INSEAD The Business School for the World®

Working Paper

2021/32/TOM

Continuity of Care Increases Clinical Productivity in Primary Care

Harshita Kajaria-Montag
University of Cambridge, hk437@cam.ac.uk

Michael Freeman
INSEAD, michael.freeman@insead.edu

Stefan Scholtes
University of Cambridge, s.scholtes@jbs.cam.ac.uk

Continuity of care, defined as an ongoing therapeutic relationship between a patient and a physician, is a defining characteristic of primary care services. However, arranging a consultation with one's regular doctor is increasingly difficult as practices seek to improve daily throughput to match rising consultation demand with an increasingly scarce clinical labor supply. The emergence of online providers accelerates this trend. We study the productivity implications of this reduction in care continuity by analyzing consultation-level data of over 10 million office consultations in 381 English primary care practices over a period of 11 years. We find that continuity of care has a significant productivity benefit. On average, the time to a patient's next visit is 13.2% (95% CI=[12.2%, 14.1%]) longer when the patient sees the doctor they have seen most frequently over the past two years, while there is no operationally meaningful difference in consultation length. The results are consistent across several model specifications that account for confounding and selection bias. The data also shows that the productivity benefit of care continuity is significantly larger for older patients, patients with multiple chronic conditions, and patients with mental health conditions. We estimate that the total demand for consultations in our sample would have fallen by 5.2% had all practices offered continuity of care at the level of the top decile of practices, prioritizing patient consultations that would yield the largest productivity benefits. We discuss operational and strategic implications of these findings for primary care practices and for third-party payers of their services.

Keywords: Healthcare; Continuity of Care; Productivity; Primary Care
History: June 16, 2021

Continuity of care increases clinical productivity in primary care

Harshita Kajaria-Montag
Judge Business School, University of Cambridge, Cambridge CB2 1AG, United Kingdom, hk437@cam.ac.uk

Michael Freeman
INSEAD, 1 Ayer Rajah Avenue, Singapore 138676, michael.freeman@insead.edu

Stefan Scholtes
Judge Business School, University of Cambridge, Cambridge CB2 1AG, United Kingdom, s.scholtes@jbs.cam.ac.uk

Continuity of care, defined as an ongoing therapeutic relationship between a patient and a physician, is a defining characteristic of primary care services. However, arranging a consultation with one's regular doctor is increasingly difficult as practices seek to improve daily throughput to match rising consultation demand with an increasingly scarce clinical labor supply. The emergence of online providers accelerates this trend. We study the productivity implications of this reduction in care continuity by analyzing consultation-level data of over 10 million office consultations in 381 English primary care practices over a period of 11 years. We find that continuity of care has a significant productivity benefit. On average, the time to a patient's next visit is 13.2% (95% CI=[12.2%, 14.1%]) longer when the patient sees the doctor they have seen most frequently over the past two years, while there is no operationally meaningful difference in consultation length. The results are consistent across several model specifications that account for confounding and selection bias. The data also shows that the productivity benefit of care continuity is significantly larger for older patients, patients with multiple chronic conditions, and patients with mental health conditions. We estimate that the total demand for consultations in our sample would have fallen by 5.2% had all practices offered continuity of care at the level of the top decile of practices, prioritizing patient consultations that would yield the largest productivity benefits. We discuss operational and strategic implications of these findings for primary care practices and for third-party payers of their services.

Key words: healthcare; continuity of care; productivity; primary care
History: June 16, 2021

Executive Summary

Problem specification: We study the relationship between continuity of care and the lengths of revisit intervals of patients in primary care practices. Continuity of care, exemplified by a patient's access to her regular doctor rather than an unfamiliar provider, is in decline as primary care practices seek to improve their clinical productivity by scaling up and standardizing their services. The adoption of online primary care consultation services, accelerated by the COVID-19 pandemic, adds further momentum to this shift from relational to transactional services. While this move to transactional services can help increase the number of consultations a doctor can handle in a day, it may have unintended negative productivity implications if patients do not receive the customized advice they need and would ordinarily receive from a doctor who is familiar with them. These patients will likely revisit sooner as their problems persist, thereby creating more demand for consultations. This paper provides evidence of this effect and estimates its size. It also explores moderating effects of patient characteristics to help practice managers identify patient segments that benefit most from care continuity. While there is much known about the benefits of primary care continuity for patient outcomes and hospital utilization, this paper is the first large-scale empirical study of the productivity effects of continuity of care in primary care practices themselves.

Practitioner audience: Managers of primary care practices are the primary audience of this study; third-party payers of primary care services are a secondary audience.

Main findings: We find that when patients have a consultation with the doctor they saw most frequently over the past two years, revisit intervals are, on average, approximately 13% longer than they are when patients see another doctor for one-off service provision. This finding is robust across multiple model specifications that account for confounding and selection bias. We also find that this effect is stronger for older patients, patients with multiple chronic conditions, and patients with mental health conditions. We do not find evidence that appointments with the regular doctor take longer. A counterfactual analysis suggests that a increase in care continuity to the top decile across practices in our sample would have led to a 5% reduction in consultations.

Practical implications: These findings have implications for primary care practice managers and third party payers of their services. First, they caution primary care managers against an overemphasis on daily throughput optimization. When throughput is increased at the expense of continuity of care, the net productivity effect can be negative, particularly for practices with older and more complex patient populations. Second, substantial productivity gains can be unlocked by focusing care continuity on patient segments that benefit most from it. Third, the strong moderating effect of patient characteristics suggests that future demand for primary care services may naturally split into two segments: (i) patients for whom transactional services are sufficient and care continuity does not provide productivity or quality benefits and (ii) patients for whom continuity is important for both service quality and productivity. The emerging online providers are likely to serve the former segment well and it will be difficult for local primary care practices to compete with these at-scale providers. A better strategy may be for local providers to double down on care continuity as their strategic priority, capturing patients who are reaching the time in their lives when such continuity becomes critical. However, if this demand segmentation becomes reality, then third-party payers need to design appropriate payment mechanisms that account for this divergence, to avoid destabilizing practices that offer the care continuity needed to manage the demand for expensive downstream healthcare resources.

1. Introduction

Primary care practices around the world are experiencing rising demand at a time when their most critical resource, primary care physician time, is becoming more scarce and more expensive. In the UK, the number of primary care physicians per 100,000 population decreased from 67 in 2009 to 60 in 2018, despite increasing demand from an aging population (Palmer 2019). The UK's Nuffield Trust estimates a shortfall of 7,000 general practitioners in the UK by 2023-24 (Beech et al. 2020), while the Association of American Medical Colleges (AAMC) estimates a shortfall of between 21,400 and 55,200 primary care physicians in the US by 2033 (Dall et al. 2020). These projections were made before the COVID-19 pandemic, and the additional pressure caused by COVID-19 will likely exacerbate these shortfalls. In response, primary care practices need to increase their clinical productivity.

Within a primary care practice, clinical productivity has two main dimensions: the number of consultations that a clinician performs in a day and the extent to which a clinician is able to extend the time to a patient's next consultation. Much emphasis is currently put on the former – increasing daily throughput – not least because it directly improves on-the-day access for patients. This focus on daily throughput has an unintended but important consequence: It becomes more difficult for patients to arrange a consultation with their regular doctor (Kajaria-Montag and Freeman 2020). In 2009, 77% of respondents to the UK's annual GP Practice Survey reported being able to see their preferred doctor at least most of the time. Ten years later, in the last survey prior to COVID-19, this proportion had dropped to 45% (Institute for Government 2019).

Traditionally, general practitioners and their patients have valued the sustained, trust-based therapeutic relationships they form over time (Liu et al. 2018). Through repeated interactions, family doctors become familiar with the patient's holistic medical needs as well as with their preferences, behaviors, and socioeconomic circumstances, and will customize their advice accordingly. Continuity of service in primary care is also known to be beneficial for both patients and the wider health system. A recent evidence review by the UK's Nuffield Trust concluded that "relational continuity of care in general practice is associated with [...] better clinical outcomes for an array of conditions; reduced mortality; better uptake of preventative services; better adherence to medication; reduced avoidable hospital admissions; and better overall experience of care among patients who prefer continuity and are able to obtain it" (Palmer et al. 2018).

Given the strong evidence of its patient health and system benefits, it is surprising that relatively little is known about the effect of care continuity on the productivity of primary care practices themselves. It is particularly important to fill this gap in the literature given the current trends in primary care provision. Many primary care practices are responding to the workforce crisis by doubling down on daily throughput improvements, e.g., through standardizing and shortening the length of patient consultations or grouping physicians into rapid access teams that serve on-the-day patient demand in a round-robin fashion akin to an emergency department. This makes care continuity harder to achieve (Sampson et al. 2008), suggesting that practices must trade off clinical productivity and care continuity as competing goals.

Yet our analysis shows that this trade-off is illusory. In particular, our data indicates that care continuity *increases* a physician's productivity by extending the interval between patient visits. Throughput-enhancing interventions will therefore have unintended productivity-reducing consequences if they decrease continuity of care. From a more positive perspective, this also means that there are untapped productivity gains to be had from a targeted increase in continuity of care.

To empirically demonstrate the productivity implications of care continuity, we use a sample containing over 10 million face-to-face consultations between over 14,000 primary care physicians and 1.8 million patients in 381 English primary care practices over a period of 11 years. For each consultation, we identify the patient's regular doctor as the doctor who had the most frequent interactions with the patient over the past two years. We then analyze whether a patient's revisit interval (i.e., the time between the focal consultation and the patient's next visit) differed when the consultation was with this regular doctor or with another doctor in the practice. Using a range of empirical methods to control for potential selection and omitted variable bias, we find robust evidence that the revisit interval is extended by 13.2% (95% CI=$[12.2\%, 14.1\%]$) if the patient sees her regular doctor. At the same time, we find no evidence that the patient's consultations with the regular doctor are longer than her consultations with other doctors in the practice. In fact, the data suggests that the consultation time with the regular doctor is shorter on average. Therefore, care continuity has a net positive overall effect on a primary care physician's productivity because it substantially extends the average time to the patient's next visit.

Having established the main effect, we then analyze the heterogeneity of the effect and show that the productivity benefit of care continuity is more pronounced for patients with more complex needs, specifically older patients, patients with chronic diseases and patients with mental health conditions. Building on these findings, we demonstrate how our estimation methods can be used as a scoring tool to enable practice managers to target relational services at those patients for whom care continuity has the most productivity-enhancing effect. We apply this scoring method retrospectively to the data to estimate the potential for demand reduction. The model suggests that if all practices in the data had offered well-targeted continuity of care for 75% of their consultations, a level similar to the 90th percentile across all practices in the data, then the overall demand for consultations would have been 5.2% lower than the realized demand.

2. Literature Review

This paper contributes to the healthcare operations and medical literature on continuity of care. These literature streams have largely focused on the consequences of care continuity on patient outcomes and secondary care utilization, such as emergency visits or hospital admissions, typically for patients with specific conditions. In contrast, we are concerned with the effect of care continuity on the clinical productivity of primary care practices themselves, averaged across all patient consultations as well as for patients with specific conditions.

2.1. Continuity of care in healthcare operations

Ahuja et al. (2020b) investigate the effect of providing continuity to patients with diabetes and find that continuity improves three important system utilization metrics: inpatient admissions, hospital length of stay, and readmission rate. In a follow-up study, Ahuja et al. (2020a) partially explain this relationship between continuity and the system utilization metrics by showing that continuity can lead to higher rates of medication adherence and consequently to lower glycemic variability. Senot (2019) also studies the effect of continuity on secondary care usage. Specifically, the study follows the journey of heart failure patients over a one-year period and finds that the continuity of the individual referring provider (along with continuity of the physical location and the accountable care organization) contributes to a reduced risk of hospital readmission for heart failure patients. Queenan et al. (2019) find that providing technology-enabled continuity coupled with increasing patient engagement in their own health reduces hospital readmissions.

The present study complements this research on system utilization effects by focusing on productivity effects within the primary care setting itself. This is important because the direction of the internal productivity effect tells us whether or not practices need to be externally incentivized to provide the continuity of care that will create the documented system utilization benefits.

The healthcare operations literature has hitherto not engaged much with primary care productivity, though Bavafa et al. (2018) is one exception. Their paper focuses on the impact of complementing office visits by e-visits on demand in primary care, and the authors show that the introduction of this new communications channel increases demand for office visits. We also address how demand for primary care services changes as a function of the service provided, with our focus being the demand-inducing effect of reduced care continuity. While Li et al. (2021) also focus on telemedicine adoption, they do so in an outpatient context. They show that the adoption of telemedicine reduces productivity by shortening the interval between patient visits in the short term, but the interval between visits increases in the long run. We also use the revisit interval as a measure of productivity in this paper, but our focus is on the productivity implication of care continuity in primary care.

This paper also relates to the healthcare operations literature on clinical decision making. In the context of an emergency department (ED) where medical decisions and patient flow decisions are handled by different decision makers, Song et al. (2015) investigate the effect of combining these responsibilities and making them the responsibility of a sole clinical decision maker. The authors show that introducing a single decision maker leads to a significant reduction in length of stay in the ED, an effect they attribute to increased ownership over patient flow and better alignment of incentives. Although our context differs in its focus on repeat visits and revisit intervals, we borrow some of the motivational lenses when developing the hypotheses.

2.2. Medical literature on care continuity

The medical literature differentiates between different types of care continuity, with Haggerty et al. (2003) providing a comprehensive review of the various forms, e.g., relational continuity, management continuity, and informational continuity. Relational continuity is the most studied of these – in fact, the terms relational continuity and continuity of care are often used synonymously in the existing literature – and is typically defined as the ongoing therapeutic relationship between a patient and one or more providers. In this paper, the focus is on this relational component of continuity, as captured by repeat consultations with the same primary care physician.

It is well documented that continuity of care in primary care is valued by patients and doctors alike, with surveys of the extant medical literature highlighting various benefits of providing care continuity (The King's Fund 2016). In terms of direct benefits to patients, the medical literature has demonstrated various health benefits and improved management of health conditions for those who receive care continuity. For instance, studies have shown improvements in quality of life outcomes (Drury et al. 2020, Chen et al. 2017, Ye et al. 2016), blood pressure for diabetic and hypertensive patients (Leniz and Gulliford 2019), mortality (Maarsingh et al. 2016, Cho et al. 2015), adherence to medication plans (Dossa et al. 2017), and the likelihood of filling risky prescriptions (Hallvik et al. 2018).

In terms of indirect benefits, a meta-analysis by Huntley et al. (2014), involving participants from OECD countries, found that unscheduled secondary care usage is highly influenced by care continuity in the primary care setting. For example, primary care continuity has been associated with reductions in emergency department visits (Pourat et al. 2015) and unplanned hospitalizations of patients with ambulatory care sensitive conditions (Barker et al. 2017). Such advantages have been consistently demonstrated across different patient populations, including patients with serious mental illness (Ride et al. 2019), dementia (Amjad et al. 2016), COPD (Lin et al. 2015), and diabetes (Worrall and Knight 2011, Dossa et al. 2017), as well as older patients (Tammes et al. 2017, Katz et al. 2015, Bayliss et al. 2015, Nyweide et al. 2013). We contribute to this stream of literature by demonstrating that care continuity also affects the need for primary care visits themselves and that this effect is particularly pronounced in older patients and patients with complex conditions, such as chronic diseases or mental illnesses.

In summary, there is rich evidence to show the benefits of relational continuity for both patients and the wider health system in terms of reduced utilization. It is therefore somewhat surprising that the effect of continuity of care on clinical productivity within primary care practices has not yet been investigated. This study expands existing knowledge of the effects of care continuity by showing that care continuity not only improves outcomes and system utilization but also enhances the productivity of the primary care physicians themselves.

3. Hypothesis Development

Before we develop the paper's hypotheses, we clarify the key variables, care continuity and clinical productivity. The concept of relational continuity of care refers to a sustained therapeutic relationship between a patient and a doctor and is epitomized by the notion of a "patient list" or "patient panel" that many primary care physicians hold, either formally or informally (Wilkin and Metcalfe 1984, Tammes et al. 2017). These are the patients for whose health and care the doctor takes personal responsibility over a prolonged period of time. The units of observation in this study are individual consultations, and we distinguish between appointments with a patient's regular doctor, who provides relational continuity and relational service, and appointments with another doctor providing one-off services, who we call a "transactional provider".

Productivity is commonly defined as a measure of output per unit input (Reinhardt 1972). While the input measure for clinical productivity is clear – the clinician's time – defining an appropriate output measure is more difficult. On the one hand, the core activity of a primary care physician is providing consultations to patients. Accordingly, the average consultation duration or, equivalently, the number of patient consultations per clinician day is one relevant measure of productivity. On the other hand, such a daily throughput metric does not account for the recurrent nature of primary care services. A fast but ineffective patient consultation may create the need for more consultations in the future, while a longer but more thorough consultation may alleviate this need. Therefore, a slower doctor may actually be more productive overall. Consequently, to account for the recurrent nature of primary care services, we must consider two productivity metrics: (i) the duration of the visit, as a measure of throughput, and (ii) the revisit interval, i.e., the time between patient visits, as a measure of productivity-relevant service quality.

We note that in some contexts it might be highly salient to explore the relationship between visit duration and continuity of care. This relationship is determined by two countervailing factors. On the one hand, a doctor has an incentive to take more time to treat her regular patients thoroughly: this will reduce her future workload by preventing revisits, which would likely be her responsibility (Jeffers and Baker 2016). On the other hand, she is more familiar with these patients than a transactional provider and can therefore be more economical in her collection of information (Hill and Freeman 2011, Rosen et al. 2020).

However, given that this study uses UK data, we do not expect much difference in the duration of patients' consultations with their regular doctor and those with a transactional provider. This is because although it is not a formally imposed guideline, 10 minutes has become the de facto standard for primary care consultations in the UK (Royal College of General Practitioners 2019). For this reason, while we do provide exploratory evidence on the relationship between care continuity and consultation duration in the data in Section 6.4, we focus the hypothesis development on the relationship between care continuity and the revisit interval.

3.1. Hypotheses

From a productivity standpoint, there are three critical differences between a patient's regular doctor and a transactional provider. First, a transactional provider shares the additional workload of a potentially avoidable revisit with all other doctors in the practice, while the patient's regular doctor will likely bear this future workload directly. As previously mentioned, this provides a stronger incentive for the regular doctor to provide service in a way that reduces the likelihood of a revisit and therefore her expected future workload. Second, the patient's regular doctor has better information about the patient than a transactional provider and may therefore be able to provide more effective customized advice and treatment that reduces the need for a revisit in the short term (Hjortdahl and Borchgrevink 1991). Third, the regular doctor has typically established a trust-based relationship with her patient, which provides rapport advantages and enables her to influence the patient more effectively (Tarrant et al. 2010, Hill and Freeman 2011). If the consultation is performed by the regular doctor, the incentive differential, information benefits, and relationship advantages interact in several ways to prolong the time to the next visit.

Specifically, the stronger incentive to reduce the likelihood of follow-up consultations makes the regular doctor more likely to diagnose the patient's problem carefully in an attempt to "get it right the first time" (Koopman et al. 2003). From a time and productivity perspective, such a root-cause diagnosis may not even cost the doctor very much if she is thoroughly familiar with the patient. By contrast, a transactional provider has no particular incentive to go beyond the basic alleviation of a patient's presenting symptoms, knowing that any follow-up work is likely to be performed by a different doctor – an effect sometimes referred to as the "collusion of anonymity" (Freeman et al. 2010, Balint 1955). Furthermore, if a diagnosis is necessary, the transactional provider, being less familiar with the patient, is likely to need more time to arrive at a careful diagnosis. Given the time pressure of a full appointments schedule, a transactional provider is therefore more likely to adopt a trial-and-error approach to diagnosis and treatment, which increases the likelihood of a return visit (Bobroske et al. 2021).

HYPOTHESIS 1. *If a patient is seen by her regular doctor, she will have a longer time interval to her next appointment than if she is seen by a transactional provider.*

A patient's regular doctor does not just have a stronger incentive to "get it right the first time" – they also have an incentive to leverage any patient encounter to explore the patient's health needs beyond the problem immediately at hand, as this may prevent an unnecessary visit in the near future (Hill and Freeman 2011). Thus, a regular doctor may check her notes and proactively deal with multiple illnesses or health issues in a single appointment. Transactional providers, by contrast, lack both the incentive to expand their scope of service beyond the immediate clinical need expressed by the patient and the holistic patient knowledge that facilitates proactive management of the patient's health. Opportunities for such proactive interventions, outside the scope of the immediate reason for the visit, are particularly salient when the patient has a chronic disease (Koopman et al. 2003, Goodwin et al. 2010). We therefore expect the productivity benefit of seeing a regular doctor to be larger for patients with multiple chronic conditions.

HYPOTHESIS 2. *If a patient is seen by her regular doctor, she will have a longer time interval to her next appointment than if she is seen by a transactional provider, and this effect will be larger for patients with multiple chronic diseases.*

A regular doctor forms an impression of a patient's health over time and will therefore respond not only to the patient's current health status but also to *changes* in their overall health. This longitudinal impression constitutes important additional information that a one-off transactional provider does not have (O'Connor et al. 1998, Ramanayake and Basnayake 2018). Indeed, the primary care physicians we interviewed for this study confirmed that when one of their regular patients enters the consultation room, they are likely to know immediately whether or not the patient is seriously ill and that they cannot do this with one-off patients. This ability to consider change in addition to state is particularly relevant for older patients, where single impressions of patient health can vary widely across patients and more information is gained from an observed change in an individual's health status. Being able to identify these changes and act early should allow the patient's regular doctor to reduce the need for future visits.

HYPOTHESIS 3. *If a patient is seen by her regular doctor, she will have a longer time interval to her next appointment than if she is seen by a transactional provider, and this effect will be larger for older patients.*

Through repeated interactions over time, the regular doctor not only gains a comprehensive understanding of the patient's needs but also forms a trust-based relationship that may help her communicate more effectively with the patient (Tarrant et al. 2010). The patient may feel more comfortable sharing information with her regular doctor and, as a consequence, the doctor will be able to design a more appropriate treatment plan. Another advantage of this rapport is that patients are more likely to comply with the doctor's instructions, making the treatment plan more effective and helping to prevent follow-up visits (Dossa et al. 2017). We expect the rapport advantage to be particularly important for patients with mental health conditions, such as anxiety, depression and schizophrenia (Biringer et al. 2017). The stigma that is still associated with mental health concerns might make it difficult for such patients to be fully transparent with an unfamiliar physician. Moreover, medication compliance is a particular problem for mental health patients, due in part to the negative side effects associated with commonly prescribed medications (e.g. weight gain, fatigue, and reduced libido) (Semahegn et al. 2018).

HYPOTHESIS 4. *If a patient is seen by her regular doctor, she will have a longer time interval to her next appointment than if she is seen by a transactional provider, and this effect will be larger for patients with mental health conditions.*

4. Clinical Setting, Data and Variables

In this section, we first provide a brief overview of the specifics of the UK primary care context that are relevant for this study. We then describe the dataset in detail and conclude with the description of the dependent variables, independent variables, and controls to be included in the analysis.

4.1. Primary care context

Although the English National Health Service (NHS) is publicly funded through taxation, primary care practices are privately owned small or medium-sized businesses, typically organized as partnerships of primary care physicians. Unlike hospitals, they are independent contractors with the NHS and therefore not under its direct control. Instead, the NHS controls their services through standard contractual arrangements. A typical practice has 8,000-10,000 registered patients, 4-5 full-time equivalent physicians, and a small number of other healthcare workers and administrative staff. Practice income is largely capitation-based, adjusted for demographic and socioeconomic characteristics of the registered practice population and geography. In 2019, a typical practice with 9,000 patients received an income of £1.3M, approximately £155 per registered patient or £32 per consultation (NHS Digital 2015).

The contract of a general practice in England stipulates a geographical catchment area for the practice. Patients who live in this area have the right to register with that practice. Patients may apply for registration with any practice but practices have discretion to accept or deny out-of-area patients. Importantly for this study, patients can only register with a single practice in England and are automatically deregistered when they register with a new practice. Since our data is practice-based, we therefore have full visibility of all primary care appointments of patients for their period of registration in the study practices.

Primary care services are free at the point of care. Patients can request to see any doctor at their practice, and practices generally try to ensure that a patient can see a preferred doctor. The NHS contract requires that each patient registered at a practice is assigned to a "named doctor," who is responsible for ensuring that the patient's needs are met (Tammes et al. 2017). However, practices may regard this as a purely administrative requirement, so the patient's "named doctor" may not be her regular doctor. This study focuses on the patient's regular doctor, the physician who has seen the patient most frequently in the past.

Consultations can be face-to-face, over the phone or video link or, in rare cases, at the patient's home. Appointments are generally booked via the phone with a receptionist. Practices have to accommodate routine and urgent appointments, the latter requiring on-the-day access. Some practices reserve a number of appointment slots for urgent services and, if those slots are booked, refer patients to the emergency department or ask them to call again the next day. Other practices accept all patients who call in before a certain cut-off time or offer unlimited access for acute care throughout the day. These practices have "duty doctors" dedicated to serving urgent care patients, typically in a round-robin fashion. Practices may also operate a phone-based triage system, with a doctor ascertaining whether a patient needs same-day service. While our data does not include the time when an appointment was made and therefore does not allow us to accurately distinguish between urgent and routine appointments, we are able to use markers, such as antibiotic prescriptions, that are more commonly associated with urgent appointments to help us distinguish the two appointment types.

4.2. Data and sample

In order to understand the effect of a patient seeing her regular doctor in one consultation on the time to her subsequent consultation, we perform a cross-sectional analysis with individual patient consultations as observations. We obtained consultation-level data from the UK Clinical Practice Research Datalink (CPRD). This database consists of anonymized electronic medical records covering over 11.3 million patients (6.9% of the population of the UK) across 674 practices in the UK; it is representative of the population in terms of age, sex and ethnicity (Herrett et al. 2015). The database encompasses a wealth of information about patients, visits, providers, diagnoses, prescriptions, referrals, treatments, immunization records, and test records. It can be linked to several other data sources such as secondary care services, enabling a fairly complete medical record to be obtained for a given patient.

We obtained data for all English practices that had consented to linkage to secondary care usage data. The restriction to English practices improves the homogeneity of the sample as the national health systems operate differently in the four constituent countries of the UK and there are differences in their standard primary care contracts. The starting data set comprised information on 370,890,526 primary care consultations corresponding to 5,475,342 patients. The analysis sample was derived from this data using the inclusion criteria described in Section EC.1 of the e-companion and summarized in Table 1.

Specifically, the sample for analysis consists of face-to-face consultations between primary care doctors and patients over the age of 18 that took place between January 2007 and December 2017 at times during which (i) a practice's data is deemed to be of research quality, (ii) the patient had sufficient past appointments to identify their regular doctor, and (iii) the patient's regular doctor was available to see them. Data was also filtered to exclude patients visiting more than once per week on average, to only include data from periods in which the patient was continuously registered at a practice, and to exclude any consultations lacking an observable revisit interval. (For more justification on the inclusion criteria, see Section EC.1 of the e-companion.)

The final sample thus consists of 11,344,065 consultations between 1,883,626 patients and 14,123 doctors in 381 practices across 11 years between 2007 and 2017.

Table 1 Data and sample inclusion criteria

Criteria	Patients	Consultations
Doctor consultations in 407 primary care practices	5,475,342	370,890,526
Face-to-face doctor consultations only	5,335,945	200,312,789
Consultations after date at which practice data is of research quality	5,037,650	161,556,335
Consultations during a patient's continuous registration period	4,921,208	139,455,412
Consultations at which the patient was over 18	3,855,445	86,399,813
Consultations with ≥ 3 and ≤ 104 consultations in the preceding 2 years	2,952,445	71,797,380
Consultations occurring after the patient's first two years following registration	2,537,781	63,087,124
Only consultations with a valid revisit interval	2,410,189	60,894,300
Only consultations between January 2007 and December 2017	2,322,773	51,711,037
Consultations occurring when the patient's regular doctor is available	2,273,571	45,376,070
Random sample of consultations from 381 remaining practices	1,883,626	11,344,065

4.3. Variable description

4.3.1. Dependent variable. The main dependent variable throughout this study is the patient's revisit interval (RI), which is defined by the time elapsed between the focal face-to-face consultation with a doctor and the next face-to-face consultation with a doctor, and is measured in days. The hypotheses posit that for a consultation with the patient's regular doctor, the revisit interval will be longer than it is for a consultation with a transactional provider. As is usually the case with durations, the distribution of revisit intervals is right-skewed. We therefore transform the variable to a logarithmic scale by taking it's natural logarithm. Figure 1 shows the distribution of the log-transformed revisit interval, and Table 2 contains summary statistics. In the infrequent event that a patient has multiple face-to-face consultations with doctors on the same day, we set the revisit interval length to 0.5 days prior to log transforming.

4.3.2. Independent variables. The main independent variable is a binary variable, which indicates whether or not the focal consultation was with the patient's regular doctor. Since this study spans an 11-year time horizon, the patient's regular doctor may change over time for various reasons. For example, a doctor may retire or leave the practice, or the patient may switch due to a change in circumstances or a positive experience with another doctor. When we determine a patient's regular doctor we therefore use a dynamic measure rather than identifying one fixed dyad for each patient.

Specifically, we consider a two-year time window over which we calculate the patient's regular doctor. For consultation i we define a patient's regular doctor as the doctor with whom the patient had the most face-to-face consultations over the preceding two year. To break ties, we choose established over unestablished doctors[1] and, if the tie persists, we choose the doctor who the patient

[1] The data allow us to distinguish between two types of physicians, *established* physicians who have a contract with

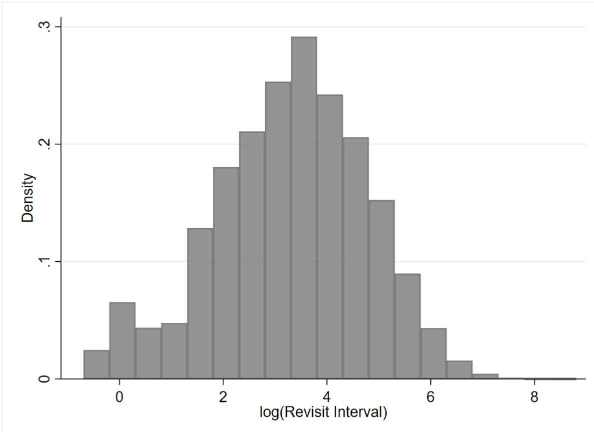


Figure 1 Distribution of ln(Revisit Interval)

	ln(Revisit Interval)
Mean	3.29
Median	3.33
Min	-0.69
Max	8.34
Std. Dev.	1.48

Table 2 Descriptive statistics for the dependent variable

saw most recently. The independent variable RD_i is thus a binary variable that is equal to 0 for consultation i when the patient does not see her regular doctor and 1 for consultation i if the patient does see her regular doctor. Overall, 50% of consultations in our analysis sample occur between a patient and their regular doctor. We provide additional information on the independent variable and the patient-level factors that affect it in Section EC.2 of the e-companion.

4.3.3. Control variables. The relationship between continuity of care and a patient's revisit interval is likely to be confounded by patient demographic factors, patient's consultation and revisit interval history, attributes of the regular doctor, temporal factors, and practice-level factors. To account for these potential confounds, we use several control variables. A summary of the controls is provided in Table 3 and an explanation for each of the controls is given in Section EC.3 of the e-companion.

5. Econometric Specifications

5.1. Ordinary least squares estimator

To identify the effect of a patient's seeing her regular doctor on her revisit interval (Hypothesis 1 (H1)) in a robust way, we estimate a series of consultation-level models using the revisit interval RI_i as the dependent variable and the indicator RD_i for the regular doctor as the main explanatory variable, where the index i refers to a consultation. We begin with a standard OLS model:

$$ln(RI_i) = \beta_0 + \beta_1 RD_i + \mathbf{X}_i \boldsymbol{\beta} + \epsilon_i \ , \tag{1}$$

where the vector $\mathbf{X}_i$ specifies the set of controls corresponding to consultation i (as defined in Table 3) and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ is the error term. We cluster standard errors at the patient level to account

the practice and *unestablished* physicians who are not permanent employees of the practice but may be self-employed or employed with an agency. The latter work on an ad-hoc basis, often in multiple practices, and are paid on an hourly basis.

Table 3 Table of controls

Variable	Type	Description
Patient demographics		
Age	Categorical (14)	Age of the patient at the time of consultation, split into age bands (18-25, 26-30, 31-35, ..., 81-85, 86+)
Number of comorbidities	Categorical (6)	Number of comorbidities at the time of consultation, calculated using the Cambridge Comorbidity Index (CCI), split into bands (0, 1, ..., 4, 5+)
Individual comorbidities	Binary (26)	For each of the 26 comorbidities as defined by the CCI, a variable to indicate whether the patient suffers from that comorbidity at the time of the consultation
Mental health	Binary	A variable to indicate whether the patient suffers from a mental health condition as defined by the CCI
Gender	Binary	A variable which is 0 if the patient is female and 1 if the patient is male
Index of Multiple Deprivation	Categorical (5)	The deprivation level assigned to the patient
Prescriptions	Categorical (10)	The number of repeat prescriptions the patient is prescribed within the 6 months preceding the focal consultation, split into bands (0, 1, 2, 3, 4-5, 6-7, 8-9, 10-12, 13-15, 16+)
Patient's past history		
Past consultation frequency	Categorical (20)	The total number of consultations in the 2 years preceding the patient's focal consultation, split into 20 bands: 3-10 consultations (8 categories of size 1), 11-20 consultations (5 categories of size 2), 21-26 consultations (2 categories of size 3), 27-30 consultations (1 category of size 4), 31-35 consultations (1 category of size 5), 36-55 consultations (2 categories of size 10), 56+ consultations (1 category).
Past revisit interval	Continuous	The 2-year past average revisit interval of the patient calculated as described in Section EC.5.3 of the e-companion.
Attributes of the regular doctor		
Established doctor	Binary	A variable to indicate if the patient's assigned regular doctor for the focal consultation is an established or unestablished doctor (see Footnote 1)
Temporal factors		
Year	Categorical (11)	Year during which the consultation took place (2007-2017)
Month of Year	Categorical (12)	Month of the year in which the consultation falls (January-December)
Day of Week	Categorical (7)	Day of the week in which the consultation took place (Monday-Sunday)
Practice-level factors		
Practice-level demand	Continuous	Total practice demand during the focal week of each consultation, standardized by a weekly average in a 52-week period around the focal week
Practice	Categorical (381)	The practice at which the consultation took place

Notes: If a variable is categorical, the number in $(\cdot)$ in the " Type" column indicates the number of levels.

for correlations of error terms when consultations are associated with the same patient. The effect of interest is captured by the coefficient β_1, where H1 posits that $\beta_1 > 0$.

5.2. Acuity subsamples

Given the observational nature of the data, we are naturally concerned about confounding. Patient acuity is an obvious example of an omitted variable that may bias the OLS results: patients with higher acuity may be unable to wait for an appointment with their regular doctor, so they may be more likely to see a non-regular doctor. At the same time, because of the acute nature of their condition, these patients may also require near-term follow-up appointments, leading to a shorter revisit interval. This could provide an alternative, non-causal explanation for a positive β_1 in (1). As we cannot directly measure patient acuity, nor do we observe when the appointment was booked, adding control variables to account for heterogeneity in acuity is not feasible in this case.

To assess the potential impact of acuity on the results, we use two subsamples of consultations with characteristics that are more likely in acute presentations: (i) the subsample of consultations where the patient was prescribed an antibiotic, and (ii) the subsample of consultations with patients who visited an emergency department (ED) in the seven-day window prior to the focal consultation. Both an antibiotic prescription and a prior ED visit are indicative of a consultation for an acute condition. Hence, if acuity confounds the effect, we would expect a substantially smaller and perhaps insignificant coefficient β_1 in (1) when estimated on these subsamples.

5.3. Instrumental variable estimators

To address general endogeneity concerns, we use two instrumental variable specifications, control functions and two-stage least squares.

5.3.1. Control functions. The control function (CF) approach is based on estimating the following selection and outcome equations

$$RD_i^* = \alpha_0 + \mathbf{X}_i\boldsymbol{\alpha} + \alpha_{n+1}IV_i + \delta_i\,,\;\; RD_i = \mathbb{1}[RD_i^* > 0]\,, \tag{2}$$

$$ln(RI_i) = \beta_0^{CF} + \beta_1^{CF}RD_i + \mathbf{X}_i\boldsymbol{\beta}^{CF} + \gamma_1^{CF}\widehat{RD}_i + \epsilon_i^{CF}\,, \tag{3}$$

where RD_i^* is a latent variable, $\delta_i \sim \mathcal{N}(0,1)$, $\epsilon_i \sim \mathcal{N}(0,\sigma^2)$, $\mathbb{1}[\cdot]$ is the indicator function, $\mathbf{X}_i$ is the $N \times n$ matrix of controls, IV_i is an instrumental variable (to be described shortly), and $\widehat{RD}_i$ is the generalized probit residual of observation i.

Following the probit estimation of Equation (2) as a first stage, $\widehat{RD}_i$ is calculated as

$$\widehat{RD}_i = \frac{\phi(\mathbf{X}_i'\boldsymbol{\alpha}')\,[RD_i - \Phi(\mathbf{X}_i'\boldsymbol{\alpha}')]}{\Phi(\mathbf{X}_i'\boldsymbol{\alpha}')\,[1-\Phi(\mathbf{X}_i'\boldsymbol{\alpha}')]}\,,$$

where $\mathbf{X}_i'\boldsymbol{\alpha}' = \alpha_0 + \mathbf{X}_i\boldsymbol{\alpha} + \alpha_{n+1}IV_i$, and $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density and cumulative distribution functions of the standard normal distribution, respectively. The estimation is then completed by estimating the second stage given in Equation (3).

The CF approach is similar to the more common two-stage least square (2SLS) method. However, in contrast to 2SLS, the CF approach estimates a probit model (i.e., Equation (2)) in the first stage and then uses the generalized probit residual $\widehat{RD_i}$ as an additional control in the outcome equation (i.e., Equation (3)). The addition of the generalized probit residual then adjusts the coefficient β_1^{CF} for unobserved confounders that affect both the endogenous regressor RD_i and the dependent variable $ln(RI_i)$ (Wooldridge 2015). The t-statistic of the coefficient γ^{CF} can be used in a straightforward manner to test for endogeneity in the CF model (Wooldridge 2002).[2]

Consistency of the CF approach requires the probit model to be a correct specification for the likelihood of seeing the regular doctor, i.e., $P(RD_i = 1|\mathbf{X}_i') = \phi(\mathbf{X}_i'\boldsymbol{\alpha}')$. In contrast, the 2SLS estimator does not impose strict distributional assumptions on $P(RD_i = 1|\mathbf{X}_i')$. However, using the standard 2SLS estimator with a nonlinear model in the first stage renders the estimates inconsistent (Wooldridge (2002) refers to this as the "forbidden regression"). An alternative is a 2SLS approach with a linear probability specification in the first stage, i.e., replacing the probit first stage with an OLS estimate. We use this method as a second IV approach to corroborate the CF estimates.

5.3.2. The instrumental variable. Both the CF and 2SLS approaches rely on the availability of an instrumental variable (IV) (Wooldridge 2002). The IV should affect the probability that the patient will see her regular doctor, so it should be significant in the first stage equation (i.e., it should be relevant), but it should not affect the dependent variable $ln(RI_i)$ except through the independent variable RD_i (i.e., it should be valid). In this study, we use an IV that captures whether the focal patient's regular doctor is relatively more accessible for her regular patients during the week of the focal consultation, compared to her long-run average accessibility. This measure is calculated using the set of patients who share the same regular doctor as the focal patient at the time of the consultation, but it excludes any visits by the focal patient herself. The formal description of the IV calculation is given in Section EC.5 of the e-companion.

We believe this is a relevant IV in our study context. A doctor who is more (or less) accessible than usual to her other regular patients is also likely to be more (or less) accessible to the focal patient. Consistent with this intuition, we find a fairly large correlation between the IV and the patient's seeing her regular doctor ($\rho = 0.16$, $p < 0.001$), indicating that the IV is likely both relevant and strong. Formal hypothesis testing for under- and weak identification, reported in Section EC.5.1 of the e-companion, provide strong evidence that the instrument is relevant and the endogenous regressor is not weakly identified.

Turning to the validity condition, there is no reason to believe that the accessibility of the regular doctor for other patients should directly affect the revisit interval of the focal patient. Yet, it is

[2] We estimate the CF model using Stata's etregress command with the two-step option and bootstrapped standard errors. We refer to Wooldridge (2002) and Wooldridge (2015) for a technically detailed explanation of the CF approach.

possible that there are unobserved factors that correlate with both the relative accessibility of the regular doctor for other patients and the focal patient's revisit interval (e.g., a flu outbreak that reduces the expected revisit interval of the focal patient and also makes it harder than normal for patients to access their regular doctors). Importantly, however, such factors should affect the revisit interval not only of the focal patient but also of the doctor's other patients. Therefore, it is possible to account for these unobservable factors by adding as a control variable the average of $ln(RI)$ of other patients who (i) share the same regular doctor as the focal patient and (ii) visit a doctor in the same week as the focal patient. When the average revisit interval of other patients changes (e.g., due to a flu outbreak), then this control variable also adjusts the expected revisit interval of the focal patient in the same direction. This control variable thus helps to account for unobserved factors correlated with both the IV and the outcome, thus strengthening the validity of the IV. A similar approach is used in Freeman et al. (2020) and Bobroske et al. (2021). The full description of this control is provided in Section EC.5.3 of the e-companion.

5.4. Propensity score-based estimation.

To validate the IV approaches, we estimate a propensity score-based matching estimator known as the minimum bias estimator (MBE). Matching itself can alleviate the impact of unobserved bias and, moreover, the MBE is an approach designed to address endogeneity bias without relying on an instrumental variable. The MBE thus provides an alternative method to corroborate the results (Rosenbaum 2005).

5.4.1. Constructing the matched sample. To implement the MBE approach, we first reduce the sample to include only one randomly chosen consultation per patient. This ensures that a patient is not matched with themselves when applying the matching procedure to the data. However, this reduced sample is not representative of the original sample. Every patient is represented only once, independently of the frequency of their use of primary care, and therefore the reduced sample is biased towards consultations with healthier and younger individuals relative to the original sample of consultations. We address this bias by drawing a sample of 25% of the reduced consultation sample using the frequencies of the patients' visits over the entire observation period as probability weights (Jann 2006). This probability weighting biases this second sample towards more frequent users of primary care (i.e., the older and less healthy population) and produces a more representative subsample of consultations.

Using this subsample of consultations, we next match consultations in the control group (those with a transactional provider for the patient's randomly chosen consultation) to the treated group (those with the patient's regular doctor for their randomly chosen consultation) using nearest neighbor matching without replacement. Matching is performed on propensity scores generated from a probit regression that is estimated based on the full set of covariates included within the control matrix $\mathbf{X}_i$ in Equation (1). We also include the condition that the maximum distance between the propensity scores of two observations chosen as potential neighbors (i.e., the caliper) is 0.001. This narrow caliper helps reduce the potential for bias when examining the difference between the average revisit intervals in the two subsamples.

The matched sample consists of 494,810 consultations, half of whom saw their regular doctor and half who did not. Summary statistics comparing the covariate profile of the full analysis sample and the matched sample are provided in Section EC.7 of the e-companion.

5.4.2. Average treatment effect estimation. Using the matched sample, we estimate the average treatment effect (ATE) in two ways. First, we report the difference in averages of $ln(RI)$ between the control and the treated groups. Second, we re-estimate the OLS regression specified in Equation (1) using the matched sample in order to control for the effect of other covariates. Generally, regression after matching is not recommended because standard errors in the regression do not correct for the fact that the matching step has already been performed. However, Austin and Small (2014) and Abadie and Spiess (2021) show that this can be addressed by correcting standard errors by clustering them at the level of the matched pair (i.e., two observations per cluster). We follow this recommendation when reporting effects.

5.4.3. Minimum bias estimation. We combine the treatment effect estimations for the matched samples with the minimum bias estimator (MBE), specified by Millimet and Tchernis (2013). This method aims to minimize the effect of unobserved bias and the impact of omitted variables without relying on instrumental variables and exclusion restrictions. The approach is motivated by the observation that any hidden bias has the biggest impact on the tails of the distribution of selection probabilities (i.e., propensity scores closer to 0 or 1, as calculated using the probit model described in Section 5.4.1). It then follows that the bias is minimized for matched observations with propensity scores equal to 0.5 (the score which is closest to the probability of random assignment) (Peel 2018). The impact of the potential hidden bias is therefore minimized by restricting the sample of matched cases to those with propensity scores within a defined narrow interval around 0.5 (Peel 2018). The direction and strength of the bias can be assessed by widening that interval. Practically, this means that ATEs are estimated on subsamples that are formed by narrowing the range of propensity scores. Black and Smith (2004) recommend a range between 0.33 and 0.67, though with more data (as in our case), this interval can be restricted further.

Table 4 Sensitivity of OLS coefficient estimates to the inclusion of different categories of controls

	OLS: Dependent variable = Natural logarithm of the revisit interval				
$RD_i = 1$	0.050*** [0.048,0.052]	0.112*** [0.110,0.114]	0.150*** [0.149,0.152]	0.150*** [0.148,0.152]	0.150*** [0.148,0.151]
Patient demographics	No	Yes	Yes	Yes	Yes
Patient's past history	No	No	Yes	Yes	Yes
Attributes of regular doctor	No	No	No	Yes	Yes
Practice-level demand	No	No	No	No	Yes
Temporal factors	Yes	Yes	Yes	Yes	Yes
Practice FE	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.045	0.132	0.153	0.158	0.160
Number of observations	11,344,065	11,344,065	11,344,065	11,344,065	11,344,065

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$; 95% confidence intervals in square brackets, with standard errors clustered at the patient level; Variables included within the categories of controls are specified in Table 3.

6. Results

Table 4 shows the OLS estimates, based on Equation (1), for various control structures. Controls are added step-wise into the model to help us understand how the inclusion of different control categories alters the main effect estimate. The step-wise introduction of factors that are correlated with poorer heath, such as consultation frequency, past average revisit interval, comorbidity controls and age and socioeconomic factors, increase the effect of seeing the regular doctor. This is suggestive of heterogeneity, which we will address directly in Section 6.3.

Examining the results, we find evidence that patients' consultations with their regular doctor are robustly associated with a longer revisit interval ($\beta_1 = 0.150$, 95% CI=$[0.148, 0.152]$, p-value $<$ 0.001 in the fully controlled model). Since the dependent variable has a log-scale, the full model suggests that the revisit interval increases by 15.0%, on average, when a patient is seen by their regular doctor.

6.1. Acuity subsamples

Estimating the fully controlled OLS model on the subsample of consultations with antibiotic prescriptions provided a smaller coefficient $\beta_1 = 0.114$ (95% CI=$[0.110, 0.119]$), suggesting that the continuity of care effect may be less pronounced for acute conditions. This is expected, as after an acute visit, a regular doctor is more likely to see the patient again for a follow-up and may have less leeway to extend the revisit interval length. However, the coefficient still remains significantly positive and the effect remains large. This observation is corroborated by the second subsample analysis, consisting of consultations with patients who had visited an ED in the seven-day window prior to the focal consultation. Again, the coefficient is lower than in the full sample ($\beta_1 = 0.098$,

Table 5 **Coefficients of RD_i (seeing the regular doctor) on $ln(RI_i)$ (log revisit interval) for different model specifications**

Dependent variable = Natural logarithm of the revisit interval								
Model	Sample	Method	Coefficient (β_1)	Std. Error	t-statistic	P> $\lvert t \rvert$	95% CI	
1a	consultations	OLS	15.0%	0.1%	182.61	0.00	14.8%	15.2%
1b	consultations	CF	13.2%	0.4%	26.61	0.00	12.2%	14.1%
1c	consultations	2SLS	12.5%	0.5%	24.04	0.00	11.5%	13.6%
2a	patients	PSM	16.1%	0.4%	39.05	0.00	15.3%	16.9%
2b	patients	PSM OLS	16.1%	0.4%	42.19	0.00	15.4%	16.9%
3a	patients	PSM 0.25<p<0.75	16.1%	0.4%	37.91	0.00	15.3%	16.9%
3b	patients	PSM OLS 0.25<p<0.75	16.1%	0.4%	42.19	0.00	15.3%	16.9%
4a	patients	PSM 0.33<p<0.67	16.0%	0.5%	35.09	0.00	15.1%	16.9%
4b	patients	PSM OLS 0.33<p<0.67	16.1%	0.4%	38.23	0.00	15.3%	16.9%
5a	patients	PSM 0.4<p<0.6	16.6%	0.5%	30.70	0.00	15.5%	17.7%
5b	patients	PSM OLS 0.4<p<0.6	16.5%	0.5%	33.12	0.00	15.5%	17.5%

Notes: Standard errors clustered at the patient level for models 1a-1c and the matched pair level for models 2a-5b.
This table reports the coefficient estimates of RD_i (seeing the regular doctor) for the taxonomy of models specified in Section 5. OLS regression refers to the ordinary least squares regression Equation 1. CF refers to the control function approach specified in Section 5.3.1 and 2SLS refers to the two-stage least squares method using the same IV as used for the CF approach. PSM corresponds to the matching-based effect estimation where we report differences in averages of $ln(RI)$ (log revisit interval) between the control and treated groups, using nearest neighbor matching without replacement and a caliper of 0.001 (Section 5.4.2). PSM OLS corresponds to an OLS regression on the matched sample that includes the covariates (Section 5.4.2). Models 3a-5b correspond to the minimum bias estimators that use subsamples determined by the propensities (p) (Section 5.4.3).

95% CI=$[0.084, 0.112]$). However, the effect remains statistically highly significant and the 9.8% estimated average extension of the revisit interval remains practically significant as well.

In summary, the subsample analyses suggest that confounding by acuity, if it occurs, is relatively small and does not explain the main effect. The analyses are fully documented in Section EC.6 of the e-companion.

6.2. Alternative model specifications

Table 5 summarizes the results from all models introduced in Section 5. The top panel reports the estimated effect size associated with seeing the regular doctor (RD_i) on the natural logarithm of the revisit interval using OLS (1a), the CF model (1b), and the 2SLS model (1c). The subsequent panels correspond to estimates using propensity score matching (PSM). All PSM coefficients are estimated using the reduced sample, where a single consultation was randomly chosen per patient, as explained in Section 5.4.1. For PSM (2a) we report differences in averages of the revisit interval, $ln(RI)$, between the control and treated groups; PSM OLS (2b) corresponds to an OLS regression on the matched sample that includes controls (see Section 5.4.2). We then repeat PSM and PSM OLS using three different propensity ranges for the MBE (3a-5b). Narrower propensity ranges correspond to a less biased estimate (see Section 5.4.3).

We find that the results are consistent across all modeling techniques employed and confirm H1. While the coefficient of the bias correction term in the CF model is statistically significant ($\hat{\gamma}^{CF} =$

0.014, 95% CI=$[0.009, 0.021]$), there is little evidence of major confounding, with all estimates close to the 15% effect size estimated by the original OLS model. The MBE estimates do not change significantly as we narrow the propensity score ranges and the direction of the change is towards higher values of the coefficient for narrower ranges.

As the CF approach is both conservative and indicated for use in contexts with a binary first stage selection equation (see Section 5.3.1), we select this approach as the main model specification going forward. The CF method estimates that patients who see their regular provider have a 13.2% longer revisit interval (95% CI=$[12.2\%, 14.1\%]$).

6.3. Moderating effects

The effect of continuity of care is likely to be heterogeneous across patient segments. In this section, we test the moderating effects of comorbidity (H2), age (H3) and mental health (H4). To do so, we re-estimate the CF model (Equations (2) and (3)) but now also include interaction terms between the main independent variable RD_i and the three patient-level moderation variables. As the moderators are naturally correlated (in particular age and comorbidity), we have included all moderators in a single model to estimate their effect net of the correlated moderating effects of the other variables.

Estimates of the average marginal effects based on the moderation results are reported in Table 6 and a graphical representation of the results is given in Figure EC.8 in the e-companion. Table 6 reports the estimated revisit interval for an average individual (within the segment specified by the first column of the row) assuming they either saw a transactional provider ($RD_i = 0$ columns) or their regular doctor ($RD_i = 1$ columns). For example, an average 18-25 year old who saw a transactional doctor is estimated to have a (natural logarithm of the) revisit interval of 3.31, as compared to 3.35 if they had instead seen their regular doctor. The difference between these two values is given by the effect size, e.g., 4.1% (95% CI=$[2.9\%, 5.3\%]$) for an average 18-25 year old patient. (The coefficients corresponding to the moderating effect estimates are reported in Table EC.8 in the e-companion.)

6.3.1. Comorbidity. The comorbidity panel in Table 6 confirms that patients who see their regular doctor ($RD_i = 1$ columns) have, on average, longer revisit intervals than those who see a transactional provider ($RD_i = 0$ columns), and this effect is independent of the number of comorbidities. The effect size columns confirms H2 by showing that the extension of the revisit interval that the regular doctor achieves is higher for patients with comorbidities. The difference between zero and one comorbidities is statistically significant. Additional differential effects beyond two comorbidities are insignificant.

Table 6 **Average marginal effects associated with seeing the regular doctor when using age, comorbidity and mental health as moderators, calculated using the control function model**

	$RD_i = 0$		$RD_i = 1$		Effect Size	
	ln(RI)	95% CI	ln(RI)	95% CI	Mfx.	95% CI
Comorbidity						
0 comorbidites	3.33	[3.32,3.34]	3.42	[3.42,3.43]	9.3%	[8.3%,10.3%]
1 comorbidity	3.24	[3.24,3.25]	3.36	[3.36,3.37]	12.2%	[11.1%,13.2%]
2 comorbidites	3.19	[3.19,3.20]	3.33	[3.32,3.33]	13.4%	[12.4%,14.4%]
3 comorbidites	3.17	[3.16,3.18]	3.30	[3.30,3.31]	13.5%	[12.4%,14.5%]
4 comorbidites	3.16	[3.16,3.17]	3.29	[3.28,3.30]	12.9%	[11.8%,14.0%]
≥ 5 comorbidites	3.15	[3.14,3.16]	3.27	[3.26,3.28]	11.9%	[10.8%,13.0%]
Age band						
18-25yrs	3.31	[3.30,3.31]	3.35	[3.34,3.35]	4.1%	[2.9%,5.3%]
26-30yrs	3.29	[3.29,3.30]	3.35	[3.34,3.35]	5.5%	[4.3%,6.7%]
31-35yrs	3.30	[3.29,3.30]	3.35	[3.35,3.36]	5.8%	[4.6%,6.9%]
36-40yrs	3.31	[3.30,3.31]	3.37	[3.37,3.40]	6.4%	[5.2%,7.5%]
41-15yrs	3.31	[3.30,3.32]	3.39	[3.38,3.40]	8.1%	[7.0%,9.2%]
46-50yrs	3.30	[3.29,3.30]	3.39	[3.39,3.40]	9.6%	[8.5%,10.7%]
51-55yrs	3.29	[3.29,3.30]	3.40	[3.39,3.41]	10.7%	[9.7%,11.8%]
56-60yrs	3.28	[3.27,3.28]	3.40	[3.39,3.41]	12.3%	[11.2%,13.4%]
61-65yrs	3.26	[3.25,3.27]	3.40	[3.39,3.40]	13.4%	[12.4%,14.5%]
66-70yrs	3.23	[3.23,3.24]	3.38	[3.37,3.38]	14.4%	[13.3%,15.5%]
71-75yrs	3.19	[3.18,3.20]	3.35	[3.34,3.36]	15.9%	[14.8%,16.9%]
76-80yrs	3.14	[3.14,3.15]	3.31	[3.31,3.32]	17.2%	[16.1%,18.3%]
81-85yrs	3.08	[3.07,3.09]	3.26	[3.25,3.27]	17.9%	[16.7%,19.0%]
86+yrs	2.97	[2.96,2.98]	3.14	[3.14,3.15]	17.6%	[16.5%,18.7%]
Mental Health						
No	3.23	[3.23,3.24]	3.35	[3.34,3.35]	11.2%	[10.2%,12.2%]
Yes	3.22	[3.21,3.23]	3.35	[3.35,3.36]	13.4%	[12.4%,14.4%]

Notes: '$RD_i = 0$' (resp., '$RD_i = 1$') columns specify the estimated natural logarithm of the revisit interval (ln(RI)) together with 95% confidence intervals (CI) for a patient who saw a transactional (resp., regular) provider; 'Effect Size' columns give the average marginal effect (mfx.), with 95% CIs, associated with a patient seeing a regular doctor versus a transactional provider, implemented using Stata's margins command.

6.3.2. Age. The age band panel in Table 6 shows the effect of patient age. As in the case of comorbidity interactions, the table confirms that patients who see their regular doctor ($RD_i = 1$ columns) have longer revisit intervals than those who see a transactional provider ($RD_i = 0$ columns). The effect size columns shows the difference and confirms H3: seeing a regular doctor is particularly productive for older patients. The extension of the revisit interval length increases from 4.1% (95% CI=$[2.9\%, 5.3\%]$) for 18-25 year-old patients to 17.6% (95% CI=$[16.5\%, 18.7\%]$) for patients over 86.

6.3.3. Mental health. Since mental health patients are more likely to be heavy users of primary care, they have a slightly shorter revisit interval on average than patients without mental health conditions. However, though the difference is small, we find that providing continuity to such patients extends their revisit interval by 13.4% (95% CI=$[12.4\%, 13.4\%]$), compared to a 11.2% (95% CI=$[10.2\%, 12.2\%]$) improvement for patients without such conditions.

6.4. Duration of consultations

As mentioned in Section 3, in addition to the length of the revisit interval, which is the focus of our analysis, consultation length is a second important productivity measure. On the one hand, consultations between a patient and her regular doctor may take longer as the doctor may take more time to avoid a potential revisit that she will have to serve. On the other hand, the regular doctor knows the patient better and may therefore be able to save the time that a transactional provider would need to spend to elicit the necessary information to obtain the same result.

We explore the duration effect empirically, using the same methods we used for revisit intervals, and report the results in Section EC.8 of the e-companion. The data suggest that the information benefits of familiarity outweigh the time taken for a more thorough service. Regular doctors spend, on average, less consultation time with their patients than transactional providers. However, the effect size is not so large that a physician could accommodate an additional patient in a typical four-hour clinical session. Nevertheless, these results suggest that increasing care continuity will not necessitate a reduction of daily clinical throughput in primary care practices.

7. Counterfactual Analysis: Targeting Continuity of Care

Our results suggest that practices could improve productivity by increasing continuity of care and that they could unlock further productivity gains by reallocating continuity to patients who benefit from it the most. In this section, we conduct two analyses to explore retrospectively what the effect would have been on the consultation demand in our data if practices had followed this recommendation in the past.

Using the insights from H2, H3 and H4, we propose a scoring system that can be used by practice managers to prioritize care continuity, targeting those patients for whom it will have the greatest productivity-enhancing effect. The scoring system ranks consultations by the estimated number of days gained if the consultation is offered by the regular doctor rather than a transactional provider. The estimate is obtained as the difference between the predicted return intervals with and without continuity of care. The prediction is based on Equation (1), which we augment by including interactions between the regular doctor variable RD_i and every other covariate in the model. Section EC.10 of the e-companion provides more details on this estimation approach.

Our first analysis does not assume that practices changed their proportion of continuity consultations. Instead, we only explore the demand reduction they would have achieved had they optimized these consultations by shifting them to the most productivity-enhancing patients, using the above scoring system. Figure 2 shows that such targeting of continuity of care has the potential

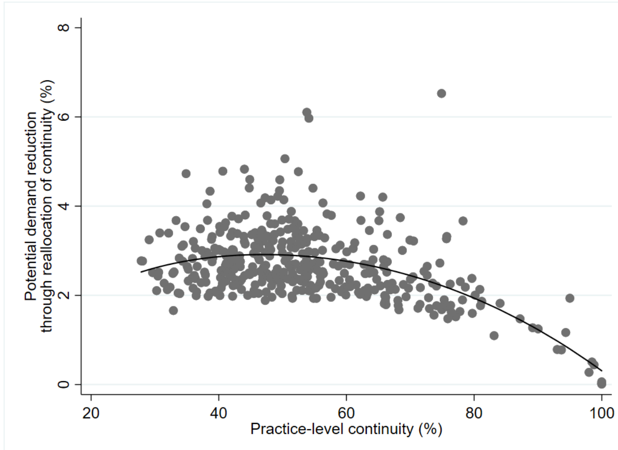


Figure 2 **Practice-level gains from reallocating consultations with the regular doctor to the most productivity enhancing patients, while keeping the proportion of the practice's consultations with regular doctors unchanged.**

of unlocking productivity gains, even without changing the overall proportion of continuity consultations. If all practices had retained continuity at the same level but better targeted continuity at the most productivity enhancing patients, then the total consultation demand in the sample would have reduced by 2.7%. In fact, as Figure 2 shows, some practices would have reduced their demand significantly more.

In our second analysis, we consider what would have happened in our sample if practices not only better targeted care continuity but also increased the level of continuity of care provided. To explore this, we first select a target proportion $0\% \leq x \leq 100\%$ of continuity consultations. We then identify all practices that offered less than $x\%$ continuity of care, and consider the impact on productivity if these practices had offered continuity to $x\%$ of their patients instead. (For those practices offering more than $x\%$ continuity, we leave their proportion of continuity consultations unchanged.) We then re-allocate, as above, the available proportion of continuity consultations in each practice to the most productivity-enhancing patients. This allows us to estimate the counterfactual demand reduction as continuity of care is increased (i.e., by increasing x) and optimally allocated.

The results are summarized in Figure 3. If continuity of care levels were increased for those practices with levels below the across-practice median ($x = 51.3\%$ continuity), 75th percentile ($x = 62.3\%$ continuity) and 90th percentile ($x = 74.1\%$ continuity), then total system demand would have reduced by 3.4%, 4.3% and 5.2%, respectively. These estimates show that there are significant productivity gains to be realized by increasing the proportion of patients seeing their regular doctors and by better targeting continuity of care at those patients who benefit from it most.

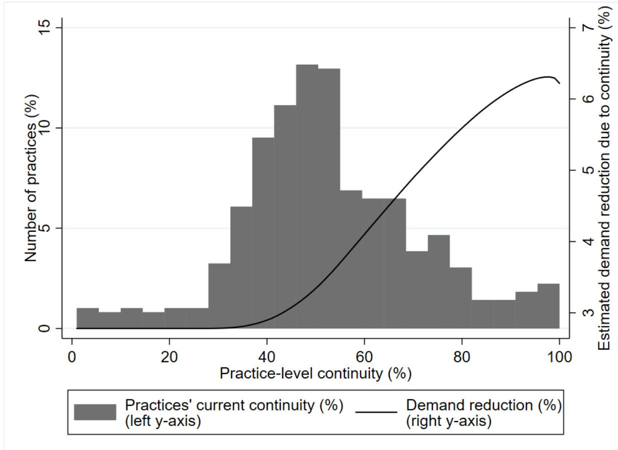


Figure 3 System-level reduction in demand if all practices offer a minimum continuity level (specified on the x-axis) to their most productivity-enhancing patients; The underlying histogram shows the current distribution of the consultations with the regular doctor across practices: if all practices were to offer continuity at the across-practice 90th percentile (74.1%), the total system-level demand would have reduced by 5.2%.

8. Managerial and Policy Implications

This paper studies the relationship between relational continuity of care and clinical productivity in primary care practices. It complements earlier work on the effect of care continuity on patient outcomes and out-of-practice resource utilization. Based on consultation-level data from 1,883,626 patients in 381 English primary practices over 11 years, we find evidence that care continuity has a significant positive effect on the productivity of primary care physicians. Specifically, the data shows that revisit intervals for patients are prolonged by an estimated 13.2% (95% CI=$[12.2\%, 14.1\%]$) when patients see their regular doctor instead of a transactional provider, while there is no evidence that consultations with a regular doctor take longer. The positive productivity effect is larger for older patients, patients with long-term medical conditions and patients with mental health problems. We use a comprehensive taxonomy of models to interrogate alternative explanations for causality and find consistent results across all model specifications, suggesting that the correlational effect is unlikely to be caused by confounding or selection bias. Our findings have important operational implications for primary care practice managers as well as strategic implications for the industry and its regulators.

On the operational side, our findings show that there is no intrinsic trade-off between providing continuity of care and maximising the productivity of a practice's clinical workforce. Doubling down on care continuity can be an effective strategy to improve clinical productivity, particularly if a practice operates in a capitation-based funding environment and serves a relatively older and more complex patient population. In fact, since the strength of the continuity-productivity relationship varies significantly across patient groups, practice managers can further improve productivity by targeting specific patients for care continuity, as we have demonstrated in our counterfactual analysis. By contrast, if practices seek to improve productivity by maximising daily throughput per clinician at the cost of continuity of care, they may find that the increase in staffing needed to serve the increased consultation demand outweighs the throughput gains.

On the strategic side, our findings are of importance for regulators and third-party payers in relation to the trend in primary care towards faster and more convenient access to general medical expertise. This trend has led to the emergence of at-scale online providers and has accelerated during the COVID-19 pandemic, when patients and doctors were forced to adapt to online consultations. Online primary care providers offer a transactional platform for online appointments, matching on-the-spot demand and supply. Much of the advantage comes from scale and the pooling of clinical time. The service is a single consultation between a doctor and a patient, not a long-term relationship. Our findings have several implications for this business model, for regulators and payers who wish to create a sustainable primary care environment, and for primary care practices that need to respond strategically to the threat of the new entrants to the market.

First, our findings provide evidence that the online business model of primary care will be particularly profitable in a fee-for-service environment, where patients or third parties pay per consultation, since transactional service provision generates higher future demand than service provision by regular doctors. By contrast, the model is less compelling in a capitation-based or subscription-based funding environment, where providers are paid a fixed fee per patient per month or year and the risk of excess demand is born by the provider.

Second, our analysis of moderators – chronic disease, age, and mental health – shows that in a capitation context, online primary care providers will have a significant incentive to design their services to be less attractive to older, more demanding or more vulnerable patients to minimize the negative productivity effect of transactional medicine for their patients. We believe this will lead to a segmentation of services, with lower-risk patients being increasingly served quickly and conveniently by transaction-focused online providers, while higher-risk patients are served by local practices that offer the relational continuity of care that these patients need and have the close relationship with the local provider network that is required to effectively serve these patients. In fact, most patients will reach a time in their lives when they move from one segment to the other.

If such a service segmentation is the future of primary care, then regulators must anticipate the destabilizing effect that this transformation will have on traditional practices during a transition period. Regulators will need to respond with adequate risk- and productivity-adjusted funding models. The scoring method used in our counterfactual analysis offers a suggestion for how such models could be developed. At the same time, local practice managers face an important strategic choice: Should they provide services for both segments in-house, managing the tensions between the transactional and relational service model, or should they out-source transactional medicine and focus their in-house services on relational medicine for patients who benefit most from continuity of care and from the close integration of their primary care provider in a local provider network that at-scale online providers will find difficult to replicate? More research is required to advise practices on this decision.

References

Abadie A, Spiess J (2021) Robust post-matching inference. *Journal of the American Statistical Association* 1–13.

Ahuja V, Alvarez CA, Staats BR (2020a) How continuity in service impacts variability: Evidence from a primary care setting, SMU Cox School of Business Research Paper No. 19-13.

Ahuja V, Alvarez CA, Staats BR (2020b) Maintaining continuity in service: An empirical examination of primary care physicians. *Manufacturing & Service Operations Management* .

Amjad H, Carmichael D, Austin AM, Chang CH, Bynum JP (2016) Continuity of care and health care utilization in older adults with dementia in fee-for-service medicare. *JAMA internal medicine* 176(9):1371–1378.

Austin PC, Small DS (2014) The use of bootstrapping when using propensity-score matching without replacement: a simulation study. *Statistics in Medicine* 33(24):4306–4319.

Balint M (1955) The doctor, his patient, and the illness. *The Lancet* 265(6866):683–688.

Barker I, Steventon A, Deeny SR (2017) Association between continuity of care in general practice and hospital admissions for ambulatory care sensitive conditions: cross sectional study of routinely collected, person level data. *Bmj* 356.

Bavafa H, Hitt LM, Terwiesch C (2018) The impact of e-visits on visit frequencies and patient health: Evidence from primary care. *Management Science* 64(12):5461–5480.

Bayliss EA, Ellis JL, Shoup JA, Zeng C, McQuillan DB, Steiner JF (2015) Effect of continuity of care on hospital utilization for seniors with multiple medical conditions in an integrated health care system. *The Annals of Family Medicine* 13(2):123–129.

Beech J, Bottery S, Charlesworth A, Evans H, Gershlick B, Hemmings N, Imison C, Kahtan P, McKenna H, Murray R, Palmer B (2020) Closing the gap: Key areas for action on the health and care workforce — The Nuffield Trust. Technical report, Nuffield Trust.

Biringer E, Hartveit M, Sundfør B, Ruud T, Borg M (2017) Continuity of care as experienced by mental health service users-a qualitative study. *BMC Health Services Research* 17(1):763.

Black DA, Smith JA (2004) How robust is the evidence on the effects of college quality? evidence from matching. *Journal of Econometrics* 121(1-2):99–124.

Bobroske K, Freeman M, Huan L, Cattrell A, Scholtes S (2021) Curbing the Opioid Epidemic at its Root: The Effect of Provider Discordance after Opioid Initiation. *Forthcoming in Management Science* .

Chen HM, Tu YH, Chen CM (2017) Effect of continuity of care on quality of life in older adults with chronic diseases: a meta-analysis. *Clinical Nursing Research* 26(3):266–284.

Cho KH, Kim YS, Nam CM, Kim TH, Kim SJ, Han KT, Park EC (2015) The association between continuity of care and all-cause mortality in patients with newly diagnosed obstructive pulmonary disease: a population-based retrospective cohort study, 2005-2012. *PloS one* 10(11).

Dall T, Reynolds R, Chakrabarti R, Jones K, Iacobucci W (2020) The Complexities of Physician Supply and Demand: Projections from 2018 to 2033. Technical Report June, Association of American Medical Colleges.

Dossa AR, Moisan J, Guénette L, Lauzier S, Grégoire JP (2017) Association between interpersonal continuity of care and medication adherence in type 2 diabetes: an observational cohort study. *CMAJ open* 5(2):E359.

Drury A, Payne S, Brady AM (2020) Identifying associations between quality of life outcomes and healthcare-related variables among colorectal cancer survivors: A cross-sectional survey study. *International Journal of Nursing Studies* 101:103434.

Freeman G, Hughes J, et al. (2010) Continuity of care and the patient experience. Technical report, The King's Fund.

Freeman M, Robinson S, Scholtes S (2020) Gatekeeping, fast and slow: An empirical study of referral errors in the emergency department. *Management Science* .

Goodwin N, Curry N, Naylor C, Ross S, Duldig W, et al. (2010) Managing people with long-term conditions. *London: The Kings Fund* .

Haggerty JL, Reid RJ, Freeman GK, Starfield BH, Adair CE, McKendry R (2003) Continuity of care: a multidisciplinary review. *Bmj* 327(7425):1219–1221.

Hallvik SE, Geissert P, Wakeland W, Hildebran C, Carson J, O'kane N, Deyo RA (2018) Opioid-prescribing continuity and risky opioid prescriptions. *The Annals of Family Medicine* 16(5):440–442.

Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Van Staa T, Smeeth L (2015) Data Resource Profile Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology* 827–836.

Hill AP, Freeman GK (2011) Promoting continuity of care in general practice. *London: Royal College of General Practitioners* .

Hjortdahl P, Borchgrevink CF (1991) Continuity of care: influence of general practitioners' knowledge about their patients on use of resources in consultations. *British Medical Journal* 303(6811):1181–1184.

Huntley A, Lasserson D, Wye L, Morris R, Checkland K, England H, Salisbury C, Purdy S (2014) Which features of primary care affect unscheduled secondary care use? a systematic review. *BMJ open* 4(5):e004746.

Institute for Government (2019) General practice — The Institute for Government. Available at `https://www.instituteforgovernment.org.uk/publication/performance-tracker-2019/general-practice` (2020/21/07).

Jann B (2006) GSAMPLE: Stata module to draw a random sample. Statistical Software Components, Boston College Department of Economics.

Jeffers H, Baker M (2016) Continuity of care: still important in modern-day general practice. *British Journal of General Practice* 66(649):396–397.

Kajaria-Montag H, Freeman M (2020) Explaining the erosion of relational care continuity: An empirical analysis of primary care in England. *SSRN Electronic Journal* INSEAD Working Paper No. 2020/47/TOM.

Katz DA, McCoy KD, Vaughan-Sarrazin MS (2015) Does greater continuity of veterans administration primary care reduce emergency department visits and hospitalization in older veterans? *Journal of the American Geriatrics Society* 63(12):2510–2518.

Koopman RJ, Mainous III AG, Baker R, Gill JM, Gilbert GE (2003) Continuity of care and recognition of diabetes, hypertension, and hypercholesterolemia. *Archives of Internal Medicine* 163(11):1357–1361.

Leniz J, Gulliford MC (2019) Continuity of care and delivery of diabetes and hypertensive care among regular users of primary care services in chile: a cross-sectional study. *BMJ Open* 9(10).

Li MM, Nassiri S, Liu X, Ellimoottil C (2021) How does telemedicine shape physician's practice in mental health? *Shima and Liu, Xiang and Ellimoottil, Chandy, How Does Telemedicine Shape Physician's Practice in Mental Health* .

Lin IP, Wu SC, Huang ST (2015) Continuity of care and avoidable hospitalizations for chronic obstructive pulmonary disease (copd). *The Journal of the American Board of Family Medicine* 28(2):222–230.

Liu N, Finkelstein SR, Kruk ME, Rosenthal D (2018) When waiting to see a doctor is less irritating: Understanding patient preferences and choice behavior in appointment scheduling. *Management Science* 64(5):1975–1996.

Maarsingh OR, Henry Y, van de Ven PM, Deeg DJ (2016) Continuity of care in primary care and association with survival in older people: a 17-year prospective cohort study. *British Journal of General Practice* 66(649):e531–e539.

MacCallum RC, Zhang S, Preacher KJ, Rucker DD (2002) On the practice of dichotomization of quantitative variables. *Psychological methods* 7(1):19.

Millimet DL, Tchernis R (2013) Estimation of treatment effects without an exclusion restriction: With an application to the analysis of the school breakfast program. *Journal of Applied Econometrics* 28(6):982–1017.

NHS Digital (2015) NHS Payments to General Practice, England 2014/15. Technical Report September, NHS Digital.

Nyweide DJ, Anthony DL, Bynum JP, Strawderman RL, Weeks WB, Casalino LP, Fisher ES (2013) Continuity of care and the risk of preventable hospitalization in older adults. *JAMA Internal Medicine* 173(20):1879–1885.

O'Connor PJ, Desai J, Rush WA, Cherney LM, Solberg LI, Bishop DB (1998) Is having a regular provider of diabetes care related to intensity of care and glycemic control? *Journal of Family Practice* 47:290–297.

Palmer W (2019) Is the number of GPs falling across the UK? Technical report, Nuffield Trust, UK.

Palmer W, Hemmings N, Rosen R, Keeble E, Williams S, Imison C (2018) Improving access and continuity in general practice. *Research Summary* .

Peel MJ (2018) Addressing unobserved selection bias in accounting studies: The bias minimization method. *European Accounting Review* 27(1):173–183.

Pourat N, Davis AC, Chen X, Vrungos S, Kominski GF (2015) In california, primary care continuity was associated with reduced emergency department use and fewer hospitalizations. *Health Affairs* 34(7):1113–1120.

Queenan C, Cameron K, Snell A, Smalley J, Joglekar N (2019) Patient heal thyself: reducing hospital readmissions with technology-enabled continuity of care and patient activation. *Production and Operations Management* 28(11):2841–2853.

Ramanayake R, Basnayake B (2018) Evaluation of red flags minimizes missing serious diseases in primary care. *Journal of Family Medicine and Primary Care* 7(2):315.

Reinhardt U (1972) A production function for physician services. *The Review of Economics and Statistics* 55–66.

Ride J, Kasteridis P, Gutacker N, Doran T, Rice N, Gravelle H, Kendrick T, Mason A, Goddard M, Siddiqi N, et al. (2019) Impact of family practice continuity of care on unplanned hospital use for people with serious mental illness. *Health Services Research* 54(6):1316–1325.

Rosen R, Massey Y, Abbas S, Hufflett T (2020) Relational continuity for general practice patients with new and changing symptoms. Technical report, Valentine Health Partnership, The Health Foundation.

Rosenbaum PR (2005) Heterogeneity and causality: Unit heterogeneity and design sensitivity in observational studies. *The American Statistician* 59(2):147–152.

Royal College of General Practitioners (2019) Fit for the future: a vision for general practice.

Sampson F, Pickin M, O'Cathain A, Goodall S, Salisbury C (2008) Impact of same-day appointments on patient satisfaction with general practice appointment systems. *British Journal of General Practice* 58(554):641–643.

Semahegn A, Torpey K, Manu A, Assefa N, Tesfaye G, Ankomah A (2018) Psychotropic medication non-adherence and associated factors among adult patients with major psychiatric disorders: a protocol for a systematic review. *Systematic Reviews* 7(1):1–5.

Senot C (2019) Continuity of care and risk of readmission: An investigation into the healthcare journey of heart failure patients. *Production and Operations Management* 28(8):2008–2030.

Song H, Tucker AL, Murrell KL (2015) The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Science* 61(12):3032–3053.

Tammes P, Purdy S, Salisbury C, MacKichan F, Lasserson D, Morris RW (2017) Continuity of primary care and emergency hospital admissions among older patients in england. *The Annals of Family Medicine* 15(6):515–522.

Tarrant C, Dixon-Woods M, Colman AM, Stokes T (2010) Continuity and trust in primary care: a qualitative study informed by game theory. *The Annals of Family Medicine* 8(5):440–446.

The King's Fund (2016) Understanding pressures in general practice — The Kings Fund. Technical report, The King's Fund.

Wilkin D, Metcalfe D (1984) List size and patient contact in general medical practice. *British Medical Journal (Clinical Research Ed.)* 289(6457):1501–1505.

Wooldridge JM (2002) Econometric analysis of cross section and panel data mit press. *Cambridge, MA* 108.

Wooldridge JM (2015) Control function methods in applied econometrics. *Journal of Human Resources* 50(2):420–445.

Worrall G, Knight J (2011) Continuity of care is good for elderly people with diabetes: retrospective cohort study of mortality and hospitalization. *Canadian Family Physician* 57(1):e16–e20.

Ye T, Sun X, Tang W, Miao Y, Zhang Y, Zhang L (2016) Effect of continuity of care on health-related quality of life in adult patients with hypertension: a cohort study in china. *BMC Health Services Research* 16(1):674.

e-companion to: "Continuity of care increases clinical productivity in primary care"

EC.1. Sample inclusion criteria

Below we provide justification for the various inclusion criteria that were used to transform the initial data set into the final sample for analysis. Each of these points is also summarized in Table 1 in Section 4.2 of the main paper.

- At most practices, patients are able to see a range of staff, such as physicians, nurses, and nurse practitioners, through various channels including face-to-face and phone consultations. In accordance with the literature on continuity of care, we include only consultations with physicians; there is currently only weak evidence of advantages of continuity between patients and non-physician care providers (Tammes et al. 2017, Barker et al. 2017). We also include only face-to-face consultations, which were the most common way patients interacted with their physicians.

- To ensure high data quality, CPRD determines when the data from a practice is considered to be of sufficient quality for research purposes. We discard any observations made before the date at which a practice's data is considered to be of research quality.

- Patients change registrations between practices over time, for example when they change their place of residency. To ensure that registration gaps with a practice do not affect the analysis, we only include the latest continuous period of registration of a patient with a practice.

- To increase the homogeneity of the sample, we have excluded babies and children as they have different medical needs to adults. The sample thus includes only consultations associated with adult patients over the age of 18.

- In line with previous literature, we exclude consultations with patients who had fewer than three consultations in the two-year window prior to the consultation since an accurate identification of a regular doctor is not possible with very few consultations (Ahuja et al. 2020b). We also exclude consultations that were preceded by more than 104 consultations over a two-year window, an average of one consultation per week. Such patients are likely to have very special needs or are on a specific care management plan that requires frequent planned visits. The relationship between care continuity and the revisit interval is less meaningful for this group of patients.

- Since we use a two-year period to calculate the patient's regular doctor, we exclude consultations that took place in the first two years following the patient's registration date at the practice. During this period, we do not have an accurate estimate of the patient's regular doctor.

- We exclude a patient's last recorded consultation as there is no subsequent consultation with which to calculate the revisit interval. We refer to the remaining consultations as consultations with a valid revisit interval.

- Our data consists of the complete medical record of all patients described above who had a contact with a primary care practice between January 2007 and December 2017[3]. However, we also have partial records on some patients outside of this date range, which we remove from the analysis.
- We wish to estimate the effect of care continuity at times when the practice is able to schedule an appointment with the patient's regular doctor. We therefore exclude consultations that take place during weeks when the focal patient's regular doctor is on leave. Our results should therefore be interpreted as being conditional on the patient's regular doctor being available in the week during which the consultation occurred.
- Finally, due to statistical estimation issues related to the large sample size, all of the models are estimated on a random sample of 25% of the remaining 45,376,070 consultations. The random samples are drawn from each of the 381 primary care practices remaining in the data and then merged, thus ensuring that each practice continues to be represented.

EC.2. What predicts continuity of care? Validation of the independent variable

To validate that we are capturing continuity correctly in our independent variable, we wish to understand the relationship between patient-level factors and the probability that the patient will see their regular doctor. We expect that the patient's past frequency of consultations and the patient's complexity (age, number of comorbidities, etc.) will be the most important predictors. To investigate this, we estimate a linear probability model of the form $RD_i = \beta_0 + \boldsymbol{\beta}_1 \mathbf{X}$, where X is a vector of control variables described in Section 4.3.3 and RD_i is a binary variable, capturing whether the patient sees her regular doctor. The results of the model are given in Table EC.1.

We find that the older the patient is, the higher the probability that they will see their regular doctor. Similarly, the more complex the patient is (higher comorbidities and higher number of prescriptions), the higher the probability that they will see their regular doctor. We also find an increasing relationship with the patient's past frequency of consultation – the probability of seeing the regular doctor increases with the increase in past frequency. However, further analysis shows us that this increase is active only up to a certain point (60 consultations), after which the probability of seeing the regular doctor stabilizes. EC.1.

As the observed effects in Table EC.1 are in line with expectations, we have confidence that the binary variable RD_i is appropriately identifying the doctor who is more likely to be a patients regular doctor.

EC.3. Description of control variables

To account for various factors that may confound the relationship between continuity of care and revisit intervals, we include different control variables.

[3] This restriction is applied by CPRD, our data provider

	Dependent Variable: RD_i	
$ln(ExpectedRI)$	-0.026***	[-0.026,-0.025]
IMD=2	-0.007***	[-0.009,-0.006]
IMD=3	-0.008***	[-0.009,-0.006]
IMD=4	-0.017***	[-0.018,-0.015]
IMD=5	-0.020***	[-0.022,-0.019]
Demand	-0.025***	[-0.027,-0.024]
Female	-0.025***	[-0.026,-0.024]
26-30 yrs	0.017***	[0.015,0.019]
31-35 yrs	0.034***	[0.032,0.036]
36-40 yrs	0.050***	[0.048,0.052]
41-15 yrs	0.070***	[0.069,0.072]
46-50 yrs	0.086***	[0.084,0.088]
51-55 yrs	0.101***	[0.099,0.103]
56-60 yrs	0.115***	[0.113,0.117]
61-65 yrs	0.127***	[0.125,0.129]
66-70 yrs	0.140***	[0.138,0.142]
71-75 yrs	0.150***	[0.148,0.152]
76-80 yrs	0.155***	[0.152,0.157]
81-85 yrs	0.150***	[0.147,0.152]
86+ yrs	0.132***	[0.130,0.135]
1 comorbidity	0.028***	[0.026,0.029]
2 comorbidites	0.037***	[0.035,0.039]
3 comorbidites	0.040***	[0.037,0.043]
4 comorbidites	0.044***	[0.040,0.048]
≥ 5 comorbidites	0.049***	[0.044,0.055]
1 prescription	0.022***	[0.021,0.023]
2 prescriptions	0.033***	[0.032,0.035]
3 prescriptions	0.043***	[0.041,0.044]
4-5 prescriptions	0.049***	[0.048,0.051]
6-7 prescriptions	0.052***	[0.051,0.054]
8-9 prescriptions	0.053***	[0.051,0.055]
10-12 prescriptions	0.052***	[0.050,0.054]
13-15 prescriptions	0.053***	[0.050,0.055]
16+ prescriptions	0.059***	[0.055,0.062]
4 consultations	0.008***	[0.007,0.010]
5 consultations	0.012***	[0.010,0.013]
6 consultations	0.007***	[0.005,0.008]
7 consultations	0.009***	[0.008,0.011]
8 consultations	0.011***	[0.009,0.012]
9 consultations	0.014***	[0.012,0.015]
10 consultations	0.015***	[0.013,0.017]
11-12 consultations	0.016***	[0.015,0.018]
13-14 consultations	0.020***	[0.018,0.022]
15-16 consultations	0.021***	[0.019,0.022]
17-18 consultations	0.022***	[0.020,0.024]
19-20 consultations	0.023***	[0.021,0.025]
21-23 consultations	0.024***	[0.022,0.026]
24-26 consultations	0.026***	[0.024,0.028]
27-30 consultations	0.026***	[0.024,0.028]
31-35 consultations	0.026***	[0.023,0.028]
36-45 consultations	0.027***	[0.024,0.029]
46-55 consultations	0.034***	[0.030,0.038]
56+ consultations	0.046***	[0.042,0.051]
Established GP=1	0.023***	[0.021,0.024]
Mental Health=1	0.020***	[0.019,0.021]
Constant	0.616***	[0.607,0.624]
Observations	11,344,065	
All other controls	Yes	

95% confidence intervals in brackets
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table EC.1 Linear probability model to validate the independent variable

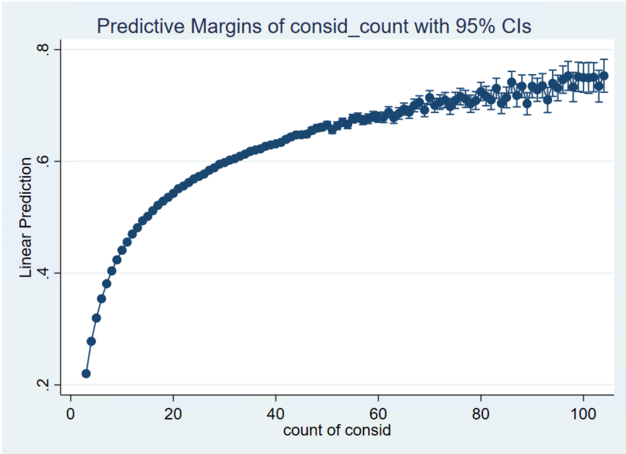


Figure EC.1 Non-linear effect of the patient's past consultation frequency on whether the patient saw her regular doctor

Patient demographics

First, differences in patient demographics may affect the patient's preference for seeing her regular doctor as well as her primary care consultation frequency and hence the revisit interval. To account for this, we control for patient age, as older patients are likely to be sicker and make more frequent contact with primary care, and they are also more likely to prefer to see their regular doctor. We include patient age at the time of consultation in categories as shown in Figure EC.2 (Left).

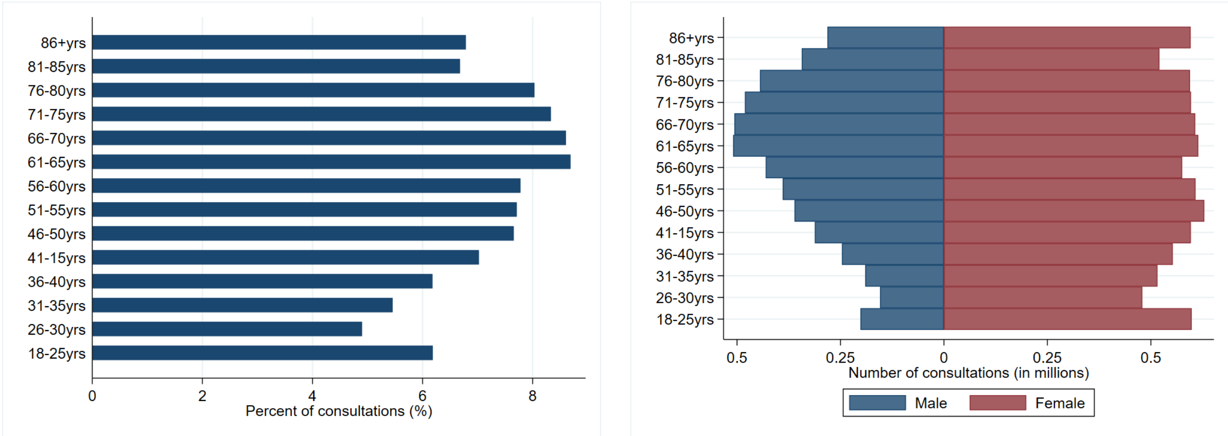


Figure EC.2 (Left) Categories of patient age at the time of consultation; (Right) Breakdown of consultations by gender and age.

Similarly, using the Cambridge Comorbidity Index (CCI), we control for (i) the total number of chronic conditions the patient suffers from at the time of consultation, (ii) a binary variable for each of the 26 individual comorbidities in the CCI and (iii) whether the patient suffers from a mental health condition (described as anxiety, depression or schizophrenia)[4]. The categories used for the total number of chronic

[4] There are a total of 38 comorbdities specified in the Cambridge Comorbidity Index, out of which we group 3

conditions the patient suffers from at the time of consultation and the prevalence of the comorbidities in our dataset are represented visually in Figure EC.3.

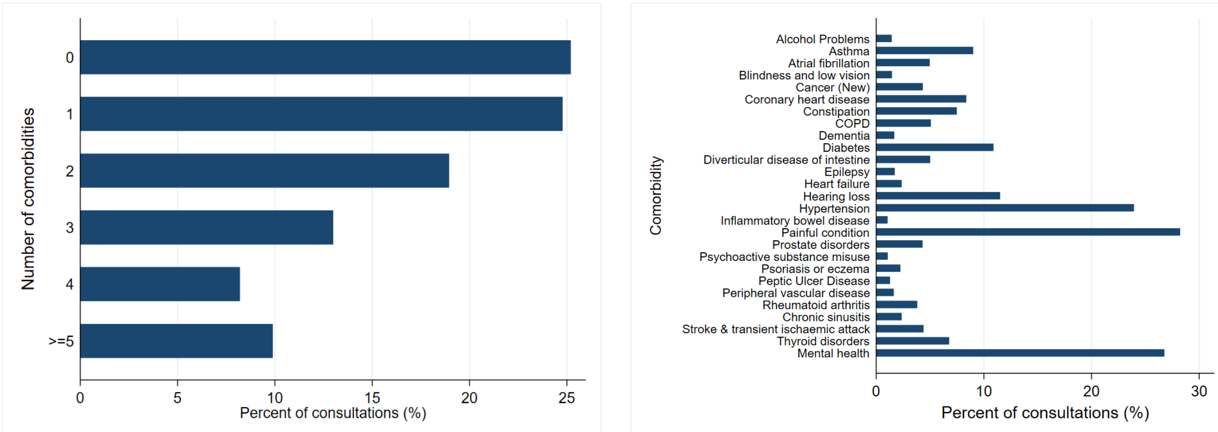


Figure EC.3 **(Left) Categories of the number of active comorbidities the patient suffers from at the time of consultation; (Right) Prevalence of each individual comorbidity in the data.**

We control for gender as it might be a potential confounder – for example, pregnant women or women who just gave birth are likely to visit a primary care provider at short intervals and also more likely to want to see a doctor they trust. The breakdown of gender by age group is shown in Figure EC.2 (Right) and suggests that for all age groups, women are heavier users of primary care than men.

To further capture the patient's severity and complexity, we calculate the total number of unique repeat medications the patient has been prescribed in the 6-month window preceding the focal consultation. The categories used for the total number of prescriptions is represented visually in Figure EC.4.

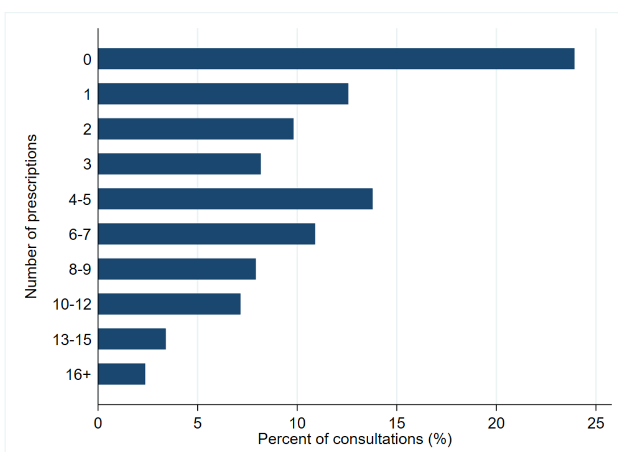


Figure EC.4 **Categories of the number of different active repeat prescriptions the patient is prescribed within a 6-month window preceding the consultation**

comorbidities as mental health condition, and only include those remaining conditions that have a higher than 1% prevalence in our dataset. The leaves us with 26 conditions that we include in the analysis (Payne et al. 2020).

Finally, we control for the socioeconomic status of the patient using the patient-level index of multiple deprivation (IMD), which is provided by CPRD in quintiles and is a relative characterization of poverty or the socioeconomic situation of the area where the patient resides. The categories of the IMD are shown in Figure EC.5.

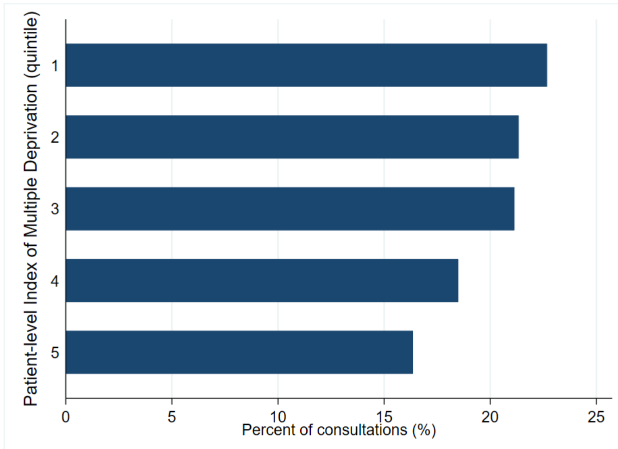


Figure EC.5 Patient's Index of Multiple Deprivation (IMD) as quintiles

Patient visit history

An important confounding factor that need to be accounted for is the patient's visit history. The more frequently a patient visited in the past, the more likely they are to also visit frequently in the future. Hence, we expect that such a patient will have a shorter revisit interval. At the same time, a patient's likelihood of a consultation with a regular doctor is affected by her consultation frequency (see Figure EC.1). Since we anticipate nonlinear effects (see Figure EC.1), we include both the patient's past consultation frequency as a categorical control as well as her average past revisit interval as a linear continuous control (see Table 3 of the main paper for details). The categories are shown in Figure EC.6.

Doctor type

Third, we note that practices are staffed with two types of GPs:

- **Established:** Partners, who are co-owners of the practice, and salaried doctors, who are permanent employees of the practice. Usually, established doctors are list-holding doctors who are accountable for the health of their patient list.
- **Unestablished:** Locums or temporary doctors who have temporary contracts with the practice and work ad-hoc shifts. Unestablished doctors are either self-employed or employed at a locum agency, and they are paid on an hourly basis.

Even though the two categories of doctors receive similar training, the incentives for the established doctors to "get it right the first time" might be stronger, which would translate to a higher quality and productivity benefit if the patient's regular provider is an established doctor. According to our regular provider assignment algorithm, established doctors are regular providers of the patient 93% of the time, whereas unestablished doctors are regular providers the rest of the time.

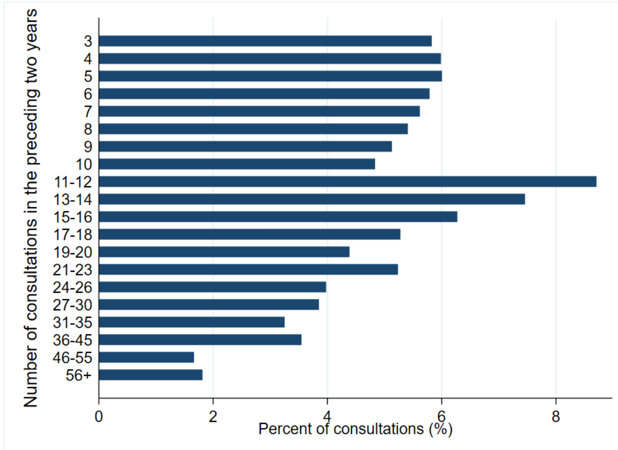


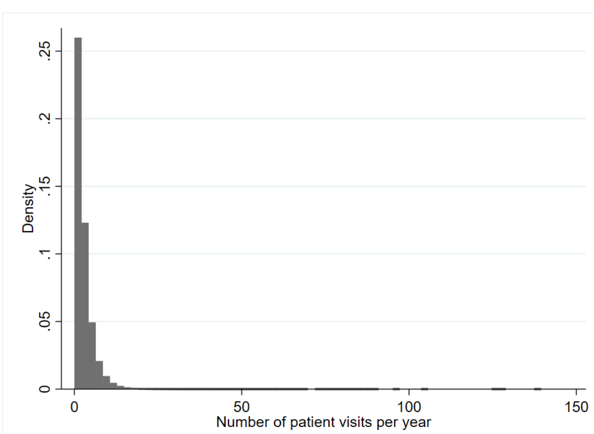
Figure EC.6 **Categories of patient's consultations in the two years preceding the focal consultation**

Practice and time-related controls

Fourth, we include practice fixed effects to capture unobserved time-invariant heterogeneity across practices. For practice-level observed time variant heterogeneity, we include a dynamic control that measures practice-level variation in demand. We calculate this as the total practice demand during the focal week of each consultation as compared to a weekly average in a 52-week period around the focal week. This measure captures fluctuations in demand, which may be correlated with the probability that the patient will see her regular doctor and with the scheduling capacity of the practice.

Lastly, time fixed effects are included to account for any factors that change over time and have a common effect on all practices. Controls are included for year to account for trends, month of year to account for seasonality, and day of the week effects to account for differences in demand and supply across different days of the week that may affect both a patient's seeing her regular doctor and the revisit interval.

EC.4. Patient visits



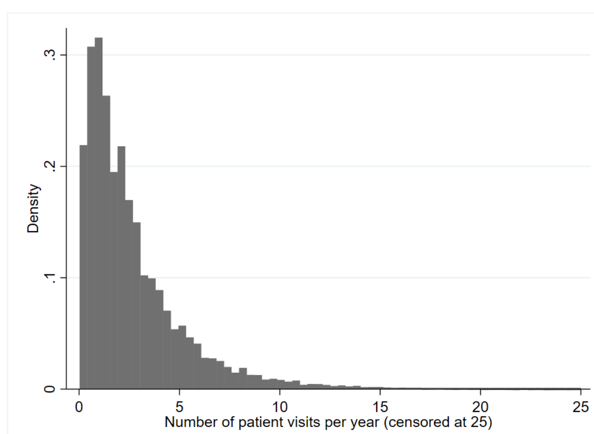


Figure EC.7 **(Left) Histogram of number of patient visits per year; (Right) Histogram of number of patient visits per year (censored at 25 visits)**

In Figure EC.7 we show the histogram of patient visits in our data. For each patient, we calculate the total number of visits in the data divided by the total number of years registered in the data. We find that the average number of visits per year is 2.7, the median number of visits is 2.0, with a standard deviation of 2.8 visits.

EC.5. Instrumental variable construction

To calculate the instrumental variable, let C_{gt} denote the set of consultations in week t for which doctor g is specified as patients' regular doctor. In other words, $|C_{gt}|$ denotes the total number of consultations in week t made by patients for whom doctor g is their regular doctor, where $|\cdot|$ specifies the cardinality of the set. Let g_c denote the actual doctor who the patient saw during consultation c. Then we define the accessibility of doctor g during week t for patients who consider doctor g as their regular doctor to be

$$WkAccessibility_{gt} = \frac{\sum_{c \in C_{gt}} I\left[g_c = g\right]}{|C_{gt}|}$$

where $I\left[\cdot\right]$ is an indicator function that takes value 1 when the condition inside the brackets is satisfied, and 0 otherwise.

Note that $WkAccessibility_{gt}$ also includes any visits by patient i, which may produce a mechanical relationship between the instrumental variable and the dependent variable in the selection equation (i.e., whether or not the patient saw their regular doctor). To prevent this, let C_{igt} be the set of consultations made by patient i in week t for which doctor g was specified as their regular doctor. Then we can define accessibility, excluding patient i, as follows:

$$WkAccessibility_{igt} = \frac{\sum_{c \in \left\{C_{gt} \setminus C_{igt}\right\}} I\left[g_c = g\right]}{|\{C_{gt} \setminus C_{igt}\}|}$$

Next, we standardize $WkAccessibility_{igt}$ by calculating the same measure over a 52-week period around week t. Specifically, let $T_t = \{t-26, \ldots, t-1, t+1, \ldots, t+26\}$ denote the set of 26 weeks prior to and 26 weeks post week t, excluding week t itself. Then the average accessibility over this 52-week period can be written as

$$YrAccessibility_{igt} = \frac{\sum_{t \in T_t} \sum_{c \in \left\{C_{gt} \setminus C_{igt}\right\}} I\left[g_c = g\right]}{\sum_{t \in T_t} |\{C_{gt} \setminus C_{igt}\}|}$$

Finally, we define the instrumental variable, IV, as:

$$IV_{igt} = \frac{WkAccessibility_{igt}}{YrAccessibility_{igt}}$$

EC.5.1. Statistical tests of the instrumental variable

To validate the instrumental variable, we perform formal hypothesis test for under- and weak identification. The underidentification test is a Lagrange Multiplier (LM) test that tests the rank of the matrix to determine whether the equation is identified i.e. whether the excluded instrument is "relevant" or correlated with the endogenous regressor in the first stage selection equation. Weak identification is when the excluded instrument is weakly correlated with the potentially endogenous regressor. Weak instruments can lead to poor performance of estimators, specifically, estimators might be inconsistent, confidence intervals can be incorrect and the tests for significance of coefficients might lead to wrong conclusions.

Though these tests are designed for a continuous endogenous variable, we proceed with testing by treating our binary endogenous variable as continuous. While this means that the critical values of the tests and the significance levels might be slightly different from those reported here, we note that the CF (endogeneous variable treated as binary) and 2SLS (endogeneous variable treated as continuous) results are nearly identical (see Table 5 in the main paper).

For the underidentification test, we use the Sanderson and Windermeijer χ^2 Wald statistic as reported using the `ivreg2` routine in Stata 16 (Sanderson and Windmeijer 2016, Baum et al. 2002). Under the null, the equation is underidentified. For our model, the SW χ^2 statistic takes a value of 150,000 with 1 d.f., which has corresponding p-value = 0.00. Hence, there is strong evidence to reject the null hypothesis of underidentification, and we conclude that the excluded instrument is "relevant".

For weak identification, we use the Sanderson and Windermeijer first stage F-statistic which is the F version of the SW χ^2 statistic. It is used as a diagnostic for whether the endogenus regressor is "weakly identified". The F-statistic from the model is compared against the critical values of the Kleibergen-Paap statistic (for cluster robust standard errors) reported in Stock et al. (2005) to determine whether the instruments are weak. The null hypothesis of the test is that the equation is weakly identified and the maximum bias of the IV estimator relative to the bias of OLS is some specified value such as 10%. For a single endogenous regressor with cluster robust standard errors, the Stock-Yogo critical values are 16.38, 8.96, 6.66 and 5.53 for maximum bias of 10%, 15%, 20% and 25%, respectively. If the estimated F-statistic is less than a particular critical value then the interpretation is that the instruments are weak for that level of bias. In our model, the estimated SW F-statistic is equal to 150,000, indicating a maximal bias of (significantly) less than 10%. Hence, this suggests that that there is no evidence to suspect that our models are affected by the problem of weak instruments.

EC.5.2. Sensitivity of the instrumental variable

Due to its construction, the instrumental variable is sensitive to low values of $\{C_{gt} \backslash C_{igt}\}$ and $\sum_{t \in T_t} \sum_{c \in \{C_{gt} \backslash C_{igt}\}}$. This leads to long tails in the distribution, though we note that 99% of the values are in the range 0 and 2, with a mean of 1.08 and a median of 1.08. Given the long tails in the IV distribution, we have also tested two alternative characterizations of the IV as robustness:

1. We use a dichotomized version of the instrumental variable following MacCallum et al. (2002), specifically, $BinaryIV = 0$ if $WkAccessibility_{igt} < YrAccessibility_{igt}$ and $BinaryIV = 1$ otherwise. The interpretation is that when $BinaryIV$ is 0, the doctor is less accessible to his patients during that week compared to his yearly average, and if $BinaryIV$ takes a value of 1, the doctor is more accessible to his patients than the yearly average.

2. We censor the instrumental variable at 2, referring to it as the $CappedIV$, and include an additional dummy variable as a control to indicate those observations for which the instrumental variable took a value of greater than 2 (1% of the observations). Similar to the interpretation of $BinaryIV$, when $CappedIV$ is less than 1, the doctor is less accessible to her patients during that week compared to her yearly average, and if the value is greater than 1, the doctor is more accessible to her patients than the yearly average.

Summary statistics for the original IV and for the two variations on the instrumental variable described above are given in Table EC.2. Using $BinaryIV$ and $CappedIV$ in our control function model, the estimated effect sizes take value 13.2% (95% CI=$[12.2\%, 14.1\%]$) and 13.2% (95% CI=$[12.4\%, 14.0\%]$), respectively. Thus, the results are insensitive to the choice of instrument.

Table EC.2 Descriptive statistics for the instrumental variable

	Mean	Median	Min	Max	SD
(1) Instrumental Variable	1.08	1.08	0.00	252.00	0.64
(2) Binary version of the IV ($BinaryIV$)	0.66	1.00	0.00	1.00	0.47
(3) Capped version of the IV ($CappedIV$)	1.07	1.08	0.00	2.00	0.32

EC.5.3. IV control

One concern is that there may be omitted variables that threaten the validity of the instrumental variable, particularly if there are unobserved factors that correlate with both the relative accessibility of the regular doctor for other patients (instrument) and the focal patient's revisit interval (dependent variable). For example, if there is a flu outbreak, the focal patient's expected revisit interval might decrease, and at the same time other patients would also find it harder than normal to access their regular provider.

However, such omitted factors should also be expected to affect the revisit interval of the doctor's other patients. We use this insight to construct a control variable to improve the validity of the instrumental variable. Specifically, we add a control that takes the focal patient's expected $ln(RI)$ and adjusts this for the average of $ln(RI)$ of other patients who (i) share the same regular doctor as the focal patient and (ii) visit a doctor in the same week as the focal patient. Intuitively, when the average revisit interval of these other patients change, then the expected $ln(RI)$ of the focal patient should change too.

More specifically, we calculate the control described above for each consultation c_i' that occurred on day t in week w for patient i. To do so, first let $C_{i[t-730,t-1]}$ denote the set of consultations that occurred between day $t-1$ and $t-730$ for patient i. The time between each consultation c_i made by patient i in that interval is then given by RC_{c_i}. Thus, the past average revisit interval (RI) for consultations that occurred over the two years $(t-730, t-1)$ prior to consultation c_i' is as follows:

$$PastAvgRI_{c_i'} = \frac{\sum_{c_i \in C_{i[t-730,t-1]}} RI_{c_i}}{\left|C_{i[t-730,t-1]}\right|}$$

where $|\cdot|$ specifies the cardinality of the set.

Next, let $g_{c_i'}$ denote the regular doctor associated with patient i at the time of consultation c_i'. Further, let $J_{c_i'w}$ specify the set of all consultations by *other* patients for which the regular doctor is the same as for consultation c_i' (i.e., is $g_{c_i'}$) and that took place in the same week w in which consultation c_i' occurred. Using the same method as described above, we can then calculate the past average revisit interval for those consultations $c_j' \in J_{c_i'w}$, where j specifies the patient associated with consultation c_j'. These revisit intervals are given by $PastAvgRI_{c_j'}$.

Next, we create a multiplier that captures the difference between the current and past revisit intervals for those consultations $c'_j \in J_{c'_i w}$, which is equal to $MultRI_{c'_j} = RI_{c'_j}/PastAvgRI_{c'_j}$. Averaging the multiplier over all consultations $c'_j \in J_{c'_i w}$ we get

$$\overline{MultRI}_{c'_i} = \frac{\sum_{c'_j \in J_{c'_i w}} MultRI_{c'_j}}{|J_{c'_i w}|}.$$

Notice that when $\overline{MultRI}_{c'_i} < 1$ it suggests that revisit intervals were shorter than normal for *other* patients who visited during the same week w in which consultation c'_i took place and who shared the same regular doctor $g_{c'_i}$. Meanwhile, when $\overline{MultRI}_{c'_i} > 1$ it suggests that revisit intervals were longer than normal.

The multiplier thus allows us to account for unobserved factors that are correlated with both the IV and the outcomes. For example, if there is a flu outbreak, we might expect the multiplier to take a value less than 1, reflecting the fact that parents are expected to return faster on average. Using this insight, we can then adjust the expected revisit interval of the focal patient by taking their past revisit interval and multiplying by this multiplier, giving us:

$$ExpectedRI_{c'_i} = PastAvgRI_{c'_i} \times \overline{MultRI}_{c'_i}$$

This variable thus gives us the expected time to the next visit of patient i based on their past revisit interval and the multiplier calculated from other patients.

Adding $ln(ExpectedRI_{c'_i})$ as an additional control variable in our model allows us to account for various unobserved factors that might be correlated with both the availability of the regular doctor for other patients (the instrument) and also the expected revisit interval of the focal patient (the dependent variable).[5] Summary statistics for this control variable are given in Table EC.3.

Table EC.3 Descriptive statistics for the control for the instrumental variable

	Mean	Median	Min	Max	SD
(1) $ln(ExpectedRI)$	4.08	4.10	-3.88	10.25	0.77

EC.6. Confounding: Subsample analysis

As outlined in Section 5.3, patient acuity is the most obvious source of omitted variable bias. Specifically, higher acuity patients may be unwilling to wait to see their regular provider, and hence are more likely to see their non-regular doctor. Moreover, these acute patients are more likely to return sooner after the focal consultation for a follow-up visit after the index consultation. This would lead to a shorter revisit interval and could provide an alternative, non-causal explanation for β_i in the OLS model.

One way we address endogeneity concerns is by subsampling the data and comparing the effect sizes associated with seeing the regular doctor across subsamples that indicate higher acuity patients. In selecting subsamples, we aim to identify potentially "acute" consultations for which patients have a higher likelihood of seeing a non-regular doctor and also a shorter revisit interval. We wish to establish that our results are consistent across acute and non-acute consultations, so we perform the following analyses.

[5] We use the natural logarithm of the control to match with the units of the dependent variable (natural logarithm of the revisit interval).

Antibiotic prescriptions

We first consider samples based on prescriptions. Specifically, we identify (i) the sample of consultations when no medication was prescribed and (ii) the sample of consultations when a new antibiotic was prescribed. Antibiotic prescriptions are identified using Gulliford et al. (2020). The rationale is that patients who are prescribed a new antibiotic during their consultation are more likely than other patients to have an acute problem. We estimate the same model as in Equation (1) (re-written below), using the same dependent variable and the same set of control variables:

$$ln(RI_i) = \beta_0 + \beta_1 RD_i + \mathbf{X}_i \boldsymbol{\beta} + \epsilon_i \,, \tag{EC.1}$$

where the vector $\mathbf{X}_i$ specifies the set of controls corresponding to observation i, as defined in Section 4.3.3, and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ is the error term.

If acuity is in fact a major confounder, we expect that the acute consultations would have a smaller value of β_1 in Equation (EC.1). The results (reported in Table EC.4) show that the effect size is indeed slightly smaller, suggesting that the effect is less pronounced for acute conditions. However, the effect remains large and significantly positive.

Table EC.4 Coefficient of RD_i across two different samples: the sample of consultations when (i) no medication was prescribed, (ii) when a new antibiotic was prescribed.

Dependent Variable = Natural logarithm of the revisit interval		
Subsample:	No New Prescription	New Antibiotic Prescription
RD_i=1	0.150***	0.114***
	[0.147,0.152]	[0.110,0.119]
Observations	6,310,933	1,354,977

Notes: 95% confidence intervals in square brackets; The total number of observations in this table is not equal to the total sample size of 11,344,065 observations, as there is a category of observations at which a non-antibiotic medicine was prescribed, which we do not include in this analysis.

Prior emergency department visit

Consultations are more likely to be acute if they are preceded by an emergency department (ED) visit. For example, a patient visiting the ED for an acute reason might be asked by the hospital doctors to follow up with a doctor at their primary care practice. Thus, we class those consultations that are preceded by an ED visit in the seven days prior to the consultation as more acute. We compare the effect of seeing the regular doctor on the revisit interval on two different subsamples: (i) when the index consultation was preceded by an ED visit within seven days and (ii) the remainder of the sample, which had no ED visits within seven days before the index consultation.

Again, we re-estimate the same model as in Equation (1). If the results are driven by acute consultations only, we would expect that the effect size of the subsample where the consultation was preceded by an ED visit would have a much larger effect size. However, as we report in Table EC.5, the effect size is large and significant in both cases.

Table EC.5 Coefficient of RD_i across two different samples: the sample of consultations that were (i) preceded by an ED visit in the 7 days prior to the consultation, and (ii) those that were not.

Dependent Variable = Natural logarithm of the revisit interval		
Subsample:	ED visit 7 days prior	No ED visit 7 days prior
RD_i=1	0.098***	0.137***
	[0.084,0.112]	[0.136,0.139]
Observations	175,123	11,168,942

Notes: 95% confidence intervals in square brackets;

EC.7. Summary statistics for the matched sample

In Table EC.6 we compare the proportions of the control variables for the full sample with those corresponding to the control group and the treatment group. (Specifically, in this table we compare the patient demographics across the different groups. Comparison across other factors (not shown) also suggests that there is no significant different between the groups.) We also report p-values corresponding to Pearson χ^2 tests for differences in the proportions between the control and treatment groups in the matched sample. As can be seen, the proportions remain the same across the three columns of Table EC.6, as desired.

Table EC.6 Comparing the balance of the full sample with the control group and treatment group of the matched sample

	Full sample (%)	Control group (%)	Treated group (%)
Age band			
18-25yrs	5.58	5.62	5.54
26-30yrs	4.68	4.68	4.67
31-35yrs	5.33	5.35	5.31
36-40yrs	6.16	6.14	6.18
41-15yrs	7.17	7.13	7.21
46-50yrs	7.82	7.78	7.85
51-55yrs	7.94	7.94	7.94
56-60yrs	8.02	8.03	8.01
61-65yrs	8.94	8.94	8.94
66-70yrs	8.60	8.63	8.57
71-75yrs	8.30	8.27	8.32
76-80yrs	7.99	8.01	7.97
81-85yrs	6.62	6.65	6.59
86+yrs	6.84	6.81	6.88
		p-value for Pearson test (χ^2)=0.913	
Comorbidity			
0 comorbidites	24.46	24.50	24.41
1 comorbidity	25.14	25.09	25.20
2 comorbidites	19.15	19.10	19.19
3 comorbidites	13.08	13.10	13.06
4 comorbidites	8.26	8.27	8.24
≥ 5 comorbidites	9.92	9.93	9.91
		p-value for Pearson test (χ^2)=0.850	
Mental Health			
No	72.80	72.73	72.86
Yes	27.20	27.27	27.14
		p-value for Pearson test (χ^2)=0.328	
Gender			
Male	37.56	37.62	37.50
Female	62.44	62.38	62.50
		p-value for Pearson test (χ^2)=0.406	
Index of Multiple Deprivation			
1	22.84	22.87	22.81
2	21.58	21.52	21.64
3	20.82	20.82	20.82
4	18.43	18.42	18.54
5	16.34	16.38	16.30
		p-value for Pearson test (χ^2)=0.813	
Prescriptions			
0	23.07	23.16	22.99
1	12.71	12.63	12.79
2	9.91	9.87	9.95
3	8.37	8.36	8.38
4-5	13.94	13.96	13.92
6-7	11.03	11.03	11.03
8-9	7.95	7.96	7.94
10-12	7.17	7.21	7.13
13-15	3.45	3.42	3.48
16+	2.40	2.40	2.40
		p-value for Pearson test (χ^2)=0.581	

Notes: The null hypothesis for the Pearson χ^2 test is that there is no difference between the two proportions. In each case, we find that we cannot reject the null at the 5% significance level.

EC.8. Results for the duration of consultations

Table EC.7 summarises the results of all the models introduced in Section 5 but using the length of consultation as the dependent variable. The results show that the regular doctors spend less time on average with their patients than transactional providers. Hence, providing care continuity will not come at a cost of reducing daily clinical throughput in primary care practices. However, the effect size is not large enough to be actionable. Physicians typically work in four-hour shifts and will be unable to accommodate an additional patient.

Table EC.7 Coefficient of RD_i across all modeling techniques

Dependent variable = Length of the consultation (minutes)								
	Sample	Model	Coefficient	Std. Error	t-statistic	P$>\|t\|$	95% CI	
1a	consultations	OLS	-0.37	0.01	-66.0	0.00	-0.38	-0.36
1b	consultations	CF-IV	-0.15	0.03	-5.01	0.00	-0.20	-0.09
1c	consultations	IV-2SLS	-0.11	0.03	-3.67	0.00	-0.17	-0.05
2a	patients	PSM	-0.32	0.02	-13.86	0.00	-0.37	-0.27
2b	patients	PSM OLS	-0.32	0.02	-14.47	0.00	-0.37	-0.28
3a	patients	PSM 0.25<p<0.75	-0.31	0.02	-13.20	0.00	-0.36	-0.27
3b	patients	PSM OLS 0.25<p<0.75	-0.32	0.02	-13.79	0.00	-0.36	-0.27
4a	patients	PSM 0.33<p<0.67	-0.31	0.03	-12.02	0.00	-0.36	-0.26
4b	patients	PSM OLS 0.33<p<0.67	-0.31	0.02	-12.57	0.00	-0.36	-0.26
5a	patients	PSM 0.4<p<0.6	-0.32	0.02	-13.86	0.00	-0.37	-0.27
5b	patients	PSM OLS 0.4<p<0.6	-0.32	0.02	-14.47	0.00	-0.37	-0.28

Notes: Standard errors are clustered at the patient level for 1a and at the pair level for 2a-5b; This table reports the coefficient estimates of care continuity for the taxonomy of models specified in this section. OLS regression refers to the ordinary least squares regression as specified in Equation 1. CF-IV refers to the control function approach specified in Section 5.3.1 and 2SLS refers to the two-stage least squares method using the same IV as used for the CF-IV approach. PSM corresponds to the matching-based effect estimation where we report differences in averages of ln(RI) between the control and treated groups, using nearest neighbor matching without replacement, and a caliper of 0.001 (Section 5.4.2); PSM OLS corresponds to an OLS regression on the matched sample that includes the covariates (Section 5.4.2); Models 3a-5b correspond to the minimum bias estimators that use subsamples determined by the probabilities (p) (Section 5.4.3). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\dagger}p < 0.10$.

EC.9. Moderation results

In Table EC.8, we report coefficient estimates corresponding to the moderation models described in the main paper. The interpretation of the coefficients in this table is described in the notes under the table. Additionally, Figure EC.8 shows a graphical representation of the estimates of the average marginal effects based on the moderation results corresponding to Table 6 in the main paper.

EC.10. Calculation of the scoring index for targeting continuity of care

Our results suggest that practices should increase their continuity offering to improve productivity. However, we do not recommend that practices offer increased continuity on a random basis, but take into account our results to design the reallocation of continuity to patients who benefit from it the most. Using the insights from Hypotheses 2, 3 and 4, we propose a scoring system that can be used by practice managers to prioritize those patients for whom care continuity has the most productivity-enhancing effect and target them for more relational services. Given the current lack of tools to monitor continuity or to measure how well practices

Table EC.8 Moderating effects of age, comorbidities and mental health, using the control function approach

CF: Dependent variable = Natural logarithm of the revisit interval				
	$RD_i = 0$		Interaction	
	Coefficient	95% CI	Coefficient	95% CI
Baseline ($RD_i = 1$)	–	–	0.011	[-0.001,0.023]
26-30yrs	-0.014***	[-0.019,-0.008]	0.010*	[0.001,0.020]
31-35yrs	-0.010***	[-0.015,-0.004]	0.018***	[0.009,0.027]
36-40yrs	0.001	[-0.004,0.006]	0.030***	[0.021,0.039]
41-15yrs	0.000	[-0.005,0.006]	0.040***	[0.032,0.049]
46-50yrs	-0.006*	[-0.011,-0.000]	0.052***	[0.043,0.060]
51-55yrs	-0.013***	[-0.019,-0.008]	0.065***	[0.057,0.074]
56-60yrs	-0.029***	[-0.035,-0.023]	0.083***	[0.075,0.091]
61-65yrs	-0.042***	[-0.048,-0.036]	0.090***	[0.082,0.098]
66-70yrs	-0.074***	[-0.080,-0.068]	0.103***	[0.094,0.111]
71-75yrs	-0.114***	[-0.120,-0.108]	0.115***	[0.107,0.124]
76-80yrs	-0.162***	[-0.168,-0.155]	0.128***	[0.120,0.137]
81-85yrs	-0.229***	[-0.236,-0.222]	0.142***	[0.133,0.151]
86+yrs	-0.337***	[-0.344,-0.331]	0.132***	[0.123,0.141]
1 comorbidity	-0.088***	[-0.091,-0.084]	0.029***	[0.024,0.033]
2 comorbidities	-0.136***	[-0.140,-0.131]	0.041***	[0.036,0.046]
3 comorbidities	-0.160***	[-0.167,-0.154]	0.042***	[0.036,0.048]
4 comorbidities	-0.168***	[-0.176,-0.160]	0.036***	[0.029,0.043]
≥ 5 comorbidities	-0.177***	[-0.187,-0.166]	0.027***	[0.020,0.033]
Mental health condition	-0.015***	[-0.018,-0.011]	0.025***	[0.021,0.028]

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$; 95% confidence intervals in square brackets, with standard errors clustered at the patient level; 'Baseline' row gives the effect of seeing the regular doctor for an 18-25 year old with no comorbidities and no mental health condition; The '$RD_i = 0$' column specifies the effect of age, comorbidity and mental health for a patient who did not see their regular doctor; The effect of age, comorbidity and mental health for a patient who saw their regular doctor can be determined by taking the baseline effect, and adding to this the sum of the coefficients from the '$RD_i = 0$' and 'Interaction' columns.

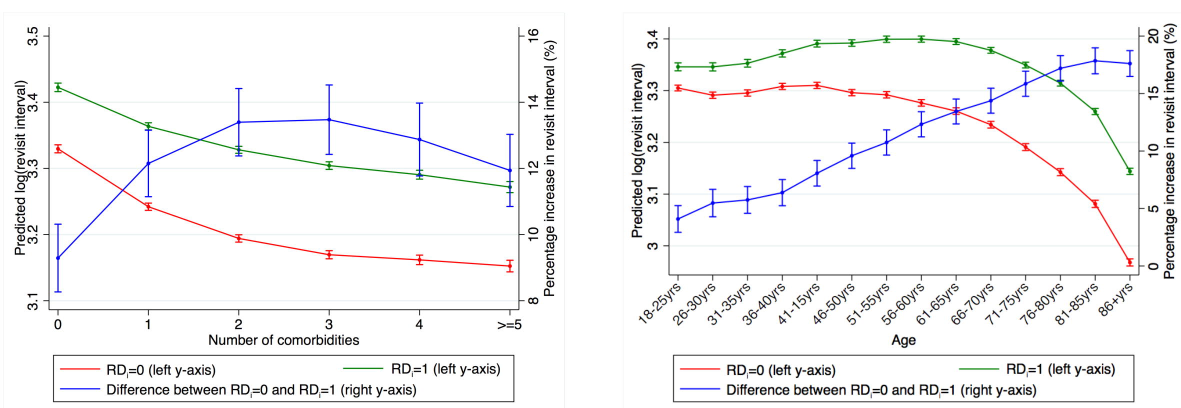


Figure EC.8 Percent increase in ln(Revisit Interval) when the patient sees a transactional provider $(RD_i = 0)$ compared to when the patient sees her regular doctor $(RD_i = 1)$, (left) for patients with multiple chronic diseases (H2) and (right) across different age groups (H3)

have been providing continuity of care to patients who benefit from it, we believe this methodology is an important practical contribution in its own right (Palmer et al. 2018, Hill and Freeman 2011).

The proposed scoring system ranks consultations by the estimated number of days gained if the consultation is offered by the patient's regular doctor as compared to another doctor. To determine the revisit interval days gained, we begin by estimating an interaction model in which we include interactions between the binary indicator for a consultation with the patient's regular doctor (i.e., RD_i) and every other independent variable, using the same method as in Equation (2).[6] For each consultation i, we then use this model to predict the revisit interval under two scenarios: (i) assuming the patient receives care continuity ($RD_i = 1$), which we denote $ln(\hat{R}I_1)_i$, and (ii) assuming the patient does not receive care continuity ($RD_i = 0$), which we denote $ln(\hat{R}I_0)_i$. In order to calculate the productivity gain in number of days rather than in logarithmic units, we transform $ln(\hat{R}I_0)_i$ and $ln(\hat{R}I_1)_i$ using Duan's smearing transformation to get $R\hat{I}_{0i}$ and $R\hat{I}_{1i}$ respectively, in days.[7] For each patient-consultation, we define $days\_gained_i = R\hat{I}_{1i} - R\hat{I}_{0i}$, which is the expected days gained or "score" in days from providing care continuity for consultation i.

Once we have calculated the scores, we can calculate the maximal productivity advantage that a practice would have gained from care continuity if it had used a specific proportion of its overall consultations as its care continuity consultation budget. The maximal number $total\_days\_gained_p$ that practice p can gain, relative to offering no continuity of care, is the sum of consultation-level $days\_gained_i$ achieved by allocating regular doctors to consultations in rank order of $days\_gained_i$, from highest to lowest, until the practice runs out of its continuity of care consultation budget.

To apply this to our data, let C_p denote the set of recorded consultations i that took place in practice p, and let N_p be the total number of these consultations. Then for all consultations $i_p \in C_p$, we can sort the $days\_gained_{i_p}$ in descending order and denote this $\text{sort}(days\_gained_{i_p})^{(a)}$ where a corresponds to the rank once the consultations are sorted using this sorting scheme. Next, let $x_p \in [0\%, 100\%]$ specify the percent of continuity provided in practice p. Then the total days gained by offering x_p% continuity to the most productivity-enhancing patients at practice p can be calculated as

$$total\_days\_gained_p^{[x_p]} = \sum_{a, a \in [0, N_p x_p]} \text{sort}(days\_gained_{i_p})^{(a)}.$$

Change in practice demand by re-allocating continuity

First, note that an extension of the average revisit interval by a factor r leads to a demand reduction of $r/(1+r)$.[8] Therefore, to estimate the effect of continuity on practice demand we can equivalently focus on estimating the impact on the revisit intervals (i.e., estimating the value of r).

To identify r, we start by observing that $\sum_{i_p \in C_p} R\hat{I}_{0i_p}$ gives the sum over the predicted revisit intervals of all consultations C_p that occurred at practice p under the assumption that *no* patient saw their regular

[6] This model gives us the best in-sample fit for the predicted revisit interval.

[7] In a model $ln(y_i) = \mathbf{x}\beta + \epsilon_i$, if the errors are normally distributed as $N(0, \sigma_\epsilon^2)$, the estimates can be recovered as follows: $E[y|\mathbf{x}] = exp(\mathbf{x}\hat{\beta} + 0.5\hat{\sigma}_\epsilon^2)$ where $\hat{\sigma}$ is the usual unbiased estimator of σ. But if the errors are not normally distributed, as it is in our case according to the Jarque-Bera test for normality, then Duan's smearing estimator is used. These estimates are recovered as: $E[y|\mathbf{x}] = exp(\mathbf{x}\hat{\beta}) * (N^{-1} \sum exp(\epsilon_i))$ where ϵ_i is the i^{th} residual from the model and $N^{-1} \sum_i exp(\epsilon)$ is the smearing factor (Duan 1983).

[8] Assuming that each patient i has a demand for x_i consultations over a period $[0, T]$, an extension of all revisit intervals by a factor r extends this demand to x_i consultations over a period $[0, (1+r)T]$ or, equivalently, to a demand of $x_i/(1+r)$ consultations over the period $[0, T]$. Demand is therefore reduced by a factor $r/(1+r)$ to the fraction $1/(1+r)$ of the original demand.

doctor. Based on the results in the main paper, we should expect intuitively that as more patients see their regular doctor these revisit intervals should increase. For example, $\sum_{i_p \in C_p} R\hat{I}_{0i_p} + total\_days\_gained_p^{[x_p]}$ gives the sum of predicted revisit intervals assuming instead that the $x_p\%$ most productivity-enhancing patients at practice p saw their regular doctor, while all other patients saw a transactional provider. Meanwhile, $\sum_{i_p \in C_p} RI_{i_p}$ (i.e., the sum of the revisit intervals as given in the raw data) gives the sum of the revisit intervals under the status quo, i.e., assuming no change in the percentage of consultations between patients and their regular doctor and no change in the allocation of continuity to consultations. Combining these two measures, we can estimate the percentage increase in the revisit interval if continuity at a practice were retained at the same level, which we denote x_p^*, but if continuity were instead re-allocated from those patients who actually received it to those patients who stood to benefit from it the most. This is given by the following expression:

$$\frac{\sum_{i_p \in C_p} R\hat{I}_{0i_p} + total\_days\_gained_p^{[x_p^*]}}{\sum_{i_p \in C_p} RI_{i_p}}$$

Subtracting 1 from this expression gives us the factor r, i.e., the increase in the revisit interval by better targeting care continuity. Observe that we should expect numerator to be greater than the denominator (i.e., $r > 0$), since in both the numerator and denominator we offer continuity to the same proportion of consultations ($x_p^*\%$) but the consultations in the numerator are allocated to the most productivity-enhancing patients.

Using r as calculated above, we can then estimate for each practice the reduction in demand that would have been expected from better allocating continuity while keeping continuity-levels unchanged. The productivity-gains for each practice are plotted in Figure 2 in the main paper.

Change in system demand by re-allocating and increasing continuity

While in the above analysis we have focused on re-allocating continuity within a practice, we can also ask what would happen if practices increased the level of continuity provided *above* their current level x_p^*. To do this, we start by estimating the total reduction in system demand if all practices had re-allocated continuity to the most productivity-enhancing patients. This is given by the expression:

$$\frac{\sum_p \left(\sum_{i_p \in C_p} R\hat{I}_{0i_p} + total\_days\_gained_p^{[x_p^*]} \right)}{\sum_p \sum_{i_p \in C_p} RI_{i_p}}$$

This is the baseline result if all practices kept continuity at their current levels x_p^*. Now let P_x denote the set of practices with current continuity levels below $x\%$, i.e., with $x_p^* < x$. Then we ask what would have happened to demand if all practices with $x_p^* < x$ had increased continuity levels to x *and* had re-allocated continuity to the $x\%$ most productivity-enhancing patients. (Note that we will assume that all practices with $x_p^* \geq x$ keep continuity at the current level x_p^* but also improve targeting of continuity). For a given minimum continuity level x, the total reduction in system demand is then given by the expression

$$\frac{\sum_{p \in P_x} \left(\sum_{i_p \in C_p} R\hat{I}_{0i_p} + total\_days\_gained_p^{[x]} \right) + \sum_{p \notin P_x} \left(\sum_{i_p \in C_p} R\hat{I}_{0i_p} + total\_days\_gained_p^{[x_p^*]} \right)}{\sum_p \sum_{i_p \in C_p} RI_{i_p}}$$

We can estimate this for all values $x \in [0\%, 100\%]$. This is shown in Figure 3 in the main paper.

Acknowledgments

This study was approved by the CPRD Independent Scientific Advisory Committee (ISAC) committee. Protocol no.: 19_004R2. We acknowledge helpful discussions on practical aspects of continuity of care with the clinical team at Granta Medical Practices, Cambridgeshire, UK.

References for the e-companion

Baum CF, Schaffer ME, Stillman S (2002) IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation. Statistical Software Components, Boston College Department of Economics, URL `https://ideas.repec.org/c/boc/bocode/s425401.html`.

Duan N (1983) Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association* 78(383):605–610.

Gulliford MC, Sun X, Anjuman T, Yelland E, Murray-Thomas T (2020) Comparison of antibiotic prescribing records in two uk primary care electronic health record systems: cohort study using cprd gold and cprd aurum databases. *BMJ open* 10(6):e038767.

Hill AP, Freeman GK (2011) Promoting continuity of care in general practice. *London: Royal College of General Practitioners* .

Palmer W, Hemmings N, Rosen R, Keeble E, Williams S, Imison C (2018) Improving access and continuity in general practice. *Research Summary* .

Payne RA, Mendonca SC, Elliott MN, Saunders CL, Edwards DA, Marshall M, Roland M (2020) Development and validation of the cambridge multimorbidity score. *CMAJ* 192(5):E107–E114.

Sanderson E, Windmeijer F (2016) A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of econometrics* 190(2):212–221.

Stock JH, Yogo M, et al. (2005) Testing for weak instruments in linear iv regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* 80(4.2):1.