Defying Distance?
The Provision of Services in the Digital Age

Amanda Dahlstrand*

February 3, 2023

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.

Keywords: Skills, Labor Markets, Health and Inequality.
JEL codes: J24, J40, I14

*LSE and Microsoft Research: amanda.dahlstrand@gmail.com. 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, Anders Fridén, Rachael Meager, Virginia Minni, Niclas Moneke, Petra Persson, Helena Svaleryd, Martina Zanella, Céline Zipfel and seminar participants at Berkeley Haas, Columbia Business School, Georgetown Economics, Harvard Business School, Institute for Fiscal Studies, IIES, LSE Economics, Microsoft Research, MIT Conference on Digital Experimentation, Northwestern Economics, OECD, Princeton Economics, Stanford GSB, Stockholm University Economics, Yale SOM and the University of Zurich Economics for helpful discussions and feedback. I gratefully acknowledge the collaboration with the digital care company and Nasim Bergman Farrokhnia, and grants from STICERD and CEP at the London School of Economics.

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 necessarily being drawn from the smaller pool of local in-person primary care providers. This setting is ideal to study the upper bounds of impacts from technology-enabled matching, as primary care is the front line of healthcare with the largest patient pool and the most heterogeneous patients and tasks. Given this heterogeneity, physician specialization and division of labor have the potential to increase output (Smith, 1776). I measure causal effects of doctors in different outcomes, and show that there is specialization even among generalist primary care providers. Hence, improvements from patient-doctor matching drawing on physician specialization could be feasible, and this is a low-cost policy when geographical distance is not a constraint.

In order to overcome the endogenous selection between in-person primary care providers and patients, which normally confounds causal effects of doctors on patients, I assemble a novel dataset of consultations, patients and doctors in digital primary care, available across an entire country – Sweden – in 2016-2018. The analysis data covers approximately 200,000 patients and 150 doctors and comes from Europe's largest digital primary care provider. The key feature of the digital care analysis sample is that the allocation of doctors to patients is random, conditional on time and date. This is a by-product of the first-come-first-served assignment procedure of patients to doctors, and neither party has the ability to intervene in this

[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.

digital process. This is in contrast to the doctor assignment process within in-person primary care, which is tightly constrained by geography.[2] To enable the analysis of healthcare outcomes in the in-person healthcare system, and to include patients' prior healthcare histories in in-person care, this dataset is merged on the individual patient level with in-person healthcare data from the universal healthcare system. In these comprehensive healthcare data, patients are followed over six years, which allows me to measure patient risks in terms of past diagnoses and healthcare utilization history. Finally, the data are matched on the individual patient level with detailed socioeconomic and demographic variables from *Statistics Sweden*, to enable 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 in-person care. I provide descriptive evidence of income-quality correlations in in-person 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 2022) and England, where contemporaneous work also shows that cardiologist mortality prevention skill for heart attack patients is lower in rural and more disadvantaged areas (Stoye 2022).[3]

I also study the *redistributional* effects of doctor skill-patient need matching policies along the patient income distribution. I provide evidence that doctor-patient matching with the aim to improve aggregate healthcare outcomes can also allow us to address healthcare inequality as a by-product, by severing the link between the quality of local area service provision and patient income. Given that I find that primary care doctors are specialized in different tasks, and high-risk patients for certain adverse outcomes have disproportionately low income, the assignment that minimizes

[2]99% of Swedish inhabitants live within 20 minutes from their closest primary care clinic (Tillväxtverket, 2011), and a majority are registered with the closest clinic.

[3]Heckman and Landerso (2021) illustrate the sorting of educated families into areas with better public school teachers in Denmark, a welfare state like Sweden where teachers are paid equal amounts across areas. According to Heckman and Landerso (2021), this results in similar intergenerational mobility as in the U.S. as more advantaged families are better able to access universally available programs. In the present paper, I quantify potential changes in total outcomes and in inequality if geographical sorting could be removed in primary care.

these adverse outcomes assigns those doctors more to low-income patients. This does not mean that higher-income patients get bad doctors – they get doctors with a different specialization.[4]

Estimating doctor ability in primary care has been challenging, as important patient outcomes are often ambiguous, rare, or delayed.[5] 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 patients who have dangerous conditions and preventing imminent adverse outcomes (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.[6] In the case of preventing imminent adverse outcomes, the negative outcome is an *avoidable hospitalization*, i.e. a hospital admission that could have been avoided with sufficient primary care. For the third doctor task, 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 risk 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 re-assignment algorithm does not use additional data.

I implement a novel empirical method that allows for both the measurement of

[4]Other risk factors, for instance the risk of having a counter-guideline antibiotics prescription are not negatively correlated with income.

[5]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.

[6]This is a slightly different type of guideline than those evaluated in recent economics literature, as it is not only intended to help the doctors make the best treatment decision for the patient at hand, but also to make the doctors factor in externalities of their treatment decisions, in this case in the form of antibiotic resistance.

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 samples: Sample 1 (a "hold-out sample") and Sample 2 (the "main sample").[7] 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[8]. This approach also has the added benefit of being plausibly implementable by healthcare providers, by first testing doctors through randomly assigned patients, and then assigning them patients suited to their skills. 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"[9]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 quantify how much physician-patient matching matters for patient outcomes, given the empirical heterogeneity in patient characteristics. Indeed, the gains from matching are driven by another fact that I establish, using a separate data set of patients' healthcare history: 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 costly to the patient and the insurer. Avoidable hospitalizations are a sign of low-quality primary care and are most common among low-income

[7]I verify the conditionally random assignment of patients to doctors in both samples.

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

[9]This could be due to different ability, for instance some are better at speaking with patients and reassuring them, while others are better at being strict with antibiotics guidelines even if a patient argues that they want antibiotics.

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.[10]

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 different questions than the common question (what would the effect be of increasing a certain input?). In particular, we can ask: how can we reallocate existing inputs to get an output improvement? This question is especially relevant in healthcare, where the lengthy and costly education of doctors means these inputs are difficult to increase in the short term. The conceptually simple framework relies on conditionally random matching to estimate an average match function (the average outcome for each doctor type when they meet 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 – as given. The framework takes into account the externality on the patient *from* whom a task-specific high-skilled doctor is moved in a reallocation. 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.

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 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, while the objective was solely to improve aggregate outcomes and not to reduce inequality, this reallocation shifts this aspect of doctor skill (risk prediction and prevention) towards lower-income patients, who are the ones most in need for this doctor task-specific skill.[11]

[10]This type of complementarity also exists for the other outcomes, which are more common and where the patient need is not correlated with income.

[11]The estimated risk of having an avoidable hospitalization, as well as the number of prior avoid-

Matching without moving people geographically is a resource-neutral policy that affects outcomes. However, its efficiency compared to resource-intense policy alternatives such as hiring and training, remains a priori ambiguous. To shed light on this, 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 considerably smaller than from doctor-patient matching policies, and would moreover be more difficult to implement. Matching has larger effects because (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 for some patients' outcomes but not to others.[12]

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.[13] Algorithmic allocation means that machine prediction is used as a complement to human skill, as opposed to substitute[14]. The algorithm allocates patients to doctors, but the doctor makes the triage, diagnosis and treatment decisions. 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.

The results on doctors' varying effects on heterogeneous patients could be gener-

able hospitalizations, are concentrated in the lower end of the income distribution. The doctors who are reallocated towards the lowest income decile are moved from 2% of patients drawn from all parts of lower to higher income patients.

[12]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, and I still find that matching has large effects for that outcome.

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

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

alizable also to in-person healthcare. 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.[15] Second, digital services can be viewed as a "lab", which helps overcome endogeneity challenges endemic in in-person primary care which have made the evaluation of causal effects of doctors challenging. This is because, at least initially and in some of digital care, doctor-patient assignment has been time-conditionally random. In regular in-person primary care, patient-doctor sorting confounds causal effects and all doctors do not meet all types of patients, meaning there is a lack of 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 outside of a pandemic[16] This is also the first study to hypothesize and test whether digital services can defy inefficient and unequal matching due to distance and location-based sorting. In addition, I bring a new source of conditionally random matching of service providers and clients 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. 2022, Bergeron et al. 2022, Fenizia 2022, Graham et al. 2021)[17]. 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 provider skill is challenging to evaluate, and where there is policy-relevant inequality in current resource allocation in many countries. I implement average reallocation effects (Graham, Imbens and Ridder 2020) in a setting without pre-existing estimates of patient need or doctor skill.

This paper also contributes to the literature on physician performance[18], by study-

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

[16]Zeltzer et al. 2021 provide an excellent study of telemedicine in Israel when it had increase due to the pandemic in 2020. Their aims are different, as they focus on providers that the patient already had a location-based relationship with.

[17]Cowgill et al. (2022) contribute with a slightly different perspective, by showing the conditions under which centralized assignment of workers is preferred compared to workers choosing positions within firms.

[18]See, e.g., Fadlon and van Parys 2020; Cutler et al. 2019; Currie and MacLeod 2017; Abaluck et

ing not only doctors' overall ability, but also specializations[19]. Alsan et al. (2019), Cabral and Dillender (2021) and Hill et al. (2018) study 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 in several dimensions.

Recent influential work (e.g., Mullainathan and Obermeyer 2022; Chan, Gentzkow and Yu 2022) has studied physician errors in decision-making. This study builds on that work in recognizing that physicians make errors, and that the error rate is heterogeneous. I add that doctors can have different skills in different dimensions, and that minimizing diagnostic errors are not the only important dimension for primary care doctors, but also minimizing externalities of treatment and making the patients informed and satisfied. The physician-patient matching model I propose incorporates potential heterogeneous physician error and assigns the patients, for whom errors are predicted to be most consequential, to the doctors who make the least errors in that dimension.[20]

2 Institutional background

2.1 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 20 regions) and private providers. Only a small share of citizens – 6% in 2017 (Glenngård 2020) – have an additional private health insurance, mainly provided by employers. Private health insurance accounts for less than 1% of health expenditures (Glenngård 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

al. 2016; Doyle, Ewer and Wagner 2010

[19]Currie and Zhang (2022) exploit the Veteran Administration's first-come first served assignment *within clinic* and find that physicians' abilities are correlated in dimensions that are closely related, such as avoidable hospitalizations vs. hospitalizations for circulatory conditions and deaths. However, they find that compliance with mental health screening guidelines is negatively associated with effectiveness in preventing hospitalizations, but in their setting the differences in screening propensity are small.

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

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

limiting waiting times are often unmet.[22] 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 the 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 (PCPs/GPs), but the variation in doctor effectiveness with different patients has been difficult to study. This is partly due to the endemic sorting between providers and patients in standard, in-person primary care – sorting and selection is more prevalent in primary care, where centers have smaller catchment areas than hospitals.[23]

Primary care physicians are institutionally positioned as a gatekeeper to access 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)[24].

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 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 in-person primary care, which is provided by a mix of private (40%) and public providers (60%), doctors

[22] In January 2019, 33% of patients could not see a doctor in person the same day across the country SKR (2022), and for some of the worst clinics, half their patients could not see a doctor within 3 days. More information is available in the Waiting times Section in the Online Appendix.

[23] Previous research has exploited plausible randomization between doctor teams and patients in hospital care (e.g., Doyle, Ewer and Wagner 2010) to evaluate doctor effectiveness. Some sophisticated designs exist in recent research on primary care, with Currie and Zhang (2022) exploiting the Veteran Administration first-come first served assignment *within clinic*, and Fadlon and Van Parys (2020) and Ginja et al. (2022) utilizing doctor exits.

[24] 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 in-person primary care setting, where in a meta-analysis, they range from 1.5% to 24.5% (O'Donnell 2000).

working in digital primary care are not paid fee for service but an hourly wage. The reimbursement level from the universal public health insurance to companies providing digital consultations has changed several times, while the fee paid 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 in-person primary care.

2.1.1 How patients choose in-person primary care providers

Regular (in-person) primary care is provided at 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 clinic.[25] 99% of Swedish inhabitants live within 20 minutes from their closest primary care clinic (Tillväxtverket 2011). 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). 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 2022).

2.1.2 In-person care sorting

In Table 1, I study in-person primary care data from the region where I have such data, Skåne. Table 1 shows that patients have a more negative experience with primary care[26] in areas with a higher deprivation index[27]. Moreover, in more deprived areas, patients are also less satisfied with the information they receive in in-person 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

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

[26] 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.

[27] 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.

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.

To make sure these relationships are similar across the entire country, I use aggregated public data. Appendix Table 5 indicates that patients across the country are less satisfied with their primary care in areas with lower income and higher share first-generation immigrants.[28] In contrast, Figure 1 shows that the shares of patients across the income deciles who meet good doctors in the 3 outcomes in digital care are similar.[29]

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

	(1) Positive experience	(2) Satisfied with information	(3) Met physician rather than other profession	(4) Recommended treatm. for diabetics
Deprivation index	-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: NPE and Region Skåne.

2.1.3 Sorting patterns into online care

I assemble and analyze proprietary data from one digital primary care provider, which is Europe's largest digital care provider in visit volume. This provider contributed with a majority of all such digital visits in Sweden 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

[28]Table 5 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).

[29]Satisfaction in the digital service actually decreases with income, opposite to the in-person results.

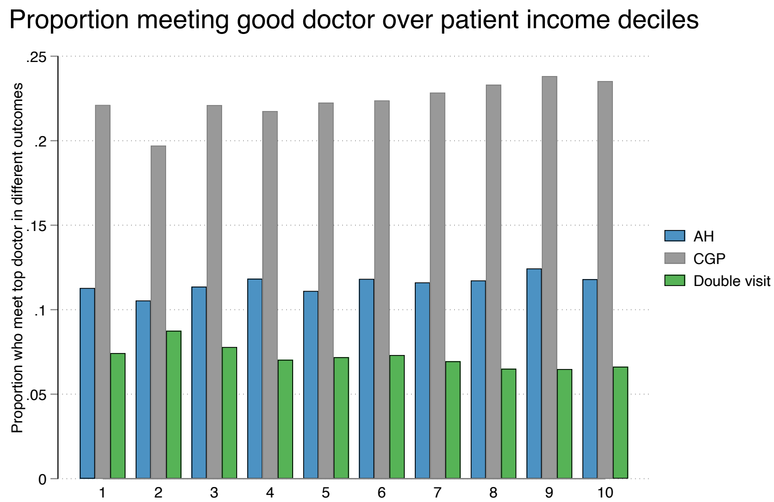


Figure 1: This figure shows what proportion of patients across income deciles who meet a doctor who is classified as top 10% in reducing avoidable hospitalizations, in following guidelines in antibiotics prescriptions, and in preventing double visits. All income deciles have different than 10% proportion of top doctors for the different outcomes, which is because doctors who are good at different things work a different amount of consultations during the sample period. The patient income is the income of adult patients in 2017.

sorting patterns into in-person primary care, I study one Swedish region where I have the universe of in-person primary care data.[30] This is Region Skåne, which is the southernmost region in Sweden, containing both rural areas and the third largest city in the country. Around 10% of the digital care users are from this region.

Using the same index of low socioeconomic status among the patients registered at the clinic as in Table 1, I find that the deprivation index is similar among digital users and non-users (Appendix Figure 8(b)) (extensive margin). However, on the intensive margin (not comparing digital and in-person anymore), individuals with higher deprivation index who use the digital service have more appointments in the digital service (Appendix Figure 7(a)). This is corroborated when looking at individual income: lower-income users use the digital service more intensively (Appendix Figure 7(b)). Figure 8(a) shows that digital care users are younger than non-users. There

[30]Primary care data is not collected by the national body (the National Board of Health and Welfare) which contributes with the rest of the in-person healthcare data to this study. To get access to in-person primary care data in the entire country, separate applications and reviews have to be made to the 20 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.

is a similar level of prior disease among digital users and non-users who are under 60 years old (Figure 9), measured by the sum of comorbidities from the Elixhauser index, a commonly used measure for summarizing disease burden (Elixhauser et al. 1998).[31] For users over the age of 60, non-users seem to have less prior disease.

Patients take up the service freely, and are not obliged to change their relationship with their regular in-person primary care clinic. Using data on in-person primary care from Region Skåne, I find that around 4% of digital care users have a nurse contact in in-person primary care the week after their digital care visit.[32]

2.1.4 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 consultations with certified medical doctors. The physicians may have different specialties, but all are acting as primary care providers/general practitioners (GP), and GP is the most common specialty. During the study period, the healthcare provider employed or contracted with around 500 doctors, but many of them were new or had not done many consultations.

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.

2.1.5 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 primary purpose of the service, but is a by-product of the aim to minimize and equalize waiting times nationally. Doctors can choose their time shifts, and often choose them around 2-3 weeks ahead. During their shifts, when they are not busy

[31] 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.

[32] 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 an in-person visit at a primary care centre within a week of using a digital care service.

with a patient or with follow-up work (such as writing prescriptions), they are in the roster of available doctors.[33] 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).

2.1.6 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 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.

	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 or in training)	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				

[33]Data from a later period may not be randomized to as large an extent since the healthcare provider after the study period started experimenting with matching, a process which this study has been informative for.

Doctors are paid per hour and there is no fee-for-service for the doctors, or bonus payments. Table 2 describes the characteristics doctors who are included in the study sample as they have worked at least 600 randomized consultations for the service.[34]

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.

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. That a patient is 'helped' means that the doctor has resolved the patients issue without redirecting them to other care. Hence, doctors have an incentive not to over-refer or redirect excessively to more care. Moreover, all doctors practicing in the country can be subject to disciplinary investigations if they engage in neglect or malpractice with adverse consequences for the patient. Hence, doctors also have an incentive to minimize adverse events for patients.

3 Data

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[35] 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

[34]Data on age, gender and employment status of these doctors are currently missing for a majority of doctors.

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

this is the sample where time-conditional randomization holds. Moreover, I remove pediatricians and the small children who are more likely to see a pediatrician (where randomization does not apply).[36]

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, consultations, hospitalizations and prescriptions from specialist, acute and inpatient care across the Swedish healthcare system in the three years preceding digital primary care, 2013-2015, and from the period concurrent to digital primary care, 2016-2018. Moreover, I match with data on prescriptions from all primary care nationwide in 2013-2018.[37] In addition, I include data on in-person primary care (2013-2019) from one Swedish region (Skåne), which matches for around 10% of the digital care sample.[38] Finally, I keep only doctors who have done >600 consultations and their patients, which leaves around 200,000 patients and 143 doctors. [39]

3.2 Measurement of outcomes

Estimating doctor performance in primary care has been challenging, as important patient outcomes are often ambiguous, rare, and/or delayed. Mortality and quality of life may be the most important outcomes, and these suffer in measurement from being delayed or rare (mortality) and ambiguous or subjective (quality of life). Other important outcomes are limiting costs to the rest of the healthcare system, as primary care physicians serve as gatekeepers, and limiting health externalities, such as the spread of contagious diseases through vaccination, and the limiting of antibiotics use leading to resistance.

Primary care physicians have multiple tasks, which opens the question of whether a single ability measure governs performance in all tasks, or whether even in *general practice* doctors are in fact specialized. I address this by creating observable output measures of doctors in three key dimensions of a primary care physician's work:

[36]These small children (born after 2012) also do not have the full set of pre-data which starts in 2013.

[37]Prescriptions data is the only data from primary care that is collected nationally.

[38]Swedish in-person primary care is devolved to 20 regions, which means that all data from primary care is not included in national registries.

[39]Many doctors are excluded as they have only done a few randomized consultations, many of them under 100. Common reasons are that they had a trial only, or were hired late in the sample period. For more details on the sample definition, see the Online Appendix.

(1) identifying risky patients and preventing serious adverse events (2) providing guideline-consistent treatment for common conditions that limit externalities, 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 risk prediction, the negative outcome is an avoidable hospitalization, defined as a hospital admission that could have been avoided with sufficient primary care. In the case of providing guideline-consistent treatment, I measure whether the patient has received a counter-guideline antibiotic. 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.

Avoidable hospitalizations (AH) The main outcome I use is defined in the medical literature since the 1990s as a hospital admission that could have been avoided with sufficient primary care, and the diagnoses for which a hospitalization is regarded as avoidable are listed by medical research independently from this study. This outcome can be seen as a proxy of mortality that is more commonly observed.[40] Avoidable hospitalizations can even be seen as a better outcome measure than mortality, as AH are more closely linked to the work of the primary care doctor. Mortality could be due to factors outside of the control of a primary care doctor, such as a car accident, while AH are defined to be preventable by primary care.

Avoidable hospitalizations are rare events: 0.2% of all patients have an avoidable hospitalization in the 3 months following the digital consultation (but 6% of patients defined as risky have an avoidable hospitalization in the same time period). 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 (Bond and Dow 2021)[41] and the economics of disasters (e.g., Barro 2009)[42]. Another reason to focus on this outcome is that one of the main tasks of a primary care doctor is to

[40] Currie and Zhang (2022) show that primary care practitioners who are better at reducing avoidable hospitalizations are also the best doctors at reducing deaths. Their choice of term is 'hospitalizations for ambulatory care sensitive conditions', which is the same concept as AH.

[41] This has also been at the forefront of public debate after the financial crisis and the pandemic.

[42] 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. 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.

sort the rare and seriously ill patients from the vast majority with minor complaints.

Bacterial pneumonia, urinary tract infection and congestive heart failure account for 77% of the AH costs in the US (Rocha et al 2020). Avoidable hospitalizations are dangerous, both because of the inherent risks when a condition has worsened unnecessarily, and because hospitalization in itself has risks such as hospital-acquired infections and risks from invasive procedures. 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) (McDermott and Jiang 2020). 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.

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 days happen quite early after the digital consultation, and the mean is 33 days. I conduct several checks to determine whether the avoidable hospitalization can actually be considered as preventable in the digital consultation, available in the Online Appendix.

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., Abaluck et al. (2021), Cuddy and Currie (2020), Finkelstein et al. (2022) and Frakes et al. (2021). While recognizing that non-adherence could be due to superior skill or access to richer information, and thus lead to better outcomes, several of these papers show that non-adherence leads to higher costs or worse outcomes for the patient at hand.

Non-adherence to antibiotics prescription guidelines is particularly interesting since excessive antibiotics prescriptions lead to the negative externality of bacterial resistance. Hence, this is an example of another of the doctors' skills in a primary care system such as the one studied, namely to minimize externalities. I have chosen this particular type of guideline for three reasons. First, it adds to the literature on guideline adherence by studying a guideline that explicitly incorporates the benefit of other people, and hence does not only serve to maximize outcomes for the patients while minimizing pecuniary costs. Second, it is a guideline where consistent non-adherence is a clear signal of lower skill, if we take the policymakers' weighting of the externality vs. patients' benefit to be the correct one.[43] Third, it is measurable in my data as other guidelines particular to online care were not yet developed, but this was one that policymakers prioritized.[44]

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 and if there is less prescription of antibiotics, it is possible to decrease the number of resistant bacterial infections (Bergman et al. 2004). Antimicrobial resistance is estimated to lead to more deaths annually worldwide than either HIV/aids or malaria (Murray et al. 2022)[45]. The non-adherence measured in my sample is quite low (4%) by international standards, as is common in Scandinavia. 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 in-person care within a week after the digital consultation This outcome will be less emphasized as it is only available in 10% of the sample, i.e., for patients in the region which delivered full in-person primary care data. It is an outcome which is important for primary care costs and for patient satisfaction. If a patient contacts an in-person primary care clinic in the week following the digital care consultation, this may indicate that they were not satisfied with the digital care

[43] Many patients want antibiotics and push for it, and the primary care physician's role here is to limit the use of antibiotics for the common good. Physicians are allowed to prescribe above the guideline in a small number of cases where they have more information, but if a physician consistently over-prescribes with a balanced set of patients, then this is a sign of low skill in resisting the patients' pushing, or low awareness of the guidelines.

[44] I code non-adherence to 16 guidelines from Swedish strategic programme against antibiotic resistance on digital care (Strama 2017, 2019). More details on the variable creation can be found in the Online Apendix.

[45] Global deaths associated with antimicrobial resistance are estimated to be 5 million/year, of which 1.2mn are deaths for which antimicrobial resistance can be held directly responsible. This is more than HIV/Aids (0.86mn) or malaria (0.64mn) (Murray et al. 2022).

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).

4 Conceptual framework

This section has two objectives. First, it presents the econometric framework for estimating match functions between patients and doctors, and counterfactual effects from reallocations. I follow Graham et al. (2020) with some modifications. This framework takes seriously that healthcare resources can be rival.[46] I take into account the "externality" on the patient from whom the a doctor, who is highly skilled in some task, is moved. I also add a consideration of opportunity costs in terms of other outcomes when doctors are multitasking and skills are potentially correlated.

The second objective of this section is to illustrate the matching problem of the healthcare planner. There are two main reasons to view this problem from the perspective of a planner. First, it could be a realistic setting not only in a public healthcare system. Cowgill et al. (2022) theoretically cover the circumstances under which centralized assignment by firm leaders leads to higher productivity, accounting for the effect on retention rates, compared to self-organized matches where worker preferences are expressed through for instance deferred acceptance. They also show empirically that within one example large firm, planner-dictated matches are more valuable than preference-based matches.

Second, healthcare is fraught with real externalities, which a planner may internalize more than in a decentralized system. However, models studying physician behavior often choose settings which are free from those to focus attention on other features.[47] I have included at least one outcome which has externalities: counter-guideline antibiotics prescriptions.I have chosen to take the perspective of a planner who has the same views as the Swedish governmental agency on antibiotics: i.e., I

[46]Many economics papers on healthcare consider covering more people under insurance or changing incentives which lead to more utilization, without explicitly recognizing that, e.g., medical doctors are a scarce resource and could be considered fixed at least in the short run.

[47]See, e.g., Abaluck et al. (2020) who compellingly study physician guideline adherence in the allocation of a drug, which has close to zero marginal cost, and whose only downsides occur within the patient themselves.

take the guidelines as striking the correct trade-off. Abaluck et al. (2020) show that there is large variation in adherence to other guidelines where there are no externalities, and physicians do not incorporate more information that is relevant to treatment effects. They also show that promoting knowledge about the guidelines does not go all the way in optimizing physician behavior. In this paper, I instead consider a planner who could reassign the doctors who are better at adhering to guidelines to the patients who need that.

This study is complementary to the literature on mechanism design in matching markets 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.

4.1 Econometric framework

Consider D doctors and N patients. Doctors have observable characteristics W_j, which measure doctor skill or effectiveness in different tasks, and patients have observable characteristics X_i which measure patients' need for different doctor inputs, and is predicted from patients' healthcare history. One of the reasons that doctors differ in skill in certain tasks could be different rates of prediction errors (Mullainathan and Obermeyer 2022). This source of difference in skill is particularly pertinent in the case of determining which patients are at risk for adverse outcomes such as avoidable hospitalizations. Another reason that doctors differ in effectiveness in some tasks could be differences in communication skill, which is particularly relevant for making the patient satisfied and not seeking unnecessary repeat care for the same issue. A third difference is how confidently doctors are able to counter patient demands for unnecessary antibiotics, or how much weight they put on the externality. 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 is:

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

The research design is based on on random assignment (conditional on time[48]) 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:

$$f_{X_i,V_i,W_i,U_i}(x, v, w, u) = f_{X_i,V_i}(x, v) f_{W_i,U_i}(w, u) \tag{1}$$

Under restriction (1) on the joint distribution of the characteristics of patients and doctors, the conditional mean of the outcome Y_{ij} is called the Average Match Function (AMF):

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

The AMF, $\beta(x, w)$, provides information on how match output varies across different types of agent pairings, when both doctor and patient are random draws from their respective subpopulations x and w. Figure 10 in the Online Appendix shows an example of how the AMF looks in this context. The AMF 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 $\tilde{W}_j$[49]:

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

which satisfies the feasibility condition (this will later be relaxed):

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

for all $w \in W$. The distribution of patients is kept fixed, i.e. $f_{X_i}(x)$ is left unmodified.

[48] 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.

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

Average healthcare outcomes under a counterfactual patient-doctor assignment equal:

$$\mathbb{E}\left[\tilde{Y}\right] = \int \left[\int \beta(x,w)\tilde{f}_{\tilde{W}_j|X_i}(w|x)dw\right] f_{X_i}(x)dx \tag{2}$$

which can be calculated with knowledge of the AMF. The Average Reallocation Effect (ARE) from the reallocation $\tilde{f}$ is $\tilde{Y}$ relative to the average outcome under the status quo allocation, $\bar{Y}^{sq}$:

$$ARE(\tilde{f}) = \mathbb{E}\left[\tilde{Y}\right] - \bar{Y}^{sq} \tag{3}$$

Since everything to the right of the equality in equations (2) and (3) is identified, so is the Average Reallocation Effect (Graham et al. 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.

4.2 Problem: Reallocation of Fixed Healthcare Resources

The objective of this problem[50] 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. As a first step, 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 of the analysis, I consider selective hiring policies where I extend the working hours of the doctors who have above median skill in several tasks.

I will make one main simplification: to focus on one outcome k at a time, e.g., reducing avoidable hospitalizations. This is reasonable as it is unclear how a planner would weigh the different outcomes against each other. Instead, I will study what happens to other outcomes when I reallocate to improve one outcome. 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.

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

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 to binary, good or bad, needy or non-needy.

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 k such as avoidable hospitalizations:

$$min_{\pi}Y^k(\pi)=\sum_{i=1}^{I}\sum_{j=1}^{J}\beta^k(x_i,w_j)\pi_{ij} \tag{4}$$

subject to feasibility constraints:

$$\sum_{j\in J}N_p\pi_{ij}=N_x \qquad \forall x\in X \tag{5}$$

(each patient gets 1 doctor)

$$\sum_{x\in X}N^{\pi}(x,w)=N^{\pi}_{SQ}(w) \qquad \forall w\in W \tag{6}$$

(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:

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

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

Outcome-maximizing 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 average reallocation effects (ARE) of assortative matchings. The 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 ARE 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 has two building blocks. The first is nationwide time-conditional random assignment between patients and doctors in digital primary care. This generates variation in patient types that each doctor meets – in geographic location, age, socioeconomic status, previous healthcare utilization, etc. The conditionally 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 to avoid overfitting and to create an implementable strategy. In particular, I evaluate doctor effectiveness using a value added method in a hold-out sample of randomized digital care (Sample 1, 40% of consultations).In Sample 2 (60% of consultations), 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 (see Figure 5). Both samples have conditional random assignment between doctors and patients. I choose each doctor's *first* 600 randomized visits because that is how the procedure could be operationalized: It gives the employer ∼3 months of work by the doctor as a sample to evaluate the doctor.[51] The employer can

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

then assign the doctor to different patients, and I show how results would look from that in a sample which does not contain the same patients as in the doctor evaluation sample.

Who should doctors skilled in a certain task be matched with? I predict patient risk factors (X_i) in another separate sample (Sample 0), which consists of pre-digital (in-person) healthcare data in 2013-2016 – the period preceding digital care. I find logical ex ante patient characteristics which indicate need for doctor input related to each outcome Y^k. For avoidable hospitalizations, I predict the risk with a simple linear method.[52]

Balance The identifying assumption both for estimating doctor skill and match effetcs 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 6 shows that characteristics are balanced. Another balance test is reported in Table 7, which shows that patient predicted risks for avoidable hospitalizations (AH) are uncorrelated with doctor AH skills in the main estimation sample.

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 are key. To overcome (1) and (2), I use the unique nationwide conditionally random patient-doctor allocation in digital primary care. 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 define several doctor tasks which stand in direct relation to measurable patient outcomes.

[52]I have also predicted risk with a random forest algorithm using much more of the data, but this does not improve much out of sample on the simple linear regression using sparse data. I therefore use the simple linear rule using only 6 variables, since would be easier to implement and also more transparent for patients.

To deal with the delayed nature of many important primary care outcomes, (4), I use a variety of shorter-term outcomes, ranging from frequent and lower-stakes, to rare and high-stakes, but all of which are measurable within 3 months. I address (5) by specifically studying the varying effectiveness of different doctors with heterogeneous patient types. The co-production of healthcare with the patient is important for 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.

$$Y_{ij} = Z_i\Pi + \lambda_t + w_j + \epsilon_{ij}$$

where $\hat{w}_j = \hat{W}_j^{EB}$ is estimated as the Empirical Bayes shrunk random effect of doctor j.[53] 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.

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.[54] $\hat{w}_j$ is unbiased due to random assignment and common support. However, $Var(\hat{w}_j)$ is positively biased due to sampling noise. 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 ratio of signal variance to signal plus noise variance. Similar shrinkage is common in studies of teacher value-added (see e.g. Kane and Staiger 2008; Chetty et al. 2014). Table 15 in the Appendix shows the regression estimating the doctor effects for avoidable hospitalization skill.[55]

Defining patient types I define patient types based on risks for the various negative events that define the outcomes. There is a tradeoff between choosing the best

[53] 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.

[54] In the sample of doctors' first 600 randomized consultations, >95% of doctors have met a patient with an avoidable hospitalization in the past 3 years.

[55] Table 15 shows that this estimation has the outcome "negative number of avoidable hospitalizations". The outcome variable is negative to ensure that the random effect is higher for a better doctor, for ease of exposition later on.

prediction of which patient is at risk (which would generate larger benefits from reallocation) and keeping the prediction simple. The benefits of keeping the prediction simple are twofold: first, the exercise becomes more realistic if we use only a small set of variables that are also available to the medical provider, which means the procedure could be implemented in practice. Second, the procedure becomes more transparent and thus politically feasible if instead of a black box sophisticated prediction, we use a simple linear rule that defines a cutoff between who will get a higher skilled doctor in each outcome. To be conservative, I have chosen the simple rule instead of a machine learning prediction that could generate larger reallocation gains.[56]

For the rare outcome avoidable hospitalizations, I create a risk score based on the lagged outcome variables from data before digital healthcare (2013-15):

$$P_i = C_i\Gamma + \upsilon_i$$

where P_i is the past number of avoidable hospitalizations and C_i are 6 demographic and healthcare-related variables. In particular, I have chosen variables that are not gameable by the patient, which minimises concerns that a patient could try to strategically affect their risk score to get assigned to another doctor.[57] I do not include any variables about the current state or symptoms, which also means that a patient would be assigned to the same type of doctor over time, and thus continuity could be achieved with patients meeting the same doctor over time. Instead, the healthcare related variables that I include in the risk prediction are for instance the Elixhauser comorbidity score, which measures the number of serious diseases that a patient has been diagnosed with over the past 6 years, a variable which is arguably not very gameable.

To define patient types X_i, I generate a prediction $\hat{P}_i$ for each i, as the patient risk variable. Table 19 in the Online Appendix reports the regression used to create the risk score for avoidable hospitalizations.[58]

[56]I have also done the prediction of patient risk with a random forest, and the prediction improvement compared to the linear regression is not very large.

[57]The variables included are Elixhauser disease index, gender, age, immigration status and number of hospitalizations 3 years before the online visit excluding avoidable. Table 20 in the Online Appendix shows versions of the regression also including other socioeconomic characteristics, and with a sparser set of regressors.

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

Creating binary types for avoidable hospitalizations To reduce reliance on the exact estimate of both patient risk and doctor skill[59], and to make fewer assumptions about the nature of complementarities in the match function, I collapse patient types to a binary variable measuring high and low risk. Since around 1% of patients have an AH each year nationally, I characterize 1% of patients as risky ($X_i = 1$) based on the rank of the risk score $\hat{P}_i$. To make a waiting time constraint less binding, I characterize 10% of doctors as highly skilled in preventing avoidable hospitalizations ($W = 1$) based on the rank of $\hat{W}_{jk}^{EB}$.[60]

Figure 12(a) illustrates that the groups created based on the risk 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 12(b) shows that the risk groups (defined only based on past healthcare records and demographics) are highly predictive of *future* avoidable hospitalizations: virtually 0% of patients who are classified as non-risky have an avoidable hospitalization within the 3 months after the online consultation, while 6% of the risky patients have it, despite the online consultation reducing some hospitalizations.

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.

I estimate the effect of a top 10% doctor on top 1% risky patient:

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

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.

Table 7 in the Appendix shows evidence of random assignment: that patient risks

[59] This is especially important for the rare outcome avoidable hospitalizations.

[60] 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.

are uncorrelated with doctor skills in the main estimation sample (Sample 2).

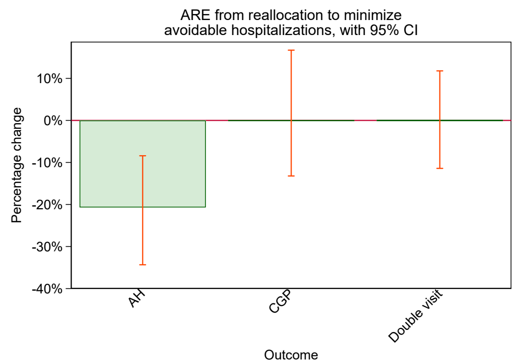
Reallocation procedures and costs The simplest reallocation procedure I carry out is to reallocate the top 10% doctors randomly to top 1% high-risk patients and let them swap doctors with some non-risky patients. The reallocation procedure where I use continuous measures of patient risk and doctor skill, is positive assortative matching (PAM): allocate the highest effectiveness doctors to the highest need/risk patients. Costs of reallocations are small in the digital setting compared to the in-person setting where geographic distances play a big role. One cost that also applies to the digital setting is longer waiting time for patients to get a more suitable 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.

6 Results

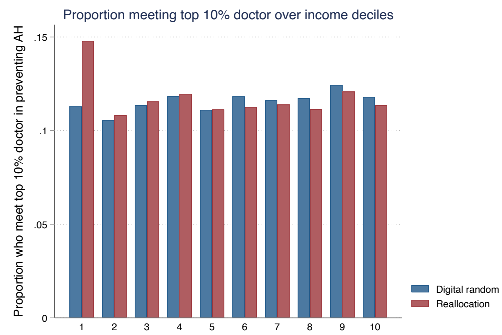
6.1 Reallocation results

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

The first set of Average Reallocation Effects are derived from the optimization problem in Section 4.2. This problem takes existing resources in terms of doctor skills and time worked as given, as it might be difficult iand costly to increase all doctors' skills at several different tasks, and there are constraints to hiring new doctors. Moreover, retraining in (and thus emphasizing) some skills may lead other skills to suffer in a multitasking setting. 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 the overall comparison Figure 4.



(a) Average Reallocation Effects with confidence intervals from a Bayesian Bootstrap of the entire doctor-patient allocation procedure.



(b) Proportion of patients across income deciles who meet a doctor who a top 10% in reducing avoidable hospitalizations.

Figure 2: Reallocation with binary match function where a good doctor is defined as a top 10% in the AH outcome. Panel (b) compares to the random allocation that actually took place in the digital service.

The first result (Figure 2 a) is that avoidable hospitalizations (AH) decrease by 20% when matching 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. For other objective functions, Figure 4 shows that reallocating the doctors who are best 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.

There are also effects on healthcare inequality from the reallocation to minimize aggregate avoidable hospitalizations. Before reallocation, the probability of meeting a top 10% doctor in risk prediction/triaging was similar across patients' income distribution (Figure 2 b).[62] After the reallocation, the chance of meeting a top 10% doctor in risk prediction/triaging increases for the bottom patient income decile. This is

[61] I.e. contacting an in-person primary care nurse the week after the digital visit.

[62] All income deciles have a slightly higher than 10% proportion of top doctors in the random allocation, which is because the top doctors work more consultations than other doctors. Patient income is the income of adult patients in 2017.

because the risk for avoidable hospitalizations is highest in the lowest income decile. More information on the correlation between avoidable hospitalization risk and socioeconomic variables can be found in the Online Appendix in Table 11.

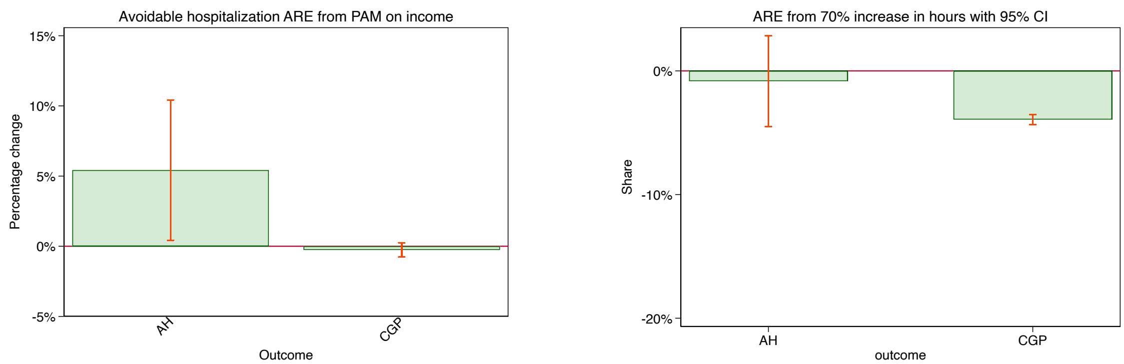
Figure 3 (a) presents another way of understanding the income-health gradient aspect of doctor-patient matching. This figure shows Average Reallocation Effects from a reallocation where the highest-skilled doctors in reducing avoidable hospitalizations are matched with the highest-income patients. This reallocation is compared to the random real-life digital assignment, and shows that aggregate avoidable hospitalizations would be around 5% worse 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 in-person care (Stoye 2022; Agency for Healthcare Research and Quality 2020). If this also applies to risk detection and triage skill for in-person care doctors, Figure 3 (a) suggests that avoidable hospitalizations after primary care could be lowered by up to 5% if patient-doctor matching changed to a random matching from an income-based sorting. Moreover, if we add together the results from Figures 2 (a) and 3 (a), they suggest that moving to an needs-based allocation on avoidable hospitalizations from an assortative matching on patient income and doctor skill could reduce the number of avoidable hospitalizations by around 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 have above median skill in all three outcome measures, and commensurately reducing the hours of the remaining doctors. However, Figure 3 (b) illustrates that when increasing these doctors' work hours by as much as 70%, there still is 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, see Figure 4). Moreover, a 70% increase in these doctors' work hours would be difficult to achieve, even if digitalization can be expected to give room for some increase in hours for the best

[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.

doctors.[64]. Hence, the gains from matching are considerably larger than the gains from improving doctor selection.



(a) Positive Assortative Matching (PAM) on doctor skill in reducing avoidable hospitalizations, and patient income.

(b) Effects of an increase by 70% in work hours for the doctors who are above median in all three measures, random matching.

Figure 3: Average Reallocation Effects, using the continuous match function.

6.2 Defying distance

To understand how much defying distance (i.e., matching patients and doctors across long distances) contributes to the estimated gains from matching doctors and patients, we would like to know how digital patients' risk and doctors' skill are distributed across the country. Figure 6 (a) in the Appendix shows the geographic distribution of avoidable hospitalization (AH) risk among digital care patients. Rural municipalities have the highest share of risky patients. Among the 10% of municipalities with highest share of AH-risky patients, the share (defined as the top 1% risky nationwide) is between 3.8% and 20%. None of these 29 municipalities with a large share of risky patients are in the municipality category 'city'.[65] Hence, the share of AH-skilled doctors would need to be 3-20 times as common in these rural municipalities, compared to the average across the country, to achieve the full matching gains without defying distance.

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

[65]10% of these high-risk municipalities are 'Dense near a city' and the other 90% are different categories of rural or remote. The six municipality categories come from Tillväxtverket, the Swedish Agency for Economic and Regional Growth.

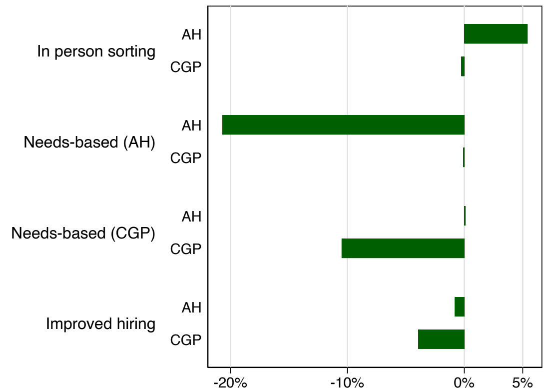


Figure 4: Percentage change in negative outcomes (avoidable hospitalizations (AH), counter-guideline prescriptions (CGP)) from reallocations compared to random matching.

How likely is such a doctor avoidable hospitalization-skill distribution to hold? First, note that I have only limited data on online doctors' location. But among the third of doctors where I have information on their region, around 50% live in Region Stockholm, the largest city, where only 20-25% of the country's population is located.[66] Regarding the in-person doctor skill distribution across municipalities, this is harder to evaluate for primary care doctors working in person, since doctors and patients are not randomly assigned in in-person care, and there is also a lack of nationwide data on in-person primary care. One could assume that the doctors working for the digital service are representative of in-person doctors working in their locations, but that might be a too strong assumption.

Instead, I use information from other contemporary work regarding specialist doctors instead of primary care providers, to shed light on how likely it is that such a doctor skill distribution holds. Stoye (2022) shows that cardiologists in urban areas are 0.1-0.3 standard deviations more skilled in preventing death after heart attack than doctors in rural areas in England. This means that the distribution of primary care doctor AH-skill across urban vs. rural areas in Sweden would have to be qualitatively opposite to the distribution of cardiologist mortality skill in England (from Stoye 2022), in order to achieve the matching gains from matching patients to doctors only within their municipality. Hence, defying distance seems key to the matching gains. Moreover, there are even larger imbalances globally in where skills and need is

[66] There are 20 regions in Sweden and 290 municipalities nested within them.

located, so the gains within Sweden are a lower bound.

6.3 What drives the gains from matching?

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 15 in the Appendix 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% .

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, Table 22[67] shows that there is no positive correlation 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 in-person nurse care the week after the digital doctor appointment). 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 in-person 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.

Large causal 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 to patients with higher estimated need/risk in outcome k.[68] 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 13). These complementarities in patient-doctor matching are illustrated

[67] For a visual representation, see Figure 14 in the Online Appendix.

[68] 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.

graphically in Figure 11. Table 16 in the Appendix shows results from the parametric version of the match regression, and includes robustness checks.

Table 3: 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

Robust SEs in parentheses. 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 of antibiotics prescriptions in the past to doctors better at following antibiotics guidelines.[69] 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 14).[70] 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[71]. 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.

[69] 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

[70] 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 pre-existing disease level. Otherwise it suggests that the patient is particularly keen on antibiotics, and potentially tries to pressure the doctor to get them.)

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

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 deviation better in the hold-out sample at following guidelines, their risk of getting a CGP is reduced by 24-27%.[72]

6.4 Mechanisms in preventing avoidable hospitalizations

Table 4: 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)
Interaction	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.

To clarify how some doctors become particularly effective at preventing avoidable hospitalizations, I study the actions that the doctors take during the digital care consultation. This is particularly important since avoidable hospitalizations could be minimized through over-referring all patients, which would not be optimal due to the costs. I show that doctors who prevent more avoidable hospitalizations do so by changing their actions for only risky patients, not by over-treating all patients. Table 4 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., an in-person primary care center, possibly one with extended opening hours, or in some cases the Emergency Department.[73] The main take-away

[72]Table 17 in the appendix uses the number of antibiotics instead of the share for the patient risk variable, and the results are similar.

[73]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

from Table 4 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 person are not redirected to that type of care, while false positives would be that patients who *do not* need additional in-person checkups *are* redirected for these checkups.

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, 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 18 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 18 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

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 in-person care setting (where in a meta-analysis, they range between 1.5% and 24.5% (O'Donnell 2000)).

design, is whether patients recognize which doctors are most appropriate for their needs. Column 3 of Table 18 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, suggesting the centralized assignment uses information that is not immediately available to patients.

7 Conclusion

The digitalization of services has several implications, three of which are especially important for the topic of this paper. First, the number of potential providers that any given user could meet has increased, as 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. Moreover, algorithmic assignment means that we can randomize allocation, or otherwise find out the exact rule for assignment, which opens the possibility of causally assessing the performance of individual providers. Taken together, at least two new possibilities are opened up: first, we could improve the selection of service providers as we can measure their performance. Alternatively, we could rethink how service providers and users are matched, taking advantage of providers' task-specific skills and users' needs.

In this paper, I have developed a framework for measuring causal skills of doctors, predicting patient risks, and quantifying effects from alternative policies than the status quo assignment. I have simulated both the above-mentioned possibilities, and found that the gains from moving to a needs- and skills-based matching are much larger than the gains from improving doctor hiring in the setting of digital primary care in Sweden. The reallocation procedure I describe is potentially cost-neutral, as opposed to training and hiring more skilled doctors. Many countries face challenges to keep down the cost of healthcare, and to deal with human capital shortages.

I have also looked at potential implications for inequality. The technology of digital services disrupts geographically-related sorting patterns between service providers and clients, which often have resulted in inequality in service quality. The stated aim of many healthcare systems, including the one studied, is to provide equal quality services for all. I have shown that we do not need to actively model the social planner's inequality aversion to get reduced inequality from matching patients to doctors: given that important healthcare outcomes such as avoidable hospitalizations are more common among lower-income patients, assigning doctors to minimize the aggregate number of those hospitalizations actually means that lower income patients get to meet more of doctors who are really good at that, reducing inequality in outcomes.

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' effects 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 move 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 doctors who are best 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)[74] and the US (USD 6.7 billion in the US for only adults in purely hospital costs), apart from lives saved.[75] Moreover, reallocating the doctors who are best 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.

[74]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 (McDermott and Jiang 2020)

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

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. When services move online, this becomes a feasible and resource-neutral, low-cost way of increasing effectiveness of service provision.

8 References

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

Abaluck, Jason, Leila Agha, David Chan, Daniel Singer, Diana Zhu (2020): "Fixing Misallocation with Guidelines: Awareness vs. Adherence," Unpublished manuscript.

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.

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 (2022): "Teacher Effectiveness and Classroom Composition: Understanding Match Effects in the Classroom," *The Economic Journal*, 132(648): 3047–3064.

Barro, Robert J (2009): "Rare Disasters, Asset Prices, and Welfare Costs," *American Economic Review*, 99(1): 243-64.

Bau, Natalie (2022): "Estimating An Equilibrium Model of Horizontal Competition in Education," *Journal of Political Economy*, 130(7).

Becker, Gary (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81(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 (2022): "Quality and Optimal Matching of Bureaucrats: Evidence from Tax Collection in the DRC," NBER Working

Paper 30413.

Bergman M, Huikko S, Pihlajamäki M, Laippala P, Palva E, Huovinen P, Seppälä H; Finnish Study Group for Antimicrobial Resistance (FiRe Network) (2004): "Effect of macrolide consumption on erythromycin resistance in Streptococcus pyogenes in Finland in 1997-2001," Clin Infect Dis. 1;38(9):1251-6.

Bond, Philip and James Dow (2021): "Failing to forecast rare events," *Journal of Financial Economics*, 2021-06.

Cabral, Marika and Marcus Dillender (2021):. "Gender Differences in Medical Evaluations: Evidence from Randomly Assigned Doctors." Unpublished manuscript.

Centers for Disease Control and Prevention (2019): "Antibiotic Resistance Threats in the United States." Link. Accessed 9-1-2021.

Chan, David C, Matthew Gentzkow and Chuan Yu (2022): "Selection with Variation in Diagnostic Skill: Evidence from Radiologists," *Quarterly Journal of Economics*, 137(2): 729–783.

Chen, Yiqun, Petra Persson, Maria Polyakova (2022): "The Roots of Health Inequality and The Value of Intra-Family Expertise," *AEJ: Applied* 14(3): 185-223.

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.

Cowgill, Bo, Jonathan Davis, Pablo Montagnes, Patryk Perkowski (2022): "Matchmaking Principals: Theory and Evidence from Internal Labor Markets," Mimeo.

Cuddy, Emily, and Janet Currie (2020): "Rules vs. Discretion: Treatment of Mental Illness in U.S. Adolescents," NBER Working Paper 27890.

Currie, Janet, and W Bentley MacLeod (2017): "Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians," *Journal of Labor Economics*, 35(1): 18977.

Currie, Janet and Jonathan Zhang (2022): "Doing More with Less: Predicting Primary Care Provider Effectiveness," NBER Working Paper 28929.

Cutler, David, Jonathan Skinner, Ariel Dora Stern, David Wennberg (2019): "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending," *AEJ: 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.

Doyle Joseph Jr, Stephen M Ewer and Wagner Todd H (2010): “Returns to physician human capital: evidence from patients randomized to physician teams,” *Journal of Health Economics*, 29(6):866-82

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.

Fadlon, Itzik and Jessica Van Parys, 2020. “Primary care physician practice styles and patient care: Evidence from physician exits in Medicare,” *Journal of Health Economics*, vol 71.

Fenizia, Alessandra (2022): “Managers and Productivity in the Public Sector,” *Econometrica*, 90(3): 1063-1084.

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 (2022): “A Taste of Their Own Medicine: Guideline Adherence and Access to Expertise,” *American Economic Review: Insights*, 4(4): 507-526.

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.”

Ginja, Rita, Julie Riise, Barton Willage, and Alexander Willén. 2022. “Does Your Doctor Matter?”. Unpublished manuscript.

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.

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." Unpublished manuscript.

Heckman, James J. and Rasmus Landersø (2021): "Lessons from Denmark about Inequality and Social Mobility," NBER Working Paper 28543.

Hill, Andrew, Daniel Jones and Lindsey Woodworth (2018): "Physician-patient race-match reduces patient mortality." Unpublished manuscript.

Kane, Thomas J. and Douglas O. Staiger (2008): "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," NBER Working Paper 14607.

Kasy, Maximilian and Alexander Teytelboym. 2021. "Learning by Matching." Unpublished manuscript.

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

McDermott K. W. and Jiang H. J. (2020): Characteristics and Costs of Potentially Preventable Inpatient Stays, 2017. Rockville (MD): Agency for Healthcare Research and Quality (US); Statistical Brief 259.

Mullainathan, Sendhil, and Ziad Obermeyer. 2022. "Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care," *Quarterly Journal of Economics*, 137(2): 679-727.

Murray, Christopher J L, Kevin Shunji Ikuta, Fablina Sharara et al. (2022): "Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis," *The Lancet*, 399(10325): 629-655.

National Board of Health and Welfare (2018): "Tillgänglighet i hälso- och sjukvården."

Nationell Patientenkät (the National Patient Survey) (2019): Link.

Neiman AB, Ruppar T, Ho M, et al. (2017): "CDC Grand Rounds: Improving Medication Adherence for Chronic Disease Management — Innovations and Opportunities." MMWR Morb Mortal Wkly Rep, 66.

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.

Rocha, J, Marques, A, Moita, B, Santana, R (2020): "Direct and lost productivity costs associated with avoidable hospital admissions," BMC Health Serv Res 20(210).

SKR (Sveriges Kommuner och Regioner) (2022): "Väntetider i vården." Link.

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."

Socialstyrelsen (2017): "Konsekvensutredning – förslag till föreskrift om belopp för vård av utskrivningsklara patienter." Link.

Statistics Sweden Statistikdatabasen, (2021): Link.

Strama (2019): "Rekommendationer för kvalitetsindikatorer vid digitala vårdmöten."

Strama (2017): "Rekommendationer för kvalitetsindikatorer vid digitala vårdmöten."

Stoye, George (2022): "The distribution of doctor quality: evidence from cardiologists in England." IFS Working Paper 22/30.

Taleb, Nicholas (2007): "The Black Swan: The Impact of the Highly Improbable". Random House, New York.

Tillväxtverket. 2011. "Tillgänglighet till kommersiell och offentlig service 2011."

Vasilik, Ashley (2021): "Triage Process in Primary Care Clinics". Link.

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." NBER Working Paper 28978.

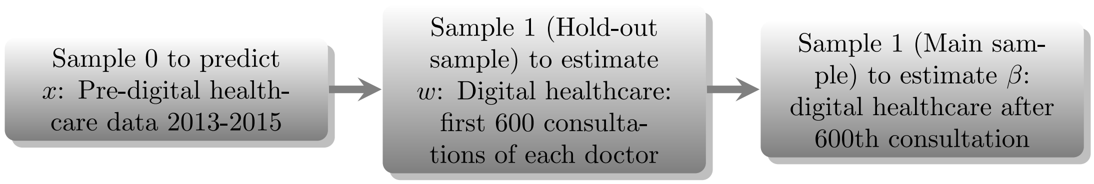


Figure 5: Illustration of the 3 samples.

Online Appendix

A Additional Tables and Figures

Table 5: In-Person 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

Robust standard errors in parentheses. The unit of observation is a 4-digit postcode 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 Nationell Patientenkät (the National Patient Survey or NPE) 2019.

Table 6: OLS of doctor on patient characteristics for dropin first visit

	Doc foreign	Only MD	Specialist	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

Table 7: Balance test in main sample: after 600th visit

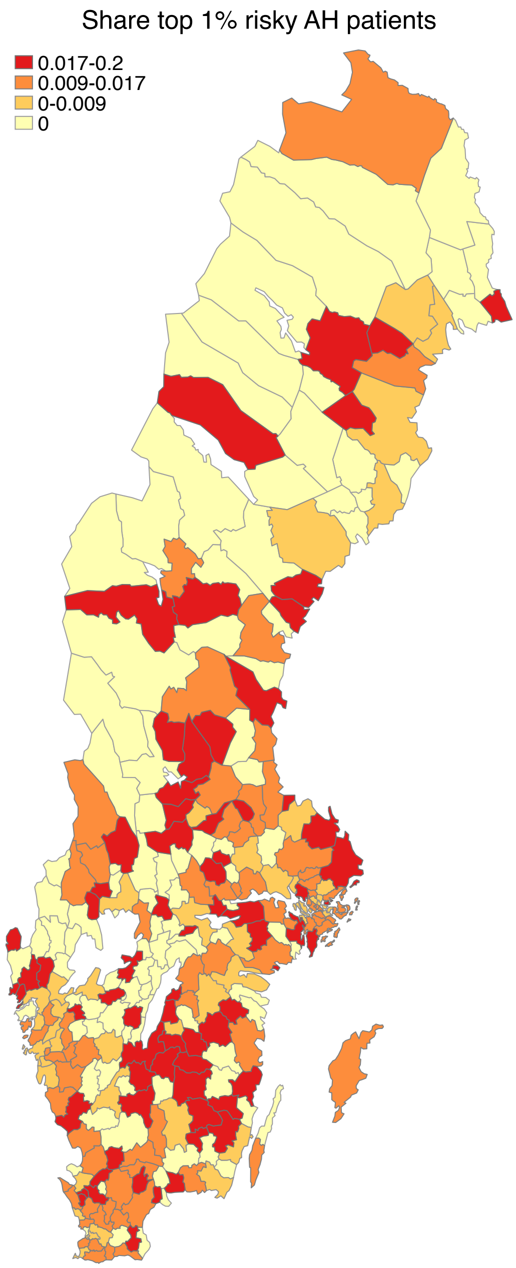
	Patient's AH risk score			Top 1 percent risky patient		
	(1)	(2)	(3)	(4)	(5)	(6)
	Clustered	Time FE	Both	Clustered	Time FE	Both
Doctor's EB AH-skill	0.000	0.000	0.000			
	(0.000)	(0.000)	(0.000)			
Top 10% doctors at AH				0.001	0.001	0.001
				(0.001)	(0.001)	(0.001)
Constant	0.016	0.016	0.016	0.010	0.010	0.010
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	95816	95816	95816	95816	95816	95816
R^2	0.000	0.000	0.000	0.000	0.000	0.000

Columns 2, 3, 5, 6 have date time fixed effects. Columns 1, 3, 4 and 6 have clustered SEs on doctors. The sample is all doctors' randomized visits after the 600th' consultation.

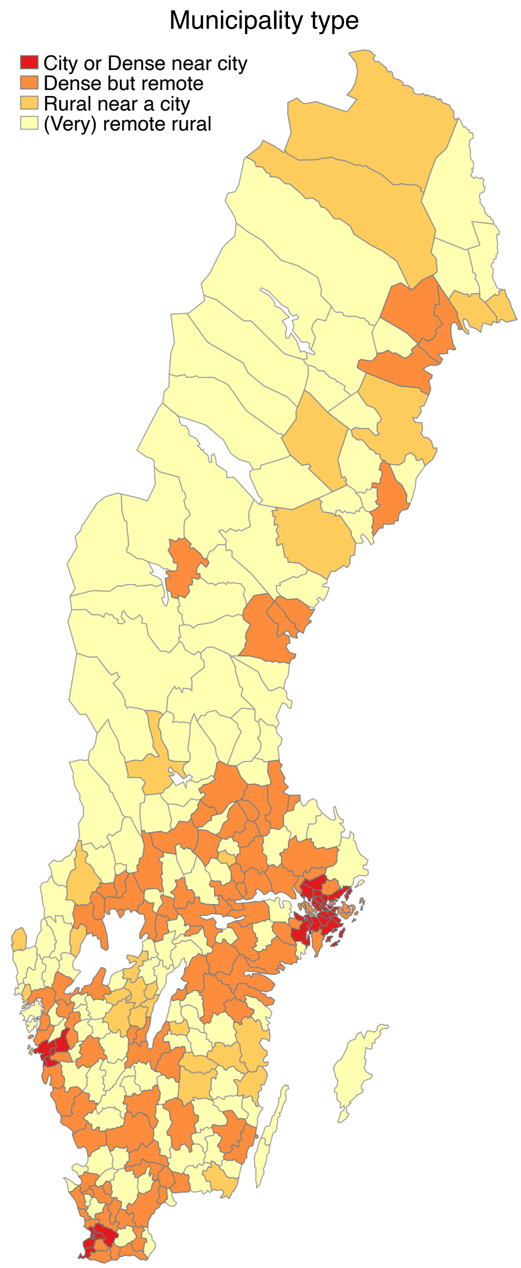
Data	**Time**
Digital care first visits	June 2016-Dec 2018
Hospital, acute & specialist	Jan 2013-Dec 2018
Prescriptions	Jan 2013-Dec 2018
Socioeconomics on adults	2013-2018
Demographics on patients	2013-2017
Primary care in 1 region	Jan 2013- Dec 2019

Table 8: Overview of data, timing and sample size.

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 in-person nurse week after	4%	11%

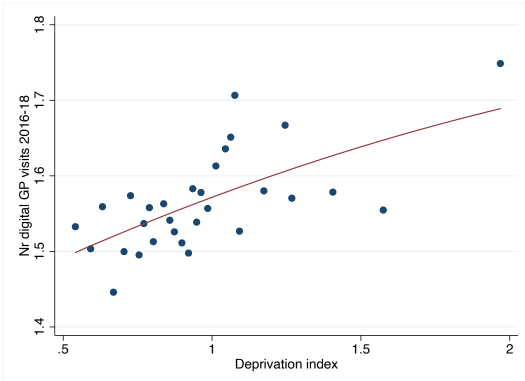


(a) Share of patients of the digital service, per municipality, who are in the top 1% of avoidable hospitalization risk.

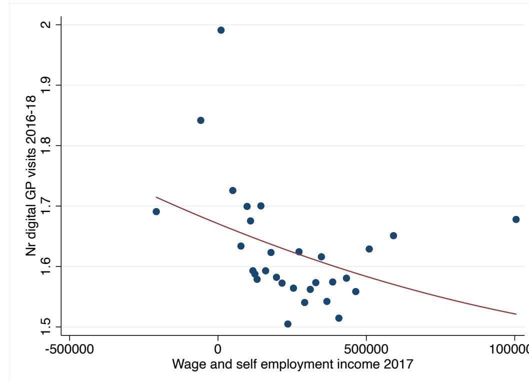


(b) Municipality type with higher type (darker) being more urban.

Figure 6: Maps of Sweden with municipalities color coded according to the subfigure captions, for a visual illustration of where the highest share of risky patients are located compared to where urban areas are located.

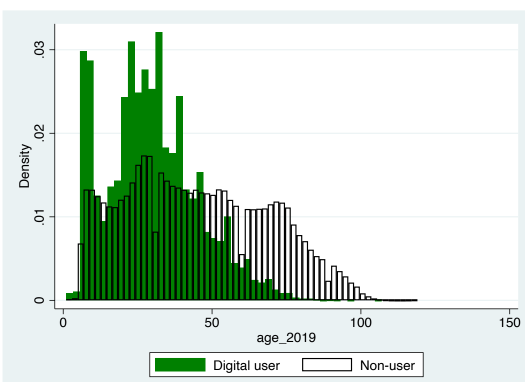


(a) Deprivation index

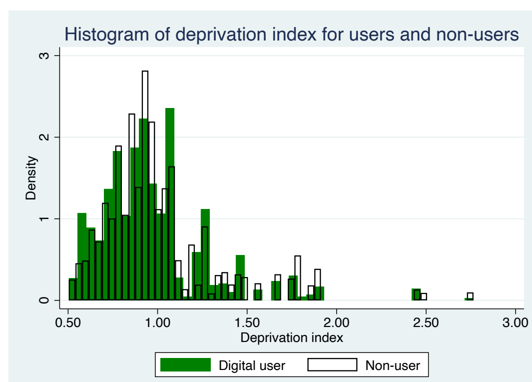


(b) Income

Figure 7: Binned scatterplot of number of digital GP visits in 2016-18 (individual level), controlling for age, against deprivation index vs. income.



(a) Histogram of ages of users and non users of digital care.



(b) Histogram of deprivation index.

Figure 8: Data from Region Skåne. Deprivation index 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.

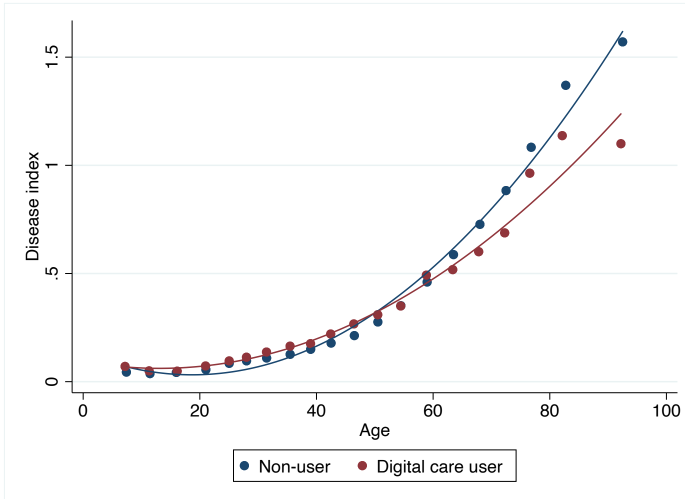


Figure 9: Elixhauser's comorbidity index using data from 2013-15 in Scania.

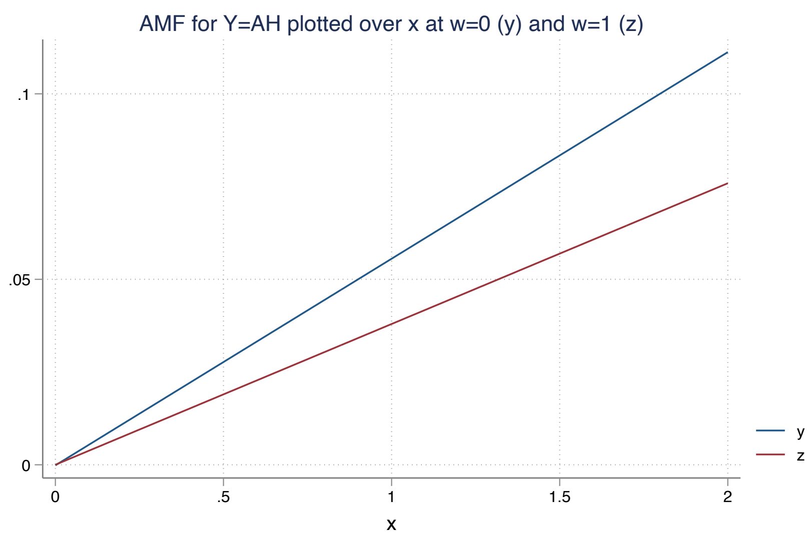


Figure 10: Illustration of the Average Match Function (AMF). The y-axis measures Avoidable Hopsitalizations (AH) and x-axis measures patient risk. W is doctor quality, where w=1 is 1 sd better than w=0, and w=0 measures the worst doctor at this outcome. The positive slopes of both graphs show that a 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.

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

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 10: 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.

Table 11: Characteristics of patients with previous avoidable hospitalizations (AH)

	(1) No AH	(2) Past AH	(3) 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 welfare benefit	0.134	0.305	0.172
	(0.340)	(0.461)	(0.006)
Disability insurance	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 y.o.)	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 and Scandinavia	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 in municipality	1,649.564	1,373.369	-276.195
	(2,062.333)	(1,951.354)	(37.415)
Observations	157,475	3,115	160,590

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.

Table 12: Disease characteristics of patients with previous avoidable hospitalizations

	(1) No AH	(2) Past AH	(3) 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

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.

Table 13: 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

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

Table 14: Counter guideline antibiotics prescription

	(1) OLS Simple	(2) Controls	(3) Bootstrap	(4) Time shift FE
Std doctor CGP-skill	-0.0028	-0.0028	-0.0028	-0.0028
	(0.0010)	(0.0010)	(0.0011)	(0.0008)
Patient's antibiotics propensity	0.0489	0.0529	0.0529	0.0526
	(0.0031)	(0.0034)	(0.0033)	(0.0033)
Interaction	-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 parentheses 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.

Table 15: Regression creating doctor AH skill

	Negative nr. avoidable hosp. 3 months after consultation
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)
Constant	0.0007
	(0.0005)
N	63576
Mean	-0.0024

With date time shift fixed effects and doctor random effects. SEs in parentheses clustered on doctors.
Sample is doctors' visits before 600th (hold-out sample). Consultations before Oct 2018 to allow 3 month follow up. Patients born before 2013, to allow 3 years pre-data. This reduces the sample from 85 000 to 64 000.
Disease index is sum of Elixhauser comorbidities.

Table 16: Nr avoidable hospitalizations 3 months after first digital visit

	(1)	(2)	(3)	(4)	(5)
	OLS Simple	Controls	Bootstrap	Time shift FE	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 parentheses 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 17: Definitive counter guideline prescription

	(1)	(2)	(3)	(4)
	OLS Simple	Controls	Bootstrap	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 parentheses 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.

Table 18: 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)
Interaction	-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 parentheses 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.

Table 19: 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 regression creates patient risk scores for AH.
Robust SEs in paraetheses.
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 20: Nr. avoidable hosp. 3 years before consultation

	(1)	(2)	(3)	(4)	(5)	(6)
Disease index	0.0733	0.0684	0.0684	0.0674	0.0666	0.0607
	(0.0070)	(0.0069)	(0.0069)	(0.0068)	(0.0068)	(0.0077)
Female	-0.0012	-0.0027	-0.0027	-0.0028	-0.0021	-0.0047
	(0.0013)	(0.0014)	(0.0014)	(0.0014)	(0.0014)	(0.0019)
Age	0.0001	0.0001	0.0001	0.0000	0.0002	0.0001
	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0001)	(0.0001)
Nr hosp 3 years before excl. AH		0.0049	0.0049	0.0047	0.0047	0.0048
		(0.0020)	(0.0020)	(0.0020)	(0.0020)	(0.0023)
2nd gen immigrant			0.0013	0.0017	0.0011	0.0011
			(0.0021)	(0.0022)	(0.0021)	(0.0028)
1st gen immigrant			0.0047	0.0044	0.0039	0.0041
			(0.0028)	(0.0027)	(0.0027)	(0.0028)
Primary school < 9 years					0.0264	0.0226
					(0.0221)	(0.0208)
Primary school 9 years					0.0020	0.0219
					(0.0031)	(0.0087)
High school					-0.0037	0.0153
					(0.0025)	(0.0063)
University < 2 years					-0.0063	0.0141
					(0.0033)	(0.0065)
University >= 2 years					-0.0101	0.0106
					(0.0027)	(0.0063)
PhD					-0.0103	0.0077
					(0.0063)	(0.0072)
Yearly income 100 000s SEK						-0.0004
						(0.0002)
Unemployed						0.0108
						(0.0070)
Constant	-0.0001	-0.0001	-0.0005	-0.0174	0.0005	-0.0125
	(0.0017)	(0.0017)	(0.0018)	(0.0036)	(0.0016)	(0.0068)
Municipality FE	No	No	No	Yes	No	No
N	95883	95883	95816	95355	95355	69200
adj. R^2	0.033	0.034	0.034	0.034	0.033	0.030

Column 3 is used to create patient risk scores for AH. Robust SEs in parantheses. Omitted education category is not finished education (child). 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 21: Explaining quality with doctor characteristics

	(1) Std CGP skill	(2) Std AH skill	(3) Std double visit skill	(4) > median in 1-3
Nr consultations (100s)	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)
Constant	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

The outcome variables in (1), (2) and (3) are standardized skill measures in preventing avoidable hospitalizations (AH), in having few counter-guideline prescriptions (CGP) and in preventing double visits. The outcome variable in column (4) is whether a doctor places above median in all three skill measures.

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

Table 22: Spearman's rank-order correlation coefficient. In parentheses: **p-value** from test of H^0: the two effectiveness measures are independent. N= 143.

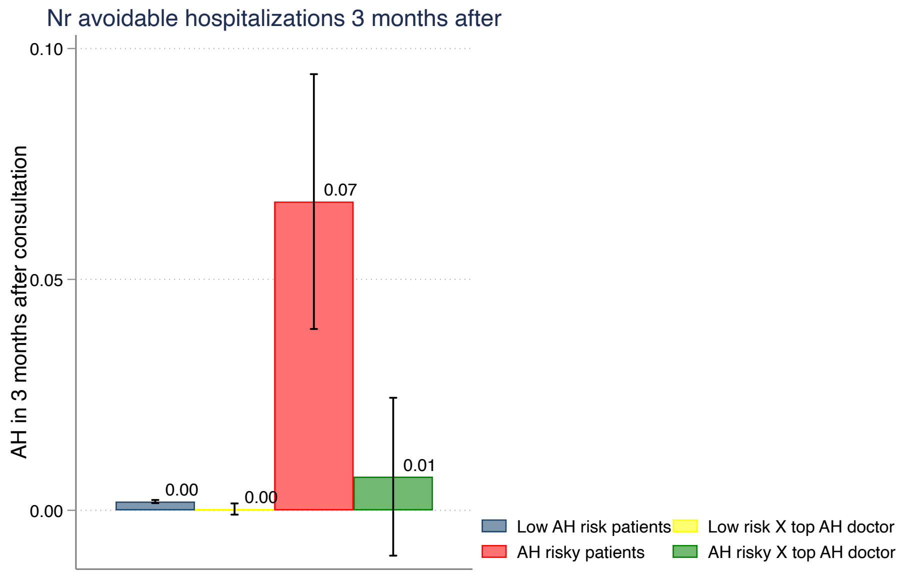
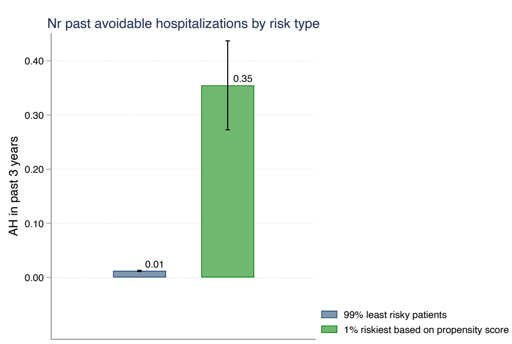
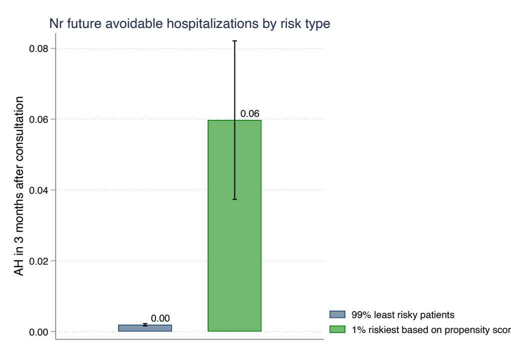


Figure 11: Bootstrap 95% confidence intervals in black.

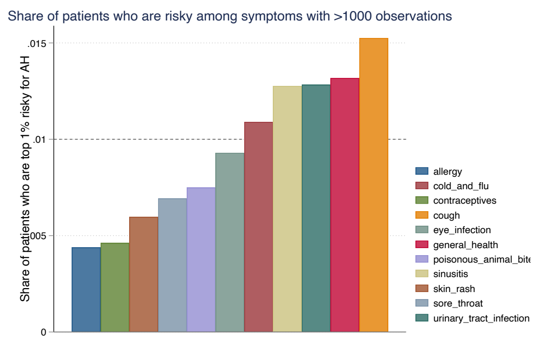


(a) Past 3 years' avoidable hospitalizations (AH).

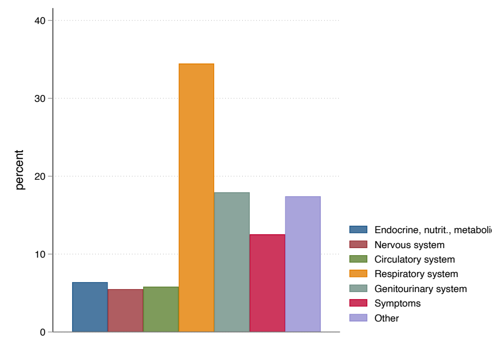


(b) Future 3 months' AH.

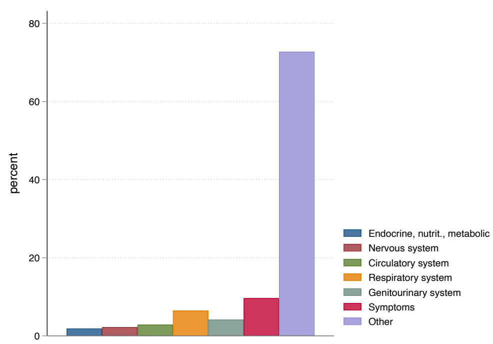
Figure 12: Patients who are predicted to be high vs. low risk for AH. 95% confidence intervals in black.



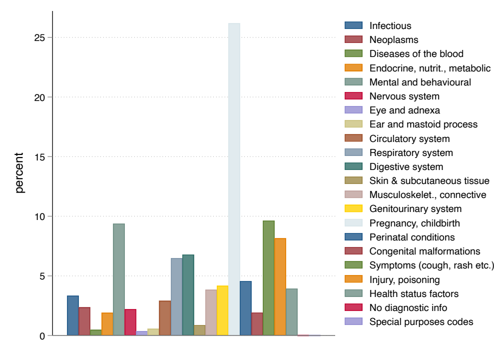
(a) Patients are asked to fill in their main symptom/reason for seeking care just before the online consultation. Exactly 1% of patients are risky for AH in the overall sample, so symptoms with over 1% risky patients are over- represented for AH-risky patients and vice versa.



(b) Grouping of the primary diagnosis code among the avoidable hospitalizations (AH) within 3 months after the digital visit.



(c) Primary diagnosis code group among all hospitalizations (not only AH) with the same diagnosis groups as in panel (b) for AH.



(d) Primary diagnosis code groups among all hospitalizations, including all diagnosis groups.

Figure 13: Comparison of which symptoms patients who are risky for avoidable hospitalizations (AH) give before an online visit (a), with the diagnosis groups that are responsible for later AH (b), compared to the diagnosis groups that are responsible for any hospitalization (c) and (d). This shows that patients who are risky for AH have symptoms in the online visit that correspond with later avoidable hospitalizations.

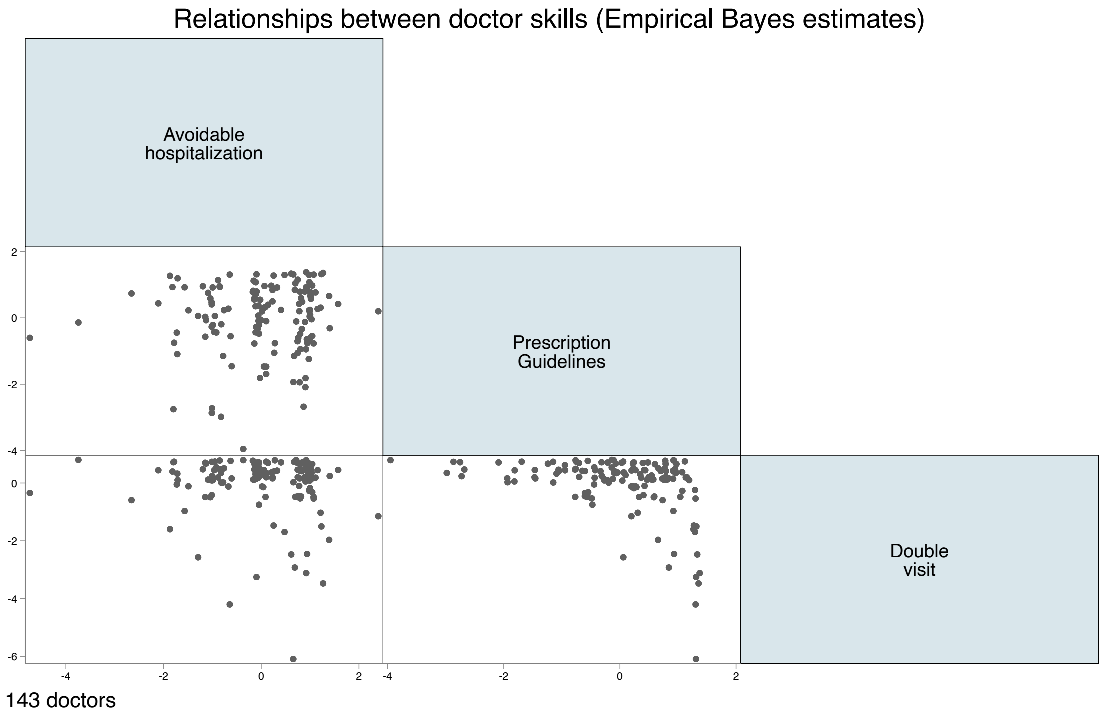


Figure 14: Scatterplots of different doctor skills.

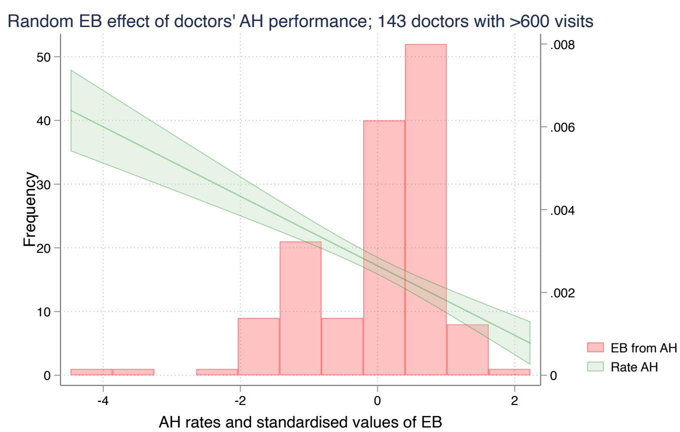


Figure 15: Histogram of the standardized (over doctors) Empirical Bayes (EB) estimates of doctor quality with the outcome *negative* avoidable hospitalizations (AH) in the 3 months after the visit (in red). Included are the 143 doctors with more than 600 randomized first visit consultations. Overlaid 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.

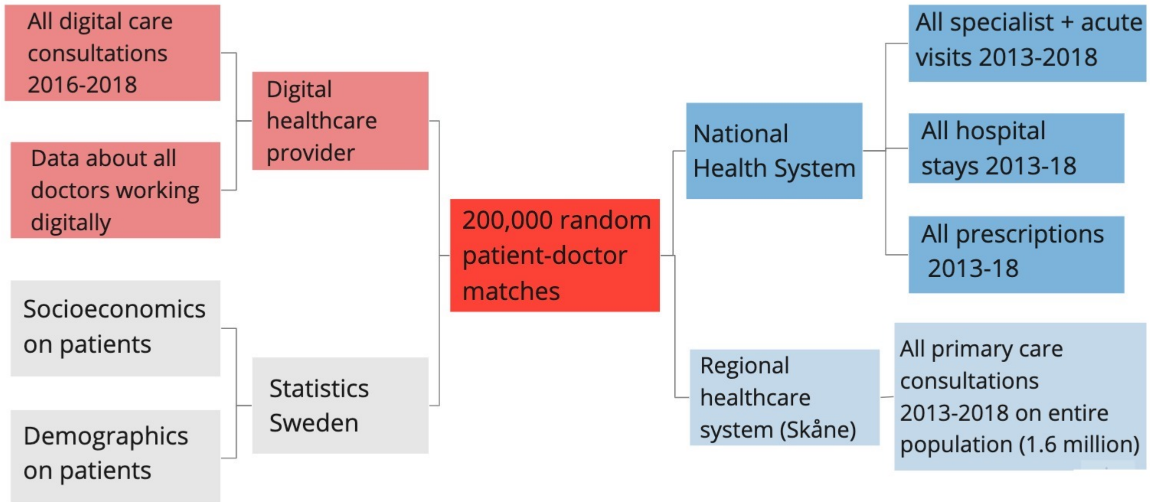


Figure 16: Illustration of data sources

B Additional Information

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.

Definition of analysis sample I start from the universe of patients who has had at least one digital consultation with one of the largest[76] providers of digital healthcare in Sweden, from the start of the service in mid-2016 to the end of 2018. There are 378,627 unique patients, who have on average has had 1.67 consultations in digital care during the sample period. There are 631,681 consultations in the dataset. I keep only the first visit for each patient, as these consultations are conditionally randomized, while there could be a concern of endogeneity in any following visits. Hence, each patient has only one observation in digital care in the analysis sample.

I match this data to official registry data from Statistics Sweden on socioeconomic

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

and demographic variables[77] 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 2013-2018. This time period covers three years before digital primary care was introduced as well as the full sample period of digital care. In this in-person healthcare dataset, there are generally many observations (consultations, hospitalizations or prescriptions etc.) per patient. In addition, I match individuals with their data on in-person primary care (2013-2019) from one Swedish region (Skåne), which matches for around 10% of the digital care sample as around 10% live in this region.

The full sample[78] now consists of all individuals (377,780) who have had an online consultation with a medical doctor at the company studied from the start of the service in 2016 until the end of 2018, and who can be matched with national registry data. I then restrict the sample to "drop in" visits, i.e., visits where the patient has no way of specifying which doctor they want to meet, but rather meet the first available doctor. This comprises 82% of the first visit sample (310,000 patients), and this is the sample where time-conditional randomization holds. Moreover, I remove pediatricians, since first-come-first-served randomization did not hold for them, as a preference was built in for pediatricians seeing small children. For the same reason, and in order to ensure I have the same amount of pre-digital care data for everyone (2013-2016), I remove small children[79] This leaves 233,489 patients and 499 doctors.

Finally, I keep only doctors who have done >600 consultations and their patients, which leaves 210,171 patients (56% of the starting sample) and 143 doctors (20% of the starting number of doctors). The reason that this reduces the number of doctors considerably 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, I need a post-digital consultation period of 3

[77]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.

[78]Going back to the national sample, not only the sample that matches with Region Skåne data.

[79]In practice, I remove all children born after 2012, since that ensures children who are in the sample are older than 3 years old at any time we observe them in digital care (which started in mid-2016). That fulfills both the condition that the remaining children do not have a pediatrician preference, and allows consistent definition of patient types according to their pre-digital in-person healthcare utilization.

months to observe whether avoidable hospitalizations happen, which means I drop all consultations which took place in October-December 2018, as the follow up data in in-person healthcare ends on 31 December 2018.

In the Statistics Sweden dataset, I 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 are all measured at the same time for all individuals, irrespective of the year when they started using the digital service.[80]

B.1.1 Important variables

The Elixhauser comorbidity score is a number between 0 and 31 which measures some important diagnoses that a patient has had in the in-person healthcare system. It counts the number of diagnoses from the following list, defined in the medical literature to be important comorbidities. They are fully listed in the Auxiliary Files, available on request. The most common of these in our sample are Obesity, Chronic Pulmonary Disease, Depression, Other Neurological Disorders and Hypertension (Uncomplicated).

B.2 Waiting times for in person primary care

In January 2019 (the closest date to my study sample available in SKR (2022)), 33% of patients could not see a doctor in person the same day across the country (in the 14 out of 20 regions which reported at this time). 19% of patients had to wait longer than the "guaranteed" maximum 3 days. The largest region, Stockholm, joined the reporting in April 2019. Looking at the distribution within this large region, the worst clinic did not fulfil the guarantee of a medical consultation with a doctor within 3 days for 48% of their meeting requests, and only 37% of patients could have an in person doctor consultation the same day they requested it (SKR 2022).

B.3 Comparison of digital care users to Swedish citizens

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).

[80] Income and employment variables are measured in 2017 and education in 2018.

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 slightly higher median income than the average citizen.

B.4 Doctors' impression of online work

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 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.

B.5 Informational requirements for reallocations

The informational requirements to carry out the reallocations consists in having access to patients' past healthcare records and some demographic data. 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 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 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.

B.6 More details on avoidable hospitalizations (AH)

Variable creation 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.

Checks to make sure the AH is related to the symptom in the online visit First, the most common diagnosis groups[81] 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 panel (b) of Figure 13 in the Online Appendix. These are conditions which are commonly treated in digital care, for instance by prescribing antibiotics for urinary tract infections.[82] Second, patients who I have determined as risky for avoidable hospitalizations based on pre-determined characteristics (i.e., not connected to the symptom at hand for the current episode) 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 panel (a) of Figure 13 in the Appendix). Moreover, I compare the diagnosis group[83] 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).

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

[82] Panels (c) and (d) in Figure 13 in the Appendix show that hospitalizations in general have a very different distribution of diagnosis groups.

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

Details on the calculation of AH costs The number of hospital days for AH was around 1 million in Sweden in 2010 (Socialstyrelsen, 2011), in a country of around 10 million inhabitants. 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. The share of AH costs out of all national health expenditures is 1.3% in Sweden[84] (and also around 1% in the United States), and the share of these (purely hospital) costs out of GDP is 0.15% in Sweden.[85]

B.6.1 More details on counter-guideline prescriptions

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 prescription.

B.7 Correlates of doctor skills

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

Female doctors are 0.7sd better at following guidelines. 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 onlyhave data on gender on 43% of doctors.

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

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

Table 23: 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)
Constant	-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

B.8 References

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.

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.

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.

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.

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

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.