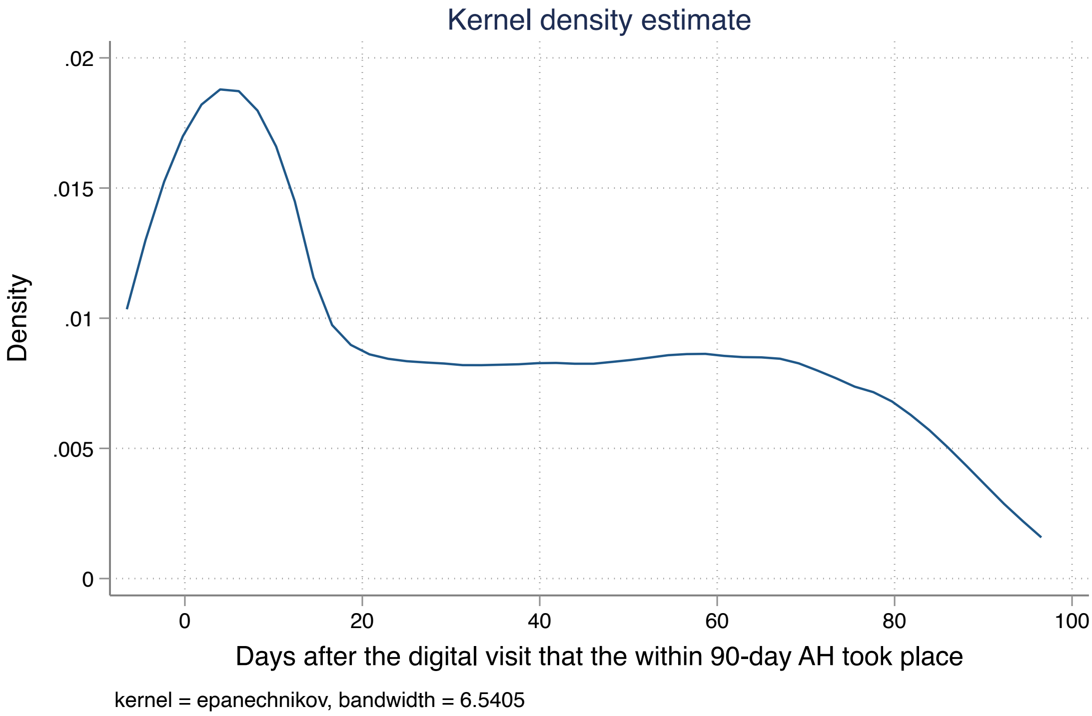


Figure 20: Distribution of days after digital visit that the AH within 90 days happened

Table 20: Neg. nr. avoidable hosp. in 3 months after consultation, re

	(1)
negative_ah	
Nr AH 3 years before	-0.0277
	(0.0196)
Disease index	-0.0076***
	(0.0020)
Female	-0.0002
	(0.0004)
Age	-0.0000*
	(0.0000)
2nd gen immigrant	-0.0005
	(0.0008)
1st gen immigrant	-0.0020**
	(0.0010)
_cons	0.0007
	(0.0005)

With date time shift FE, doctor RE.
SEs in parantheses clustered on doctors.
Sample born before 2013, doctors' visits before 600th.
Sample before Oct 2018 to allow 3 month follow up.
Disease index is sum of Elixhauser comorbidities.

Table 21: Nr AH 3 months after first digital visit

	(1) OLS Simple	(2) Controls	(3) Bootstrap	(4) Time shift FE	(5) ZI Poisson
Std doctor FE	0.0002	0.0001	0.0001	0.0002	-0.0275
	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0644)
Nr AH 3 years before	0.0586***	0.0551***	0.0551***	0.0552***	-0.0157
	(0.0103)	(0.0101)	(0.0099)	(0.0102)	(0.0440)
Std doctor FEX AH 3yrs before	-0.0188**	-0.0190**	-0.0190**	-0.0193**	-0.0602**
	(0.0088)	(0.0087)	(0.0091)	(0.0088)	(0.0247)
Inflation for the ZIP:					
Nr AH 3 years before					-1.6412***
					(0.4146)
Controls	No	Yes	Yes	Yes	Yes
N	122662	122564	122564	122564	122564
R^2	0.051	0.056	0.056	0.056	
Mean	0.0023	0.0023	0.0023	0.0023	0.0023
Mean_risky	0.0589	0.0589	0.0589	0.0589	0.0589

All columns: SEs in parantheses clustered on doctors. Col 3 has booststrapped SEs.

Sample is dropin first visits after doctor's 600th such visit. Only patients born before 2013.

Controls are Elixhauser sum of comorbidities, female, age, first- and second-generation immigrant.

The 6th column shows results from a Zero-Inflated Poisson model.

Table 22: Definitive counter guideline prescription

	(1) OLS Simple	(2) Controls	(3) Bootstrap	(4) Time shift FE
Std doctor FE based on no CGP	-0.0035***	-0.0035***	-0.0035***	-0.0035***
	(0.0008)	(0.0008)	(0.0008)	(0.0006)
Nr antib filled 3yrs before	0.0021***	0.0022***	0.0022***	0.0021***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Std doc FE X nr antib 3yrs b4.	-0.0005**	-0.0005**	-0.0005**	-0.0005**
	(0.0002)	(0.0002)	(0.0003)	(0.0002)
Controls	No	Yes	Yes	Yes
N	116396	116391	116391	116391
R^2	0.006	0.006	0.006	0.006
Mean	0.0172	0.0172	0.0172	0.0172

All columns: SEs in parantheses clustered on doctors. Col 3 has booststrapped SEs.
Sample is dropin first visits after doctor's 600th such visit. Only patients born before 2013.
Controls are Elixhauser sum of comorbidities, female, age, first- and second-generation immigrant.

A.1 Additional Results

Additional results: Good doctors at all three measures are less senior and have worked less in the service.

Table 23: Explaining quality with doctor characteristics

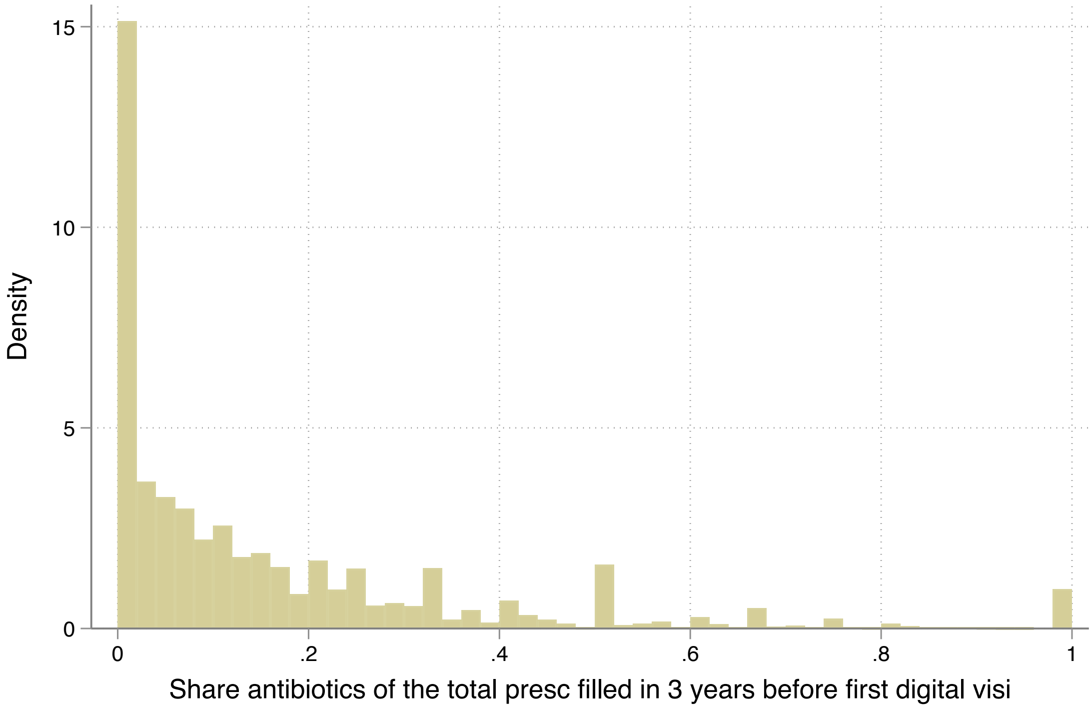
	(1) Std CGP skill	(2) Std AH skill	(3) Std double visit skill	(4) Over median at all
100s randomized consultations	0.00	-0.00	-0.02***	-0.00*
	(0.00)	(0.00)	(0.01)	(0.00)
In specialty training	-0.02	-0.39**	-0.14	-0.14**
	(0.22)	(0.19)	(0.20)	(0.07)
Specialist	-0.03	-0.38	-0.11	-0.15**
	(0.22)	(0.25)	(0.19)	(0.07)
Non EU15 language	-0.35*	-0.20	0.25*	0.01
	(0.20)	(0.19)	(0.15)	(0.05)
_cons	0.02	0.36*	0.57**	0.24***
	(0.19)	(0.19)	(0.22)	(0.07)
N	143	143	143	143
R^2	0.03	0.04	0.20	0.06

This corroborates studies e.g. Newhouse et al. (2017) showing that younger hospital doctors have lower mortality and costs than older doctors. Older doctors have more experience, but are less up to date with recent medical knowledge.

Additional result: Female doctors are 0.7sd better at following guidelines.

Table 24: Gender and doctor characteristics

	(1) Std CGP skill	(2) Std AH skill	(3) Std double visit skill	(4) Over median at all 3
Female doctor	0.71***	0.11	0.23	0.09
	(0.24)	(0.27)	(0.31)	(0.07)
_cons	-0.30	0.03	-0.33	0.03
	(0.19)	(0.19)	(0.25)	(0.03)
N	61	61	61	61
R^2	0.12	0.00	0.01	0.03



This corroborates studies e.g. Kim et al. 2005; Berthold et al. 2008; Baumhäkel et al. 2009, which show that female doctors adhere more to other guidelines. Note that I have data on gender on only 43% of doctors.

A.1.1 Match effects

A person who has had 1 more AH in the past 3 years has 6 percentage points higher risk of getting one again within 3 months, but only 4 percentage points higher risk if they meet a 1 standard deviation better doctor (a reduction of 29%).

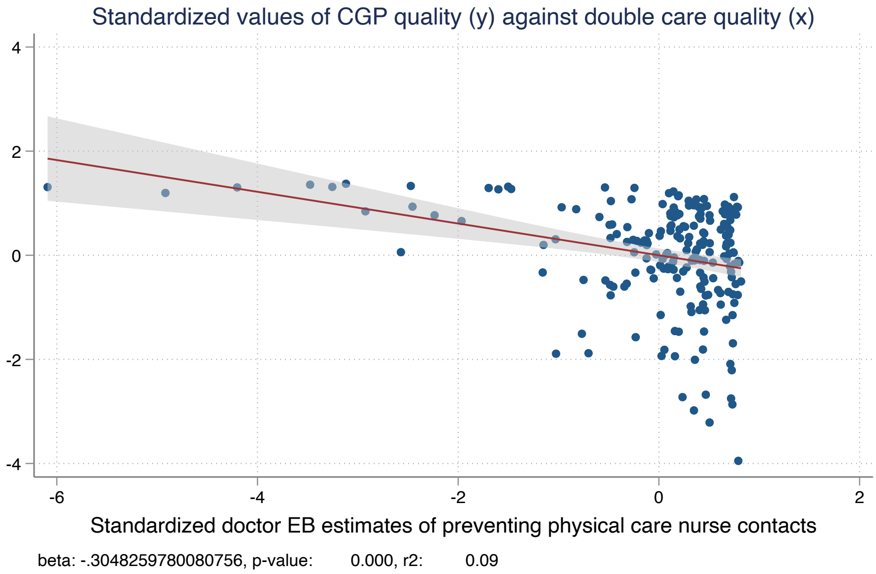


Figure 21: Correlation between two different quality measures within doctors.

Table 25: Number Avoidable Hospitalizations within 3 mo. after visit

	(1) Clustered SEs	(2) Bootstrapped SEs
Top 10% doctor X top 1% risky patient	-0.060***	-0.060***
	(0.014)	(0.016)
Top 10% doctor on AH	0.000	0.000
	(0.001)	(0.001)
Riskiest 1% patient in AH	0.067***	0.067***
	(0.012)	(0.014)
N	95816	95816
Mean	0.003	0.003
Mean_risky	0.062	0.062

All columns have date-time shift fixed effects. Sample is all doctors' randomized visits after the 600th' consultation

A.2 ARE for Counter-Guideline Antibiotics Prescriptions

A.2.1 Reallocation on patients' previous share of antibiotics and doctor CGP effect

	Status quo	PAM	PAM-SQ
CGP	3417	3059	-358
SE		(2.722)	
N	199,867	199,867	199,867

Table 26: Average Reallocation Effects for Counter-Guideline Antibiotics Prescriptions (CGP)), with Positive Assortative Matching (PAM) on patient previous antibiotic share out of total prescriptions and doctor CGP quality. Standard errors are bootstrapped in the regression computing the average match function, which is calculated over the appointments after the 600th, then using 'predict' for a counterfactual allocation and then aggregating all individual effects using 'total'. The measure of doctor CGP quality here is the standardized shrunk Empirical Bayes estimate calculated over doctors' first 600 appointments.

A.2.2 Reallocation on patient income and doctor CGP effect

When patients with lower income are matched with doctors with higher quality on CGP (i.e. who have been more likely to follow guidelines in the auxiliary sample), the result is more counter-guideline prescriptions, but only by a very small amount, 46, corresponding to a 1.6% increase. However, we also see an increase in CGP when patients with higher income are matched with doctors of higher quality, but by an even smaller amount.

	Status quo	PAM	PAM-SQ	NAM	NAM-SQ
CGP	2792	2807	15	2838	46
SE		(4.776)		(5.329)	
N	163,443	163,443	163,443	163,443	163,443

Table 27: Average Reallocation Effects for Counter-Guideline Antibiotics Prescriptions (CGP)), with Negative Assortative Matching (PAM) on patient income and doctor CGP quality.

B Appendix II: Data Appendix

B.1 Datasets

All datasets are proprietary and confidential, and were accessed after applications to the Stockholm Regional Ethics Council (2018, number 2108/2318-31 and Swedish Ethics Authority (2019, number 2019-06062) had been approved. Additionally, Statistics Sweden and the other entities carried out their own confidentiality assessments before approving the sharing of data. Statistics Sweden anonymized the personal identifiers and matched with other datasets, and then shared only an anonymized version of the data with the researcher.

B.1.1 Definition of analysis sample

I start from the universe of patients who has had at least one digital consultation with one of the largest[73] providers of digital healthcare in Sweden, from the start of the service in mid-2016 to the end of 2018. There are 631,681 consultations in the dataset. I keep only the first visit for each patient, as these consultations are conditionally randomized, but there is a bigger concern of endogeneity in any following visits. Hence, each patient has only one observation in digital care.

There are 378,627 unique patients, who have on average had 1.67 consultations. We match this data to official registry data from Statistics Sweden on socioeconomic and demographic variables[74] and data from the National Board of Health and Welfare (NBHW / *Socialstyrelsen*) on diagnoses of chronic conditions from specialist, acute and inpatient care across the Swedish healthcare system in the three years preceding digital primary care, 2013-2015. In this physical healthcare dataset, there are many observations per patient. In addition, we match with data on physical primary care (2013-2019) from one Swedish region (Skåne), which matches for around 10% of the digital care sample.[75]

[73]In terms of patient volumes in 2016-2020.

[74]In total 847 people (0.22% of the initial sample) could not be matched to the Statistics Sweden or NBHW records. Of these, there are 262 individuals with an incorrect personal identification number (PIN) according to Statistics Sweden. In addition, there are 112 people with a re-used PIN, which are dropped. An additional 473 people could not be matched for other reasons.

[75]Swedish physical primary care is devolved to 21 regions, which means that data from primary care is not included in the National Board of Health and Welfare data. Receiving primary care data in Sweden across all regions has eluded researchers, as policies, codings and applications vary across the country.

The sample now consists of all individuals (377,780) who have had a digital video consultation with a medical doctor from the start of the service in 2016 until the end of 2018, and can be matched with registry data. For most of the analysis, we restrict the sample to “drop in“ visits, that is visits where the patient had no way of specifying which doctor they want to meet, but rather meet the first available doctor. This is 82% of the first visit sample (310,000 patients), and this is the sample where conditional (on time) randomization holds. Moreover, we remove pediatricians and those children who are more likely to see a pediatrician (where randomization did not hold), which leaves 302,883 patients and 511 doctors.

For consistent definition of patient types according to their pre-digital physical healthcare utilization, we drop patients for whom we do not observe the full pre-period 2013-2016, i.e. patients who were born in or after 2013. This leaves 233,489 patients and 499 doctors. Finally, we keep only doctors who have done >600 consultations and their patients, which leaves 210,171 patients (56% of original N) and 143 doctors (20% of original D). The reason is that many doctors were hired late in the sample period, since the service was expanding. These doctors have only done a few randomized consultations, many of them under 100. This is not a sufficient sample to base the analysis on. For the outcome avoidable hospitalization, we need a post-digital consultation period of 3 months, which means we drop all consultations which took place in October-December 2018, as our follow up data in physical healthcare ends on 31 December 2018, just as the digital care dataset does.

B.1.2 Statistics Sweden dataset

We can measure most socioeconomic characteristics for adults only, since the variables on e.g. income and education do not exist for minors. The socioeconomic variables from Statistics Sweden reported here are all measured at the same time for all individuals, irrespective of the year when they started using the digital service (income and employment variables in 2017, education in 2018).

B.1.3 Digital care dataset

This dataset was created from the internal logs of the digital healthcare company.

Sample definition:

- Keeping only the first visits: from ∼630,000 visits (1.67 visits/ person) to ∼378,000;

715 doctors.

- Randomly assigned (“drop in“): 310,000 patients (82%), 526 doctors.
- Removing pediatricians and those children who are more likely to see a pediatrician: 302,883 patients, 511 doctors.
- Dropping patients born before 2013 (when pre-data starts): 233,489 patients (62% of original N), 499 doctors (70% of original D).
- Keeping only doctors who have done >600 consultations **210,171 patients** (56% of original N), **143 doctors** (20% of original D).

C Appendix III: Additional Information

C.1 Literature review

Reallocation and matching
Methodologically, this paper relates to the nascent empirical literature on reallocation and matching as mechanisms to improve outcomes instead of input augmentation (Aucejo et al. 2021, Bergeron et al. 2021, Fenizia 2020, Graham et al. 2021). The present paper is the first to apply this to digital services, where the costs to reallocation are smaller as the constraint of shared location is relaxed and only access to data and a matching algorithm would be needed to achieve reallocation. Moreover, gains are large due to the size of these service markets. It is also the first paper to apply the methods to the healthcare setting, where cost and effectiveness challenges are pressing. Assignment and matching problems have been an interest in economics for a long time; for early sources see Koopmans and Beckmann (1957), Shapley and Shubik (1971) and Gale and Shapley (1962). These authors studied the theoretical problem of finding an optimal match or assigning each match a utility (and transfer). Graham, Imbens and Ridder (2007) take on reallocation problems statistically and show that objects of interest such as average reallocation effects can be identified given the joint distribution of match outcomes and partner characteristics.[76]

Bonhomme et al. 2019 compute the effects of reallocation in another direction: from equilibrium sorting between employers and employees to a random allocation, and conclude that workers and firms are not additively separable as they find a mean impact from the

[76]For an overview of this literature, see Graham (2011).

reallocation.

Fenizia (2020) uses a two-way fixed effects model to calculate how much managers matter in the public sector in Italy, and finds that they account for 10% of productivity when workers have mechanical tasks with no multitasking. Moreover, she conducts a reassignment exercise where she finds that reallocation of the best managers to the most productive offices would increase total output by around 7% but also increase inequality between offices. In contrast, in the primary care setting, it is possible that we have no trade-off between efficiency/total output and equality, such that increasing output could also *decrease* inequality.

One of the most related papers to ours is Bergeron et al. (2021). This paper considers tax collectors in the Democratic Republic of Congo, but with an approach similar to the present in the sense of first calculating fixed effects of tax officers, and later using these to estimate the effects of reassignment of tax collectors to households. Preliminary results say that the effects of reallocation are large, and also depend on peer effects between teams of tax collectors.

Graham et al. (2021) calculate the optimal counterfactual assignment between teachers and classrooms, for the outcome variable student test scores. They find a meaningful effect from the potential reallocation of teachers within school district, of 0.02sd compared to the random allocation. Yet, test scores may not be the variable the policymaker cares about in the long run, rather it may be students' actual learning and life outcomes. But compared to the primary care setting, the teaching setting at least has the variable of grades which are used in many settings as a proxy for what we really care about. In the primary care setting, the patients are diverse and seek for many different conditions, and the role of the doctor can be both to assign treatment, not to assign too many antibiotics, to detect really serious diseases and refer them to avoid negative health events, and to educate and calm patients.[77]

Matching and mechanism design
This study is complementary to the literature on mechanism design in matching markets (for an overview, see Roth 2012), where strategic incentives of agents are taken into account when studying matching problems. In this paper, I do not study strategic incentives

[77] We could use patient grading of doctor as a proxy for the outcomes similar to test scores in the teaching literature, but this would be more comparable to a teaching evaluation completed by the students. This has the issue of patients being less medically informed than the patient, and might not know what good quality care is, due to this asymmetric information.

of patients and doctors over whom they match with. There are two main reasons for this. First, in some settings (such as the new digital assignments in several markets), agents have little control over who they match with. Second, as Graham (2011) points out, the study of the effects of alternative assignments is the first step in a more complete policy formulation - before deciding if mechanism design of a decentralized system to implement a desired outcome is relevant, we need to know if there are large benefits to alternative allocations.

Teacher value added and other fixed effects strategies
I attempt to conceptually speak to the issue of using agent fixed effects in the presence of multiple tasks and different subjects, which has been an issue both in the tax auditor literature (e.g. Reck et al 2021) and in the teacher literature (e.g. Condie et al 2014). In most of the teacher value added literature, each teacher gets only one fixed effect, which is a weighted average of their effectiveness in different subjects and across heterogeneous students. This paper shows that, in a field which in some ways is similar to education, there are important heterogeneities across both dimensions: (1) healthcare tasks (which in teaching would correspond to subjects) and (2) patients (students). This is important not least in light of evidence from papers such as Aucejo et al. (2021) who show substantial complementarities between teachers and classroom composition (not individual students).

Homophily in doctors' and patients' attributes
A recent literature on complementarities between patient and physician characteristics focuses on homophily in a single trait of patients and physicians, either gender (Cabral and Dillender, 2021) or race (Alsan et al., 2019; Hill et al., 2020).[78] I contribute to this literature by measuring complementarities between a different set of doctor and patient attributes such as doctor skill and patient risk. This exercise suggests to different channels for complementarities than those suggested by the homophily literature (communication and trust). My results imply that reallocations that match doctors specialized in preventing adverse outcomes to lower-income patients (since high-risk patients turn out to be lower income) could improve outcomes, through the mechanism of doctor receptiveness to risk factors. This could be an alternative policy to matching patients and doctors on the characteristics suggested in papers examining the effects of doctor-patient homophily on gender or race.

[78] There are also studies on diversity vs. homophily of workers and clients outside the medical setting. E.g., Marx, Pons and Suri (2021) find that this type of "external" diversity has no effect on NGO canvassers' performance in Kenya.

Physician practice style and skill

Several papers in economics suggest that physician practice style and skill heterogeneity could be one potential explanation for the wide variation in healthcare spending and productivity between units (hospitals, regions etc.). To identify physician heterogeneity separately from patient heterogeneity requires at least conditionally random allocation between doctors and patients. Since plausible random allocation is more frequent in emergency departments[79] or with specialists[80] than in primary care, few papers focus on primary care physicians where patient-doctor sorting is endemic. Moreover, datasets comprising both primary care physicians and their patients followed over time are rare. There are a few recent exceptions: Kwok (2019) and Fadlon and Van Parys (2020), both of which study patients who are over 65 years old and use Medicare, the former using a two-way fixed effects design and the latter using physician exits for identification. I contribute by measuring complementarities between patients and physicians, with a diverse set of patients of all ages.[81]

Digital services

The paper also relates to a nascent literature on the potential effects of digitalization on outcomes and inequality in public services. Zeltzer et al. (2021) study telemedicine in Israel during a lockdown, and find that increased access results in a 3.5% increase in primary care visits, but a 5% lower per-episode cost, implying lower overall resource utilization. Tucker and Yu (2019) find that IT technology can mitigate the effect that customers with better education get better treatment because they are better able to advocate for themselves. The suggested mechanism is that mobile application standardizes communication. Gresenz et al. (2016) find that increased primary care IT adoption can lower avoidable hospitalizations.[82] A review on digital economics (Goldfarb and Tucker, 2019) suggests that digital technology can lower search, replication, transportation, tracking and verification costs. This paper adds another dimension which digital technology can enable on top of these cost-cutting aspects: effectiveness improvements due to data-driven centralized matching, enabled both by the data and the algorithms that can assign people irrespective of physical location.

Theoretical matching

[79]See e.g. Abaluck et al. (2016) and Chandra and Staiger (2011). Both papers use emergency department settings, but the latter argues that controlling for patient characteristics is sufficient to account for patient heterogeneity.

[80]See e.g. Currie and MacLeod (2017) who study obstetricians and Molitor (2018) who study cardiologists.

[81]Older studies on primary care physicians are e.g. Grytten and Sorensen (2003) in a Scandinavian context.

[82]The term used in the paper is Ambulatory Care Sensitive Condition hospitalizations, and this is equivalent to Avoidable Hospitalizations.

Theoretically, this paper's focus on doctor-patient matching relates to theories of stable matching - most closely to the area between one- and two-sided matching, where a set of agents (patients) should be matched to a set of non-strategic units (doctors). Doctors can be seen as having priorities over patients, such as in the school choice literature (for an overview, see e.g. Abdulkadiroglu and Sonmez 2003), but do not act them out strategically. Priorities can reflect fairness or public policy considerations, for instance severity of disease, risk factors such as age, or waiting time. Patients, on the other hand, would have strategic preferences over doctors if the institutional features allowed for doctor choice. Gale and Shapley (1962)'s deferred-acceptance (DA) and Gale's top trading cycles (TTC) (Shapley and Scarf 1974) are common strategy-proof mechanisms for priority-based matching, of which DA eliminates justified envy and TTC is Pareto efficient - for a recent comparison see Abdulkadiroglu et al. 2020.

Physician practice style in primary care
There are only very few papers studying physician performance or practice style in primary care. Kwok (2019) studies primary care spending/utilization outcomes, using a two-way fixed effects model to study patients who switch physicians, and find that spending styles of primary care physicians explain 2-3% of total long-run healthcare spending and 13% of primary care long-run spending.

Fadlon and Van Parys (2020) use physician exits in Medicare as a source of identification. They find large and long-lasting effects of primary care physicians on patient utilization, and similar to this paper, they find it necessary to categorize practice style across several dimensions, including avoidable hospitalizations and guideline-consistent care, but also utilisation quantity. Fadlon and Van Parys' study (2020) is restricted to patients over 65 years old, whose physician moves or retires. The present study is complementary to Fadlon and Van Parys' study, as we measure complementarities between doctors and patients unlike them, while they show that there are long-lasting effects (at least up to 6 years) of physicians, while we do not have as long follow up time. Moreover, a key difference between this study and Fadlon and Van Parys (2020) is that their measures of physician differences include both (what they call) "extrinsic" and "intrinsic" factors - physicians may have different colleagues and settings in their case, but in this paper all work in the same company but still meet patients from all over the country. Therefore it is more likely that we capture only "intrinsic" variation between physicians.[83]

[83]Of course, even intrinsic factors could have arisen because of some earlier extrinsic factor, such as

Patient-physician homophily

In an experimental setting, Alsan et al. (2019) showed that patient-doctor race match can lead to more preventive health investment, and Hill et al. 2018 showed that in an emergency hospital setting, race match can lead to lower mortality.[84] Cabral and Dillender (2021) find that gender match with doctors matters for female patients' disability evaluations but not for male patients' evaluations. It is still unclear exactly through what mechanism homophily improves healthcare outcomes, though improved communication is suggested. My paper takes a different angle by evaluating the task-specific skill of doctors, and relating this to characteristics of doctors, as well as evaluating the effects of matching on several characteristics of patients.

schooling and different colleagues/peer effects at earlier workplaces. But then that goes back all the way to nature and nurture in childhood, and we would need to draw the line somewhere.

[84]There are also studies on diversity vs. homophily of workers and clients outside the medical setting. E.g., Marx, Pons and Suri (2021) find that this type of "external" diversity has no effect on NGO canvassers' performance in Kenya.