Defying Distance
The Provision of Services in the Digital Age

Amanda Dahlstrand

June 14, 2023

bstract

Digital platforms are transforming services by making the physical distance between provider and user less relevant. I quantify the potential gains this flexibility o ers in the context of digital primary care in Sweden, harnessing nationwide conditional random assignment between 200,000 patients and 150 doctors. Central to the extent of gains from service-provider reallocation is the extent of skill-risk complementarity between providers and users, which has been challenging to evaluate due to sorting. I evaluate skill and risk variation as well as causal e ects 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 a ecting 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 could sever the link between local area income and service quality in favor of a needs-based assignment, potentially improving the e ectiveness and equity of service provision.

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

LSE, Microsoft Research New England, University of Zurich: amanda.dahlstrand@gmail.com. I thank Dan Zeltzer for insightful comments as discussant at NBER Health Care. I am grateful for advice and support from Nava Ashraf, Oriana Bandiera, Robin Burgess and Guy Michaels. I also thank seminar participants at Berkeley Haas, Columbia Business School, Georgetown Economics, Harvard University, 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 healthcare 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 could be *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. This will involve estimating the extent of skill-risk complementarities in healthcare, which is of independent interest. 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 less of a constraint.

In some settings, geographical distance still matters, as patients at times need to visit in person or take a test. In many settings however, including the one studied here, tests and blood work can be done at a local pharmacy or even at home (even in an in-person setting, testing is usually not carried out by your doctor but by another member of staff). Moreover, even in the cases where an in-person visit is required, this paper can shed light on the improved within-region matching that could be achieved, instead of the status quo matching based mostly on shortest local distance.

In order to overcome the endogenous selection between in-person primary care

[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.

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 sets cover approximately 200,000 patients and 150 doctors and come 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 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 rest of the (in-person) healthcare system, and to include patients' prior healthcare histories in in-person care, the data above are merged at the individual patient level with in-person healthcare data from the Swedish universal healthcare system. In these comprehensive healthcare data, patients are followed over six years, both before and after their utilizaiton of digital care, which allows me to measure both outcomes and patient pre-existing risks in terms of past diagnoses and healthcare utilization history. Finally, the data are matched at 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.

I compare counterfactual skill-need matching policies to the most relevant other policies. These benchmarks are, first, the status quo of online time-conditional random matching between doctors and patients. Second, I simulate a second benchmark of positive assortative matching on patient income and one dimension of 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]

[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 available, which is the default in many regions.

[3]Heckman and Landerso (2021) illustrate the sorting of educated families into areas with better

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 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 have different task-specific skills, and high-risk patients for certain adverse outcomes have disproportionately low income, the assignment that minimizes 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 task-specific skill.[4]

Estimating doctor ability in primary care has been challenging, as important patient outcomes are often ambiguous, rare, or delayed. 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 primary care physicians in three key dimensions of their work: (1) identifying patients who have dangerous conditions and preventing imminent adverse outcomes (2) providing guideline-consistent treatment for common conditions to minimize externalities and (3) leaving the patient informed and satisfied so that they do not seek additional, costly, care more than necessary. The first of these is the main outcome and is measured through (less) avoidable hospitalizations, which is defined in the medical literature as hospitalizations that are preventable by primary care.[5]

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] For the third doctor task, I measure whether the

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.

[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 can be seen as a preferable outcome to mortality as it is also more closely linked to the work of the primary care doctor.

[6]This is a slightly di erent type of guideline than those evaluated in recent economics literature,

patient has sought additional in-person primary care in the week following the digital care visit, for a subsample where this is measurable. 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 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 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. The doctor ability was estimated on different patients, so that it can be verified. This approach has the important benefit of being plausibly implementable by healthcare providers, by first testing doctors through a relatively small set of 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. For instance, a 1 standard deviation better doctor at following guidelines prescribes 28% less counter-guideline antibiotics. 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".[8]These specializations are usually not taken into account in the organization

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. All doctors are allowed to deviate from the guidelines in rare cases, and I compare doctors who deviate a little to those who deviate a lot with similar sets of patients, and consider the latter to be worse.

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

[8]This could be due to di erent 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.

of primary care, as a primary care doctor is expected to deal with all types of tasks. As parts of primary care is now reorganized in the digital age, these task-specific skills can be taken into account.

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%. At the same time, patients who are not predicted 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. Avoidable hospitalizations are a sign of low-quality primary care and are most common among low-income individuals.[9]

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 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

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

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 we can accurately predict who the patients at risk for avoidable hospitalizations are using a limited set of past healthcare data, and they are a small fraction of all patients. Reducing avoidable hospitalizations on aggregate reduces costs to the healthcare system or insurer, and saves lives among the risky patients, who are mostly low-income, thus reducing health inequality as a by-product.

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.[1]

Matching of service providers to users is an under-utilized policy tool, which could be welfare-improving at low cost if distance is defied by digital services. The costs would be a small increase in waiting time for some patients, and the costs of importing data and developing the matching algorithm. Matching across distances may not be feasible across all types of services, or across any distances, but this paper shows that in a setting where it is feasible across an entire country, there are large potential gains.

[1] 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 e ects (10% reduction on aggregate) for that outcome.

Other settings where it could be feasible include treatment of chronic conditions or one-off simpler primary care needs in settings such as Amazon Health (a prescription service) in the U.S., or in many integrated health systems in Europe and Asia. Algorithmic allocation means that machine prediction is used as a complement to human skill, as opposed to substitute.[11] 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 have been of independent interest also in 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.Moreover, in digital care matching can be done at an instant by algorithms that quickly access patient and doctor data. Finally, 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 of services has become widespread in many sectors. This is the first paper to study nationwide digital service provision outside of a pandemic[12] This is also the first study the potential new types of matching that digital services could

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

[12] Zeltzer et al. (2021) study the e ects of telemedicine adoption on costs and quality in Israel during the pandemic in 2020. Their aims are di erent, as they focus on providers that the patient already had a location-based relationship with.

enable. In addition, I bring a new source of conditionally random matching of service providers and clients to the literature of estimating provider skills and provider-user match effects. 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).[13] 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 medical decision making and physician performance, by studying not only doctors' overall ability, but also task-specific skill.[14] Moreover, the focus on skill-risk complementarities between doctors and patients that I evaluate here is also new to this literature. 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 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' error rate may be heterogeneous, both across physicians, and across tasks within physician. 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. That is, I focus on physician-patient skill-risk

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

[14]See, e.g., Fadlon and van Parys 2020; Cutler et al. 2019; Currie and MacLeod 2017; Abaluck et al. 2016; Doyle, Ewer and Wagner 2010. Currie and Zhang (2022) also focus on specializations by exploiting 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 e ectiveness in preventing hospitalizations, but in their setting the di erences in screening propensity are small.

complementarities.

Matching on platforms becomes more and more important. In many contexts, such as online retail, reviews are the only source of information on match quality, but they are at best a very noisy signal, and often missing or open to manipulation (Fradkin 2017). On online job boards, it is often hard to track even if a hire occurred, and even harder to follow tenure and satisfaction at the job. In this paper, I contribute to the platform matching literature by linking the matches that occurred on the platform to Swedish administrative data on medical outcomes. This means that match quality can be evaluated by the econometrician, and in the long run optimized by firms as data on outcomes can be tracked with linked electronic health records. Another way to think about the results is in terms of how much a society would loose from not linking administrative and private data, as well as not allowing matching across a larger market.

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.[15] 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 limiting waiting times are often unmet.[16] 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).

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

[16]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.

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.[17]

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).[18]

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 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.

The results of meeting an online and an in-person doctor are compared using

[17]Previous research has exploited plausible randomization between doctor teams and patients in hospital care (e.g., Doyle, Ewer and Wagner 2010) to evaluate doctor e ectiveness. 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.

[18]The primary care institutional setting 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).

an instrumental variables design, in a similar institutional setting as this paper, in Dahlstrand et al. (2023). Online care does not result in significantly more avoidable hospitalizations or prescriptions.

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.[19] 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 in areas with a higher deprivation index.[2] 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 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.

[19]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.

[2] 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. 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.

To make sure these relationships are similar across the entire country, I use aggregated public data. Appendix Table 6 indicates that patients across the country are less satisfied with their primary care in areas with lower income and higher share first-generation immigrants.[21] 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.[22] The reason that the shares of patients who meet doctors in the top 10% in the three skills is different from 10% is that they work different total hours. Patient income is the income of adult patients in 2017.

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 sorting patterns into in-person primary care, I study one Swedish region where I have

[21]Table 6 covers most of Sweden, resulting from 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).

[22]Satisfaction in the digital service actually decreases with income, opposite to the in-person results. Result available on request.

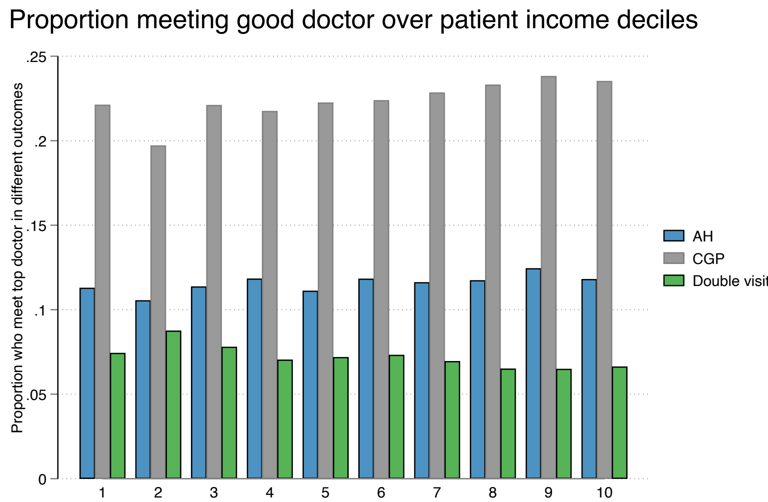


Figure 1: This figure shows what proportion of patients across income deciles who meet a doctor who is classified as top 10% in reducing avoidable hospitalizations (AH), in reducing counter-guideline prescriptions (CGP) and in preventing double visits. All income deciles have different than 10% proportion 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 total labor and self employment income of adult patients in 2017.

the universe of in-person primary care data.[23] 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 is a similar level of prior disease among digital users and non-users who are under

[23]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.

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).[24] 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.[25]

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 with a patient or with follow-up work (such as writing prescriptions), they are in

[24] 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.

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

the roster of available doctors.[26] 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

	(1)				
	mean	sd	min	max	count
Specialist	0.31	0.47	0	1	143
In specialty training	0.36	0.48	0	1	143
MD + residency only	0.33	0.47	0	1	143
doc_speaks_noneu15_lang	0.36	0.48	0	1	143
GP specialist	0.40	0.49	0	1	143
Age	36.9	7.25	28	57	61
doctor_female	0.38	0.49	0	1	138
Employed rather than contractor	0.38	0.49	0	1	52
Observations	143				

[26]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 **??** describes the characteristics doctors who are included in the study sample as they have worked at least 600 randomized consultations for the service.

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 provider of digital healthcare in Sweden, from the start of the service in mid-2016 to the end of 2018.[27] 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,

[27] It is by far the largest in terms of patient volumes in 2016-2020. By its own account it is also the largest digital care provider in Europe, operating in several other countries, often with national health insurance contracts.

but rather meet the first available doctor. This is 82% of the first visit sample, and 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).[28]

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.[29] 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.[3] Finally, I keep only doctors who have done >600 consultations and their patients, which leaves around 200,000 patients and 143 doctors. [31] Table 5 in the Online Appendix shows summary statistics of the two main samples. Patients are on average 30 years old, and 60% are female. Around 10% are born outside of Sweden, compared to almost 20% nationally. Around 6% are second-generation immigrants, which is similar to the share in the country as a whole.

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

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

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

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

[31] 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.

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: (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.[32] 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

[32]Currie and Zhang (2023) 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.

and Dow 2021) and the economics of disasters (e.g., Barro 2009).[33] Another reason to focus on this outcome is that one of the main tasks of a primary care doctor is to 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

[33] 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.

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.[34] 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. 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 Appendix.

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).[35] 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.

[34]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.

[35]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).

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 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.[36] 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

[36] A number of studies consider covering more people under insurance or changing incentives which could 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.

behavior often choose settings which are free from those to focus attention on other features.[37] 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 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

[37]See, e.g., Abaluck et al. (2020) who 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.

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) of patients to doctors.[38] 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:

$$f_{X,V,W,U}(x, v, w, u) = f_{X,V}(x, v) f_{W,U}(w, u) \tag{1}$$

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):

$$[Y_{ij}|X_i = x, W_j = w] = \iint [g(x, w, v, u) f_{V|X}(v|x) f_{U_j|W_j}(u|w)] dv du \equiv \beta(x, w)$$

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 $\tilde{W}_j$:[39]

[38] 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.

[39] $\tilde{W}$ has an equal marginal distribution to W (due to the feasibility condition) but the distribution conditional on patient attributes will di er.

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

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

$$\int \tilde{f}_{\tilde{W}_j|X}\,(w|x) f_X\,(x) dx = f(w)$$

for all $w \in W$. The distribution of patients is kept fixed, i.e. $f_X\,(x)$ is left unmodified. Average healthcare outcomes under a counterfactual patient-doctor assignment equal:

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

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, $\bar{Y}^{sq}$:

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

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 (2). I then average over the status quo distribution of X_i, which is left unchanged (the outer integral in equation (2)). 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 is to improve healthcare outcomes, under the constraint that resources are fixed.[4] 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.

[4] 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.

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.

To be realistic, I assume that the planner does not observe U_j or V_i, hence I am 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:

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

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)

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

(same workload as in Status Quo (SQ)).

where N_p = total number of patients, N_x = number of patients of type x, $N\ \ (x, w)$ = number of patients of type x that doctor w meets in any assignment π, $N_{SQ}(w)$ = 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:

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

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 $[\beta(w_J, x_I) + \beta(w_j, x_i)] \quad [\beta(w_J, x_i) + \beta(w_j, x_I)] > 0$ (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 separate 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 rst 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.[41] The employer can 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.[42]

Balance The identifying assumption both for estimating doctor skill and match effects 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 7 shows that characteristics are balanced. Another balance test is reported in Table 8, 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

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

[42] 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.

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.

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.[43] 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 conditional on time.[44] 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). It introduces a bias but this is preferred the main goal is to study the distribution of doctor skills, for instance in counterfactual reallocations, rather than the individual doctor effects. Table 15

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

[44] Year by month by day by time shift fixed e ects are always included. 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.

in the Appendix shows the regression estimating the doctor effects for avoidable hospitalization skill.[45]

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 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.[46]

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

$$P_i = C_i \quad + \upsilon_i$$

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.[47] 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

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

[46]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.

[47]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.

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.[48]

Creating binary types for avoidable hospitalizations To reduce reliance on the exact estimate of both patient risk and doctor skill, 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 $\hat{W}_{jk}^{EB}$.[49]

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 e ects: 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:

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

[49] This will give a lower e ect of the interaction e ect 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.

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

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_2 measures how different the patient group as I defined it is in the outcome variable on average.

Table 8 in the Appendix shows evidence of random assignment: that patient risks are uncorrelated with doctor skills in the main estimation sample (Sample 2).

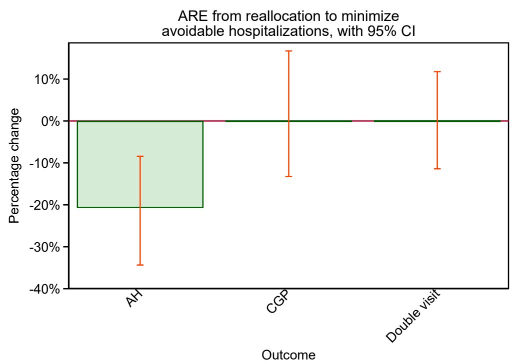
Reallocation procedures and costs The simplest reallocation procedure I carry out is to reallocate the top 10% doctors randomly to top 1% high-risk patients and let them swap doctors with some non-risky patients. The reallocation procedure where I use continuous measures of patient risk and doctor skill, 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

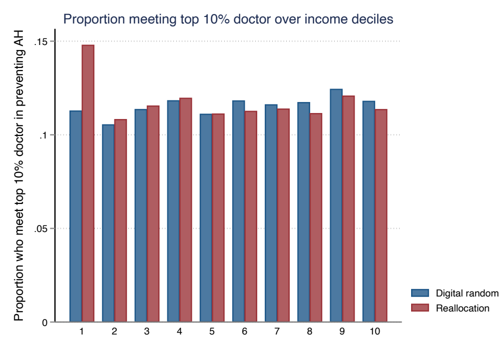
6.1 Reallocation results

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 and 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 E ects 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 (contacting in-person primary care the week after the digital visit) 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).[5] After the reallocation, the chance of meeting a top 10% doctor in risk prediction/triaging increases by 31% for the bottom patient income decile (from 11.3% to 14.8%). This is 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 12.

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.[51]

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

[5] All income deciles have a higher than 10% proportion of top doctors in AH prevention 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.

[51] 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.