Working Paper in Economics No. 808

An App Call a Day Keeps the Patient Away? Substitution of Online and In-Person Doctor Consultations Among Young Adults

Lina Maria Ellegård, Gustav Kjellsson and Linn Mattisson

Department of Economics, June 2021 and rev. May 2022

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

An App Call a Day Keeps the Patient Away? Substitution of Online and In-Person Doctor Consultations Among Young Adults

Lina Maria Ellegård, Gustav Kjellsson and Linn Mattisson*

Abstract

The emergence of markets for on-demand online physician consultations –direct-to-consumer telemedicine (DCT) – is currently transforming many health-care settings. DCT may be a cost-effective substitute for ordinary consultations, but the convenience of seeking DCT may increase demand and costs for health insurers. To causally assess to which degree DCT consultations substitute for in-person consultations, we exploit exogenous changes in patient fees in a fuzzy difference-in-discontinuities analysis of young adults in Sweden. We estimate a degree of substitution of 45%, implying an increase in the consultation volume. Characteristics of the additional demand raise concerns related to healthcare equity, efficiency and costs.

*Ellegård: Department of Economics, Lund University & Kristianstad University, lina-maria.ellegard@nek.lu.se. Kjellsson: Department of Economics, gustav.kjellsson@economics.gu.se. Mattisson: Department of Economics, Lund University, linn.mattisson@nek.lu.se. Acknowledgements: We are thankful to Otàvio Bartalotti, Jan Bietenbeck, Carl Bonander, Libertad González, Yang He, Annika Herr, Naimi Johansson, Filippo Tassinari and seminar participants at the EuHea seminar series, Universitat Pompeu Fabra, University of Gothenburg, and Lund University for helpful comments. Belit Ay provided excellent research assistance. This work was funded by Jan Wallanders och Tom Hedelius stiftelse and Tore Browaldhs stiftelse (grant no. P19-0064), the Swedish Research Council for Health, Working Life and Welfare (FORTE, grant no. 2019-00123), Adlerbert Research Foundation, and Foundation for Economic Research in West Sweden. None of the authors have any conflicts of interest to disclose, and no third party had the right to review this work prior to its circulation. The study has been approved by the regional ethics committee (i.e. IRB) in Gothenburg, Sweden (Dnr 068-18).

1 Introduction

As in many parts of the economy, the outbreak of the Covid-19 pandemic greatly reinforced the ongoing digitalisation of health care services (Mehrotra, Bhatia and Snoswell, 2021). But even before the pandemic, the traditional way of providing care in-person at the physician's office was challenged by fast-growing tech start-ups offering online consultations on-demand, so-called *direct-to-consumer telemedicine* (DCT), in many countries.[1]

The usual virtues of online services, to which consumers have become accustomed from experiences with online retail, carry over to DCT: Around the clock, DCT platforms are ready to respond to demand by matching patients with available providers (physicians), who may be located anywhere geographically. The online format virtually eliminates the need to spend time travelling or waiting for appointments, time that in the in-person setting easily may exceed the duration of the consultation itself. This indicates that there are substantial economic gains to be made by substituting DCT for regular in-person consultations,[2] whenever medically appropriate.

Yet, due to the novelty of DCT, little is known of how patients navigate the new landscape enabled by digitalisation. One central question is to what extent the rapidly growing number of DCT consultations represents substitution of in-person visits versus new demand, spurred by the convenience of DCT.[3] In settings where patients do not bear the full cost of DCT,[4] the

[1]Examples of DCT companies include Teladoc and K-Health in the US, Babylon GP at Hand in the UK, Ping An Good Doctor in China, and Kry in Sweden (Salisbury et al., 2020). Even before the pandemic outbreak, Teladoc and Kry were doubling the number of consultations annually (SALAR, 2020; Li et al., 2021)

[2]For instance, the value of the time saved by Swedish patients if half of all primary care consultations were to be provided online has been estimated at over 500 million USD (Ekman, 2018).

[3]DCT may induce new utilisation through several channels. First, the convenience of DCT may lower the bar for when patients choose to seek care rather than to wait and see. Second, some DCT consultations will inevitably result in referrals to in-person appointments, as patients lack all information needed to perfectly judge the need for a physical examination. Third, the inability to perform physical examinations in DCT (or remote care more generally) may have to be compensated for by more consultations to reach the same level of clinical certainty an in-person setting.

[4]For instance, in the US, virtual care was covered by the health plans of a vast majority of large employers already in 2019 (AHIP, 2019). In the UK and Sweden, the main DCT providers have access to public funding.

degree of substitution relates to issues of efficiency as well as health system sustainability and equity. Efficiency-wise, the accessibility of DCT might aggravate already existing moral hazard problems of insurance. From a health system perspective, the additional demand might lead to increasing costs or a redistribution of health care resources away from patients with more severe problems.

This study provides causal estimates of the degree of substitution, using detailed administrative data on a key group of early adopters – young adults – in the two largest regions in Sweden. The institutional setting provides opportunities to obtain plausibly causal estimates, due to the combination of a rapidly emerging DCT market and exogenous variation in regulated user fees. We are able to exploit these features in a pre-Covid-19 context, as the growth of the DCT market took off already in mid-2016. By 2018, DCT consultations accounted for 5% of all primary care physician consultations, but as much as 20% in the age groups we study (SALAR, 2020).

During our study period, patients could consult DCT providers for free until they turned 20; after the 20th birthday, the consultation fee was around EUR 25/USD 30. We use this exogenous change to identify the effect of the fee on the demand for DCT and thereby infer the degree of substitution. To account for other potential demand effects around the age threshold, we use a counterfactual provided by the change in in-person visits around the 20-year birthday of older cohorts, who turned 20 before the DCT market emerged.[5]

We find that the demand for these services is very sensitive to price: After the onset of the fee, the number of DCT consultations falls by half.[6] Further, our estimate of the substitution rate implies that 45% of online consultations replace in-person physician visits. Conse-

[5]We obtain similar estimates with an alternative estimation strategy that does not rely on similarity of different cohorts.

[6]This is a large decrease in comparison both to the literature on how co-payments affect care utilisation (e.g., Manning et al., 1987; Newhouse, Group and Staff, 1993; Bundorf, 2012; Chandra, Gruber and McKnight, 2014; Brot-Goldberg et al., 2017), and in particular to studies using similar age-based threshold strategies to estimate price sensitivity for in-person consultations in Sweden (Nilsson and Paul, 2018; Johansson, Jakobsson and Svensson, 2019).

quently, around half of all DCT consultations represent additional demand, i.e., consultations that would not have taken place, or would have been dealt with by other professions (e.g., nurses, midwives or psychotherapists), in the absence of DCT.

A decomposition exercise suggests that slightly less than half of the additional consultations are one-off encounters with no follow-ups within 14 days. The rest are thus related to care episodes spanning more than one visit; of these consultations, one quarter initiate new episodes, and three quarters represent increases of the number of follow-ups per episode. Importantly, we do not find that the substitution rate is limited to a great extent by DCT consultations that only function as an additional step before an in-person visit (i.e., cases when physical examinations or lab tests are necessary). Thus, the additional consultations stem primarily from the low barriers of access to DCT services.

Both demand- and supply-side factors may explain why the barriers are lower in DCT than in regular care. On the demand side, the convenience of DCT may increase the propensity to seek care for less severe conditions. That almost half of the additional consultations are one-off encounters is in line with such an interpretation. On the supply side, DCT providers, which are funded fee-for-service, may be willing to meet care needs that regular care providers would have assigned low priority or delegated to other professions. In this regard, it is notable that we find particularly low substitution for diagnoses related to skin conditions and reproductive health, which in the regular Swedish primary care setting typically would be handled by referring the patient to self-care, or by other professions (nurses or midwives).

Our results speak to the present challenge of designing incentives for the continued use of telemedicine in the post-pandemic era (Cutler, Nikpay and Huckman, 2020; Mehrotra, Bhatia and Snoswell, 2021). Despite the additional demand, our back-of-the-envelope calculations suggest that the production cost of a DCT consultation is low enough to make the availability of DCT close to resource neutral. Although the estimates are surrounded by uncertainty, this result is remarkable given the institutional structure, which gives DCT providers very strong incentives to schedule additional consultations (i.e., to reduce the degree of substitution). For health insurers considering to cover DCT services, our results thus indicate that such coverage need not result in a cost explosion, provided that the reimbursement rate is set at a sensible level.[7]

Nevertheless, our study setting highlights the importance of designing incentives to address distributional concerns – how to distribute the gains from DCT. With a reimbursement rate amounting to almost twice the marginal cost of a DCT consultation, third-party payers in our setting are likely paying in excess given current patient volumes (over 2 million consultations annually). The economic profits of DCT providers come at the expense of the insured population (tax payers), who may need to pay higher premiums (taxes), or, to the extent that the DCT profits are financed within a given health budget, patients whose health problems cannot be dealt with in a DCT setting. Arguments related to both efficiency (deadweight loss of taxes) and equity (larger needs of patients unsuitable for DCT (Roland, 2019)) may justify another distribution of the gains from DCT, which at present fall exclusively on DCT providers and on DCT users.

Our paper adds to a small number of previous studies on the substitution between DCT and regular care. Survey data from the US indicate that the degree of substitution may be even higher than 45% (Martinez et al., 2018; Nord et al., 2018). However, these estimates may be exaggerated, if respondents rationalise their decision to seek care. Indeed, the few previous studies looking at actual care utilisation have obtained smaller estimates. Ashwood et al. (2017) studied acute respiratory infections in a sample of Californian public employees. Their matched difference-in-differences (DiD) analysis indicated that only 12% of DCT consultations replaced in-person visits.[8] Ellegård and Kjellsson (2019) found no evidence

[7]Coverage would imply large gains for the individuals in terms of time saved and decreased worry.

[8]Related, two studies of DCT visits for acute respiratory infections in the US find that the number of follow-up visits were greater for patients whose first contact was in a DCT setting than for patients whose first contact

of substitution in a DiD analysis of a representative sample of the population in a Swedish region (Skåne).

A limitation of the previous studies is that their estimation strategies may fail to account for time-variant unobserved heterogeneity. DCT providers typically address sudden and transitory health problems, such as respiratory infections or skin conditions; i.e., issues that are neither subsumed by time-invariant group-specific characteristics (fixed effects) nor possible to account for by matching on previously observed conditions. Consequently, DiD analyses may underestimate the degree of substitution. In line with this, the present study, which uses an alternative identification strategy that avoids this limitation, yields a substantially higher degree of substitution.

Outside the DCT setting, a related strand of literature considers telemedicine in traditional healthcare organisations, e.g., communication with the patient's regular physician via video calls or asynchronous messaging systems (North et al., 2014; Shah et al., 2018; Bavafa, Hitt and Terwiesch, 2018; Zeltzer et al., 2021).[9] Although the estimated magnitudes vary between telemedicine applications and contexts, the results from this literature are overall in line with the evidence from DCT, indicating that telemedicine on net leads to more consultations. Methodologically, most of the studies suffer from the same selection problem as the earlier DCT literature. A recent exception is the study by Zeltzer et al. (2021), who exploit variation in telemedicine adoption among Israeli GPs during the Covid-19 lockdown in a DiD setting. They find that increased access to telemedicine leads to a small increase in primary care visits, but 5% lower health care costs due to lower costs of drugs, lab tests, and in- or outpatient services further downstream.

was an in-person visit (Shi et al., 2018; Li et al., 2021). In the regular (i.e., non-DCT) Swedish primary care setting, Entezarjou et al. (2022) did not find differences between patients contacting a provider online first vs in-person first with respect to the number of in-person follow-up visits within 2 weeks for respiratory or urinary symptoms.

[9] Another small literature shows that states adopting telemedicine parity laws experience increases in primary care but reductions in hospital use (Dills and Chen, 2018; Grecu and Sharma, 2019). Due to data limitations, these studies cannot definitively attribute these effects to changes in telemedicine use.

The paper proceeds as follows. The next sections provide a background to the institutional setting (Section 2) and describe the data and our empirical strategy (Section 3). Section 4 provides results and Section 5 concludes.

2 Institutional background and conceptual framework

2.1 Traditional primary care

The financing and provision of health care services in Sweden is delegated to 21 independent regions. Health care is mainly financed by regional proportional income taxes (71%), central government grants (20%) and patient fees (5%) (SALAR, 2017). Primary care handles health issues that can be treated outside hospitals, and is organised in group practices – *primary care centres* (PCCs). PCCs may be public or private, and the staff (typically general practitioners (GPs) and nurses) are salaried employees (Anell, Glenngård and Merkur, 2012). Patients are free to contact any PCC, which typically are open during office hours. During evenings and weekends, care is supplied by a subset of the PCCs.

The PCCs are reimbursed by a mix of (risk-adjusted) capitation, i.e., a fixed monthly sum per registered patient, and fee-for-service based on the number of visits. In one of our study regions (Region Västra Götaland), the reimbursement is almost entirely based on capitation, whereas in the other region (Region Stockholm), capitation accounts for roughly 60% of the reimbursement.

2.2 The Swedish DCT market

Since 2016, the PCCs face competition from private DCT providers. The emergence of the Swedish DCT market was an unintended consequence of the Patient Right Law, enacted in 2015, which gave patients the right to seek care outside their region of residence. DCT entrepreneurs realised that they could locate a DCT company in one region, treat patients in other regions, and then bill the patients' home regions. Notably, this arrangement means that the DCT providers operate outside the regular reimbursement systems, and are instead subject to the regulation for inter-regional reimbursement. This regulatory framework is negotiated by the Swedish Association of Local Governments and Regions (SALAR) and designed as a fee-for-service scheme using visits as the base for payment.

Patients contact a DCT provider by describing their symptoms in a smartphone app or on the company's website, and are then contacted by the DCT physician.[10] Around 80% of the patients at the largest DCT provider are scheduled for an appointment within 30 minutes (Kry, 2019). The DCT consultation may be asynchronous (chat) or synchronous (video call). Notably, few traditional primary care providers offered those care modes in our study period, though telephone consultations were common (especially for follow-up visits).[11] DCT physicians have the same authorities as physicians working in PCCs; i.e., they may set diagnoses, prescribe drugs, and write referrals to other providers.

The patient population and case-mix differs between DCT and traditional primary care. Infants and adolescents are over-represented among DCT patients, and skin conditions and respiratory infections account for a higher share of consultations in DCT settings (Dahlgren et al., 2021). Common conditions in regular primary care such as hypertension are much more rare.

2.3 User fees

Each regional health authority regulates the user fee for consultations supplied by providers in the region. Throughout our study period, the fee level in Region Stockholm (RS) was

[10]The third largest company during our study period, *Doktor.se*, which only served a small fraction of the market, had an initial nurse-led triage step.

[11]The traditional provider Capio and the public providers in Region Västra Götaland launched online platforms in 2018, but the outreach was negligible.

EUR 20, while patients in Region Västra Götaland (RVG) paid EUR 10 for visits at the PCC where they were registered and EUR 30 for visits at other providers. Children and adolescents were (and still are) exempt from paying fees; the lower age limits for paying the fee was 18 in Region Stockholm and 20 in Region Västra Götaland. In a given year, patients never have to pay more than EUR 100 in user fees.

Since the user fee is based on the location of the provider, patients in our study regions did not pay the same amount for DCT consultations as for visits in their region of residence. During our study period, the largest DCT providers *Kry* and *Min Doktor*, accounting for almost 90% of consultations in 2018, were located in Region Jönköping, where patients aged 20 or above paid a user fee of EUR 25. A minor provider (*Doktor.se*) was located in Sörmland, where there was no user fee.[12] Table 1 summarises the user fee and reimbursement systems in the traditional primary care and DCT sectors.

[Table 1 about here.]

2.4 Conceptual framework

To clarify the study setting, Figure 1 provides a framework illustrating the care paths for an individual who experiences a care need. The individual might try do address the issue by herself (self care). If she chooses not to, or if self care does not resolve the problem, she may contact a DCT provider or a primary care centre. In the first case, the patient will be offered an appointment with a physician contracting with the DCT company. In the second case, a triage nurse will examine the need for further care and either schedule an appointment with a physician or another health professional at the PCC, or deny access to further care (typically combined with some self care advice). If, after this step, the patient perceives that the problem has not been resolved, she may again choose between self care or to contact a

[12] In 2019, *Kry* and *Min Doktor* moved to Sörmland.

DCT provider or a PCC (possibly another one than the initial PCC) and so on.[13]

The framework is helpful to understand why the DCT providers may not only serve as substitutes, but also increase the number of consultations. One example is that patients who are dissatisfied with their initial consultation (be it a DCT or an in-person consultation) may contact the other type of care provider for a second opinion. Further, DCT consultations may substitute for self care – partly by virtue of its convenience, partly because DCT providers may schedule appointments in situations where the PCC triage nurse would have denied the patient access to further care.[14] Additional consultations may also be due to differences in the quality of the two care modes, relating to the impossibility of performing physical examinations and lab tests in an online setting. In such cases, the DCT consultation may end with the physician referring the patient to a PCC or scheduling a follow-up appointment to monitor the development of the situation. In both cases, total consultations would increase.

The framework also illustrates that physician consultations in the DCT setting may be substituting for consultations with other professions at PCCs. To give a relevant example for our study population, contraceptive prescriptions are by default handled by midwives in traditional primary care, but by physicians in DCT.

[Figure 1 about here.]

[13]The framework abstracts from care seeking outside the primary care setting, i.e., emergency unit attendance, which is outside the scope of the paper due to data limitations.

[14]The opposite situation – DCT rejecting patients in situations when nurses would not – is less plausible, as DCT providers but not PCCs obtain their entire funding from fee-for-service and are heavily profiled towards providing accessible care.

3 Data and empirical strategy

3.1 Study population and data sources

Employed with data from the Swedish population register (held by Statistics Sweden, SCB), we define a study population consisting of all individuals who belonged to the 19-20 year age group in any of the years 2012-2018, who resided in either Region Stockholm or Region Västra Götaland two consecutive New Year's Eves,[15] and who had lived in Sweden at least since they were 15 years old.[16]

These data are linked to daily care utilisation from regional administrative registers for 2012-2018 covering the universe of consultations with primary care physicians and nurses with providers within the patients' region of residence (i.e. Region Stockholm or Region Västra Götaland) and in the two regions where the DCT providers where located during the study period.Consultations with providers in the region of residence include diagnosis codes (ICD-10). To obtain diagnoses for DCT consultations the data are complemented with information from registers from Region Jönköping, where the two main DCT providers were located.[17]

These care data are also linked to annual data on demographic and socioeconomic characteristics of the study population and their parents obtained from SCB. As the data include the exact date of birth, we can construct a daily panel, where the time dimension is defined relative to the 20th birthday.

[15] Our annual data on place of residence is measured on December 31st.

[16] We employ this restriction to avoid compositional changes driven by the immigration wave in 2014-15.

[17] These data are also linked to daily information on antibiotic prescriptions from pharmaceutical register held by the National Board of Health and Welfare. This information is used in further analyses presented in the appendix only.

3.2 Fuzzy difference-in-discontinuity design

Our objective is to estimate the causal effect of an online consultation, *DCT*, on the number of in-person consultations, y. The identification problem is that an individual's decision to contact a DCT provider may correlate with unobservable characteristics that in turn influence the decision to make an in-person visit at a regular PCC. To eliminate the influence of such omitted variables, we need to find factors that exogenously alter the individual's incentives to contact DCT providers, while not directly changing the incentives to contact traditional providers.

During our study period, the incentives to contact a DCT provider changed exogenously at the 20th birthday, due to the onset of the DCT user fee. A natural starting point for the estimation of the degree of substitution is therefore to consider a fuzzy regression discontinuity (RD) design, using the discontinuity at the 20th birthday as an instrument for the number of DCT consultations. The problem with such a strategy is that the incentives to contact traditional health care may also change at the 20th birthday. As already noted, one of our study regions (Region Västra Götaland) used the same age limit for in-person consultation fees.[18] To purge our estimates of other effects of turning 20, we use an older cohort to estimate discontinuities around the 20th birthday in the period before DCT was available – a differences-in-discontinuities (diff-in-disc) strategy.

Before we introduce our fuzzy diff-in-disc estimand of the degree of substitution, it is instructive to first express a *sharp* diff-in-disc estimand for any random variable Z:

$$\tau_Z = (Z_1^+ - Z_1^-) - (Z_0^+ - Z_0^-) \tag{1}$$

Here, Z_c^+ (Z_c^-) denotes the upper (lower) limit of the regression function $E(Z_c|age_c = 20)$

[18] There may also be other changes at 20. For example, this is the lower age limit for being allowed to buy strong liquor in Sweden, which has been shown to increase the risk of being hospitalised (Heckley, Gerdtham and Jarl, 2018).

of cohort $c \in (0,1)$ as it approaches the age threshold. Thus, the sharp diff-in-disc estimand compares discontinuities at the 20th birthday of two cohorts: a young cohort with access to DCT services both before and after they turned 20 (cohort 1), and an old cohort, who turned 20 before DCT emerged (cohort 0).[19]

Grembi, Nannicini and Troiano (2016) provide assumptions under which the sharp diff-in-disc identifies the causal effects of the new treatment in a setting where other, pre-existing, treatments are assigned at the same threshold. To identify the treatment effect on the cohort affected by both the new and the confounding treatments, two assumptions have to be satisfied. The first is the standard RD assumption that the conditional expectations of all potential outcomes must be continuous around the threshold (for both cohorts). Second, the effects of the confounding treatments must be time-invariant. The assumption implies that the only reason why the discontinuity at the 20th birthday would look different for the two cohorts is that the younger cohort had access to DCT.

Under these assumptions, the sharp diff-in-disc identifies the effect of becoming subject to the DCT consultation fee: When $Z = y$, Eq. (1) describes the effect of the DCT fee on the number of in-person consultations, and when $Z = DCT$, the equation describes the effect of the DCT fee on the number of DCT consultations. In principle, the second term of Eq. (1) is zero when $Z = DCT$, as the DCT market did not yet exist for the older cohort.[20]

In order to estimate the degree of substitution between online and in-person consultations, we turn to a fuzzy diff-in-disc framework. Analogous with the standard fuzzy RD, we construct the fuzzy diff-in-disc estimand as the ratio of the sharp diff-in-discs of y and

[19] Specifically, the old cohort comprises individuals turning 20 before July 1 2016.

[20] In practice, the data from the non-home regions used to identify the DCT consultations includes a small number of in-person consultations and so the term differs from zero.

DCT:[21]

$$\theta = \frac{\tau_y}{\tau_{DCT}} = \frac{(y_1^+ - y_1^-) - (y_0^+ - y_0^-)}{(DCT_1^+ - DCT_1^-) - (DCT_0^+ - DCT_0^-)} \tag{2}$$

The fuzzy diff-in-disc identifies a local average treatment effect that can be interpreted as the degree of substitution for compliers – i.e., individuals who consult DCT providers less often only because of the fee – under the assumption of monotonicity (Millán-Quijano, 2020). In our context monotonocity implies an assumption of no one responding to the DCT fee by consulting DCT providers more often. Monotonicity thus rules out that DCT services are Giffen goods, which seems a plausible assumption to make.

The first assumption of continuous conditional expectations around the threshold warrants some extra discussion. On one hand, the assumption fits well to a context using age as the running variable: Individuals will age and eventually be observed the other side of the age threshold. On the other hand, individuals may anticipate the onset of user fees, and adjust by scheduling care appointments before rather than after the 20th birthday. A strength of the diff-in-disc approach is that, as seen from the nominator of Eq. (2), intertemporal substitution of in-person consultations due to anticipation effects would be purged by the difference of the two RDs, assuming that the incentives for intertemporal substitution are the same for both cohorts (which is likely, as the user fees were constant).

For the online consultations, it is by definition impossible to use the old cohort to net out ”usual” intertemporal substitution. In section 4.1, we instead examine if intertemporal substitution is an issue by checking if the estimated τ_{DCT} is sensitive to removing observations close to the 20th birthday. As seen from the denominator of Eq. (2), intertemporal substitution of online consultations would imply that we underestimate of the degree of substitution.

[21] Another example of a fuzzy diff in disc is Galindo-Silva, Some and Tchuente (2019). These authors discuss a special case in which the treatment of interest – buying insurance – is affected by multiple policies in a young cohort, but only by one policy in an old cohort. This setup differs from our setting, where the treatment of interest – the number of DCT consultations – is affected by one policy in a young cohort, but not available at all to the old cohort.

3.3 Estimation

It is standard to estimate the parameters of a RD model using a local linear (first order) polynomial regression for a given bandwidth (e.g. Calonico, Cattaneo and Titiunik, 2014). We follow the same route to estimate our diff-in-disc model. We apply a uniform kernel throughout. Estimating a fuzzy diff-in-disc in this way is equivalent to estimating a two stage least square model. The first-stage and the reduced form equations for the number of DCT and in-person consultations made by observation i in age-bin j in cohort c is specified as follows:

$$Z_{ijc} = \beta_1^Z + \beta_2^Z\,\mathbf{I}(20)_{ijc} + \beta_3^Z C_{ij} + \beta_4^Z \mathbf{I}(20)_{ijc} \times C_{ij} + f(age_{ijc}, \mathbf{I}(20)_{ijc}, C_{ij}) + \gamma_{ijc}^Z \tag{3}$$

where $\mathbf{I}(20)_{ijc}$ is a dummy for being at least 20 years old, $C_{ij} \in 0,1$ is a cohort dummy, and γ_{ijc}^Z is an error term. $f(age_{ijc}, \mathbf{I}(20)_{ijc}, C_{ij})$ is a function of the running variable age_{ijc} (normalised to 0 at the 20th birthday) and the age and cohort thresholds. In our main specification, this function equals

$$f(age_{ijc}, \mathbf{I}(20)_{ijc}, C_{ij}) = age_{ijc}\left(\beta_5^Z + \beta_6^Z \mathbf{I}(20)_{ijc} + \beta_7^Z C_{ij} + \beta_8^Z \mathbf{I}(20)_{ijc} \times C_{ij}\right) \tag{4}$$

The coefficient of main interest in Eq. (3) is β_4^Z, the diff-in-disc estimate. As we rescale all care utilisation variables to reflect annual averages, the sharp diff-in-disc coefficient provides an estimate of the effect of the DCT fee on the number of consultations per year (e.g., a value of 1 implies one additional consultation annually per capita).

The second stage equation for the number of in-person visits y_{ijc}, in which the endogenous DCT_{ijc} is replaced by the prediction from the first stage equation, can be expressed as follows:

$$y_{ijc} = \alpha_1 + \alpha_2 \mathbf{I}(20)_{ijc} + \alpha_3 C_{ij} + \alpha_4 \hat{DCT}_{ijc} + f(age_{ijc}, \mathbf{I}(20)_{ijc}, C_{ij}) + \varepsilon_{ijc} \quad (5)$$

where ε_{ijc} is an error term and all other variables are defined as above. The fuzzy diff-in-disc estimate $\alpha_4 = \beta_4^y / \beta_4^{DCT}$ can be interpreted as the degree of substitution between online and in-person consultations. $\alpha_4 = -1$ implies that each DCT consultation replaces exactly one in-person visit. If $\alpha_4 < -1$, then each online consultation replaces more than one in-person visit; this might occur for problems for which regular PCCs, but not DCT companies, would provide both an initial and a follow-up consultation. $\alpha_4 \in (-1, 0)$ implies that each DCT consultation offsets less than one in-person visit. In this case, the net effect of the availability of DCT is an increase in the total number of physician consultations (DCT + in-person).

In our main estimations we use a fixed bandwidth of 120 days each side of the 20th birthday. To ensure that our results are not dependent on the bandwidth, we also estimate the model across a range of alternative bandwidths as well as using an optimal bandwidth that is chosen by a data-driven procedure minimising the mean square error of the reduced form equation of in-person visits y (Calonico, Cattaneo and Titiunik, 2014).

We cluster standard errors on the running variable (age in days relative to 20th birthday), using separate clusters for the young and old cohorts (Lee and Card, 2008).[22] With standard errors clustered at the daily level, we may greatly save computational time – without affecting the point estimates or standard errors – by estimating the model on aggregated data. We therefore collapse the individual-day-level data to cells defined by age (in days relative to the 20th birthday), gender, region, and time period,[23] and include frequency weights (=

[22]A recent literature discusses methods to obtain bias-corrected estimates and robust confidence intervals for settings with data-driven bandwidth choices standard RD settings (Calonico, Cattaneo and Titiunik, 2014; He and Bartalotti, 2020). Such methods are yet to be developed for the diff-in-disc setting. However, in a robustness check we modify the wild bootstrap procedure of He and Bartalotti (2020) to fit our fuzzy diff-in-disc setting.

[23]Time periods are equivalent to calendar year for the younger cohorts (who are 20 ($\pm$ 365 days) years old at any point in 2017 or 2018). For the old cohort (who are 20 ($\pm$ 365 days) years old at any point in the pre-DCT

the number of individuals in each cell) in the estimations. To examine if auto-correlation in the age dimension is a problem in our main specification, we also estimate a model on individual-level data in which we cluster standard errors by individual.

3.4 Variable definitions

3.4.1 DCT consultations and in-person visits

Our main outcome variable counts the daily number of in-person physician consultations (visits) at a PCC in the patient's region of residence. To interpret the measure as the annual number of visits, the number of consultations are multiplied by 365. To measure the number of DCT consultations per day, we count the daily number of contacts with PCCs in the two regions where DCT providers were located (Jönköping and Sörmland).[24] This approach slightly overestimates the number of online physician consultations, as the out-of-region data does not allow us to distinguish between online and in-person consultations. Auxiliary analyses in Appendix A show that the overestimation is completely inconsequential, which is expected as the difference should by and large be purged out by the diff-in-disc estimator.[25]

In our main analysis, we include all consultations regardless of diagnosis. In sub-analyses, we use data from Region Jönköping to look specifically at diagnoses that are commonly set by DCT providers. Our definition of common DCT diagnoses cover roughly 90% of all DCT consultations with a physician.[26] We also divide the set of common diagnoses into

period July 1 2012 to June 30 2016), we define four 365-day time periods, each running from July 1 in year t to June 31 in $t+1$ for $t \in$ (2012 to 2015).

[24]The measure also includes a small number of DCT contacts with a provider that was located in a third region (Region Skåne) before it moved to Jönköping, and with the public online service in Region Västra Götaland. The age threshold for the user fee in these Regions are the same as in Region Jönköping.

[25]Using additional data obtained from Region Jönköping, where the age based user fee was applied, we note that online consultations in this region account for almost 90% of the consultations in our preferred measure, and that 9 out of 10 online consultations was with a physician (rather than a nurse etc). Our first stage estimates are practically the same when we use data on out-of-region contacts from registers held by Region Stockholm and Region Västra Götland and when we use data from Region Jönköping (see Appendix A)

[26]Common diagnoses = ICD-codes (on a three-digit level) that cover 80% of the registered diagnoses for 19– and 20– year-olds during online consultations with private providers, including diagnoses within the same

four subsets: upper respiratory infections, skin conditions, diagnoses related to genital and reproductive organs, and a residual category (other), enabling us to estimate the degree of substitution within diagnosis type.

In further analyses, we decompose the consultations by their stage in the care episode (initial vs follow-up), and study outcomes such as the daily number of other in-person visits (nurse visits at a PCC; visits at a midwife in primary care, or a visit at a youth clinic or a clinic specialised in sexually transmitted diseases (STD-clinic)).

3.4.2 Background variables

We use the following predetermined background variables: Father (mother) with university education (Y/N), father (mother) with income above national median (Y/N), at least one parent born outside Scandinavia (Y/N), number of physician visits at age 18 (0/1/>1), rurality of municipality of residence (sparsely populated / densely populated / metropolitan)[27]. Descriptive statistics summarising these characteristics as well as the physician visits are provided in Appendix Table C.1.

Individuals who used DCT in 2018 differ from persons that did not (non-users) in terms of these background variables. For instance, when they were 18 years old, the DCT users made almost twice as many in-person visits as non-users in their birth cohort. Women and city residents are clearly overrepresented among DCT users. Further, a slightly larger share of DCT users have parents that were born in Sweden, have high income or high education.

To say something about the generalisability of our analysis, we compare our study population to other parts of the age distribution in Appendix D. The 19-20 year-olds are quite similar to other adolescents and young adults (<35) in terms of expected health care costs. The share of individuals with a DCT-relevant diagnosis is similar for a much wider age range,

ICD-block with more than 10 registered episodes. See Appendix B.

[27] The rurality variable follows Statistics Sweden's definition. The metropolitan category includes the city of Stockholm, the city of Gothenburg, and municipalities close to these cities.

up to 49 years.

4 Results

4.1 Main results and robustness

4.1.1 Graphical evidence

The first step in the analysis is to examine if the use of DCT services changes discontinuously at the 20th birthday, when the patient starts paying a user fee. Figure 2 shows the annual number of DCT consultations per capita in different years,[28] sorted by the day of the online contact relative to the individual's 20th birthday (day ”zero”). The subgraphs illustrate the evolution of the DCT market. Starting from a situation with virtually no consumption prior to July 2016, the market started to take off and expanded further in 2017 and 2018.[29]

[Figure 2 about here.]

Overall, the figures support that the onset of the DCT fee at the 20th-birthday reduced the demand for DCT services. The black regression lines are estimated using 120 days on each side of the user fee threshold (reflecting how the data portrayed in the figure are used to obtain diff-in-disc estimates in our formal estimations). The regression lines illustrate a small jump at the 20th birthday in both 2017 and 2018. In 2018, when the market had gained some size, the drop at the threshold corresponded to about 50% of the annual average of .3 DCT consultations per capita for individuals aged below 20. The rapid growth of the market makes it difficult to pool data for 2017 and 2018. In our further analyses, we therefore focus

[28] For each day relative to the 20th birthday, the number of consultations is multiplied by 365 to give an annual interpretation.

[29] The definition of DCT consumption includes physicial visits at PCCs in Region Jönköping made by our study population. Such visits account for the non-zero consumption in the preperiods.

on the comparison of the pooled pre-period and 2018. (An analysis for 2017 is presented in Appendix E.)

Next, we examine if the reduction in DCT consultations at the 20th birthday is complemented by an increase in in-person consultations, i.e., if there is evidence of substitution. Figure 3 plots annual in-person physician visits per capita,[30] by the day relative to the 20th birthday. Subtracting the regression discontinuity in the left graph (pre-period) [31] from the discontinuity in the right graph (post period, 2018) yields the sharp diff-in-disc of interest.

The graphs suggest that DCT partially substitutes for in-person consultations. In the pre-period, the 20th birthday was associated with a drop in the number of in-person visits. In 2018, when the DCT market had exploded, there was no longer a notable drop in in-person visits at the age threshold.[32] Combined with the distinct drop in the number of DCT consultations documented in Figure 2, this difference in the discontinuities in 2018 and the pre-DCT period suggests that the PCCs absorb some of the demand served by DCT companies in the absence of a fee.

[Figure 3 about here.]

4.1.2 Main results: Fuzzy diff-in-disc

The first column in Table 2 presents the main results from the fuzzy diff-in-disc estimation using a bandwidth of 120 days. In line with the graphical analysis, the precisely estimated sharp diff-in-disc in Panel A confirms that the onset of the user fee reduces the number of DCT consultations by .15 visits per year, corresponding to a decrease of about 50% compared to the pre-20 mean. The F-statistic of 163.8 suggests that the effect of the user fee on DCT

[30] As before, the statistics are scaled by 365 to allow an annual interpretation.

[31] Appendix Figure H.5 shows that there were similar jumps in each of the years in the pooled pre-period.

[32] Appendix Figure E.1 shows that the decrease of in-person visits at the 20th birthday was similar in 2017, when the DCT market still had limited outreach, as in the pre-period. In same the Appendix, we also provide a formal analysis for 2017.

consultations is strong enough to allow us to use it to study the substitution between DCT and in-person care.

Panel B captures the effect of the DCT user fee on the number of in-person visits at the 20th birthday (i.e., the reduced form estimate). The coefficient equals 0.067 and is statistically significant at the 5% level. The third row presents our fuzzy diff-in-disc (IV) estimate, which can be interpreted as the degree of substitution. The estimate of -.45 suggests that roughly every other DCT consultation replaces an in-person visit. The corresponding 95% confidence interval covers neither zero (0) nor full (-1) substitution.

[Table 2 about here.]

4.1.3 Heterogeneity

The descriptive statistics (Appendix C) as well as further analyses of compliers and first stage heterogeneity (Appendix F) show that the difference between women and men is the most notable of all subgroups. The second and third columns of Table 2 show estimates by gender. The first stage estimates (Panel A) suggest that women's DCT use decreases much more than men's following the onset of the DCT user fee. This reflects that women use DCT more in general; in relative terms, both genders react similarly to the fee. As the sharp diff-in-discs for in-person visits (Panel B) are similar for both genders, the estimated degree of substitution (Panel C) is considerably lower for women than for men. For women, we cannot reject the null hypothesis of no substitution. For men, the results indicate that all DCT consultations represent substitution from in-person visits, although the standard error of the estimate is very large.

Apart from the gender differences, we find that the groups that are overrepresented among DCT users — individuals residing in urban areas, and individuals who visited a physician relatively often in the past – are also overrepresented among the compliers. That is, these groups aggravate the (negative) estimate of the effect of the fee on the number of

DCT consultations. Notably though, all subgroups display similar *relative* decreases of DCT consultations due to the fee (about 50%).

Appendix G presents analyses of the degree of substitution for the subgroups with larger shares of compliers. The results suggest that there is no qualitative difference in the substitution for individuals with different levels of previous care utilisation. The degree of substitution is relatively higher (and more precisely estimated) for individuals living in urban areas than for individuals in rural areas. Thus, the overall degree of substitution largely reflects the degree of substitution among urban residents.[33]

4.1.4 Robustness and precision

Revisiting the scatter plots in Figure 2 and 3, we note that while the relationships between age and the outcome variables are also almost flat over the two years surrounding the 20th birthday, the black regression lines within the 120-day bandwidths may be influenced by temporary fluctuations. We therefore examine the stability of the results with alternative specifications using various bandwidths and a 0-degree polynomial in the running variable (i.e., comparing difference in mean level of care utilisation each side of the threshold for the two cohorts, see Appendix H.1).

[Figure 4 about here.]

Bandwidth: In line with the scatter plots in Figure 2, the baseline estimate of the effect of the onset of the user fee on DCT consultations (first stage) is similar to the ones obtained using other fixed bandwidths up to 365 days (see the upper left graph in Figure 4). The baseline estimate of the substitution (IV) is very similar to the estimates obtained using bandwidths between 90 and 180 days (shown in the lower left graph). Both the first stage

[33] An analysis by region (also in Appendix G) suggests that the degree of substitution is larger in Region Stockholm than in Region Västra Götaland, but that it is mainly due to the large share of urban residents in Region Stockholm.

and IV estimates are similar when using a data-driven procedure to choose an MSE-optimal bandwidth (Calonico, Cattaneo and Titiunik, 2014). We allow the MSE-optimal bandwidth to vary each side of the threshold and in the pre- and post period, but all lie within the range of 105 and 122 (i.e., similar to the bandwidth in our main specification). The degree of substitution is slightly smaller for bandwidths outside of this window.

Intertemporal substitution of care consumption is a potential threat to our identification strategy. Individuals who anticipate the fee increase may decide to contact DCT providers before their 20th birthday rather than afterwards. Such behaviour would lead us to overestimate the effect of the user fee on DCT consultations, and consequently to underestimate the degree of substitution. The upper right graph in Figure 4 presents coefficients from donut estimations of the first-stage relationship, showing that removing up to 28 days on each side of the threshold changes the coefficient little if at all. Given the dominance of sudden and acute (though not necessarily severe) conditions among the DCT consultations, it is arguably unrealistic that individuals to a large extent would reschedule physician appointments over a period of more than four–eight weeks to avoid a fee.[34] Thus, there is, at worst, very limited intertemporal substitution of DCT.

For in-person visits, our diff-in-disc strategy accounts for intertemporal substitution if such behavior is stable over time. The lower right subgraph in Figure 4 shows that removing a donut of up to 28 days on each side of the threshold has only small effects on the size of the fuzzy diff-in-disc estimate, although the precision decreases.[35]

Simple mean comparison: Our main estimates may be sensitive to the parametric spec-

[34] In Appendix H.2, we show that for larger outer bandwidths the first stage coefficient is also robust to removing much larger donuts. Looking specifically at skin conditions, a conditions for which intertemporal substitution would more be likely, the data also suggest that if intertemporal substitution occurs, it is limited to the weeks closest to the threshold (see Appendix H.3)

[35] Extending the donut further within the 120-day bandwidth leads to small sample sizes and noisy estimates. In Appendix H.2, we show that for larger bandwidths (180/365 days), the results are stable when removing up to seven weeks on each side of the 20th birthday. Results are stable for even larger donuts when using a zero-degree polynomial instead of a linear specification.

ification of the regression lines on each side of the threshold. Given the graphical evidence suggesting no clear trends over the two years closest the 20th birthday, it seems most plausible to specify a zero-degree polynomial in the running variable (see *Robust inference* below for higher-order polynomials). Across different bandwidths, this exercise produces point estimates that are in the similar range as the main estimate, but more precisely estimated (Appendix H.1). Notably, these estimates should be more robust to intertemporal substitution than our baseline estimates, as fluctuations within the bandwidth are smoothed out in this specification.

Conditional expectations: As our sample consists of a panel of individuals followed (up to) 240 days, we do not expect any discontinuities in the distribution of predetermined background variables around the threshold. The balance test presented in Appendix H.4 confirms that there are no differences of meaningful size at the threshold. Including these covariates in the diff-in-disc model does not affect neither the size nor the precision of the estimates (see Appendix H.5).

Trends and time invariant confounding: As we use the discontinuity of the pre-DCT cohorts to purge out effects of any confounding treatments, a crucial assumption is that the effects of confounding treatments are time invariant, which in turn relies on the pre- and post-cohorts being largely similar. A placebo analysis using other birthdays as the threshold in a sharp diff-in-disc is a strong test of this assumption. In Appendix H.6, we show that sharp diff-in-discs estimated around the 19th or 21st birthdays are zero for both DCT and in-person consultations. This indicates that our main estimate is not a result of general differences between the cohorts. Figure H.5 shows that the discontinuity at the 20th birthday is of similar size when we split the pre-period into four 12 month windows, lending further support to our strategy. In Appendix table H.4, we also re-estimate the fuzzy diff-in-disc after subsequently deleting up to three out of the (baseline) four pre-cohorts. In all cases, the estimate suggests a substitution rate of around 50%. Finally, in appendix H.8, we also estimate the impact of the potential bias from the secular decrease in the number of in-person visits between the pre- and the post-period. We find that if there is such bias, it would likely be small (reducing the degree of substitution from 45 to 42%).

An alternative strategy that relaxes the assumption of time-invariant effects of confounding policies would be to focus exclusively on Region Stockholm, where there was no confounding change of in-person user fees at age 20. Assuming there are no other confounding policies at age 20, we can then use an ordinary fuzzy RD specification (instead of a diff-in-disc). Appendix Table H.5 shows that such a specification yields a similar estimate (-.48).

Robust inference: A recent literature discusses methods to obtain bias-corrected estimates and robust confidence intervals for settings with data-driven bandwidth choices in standard RD settings (Calonico, Cattaneo and Titiunik, 2014; He and Bartalotti, 2020). Such methods are yet to be developed for a fuzzy diff-in-disc setting. However, we obtain similar confidence intervals for the optimal bandwidth when modifying the wild bootstrap procedure of He and Bartalotti (2020) to fit our setting (Appendix H.10). Notably, this approach uses a second-order polynomial to assess bias, and thus also mitigates concerns that the first-order polynomial in our baseline specification is not flexible enough.

Autocorrelation: To address worries that the standard errors of the preferred model disregard autocorrelation in the age dimension, we estimate the main model on individual-level data with standard errors clustered by individual (Table H.7). This leads to a small increase in the standard error of the point estimate (.22, compared to .20).

approximately half

In sum, our robustness checks support the conclusion that approximately half of all DCT consultations reflect substitution from in-person visits in regular primary care.

4.2 Extensions

4.2.1 Substitution at various stages of the care episode

The estimated degree of substitution of 45% implies that the remaining 55% of the DCT consultations in the no user fee scenario represent additional, induced consumption. To examine where in the care episode these induced consultations take place, Figure 5 displays sharp diff-in-disc estimates of the DCT user fee on DCT and in-person consultations, decomposed by whether the consultation is the initial contact in a care episode or a follow-up. An initial consultation is defined as a consultation with no DCT or in-person consultation during the preceding 14 days.[36] The remaining consultations are defined as follow-ups (although some of these in practice may not belong to the same care episode). The onset of the user fee leads to a reduction in *initial* DCT consultations – the effect on one-off encounters equals -.093, and the effect on initial DCTs with a follow-up within 14 days equals -.021 – as well as in DCT *follow-ups* (-.035). While these reductions are complemented by increases in *initial* in-person visits (of 0.066 and 0.010), the effect on *follow-up* in-person visits is practically zero.

Of the (approximately) 55% induced consultations, we thus find that 24 percentage points are one-off encounters with no follow-up, and so the remaining part belong to multiple-encounter episodes. Initial consultations in an episode (with a follow-up within 14 days) account for 7 percentage points and follow-up consultations for the remaining 23 percentage points.

Notably, the zero effect on in-person follow-ups is the net effect of an increase in follow-ups that occur after an initial in-person visit and a decrease in follow-ups that occur after an initial DCT.[37] The decrease in the latter type of follow-ups implies that the overall de-

[36] See Appendix I for wash-out periods of 7 and 28 days.

[37] Appendix I show that the zero net effect is a result of heterogeneous effects across genders. The increase in follow-ups after an initial in-person visit is driven by men, and the decrease in follow-ups after an initial DCT consultation is driven by women.

gree of substitution is limited by initial DCT consultations that later are followed by one or more in-person visits, potentially reflecting cases where a physical examination is necessary. However, this effect is rather limited: such consultations account for about 8% of the additional consultations in the absence of a user fee. Thus, most of the additional consultations do *not* reflect extra steps on the paths towards an unavoidable in-person visit. Consequently, other explanations, such as the lower threshold for seeking care in the DCT setting, are more important factors to explain why the total number of consultations increase.

[Figure 5 about here.]

4.2.2 Substitution within diagnosis types

We next focus on consultations with diagnoses that are commonly set during DCT consultations. The 80% most common diagnoses cover 89% of all DCT consultations in the estimation sample. In a specification including only these diagnoses, shown in the leftmost column of Table 3, the estimated degree of substitution is very similar to our main estimate. The other columns of Table 3 show estimates from models in which we relate DCT consultations to in-person visits within the same diagnosis category: Upper respiratory infections (*Resp*), skin related problems (*Skin*), genital and reproductive organs *Gen/Rep*, and a residual category capturing the *Other* common diagnoses set by DCT providers. These categories cover 19% 24%, 21%, and 27% of all DCT consultations in the post period.[38] The first stage coefficients stand in proportion to each category's share of all DCT consultations, but the estimated degree of substitution varies between categories.[39]

The highest degree of substitution is obtained for upper respiratory infections. The estimate suggests that DCT consultations for respiratory infections (more than) fully replace in-person consultations. The estimates for skin and genital/reproductive diagnoses are small

[38]The subsets are overlapping to some extent, because more than one diagnosis may be set during a given consultation.

[39]These results are similar when using longer bandwidths, see Appendix J.

and not statistically significant.[40] Although we cannot rule out partial substitution of similar magnitudes as before, we believe it is plausible to find more induced consultations for this type of diagnoses. First, the convenience of DCT may tip the balance for individuals that would not have bothered to have a rash or eczema examined, if they would have had to travel to the PCC for the examination. Further, for minor skin conditions in this category, the triage nurse at the PCC might refuse to schedule a physician appointment (possibly offering a nurse examination instead, or give self-care advice on the phone). Second, the genital/reproductive diagnoses include many contraceptive prescription renewals. In Sweden, contraceptive management is normally handled by midwives in traditional primary care, so there are few physician in-person visits that the DCT consultations may replace.

Notably, the *gen/rep* category mainly includes diagnoses that are only relevant for women. For this reason, the table also presents estimates by gender. We note that, as expected, the first stage estimate for men for conditions in the *gen/rep* category is very weak. We further note that in the specification for the most relevant population (women), there is still no evidence of substitution in this category.

For *other* DCT-relevant diagnoses (including, e.g., vague symptoms, mental problems, and renewal of prescriptions) the degree of substitution is similar to that for the overall category (*Common*), but the estimate is not statistically significant at conventional levels.

[Table 3 about here.]

[40] Given the prevalence of upper respiratory infections, skin conditions (acne) and genital and reproductive health (cystitis) among DCT diagnoses, it is possible that DCT affects antibiotic use. Appendix K shows that the onset of the user fee is *not* associated with a *decrease* in antibiotic prescriptions. Thus, DCT physicians seem to be at least as restrictive as other physicians in terms of antibiotics prescriptions.This is consistent with recent descriptive evidence (Entezarjou et al., 2021; Shi et al., 2018), though earlier studies indicated otherwise (Uscher-Pines et al., 2015). Notably, the result holds also for the subset of antibiotics prescribed against respiratory infections. This is consistent with the medical guidelines, that advocate against prescriptions without a physical examination.

4.2.3 Other consultations and antibiotic prescriptions

Acknowleding the role of other professional categories (nurses and midwives) in Swedish primary care, Table 4 presents estimates of the degree of substitution for various types of other visits: Nurse visits at a primary care centre, visits at a midwife, youth or STD clinic, and the sum of these two types of visits. To grasp the total degree of substitution across all categories, we lastly sum of these two types with in-person physician visits at a PCC.

The results do not indicate that the additional DCT consultations found in the main analysis reflect visits that would otherwise have taken place in another part of the traditional primary care setting. The estimated substitution is low for both nurse visits and visits at midwife/youth/STD clinics. The specification including all professions together (*All consultations*) indicates that the total substitution rate is slightly higher than in the main analysis.

Splitting the sample by gender, the point estimates suggest that women to some extent substitute across professional categories (DCT physician vs midwife in primary care), although the estimate is statistically significant. In Appendix L, we explore this further and find that the estimate is closer to -1 when including only DCT consultations related to genital/reproductive issues. Another interesting patterns is that DCT consultations appear to induce subsequent consultations at youth/STD clinics among men.[41]

[Table 4 about here.]

4.2.4 Cost analysis

Our main analysis suggests that the availability of DCT leads to more physician consultations, making it interesting to examine whether DCT also leads to increasing health care costs.[42] Given a substitution rate of approximately 45%, DCT would not increase costs if

[41] Although imprecise, these patterns are robust to using other bandwidths, see Appendix L.

[42] A full economic evaluation of the effects of DCT would not only include a comparison of costs, but also of benefits. In particular, the two services may differ in quality. However, we have limited opportunities to

the unit costs in the DCT setting are at most 45% of those in the in-person setting.

In Appendix M, we compare the unit costs of a visit using the best available estimates of costs and time use from the different settings. This implies lower wages and shorter consultation times in the DCT setting, although we also examine costs under assumptions of similar wages and consultation times.[43] The cost analysis abstracts from all fixed costs except for office space.[44]

Our baseline estimate indicates that the direct cost of a DCT consultation is 36% of that for an in-person consultation, i.e., low enough not to add to total health care costs despite the induced demand. The ratio reduces to 30% when we take into account the value of the time saved by patients in DCT. In more conservative scenarios that assume similar wages and/or consultation times in the two modes of care, the health care cost ratio rises to 42-52%. The only scenario in which the additional consumption implies increasing health care costs is the one with equal wages and consultation times. Notably, DCT is still resource saving from a societal perspective (i.e., when adding the value of patients' time savings) in this scenario.

We also note that in our study setting, the reimbursement per DCT consultation from the public third-party payer is around twice as high as the cost for the physician time. Although the reimbursement clearly must be high enough to cover fixed costs, this large differential

address the less tangible benefits, such as the reduced worry of DCT patients who would not otherwise see a doctor at a PCC, or the potentially higher quality of in-person consultations for patient who would.

[43]The wage differential reflects that DCT firms employ less senior staff, which is not so interesting in a general equilibrium perspective. Further, for patients who do not need to be physically examined (i.e., patients whose care need is well suited for online treatment), there are no strong reasons to believe that the consultation time would differ considerably between the two modes of care. Our choice of using a longer time in PCC in the baseline calculation reflects that the in-person setting may by itself prompt physicians to physically examine patients in borderline cases when it may not be strictly necessary, and to spend some minutes accompanying patients to/from the waiting room.

[44]To realize savings from costs for management, lab equipment and a receptionist in the in-person setting, the PCC would have to shut down. In our setting, it would be politically infeasible to realize such savings on a large scale. It thus makes sense to abstract from these costs. When it comes to fixed costs that are specific for the DCT setting, the key costs stem from the development and maintenance of the consultation platform. For the purpose of comparing the costs to produce care in the two settings, it is less consequential than it may seem to omit these costs, as PCCs nowadays also offer online consultations using similar platforms. Although, of course, the overall costs would have been lower if there was only one platform.

suggests that the health care authorities would be able to reduce the reimbursement and still make it worthwhile for the larger DCT companies to stay in the market.

5 Concluding remarks

We show that the demand for DCT consultations falls by half as individuals reach the age at which they start to pay a consultation fee. Exploiting this exogenous variation in demand, we estimate the degree to which online consultations substitute for in-person physician consultations at the primary care practice. Our estimates imply that slightly less than 50% of all online consultations replace in-person visits. Consequently, the availability of a direct-to-consumer telemedicine market increases the total number of physician consultations (online and in-person). This conclusion is robust to different bandwidths, functional form assumptions, and methods for estimating standard errors.

We do not find that the substitution rate is limited to a great extent by DCT consultations that only function as an additional step before an in-person visit. This suggests that the additional demand primarily stems from the lower barriers of access to DCT services. The fact that half of the additional consultations are one-off encounters with no follow-up within 14 days aligns with the notion that the convenience of DCT may increase the propensity to seek care for less severe conditions. The particularly low substitution for diagnoses related to skin conditions and reproductive health, which would normally be dealt with by other professions in Swedish primary care (or by referring the patient to self-care), is also indicative of lower barriers to access.

Compared to previous related research, an important strength of the study is that our identification strategy is better equipped to deal with time-variant unobserved heterogeneity such as transitory health issues. In line with this, we also estimate a relatively high degree of substitution. However, the differences may also relate to the external validity of our analysis.

Here, the study population and the study context are of particular concern.

With regards to the study population, our identification strategy captures causal effects for individuals around 20 years of age that respond to the DCT user fee. A complier analysis suggests that groups that are more frequent users of DCT in the first place (women, urban residents, and users of in-person care), are also more likely to respond to the onset of the user fee. As shown in Appendix D, the studied age group is similar to the (rest of the) 15-34 age group – a large fraction of DCT users – in terms of the proportion with a DCT-relevant diagnosis and expected health care spending. It thus appears reasonable to generalise our results at least to other young adults.

The Covid-19 pandemic has increased the tendency of both patients and health care providers to adopt a 'digital-first' approach, and might thus have affected the degree of substitution. However, our results are likely relevant also in the post-pandemic era, as the young individuals we study were among the most prominent early adopters already before the pandemic.

Across contexts, differences in incentives facing patients and provider may affect the degree of substitution. On the patient side, the setting with heavily subsidised access to online as well as traditional care is descriptive of many health systems. On the provider side, the Swedish health care system is characterised by relatively few primary care physicians, task shifting towards nurses, and a strong reliance on fixed payments rather than fee-for-service. The scope for substitution might be greater in contexts where such factors, which limit the access to physician consultations, are less pronounced. Notably though, we do not find evidence of much more substitution when considering nurse and physician consultations together, and the results are similar in our two study regions, which have different incentive structures. These results suggest that our main findings are relevant outside the particular study context.

Our preferred estimate of the substitution rate implies that if the unit cost of a DCT consultation is at most 45% of the unit cost of an in-person visit, then the DCT market has the potential to increase the care volume while leaving total costs unaffected. Our back-of-the-envelope estimates suggest that the production cost of a DCT consultation is likely low enough to enable this criterion to be fulfilled, and almost surely so when taking into account the value of the time saved by patients.

However, due to the (positive) gap between the reimbursement and production costs in our study setting, third-party payers likely pay more than they need for DCT consultations at current care volumes. The overpayments to DCT providers have thus come at the expense of individuals with greater health needs, for whom online consultations are inadequate (Roland, 2019), and, to the extent that total costs have been allowed to increase, at the expense of the insured population (tax payers). Arguments related to both efficiency (deadweight loss of taxes) and equity (larger needs of patients unsuitable for DCT (Roland, 2019)) might justify another distribution of the gains from DCT, which at present fall exclusively on DCT providers and on DCT users.

From a health policy perspective, it may also be a concern that about half of all spending on DCT goes to patients who would not otherwise have seen a doctor.[45] A potential policy implication is that the regulation of the DCT market should incorporate similar incentives for nurse triage as in traditional primary care. A further step in that direction might be to apply higher reimbursement rates for consultations with higher priority, or with the cheapest equivalent staff category.

References

AHIP. 2019. "Virtual Care Delivers Value." https://www.ahip.org/virtual-care-delivers-value/ Accessed: 2021-05-10.

[45] In the Swedish case, this would imply that 1.5% of the expenditure on primary care physician consultations in 2019 was redirected in this fashion.

Anell, Anders, Anna H Glenngård, and Sherry M Merkur. 2012. "Sweden: Health system review." *Health systems in transition*, 14(5): 1–159.

Ashwood, J. Scott, Ateev Mehrotra, David Cowling, and Lori Uscher-Pines. 2017. "Direct-To-Consumer Telehealth May Increase Access To Care But Does Not Decrease Spending." *Health Affairs*, 36(3): 485–491.

Bavafa, Hessam, Lorin M. Hitt, and Christian Terwiesch. 2018. "The Impact of E-Visits on Visit Frequencies and Patient Health: Evidence from Primary Care." *Management Science*, 64(12): 5461–5480.

Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad. 2017. "What does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics*." *The Quarterly Journal of Economics*, 132(3): 1261–1318.

Bundorf, M Kate. 2012. "Consumer-directed health plans: Do they deliver?" The Robert Wood Johnson Foundation Research Synthesis Report 24.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica*, 82(6): 2295–2326.

Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs: Foundations.* Cambridge:Cambridge University Press.

Chandra, Amitabh, Jonathan Gruber, and Robin McKnight. 2014. "The impact of patient cost-sharing on low-income populations: Evidence from Massachusetts." *Journal of Health Economics*, 33: 57–66.

Cutler, David M., Sayeh Nikpay, and Robert S. Huckman. 2020. “The Business of Medicine in the Era of COVID-19.” *JAMA*, 323(20): 2003.

Dahlgren, Cecilia, Margareta Dackehag, Per Wändell, and Clas Rehnberg. 2021. “Determinants for use of direct-to-consumer telemedicine consultations in primary healthcare—a registry based total population study from Stockholm, Sweden.” *BMC family practice*, 22(1): 1–10.

Dills, Angela K, and Jiajia Chen. 2018. “Does Telemedicine Save Lives? Evidence on the Effect of Telemedicine Parity Laws on Mortality Rates.” Working Paper, Available at SSRN: https://ssrn.com/abstract=3272821.

Ekman, Björn. 2018. “Cost Analysis of a Digital Health Care Model in Sweden.” *PharmacoEconomics - Open*, 2(3): 347–354.

Ellegård, Lina Maria, and Gustav Kjellsson. 2019. “Nätvårdsanvändare i Skåne kontaktar oftare vårdcentral och gör inte färre akutbesök.” *Läkartidningen*, 116.

Entezarjou, Artin, Maria Sjöbeck, Patrik Midlöv, Veronica Milos Nymberg, Lina Vigren, Ashkan Labaf, Ulf Jakobsson, and Susanna Calling. 2022. “Health care utilization following “digi-physical” assessment compared to physical assessment for infectious symptoms in primary care.” *BMC Primary Care*, 23(1): 4.

Entezarjou, Artin, Susanna Calling, Tapomita Bhattacharyya, Veronica Milos Nymberg, Lina Vigren, Ashkan Labaf, Ulf Jakobsson, and Patrik Midlöv. 2021. “Antibiotic Prescription Rates After eVisits Versus Office Visits in Primary Care: Observational Study.” *JMIR Medical Informatics*, 9(3): e25473.

Galindo-Silva, Hector, Nibene Habib Some, and Guy Tchuente. 2019. “Does Obamacare Care? A Fuzzy Difference-in-discontinuities Approach.”

Grecu, Anca M, and Ghanshyam Sharma. 2019. "The effect of telehealth insurance mandates on health-care utilization and outcomes." *Applied Economics*, 51(56): 5972–5985.

Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano. 2016. "Do Fiscal Rules Matter?" *American Economic Journal: Applied Economics*, 8(3): 1–30.

Heckley, Gawain, Ulf-G. Gerdtham, and Johan Jarl. 2018. "Too Young to Die: Regression Discontinuity of a Two-Part Minimum Legal Drinking Age Policy and the Causal Effect of Alcohol on Health." Working Paper, Department of Economics, Lund University 2018:4.

He, Yang, and Otavio Bartalotti. 2020. "Wild Bootstrap for Fuzzy Regression Discontinuity Designs: Obtaining Robust Bias-Corrected Confidence Intervals." *Econometrics Journal*, 23(2): 211–231.

Johansson, Naimi, Niklas Jakobsson, and Mikael Svensson. 2019. "Effects of primary care cost-sharing among young adults: varying impact across income groups and gender." *The European Journal of Health Economics*, 20(8): 1271–1280.

Kry. 2019. "Kvalitetsrapport januari–september 2019 Sverige." https://www.kry.se/medicinsk-kvalitet/kvalitetsrapport-q1-q3-2019/ Last accessed: 2021-06-04.

Lee, David S., and David Card. 2008. "Regression discontinuity inference with specification error." *Journal of Econometrics*, 142(2): 655–674.

Li, Kathleen Yinran, Ziwei Zhu, Sophia Ng, and Chad Ellimoottil. 2021. "Direct-To-Consumer Telemedicine Visits For Acute Respiratory Infections Linked To More Downstream Visits." *Health Affairs (Project Hope)*, 40(4): 596–602.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz. 1987. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." *The American Economic Review*, 77(3): 251–277. Publisher: American Economic Association.

Martinez, Kathryn A., Mark Rood, Nikhyl Jhangiani, Lei Kou, Susannah Rose, Adrienne Boissy, and Michael B. Rothberg. 2018. "Patterns of Use and Correlates of Patient Satisfaction with a Large Nationwide Direct to Consumer Telemedicine Service." *Journal of General Internal Medicine*, 33(10): 1768–1773.

Mehrotra, Ateev, R. Sacha Bhatia, and Centaine L. Snoswell. 2021. "Paying for Telemedicine After the Pandemic." *JAMA*, 325(5): 431.

Millán-Quijano, Jaime. 2020. "Fuzzy difference in discontinuities." *Applied Economics Letters*, 27(19): 1552–1555.

Newhouse, Joseph P., Rand Corporation Insurance Experiment Group, and Insurance Experiment Group Staff. 1993. *Free for All?: Lessons from the Rand Health Insurance Experiment.* Harvard University Press. Google-Books-ID: SVUJ4W9Lk5IC.

Nilsson, Anton, and Alexander Paul. 2018. "Patient cost-sharing, socioeconomic status, and children's health care utilization." *Journal of Health Economics*, 59: 109–124.

Nord, Garrison, Kristin L. Rising, Roger A. Band, Brendan G. Carr, and Judd E. Hollander. 2018. "On-demand synchronous audio video telemedicine visits are cost effective." *The American Journal of Emergency Medicine*.

North, Frederick, Sarah J. Crane, Rajeev Chaudhry, Jon O. Ebbert, Karen Ytterberg, Sidna M. Tulledge-Scheitel, and Robert J. Stroebel. 2014. "Impact of patient portal secure messages and electronic visits on adult primary care office visits." *Telemedicine Journal and E-Health*, 20(3): 192–198.

Roland, Martin. 2019. "General practice by smartphone." *BMJ*, 366.

SALAR, (Swedish Association of Local Authorities and Regions). 2017. "Sveriges Kommuner och Landsting – Sektorn i siffror." http://skl.se/ekonomijuridikstatistik/ekonomi/sektornisiffror.1821.html Last Accessed 2021-06-04.

SALAR, (Swedish Association of Local Authorities and Regions). 2020. "Digitala vårdtjänster i primärvården." https://skr.se/halsasjukvard/ehalsa/digitalavardtjansteriprimarvarden.28301.html Last Accessed: 2021-06-03.

Salisbury, Chris, Anna Quigley, Nick Hex, and Camille Aznar. 2020. "Private Video Consultation Services and the Future of Primary Care." *Journal of Medical Internet Research*, 22(10): e19415.

Shah, Sachin J., Lee H. Schwamm, Adam B. Cohen, Marcy R. Simoni, Juan Estrada, Marcelo Matiello, Atheendar Venkataramani, and Sandhya K. Rao. 2018. "Virtual Visits Partially Replaced In-Person Visits In An ACO-Based Medical Specialty Practice." *Health Affairs*, 37(12): 2045–2051.

Shi, Zhuo, Ateev Mehrotra, Courtney A. Gidengil, Sabrina J. Poon, Lori Uscher-Pines, and Kristin N. Ray. 2018. "Quality Of Care For Acute Respiratory Infections During Direct-To-Consumer Telemedicine Visits For Adults." *Health Affairs*, 37(12): 2014–2023.

Uscher-Pines, Lori, Andrew Mulcahy, David Cowling, Gerald Hunter, Rachel Burns, and Ateev Mehrotra. 2015. "Antibiotic Prescribing for Acute Respiratory Infections in Direct-to-Consumer Telemedicine Visits." *JAMA internal medicine*, 175(7): 1234–1235.

Zeltzer, Dan, Liran Einav, Joseph Rashba, and Ran D Balicer. 2021. "The Impact of Increased Access to Telemedicine." NBER Working Paper 28978.

Table 1: Summary of Institutional Setting

		RS	RVG
TRADITIONAL PCC			
	Reimbursement:	Capitation + Per visit	Capitation
	User fee age limit:	18	20
	User fee:	€20	€10/30*
DCT (ONLINE) MARKET			
	Rembursement:	Per visit	Per visit
	User fee age limit:	20	20
	User fee:	€25	€25

Note: The table describes the reimbursement system and user fees for in-person visits and DCT consultations in the two study regions – Region Stockholm (RS) and Region Västra Götaland (RVG) – during the study period. The DCT user fee was set by a third region, Region Jönköping, where the two largest DCT providers were located during the study period. * The lower fee applied if the patient visited the PCC at which (s)he was registered as a patient.

Table 2: Fuzzy difference-in-discontinuity results

	All	Men	Women
A. First stage (sharp diff-in-disc)			
Online consultations	-0.149***	-0.0700***	-0.234***
	(0.0116)	(0.0114)	(0.0231)
F-stat	163.8	37.76	103.4
B. RF (sharp diff-in-disc)			
In-person visits	0.0670**	0.0718*	0.0611
	(0.0298)	(0.0380)	(0.0463)
C. IV (fuzzy diff-in-disc)			
In-person visits	-0.450**	-1.026*	-0.260
	(0.200)	(0.549)	(0.196)
Avg. ind/day pre	151525	78130	73395
Avg. ind/day post	31490	16302	15188

Note: Variable names in left column represent the dependent variable. Each model uses a fixed bandwidth of 120 days on either side of the threshold (20th birthday). Avg. ind/day = average number of individuals per day in cells, shown separately for the pre cohorts (which include several birth cohorts) and the post cohort (which only includes individuals turning 20 in 2018). Data are collapsed by region, gender, birth year and day relative to 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Table 3: Decomposition by type of diagnosis, (bw=120)

	A. DCT consultations (sharp diff-in-disc)				
	Common	Resp	Skin	Gen/Rep	Other
All	-0.131***	-0.0260***	-0.0424***	-0.0244***	-0.0457***
	(0.0106)	(0.00553)	(0.00571)	(0.00555)	(0.00696)
Men	-0.0603***	-0.00689	-0.0214***	-0.00347*	-0.0308***
	(0.00988)	(0.00605)	(0.00578)	(0.00197)	(0.00639)
Women	-0.208***	-0.0467***	-0.0650***	-0.0470***	-0.0619***
	(0.0198)	(0.00980)	(0.00949)	(0.0113)	(0.0119)
	B. Substitution (fuzzy diff-in-disc)				
	Common	Resp	Skin	Gen/Rep	Other
All	-0.436**	-1.589***	-0.0426	0.0846	-0.504
	(0.193)	(0.553)	(0.259)	(0.339)	(0.407)
Men	-0.693	-2.336	-0.223	-1.418	-0.601
	(0.526)	(3.216)	(0.656)	(2.036)	(0.716)
Women	-0.351*	-1.465***	0.0223	0.208	-0.446
	(0.186)	(0.490)	(0.241)	(0.319)	(0.462)

Note: Table shows results from the first stage equation (sharp diff-in-disc) and the IV-model (fuzzy-diff-in-disc) by diagnosis groups. The first column shows all *Common* diagnoses set by DCT providers. In the second to fourth column, these common diagnoses are decomposed into subgroups: upper respiratory infections (*Resp*), skin related diseases (*Skin*), genital and reproductive organs *Gen/Rep*, and *Other* common diagnoses. Each row presents the diff-in-disc estimates for the diagnosis groups for the given estimation sample (All, Men, Women). Each model uses a fixed bandwidth of 120 days each side of the age cutoff. Estimates using data collapsed by region, gender, year and day relative to 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Table 4: Visits to other health care professionals

	Fuzzy Diff-in-disc		
	All	Men	Women
Nurse visits at a PCC	-0.0545	-0.605*	0.124
	(0.118)	(0.312)	(0.120)
Visits at a midwife/youth/STD clinic	-0.129	0.352*	-0.274
	(0.176)	(0.208)	(0.219)
Nurse+midwife/youth/STD	-0.183	-0.253	-0.150
	(0.211)	(0.350)	(0.249)
All consultations	-0.633**	-1.279*	-0.410
	(0.320)	(0.694)	(0.339)

Note: The table shows fuzzy diff-in-discs estimates of the effect of online consultations on in-person visits with other health care professionals than physicians at primary care centers: nurse visits at a primary care center; midwife/nurse/physician visits at a midwife/youth/STD clinic, the sum of these two types of consultations; and the sum of all consultations, including the physician visits at a primary care center we use in our main analysis. Each model uses a fixed bandwidth of 120 days each side of the threshold. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * $P<0.1$, ** $P<0.05$, *** $P<0.01$.

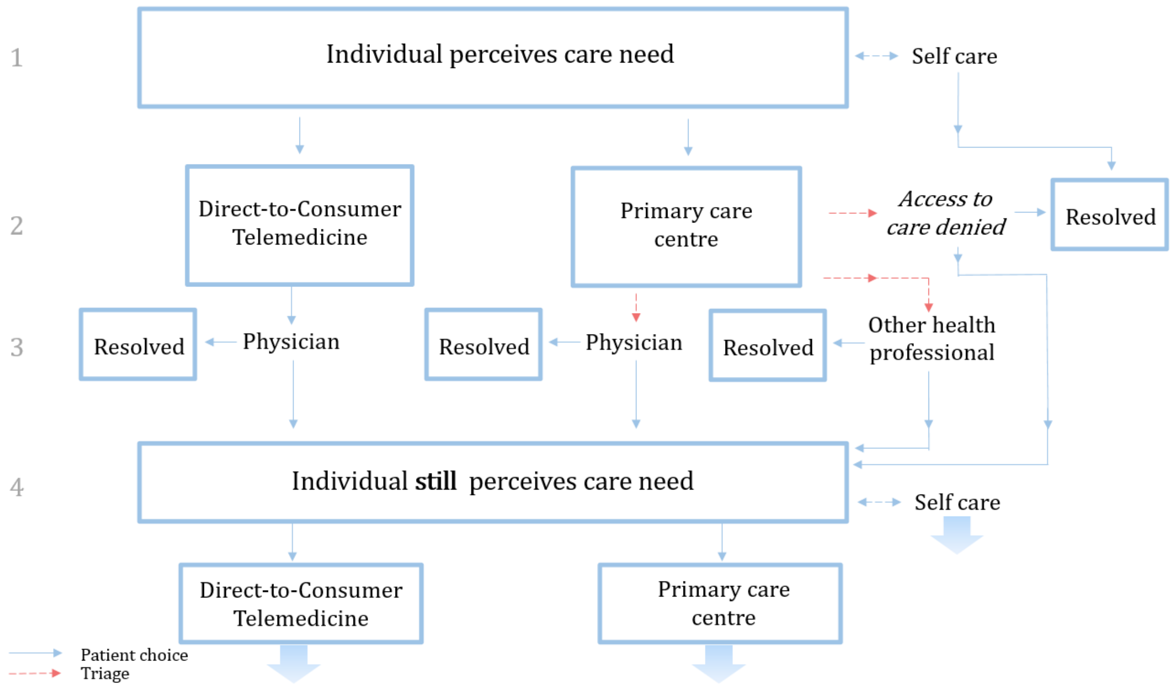


Figure 1: Conceptual framework

Note: This figure illustrates the available care paths for an individual who experiences a care need (1). The individual might try do address the issue by herself (self care). If she chooses not to, or if self care does not resolve the problem, she may contact a DCT provider or a primary care centre (2). In the first case, the patient will be offered an appointment with a physician contracting with the DCT company. In the second case, a triage nurse will examine the need for further care and either schedule an appointment with a physician or another health professional at the PCC, or deny access to further care (typically combined with some self care advice) (3). If, after this step, the patient perceives that the problem has not been resolved, she may again choose between self care or to contact a DCT provider or a PCC (possibly another one than the initial PCC) (4) and so on.

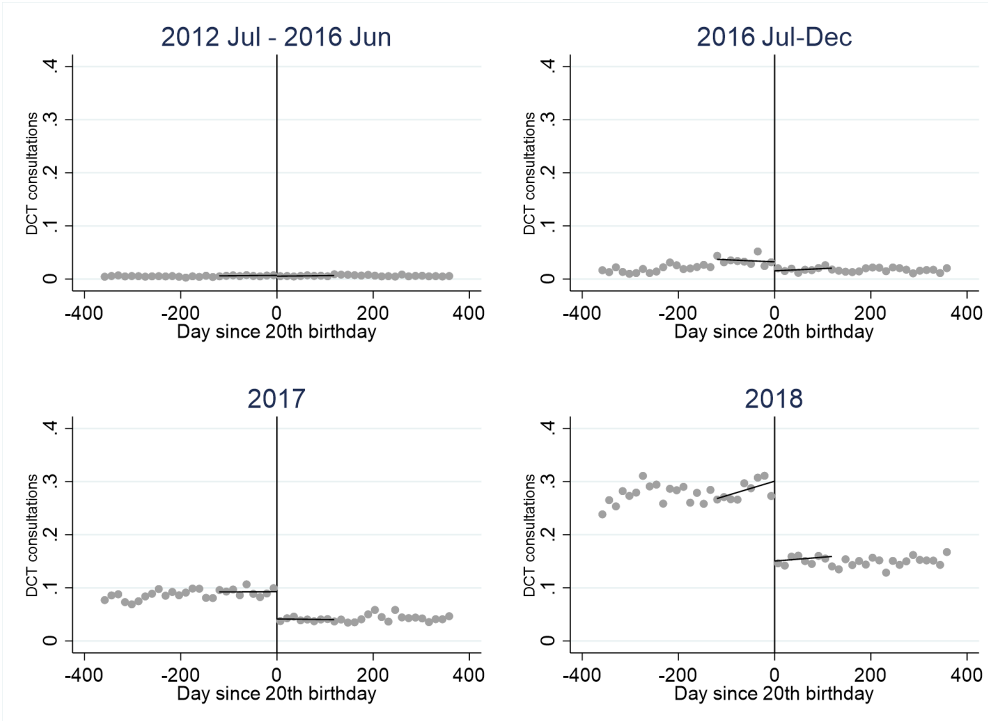


Figure 2: DCT consultations over time

Note: The graphs display the average number of DCT consultations by age for four periods. The age variable is measured as days in relation to the 20th birthday. The number of DCT consultation is rescaled to reflect annual averages, and is plotted for 730 days grouped in 52 bins.

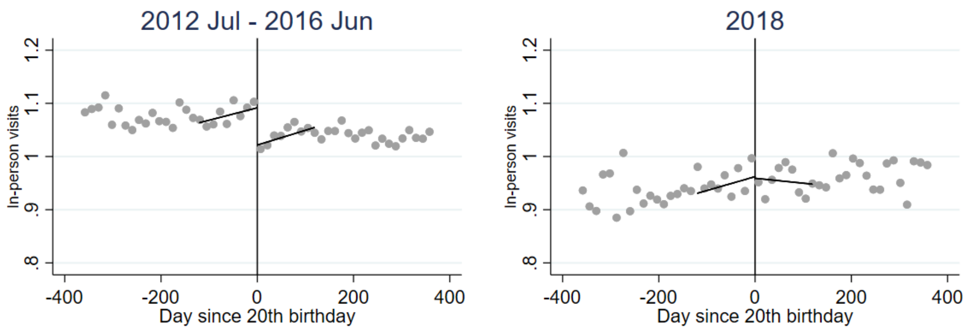


Figure 3: In-person physician visits

Note: The graphs display the average number of in-person physician visits by age. The age variable is measured as days in relation to the 20th birthday. The number of in person visits is rescaled to reflect annual averages, and is plotted for 730 days grouped in 52 bins.

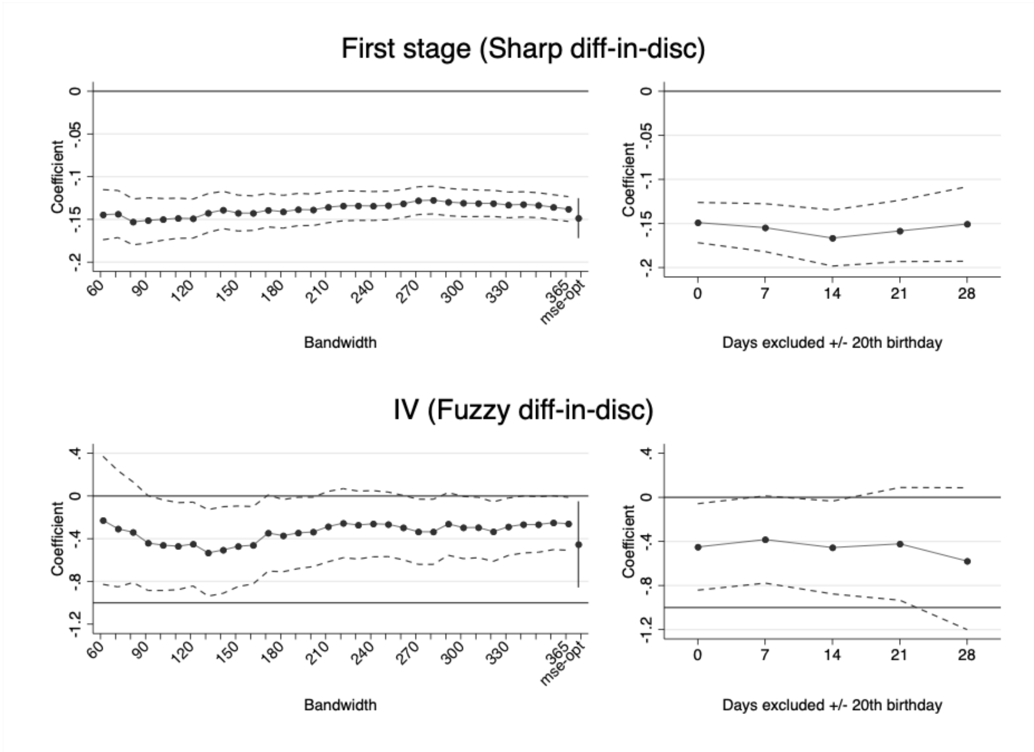


Figure 4: Robustness analyses – bandwidth and donut

Note: The upper panel shows coefficients from sharp diff-in-disc models of the effect of the user fee on DCT consultations (the first stage). The lower panel shows coefficients from fuzzy diff-in-disc models of the relationship between DCT consultations and in-person visits (IV). The graphs on the left show how the estimates vary over different bandwidhts (60–365 days + MSE-optimal bandwidth (Cattaneo, Idrobo and Titiunik, 2019)). The coefficient at ”120” is equivalent to the main specification in Table 2). The graphs on the right shows how the estimates vary when successively removing more observations around the 20th birthday (“ donut” estimations), in all cases using a bandwidth of +/- 120 days relative to the 20th birthday. Here, the leftmost coefficients (0 days excluded) are equivalent to the main estimates in Table 2).

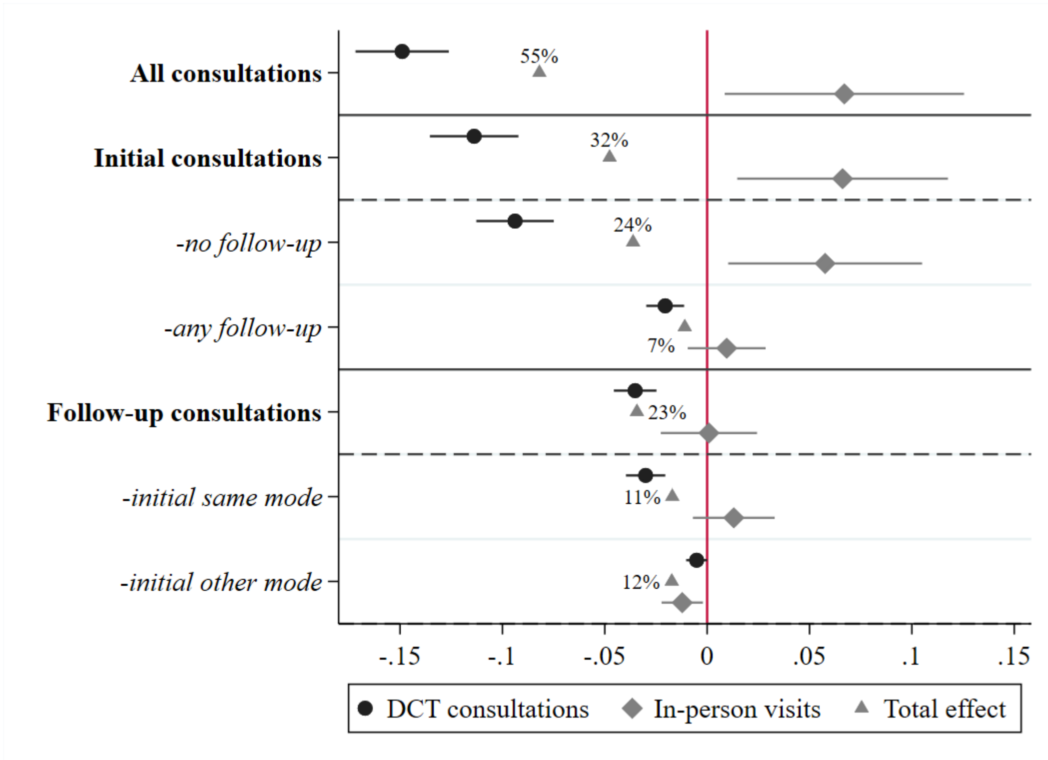


Figure 5: In-person physician visits

Note: The graph displays coefficients and 95% confidence intervals from sharp diff-in-disc models decomposing the main effect. Each model uses a bandwidth of 120 days before and after the 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. The outcome variables are All DCT (in-person) consultations, All DCT (in-person) *initial* consultations and All DCT (in-person) *follow-up* consultations. An initial consultation is a consultation for which the individual had no prior consultation during the preceding 14 days. The remaining are defined as follow-ups. Initial consultations are decomposed into consultations with *no follow-ups* and consultations with *any follow-ups*. Follow-up consultations are decomposed by the mode of the initial consultation of the episode (same/other).

Online Appendix to An App Call a Day Keeps the Patient Away?

Contents

Appendix A Comparison of DCT measures

The main data sources of our DCT measure are Stockholm and Västra Götaland's registers of billing information for their residents' care consumption in other regions. (Regions are financially responsible for their residents and get billed when their residents are treated by providers in other regions.) The DCT variable in the main analysis includes

- all primary care contacts in the Sörmland and Jönköping regions.
- contacts with private DCT providers in the Skåne region (Capio Go, until May 2018) and a public DCT provider in Västra Götaland (Närhälsan Online).

The vast majority of the DCT contacts come from Region Jönköping, where the largest DCT providers were located at the time (Min Doktor, Kry, Doktor 24, Medicoo, Accumbo, and from June 2018 Capio Go). Doktor.se was the only provider located in Region Sörmland before 2019.

In Tables A.1, we examine how sensitive the estimated first stage sharp diff-in-disc is to various definitions of the DCT variable. In column 1, the definition of the DCT variable is the same as in the main analysis. This variable is based on the billing information, and includes all primary care visits in Sörmland and Jönköping. In the next two columns, we use the same definition as in column 1 but include only consultations registered in Region Sörmland (column 2) or Region Jönköping (column 3). In column 4, we apply the same definition of DCT contacts, but use information from register data from Region Jönköping (instead of billing information from Västra Götaland and Stockholm). In column 5, the DCT definition only includes registered remote contacts at the private PCCs in Jönköping that have agreements with DCT providers.[1] In column 6, we further restrict this definition to only include remote consultations with physicians.

The estimates are similar across outcomes, except when we look at contacts in Sörmland only (column 2). The similarity between the other coefficients implies that the discontinuity at age 20 primarily affected the number of DCT contacts with providers located Region Jönköping. This pattern is explained by the age differentiated DCT user fee in Region Jönköping; in Region Sörmland the user fee was zero for all ages groups. Notably, only one minor provider had an agreement with a primary care center in Region Sörmland at this time period, and this

[1] It was the agreements with these PCCs that enabled the DCT get public funding via the inter-regional agreement.

provider had a nurse triage system in place making the patient less likely to see a physician. The positive coefficient in column 2 indeed suggests that the user fee in Region Jönköping, if anything, had a reverse (although small) effect on the number of DCT contacts in Region Sörmland. This is also supported by the coefficients in column 3 and 4 being larger than in column 1.

The similarity between columns 1, 3, 4, and 5 suggests that the billing information from the home regions own registers provides an accurate measure of the DCT contacts and a reliable estimate of the effect of the reduction of the DCT user fee at age 20. (The estimated degree of substitution (fuzzy diff-in-disc) is also very similar when using the data from Jönköping to measure DCT contacts) The small difference between column 5 and 6 further suggests that only about 10% of the first stage coefficient relates to changes in consultations with other health care professionals than physicians.

Table A.1: Variation of DCT definitions

	DCT	DCT-Smld	DCT-Jkpg1	DCT-Jkpg2	DCT-digpr	DCT-phys
	BOTH REGIONS (2018)					
Both	-0.149***	0.00818**	-0.158***	-0.163***	-0.161***	-0.151***
	(0.0116)	(0.00392)	(0.0108)	(0.0113)	(0.0111)	(0.0106)
Men	-0.0700***	0.00124	-0.0693***	-0.0674***	-0.0668***	-0.0622***
	(0.0114)	(0.00481)	(0.0100)	(0.0100)	(0.0100)	(0.0101)
Women	-0.234***	0.0156**	-0.253***	-0.266***	-0.262***	-0.247***
	(0.0231)	(0.00757)	(0.0213)	(0.0220)	(0.0215)	(0.0206)
	REGION STOCKHOLM (2018)					
Both	-0.158***	0.0128**	-0.170***	-0.177***	-0.176***	-0.163***
	(0.0174)	(0.00593)	(0.0168)	(0.0175)	(0.0174)	(0.0161)
Men	-0.0783***	-0.000230	-0.0781***	-0.0773***	-0.0804***	-0.0710***
	(0.0166)	(0.00792)	(0.0151)	(0.0155)	(0.0156)	(0.0151)
Women	-0.243***	0.0266***	-0.270***	-0.284***	-0.278***	-0.262***
	(0.0329)	(0.0102)	(0.0312)	(0.0319)	(0.0314)	(0.0294)
	REGION VÄSTRA GÖTALAND (2018)					
Both	-0.138***	0.00240	-0.142***	-0.145***	-0.143***	-0.136***
	(0.0174)	(0.00512)	(0.0152)	(0.0158)	(0.0154)	(0.0136)
Men	-0.0596***	0.00308	-0.0585***	-0.0551***	-0.0501***	-0.0513***
	(0.0163)	(0.00584)	(0.0138)	(0.0139)	(0.0137)	(0.0137)
Women	-0.224***	0.00166	-0.232***	-0.244***	-0.243***	-0.228***
	(0.0317)	(0.00894)	(0.0278)	(0.0287)	(0.0274)	(0.0251)

Note: Table shows sharp diff-in-disc results for various definitions of DCT consultations using a bandwidth of 120 days. Each column presents the sharp diff-in-disc estimates for the a given outcome for each estimation sample (All, Men, Women) for both regions jointly, and separately. Each model uses MSE-optimal bandwidths for in-person visits used in the main estimations for both genders and regions jointly. Estimates using data collapsed by region, gender, year and day relative to 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Appendix B List of diagnosis categories

We define the most common diagnoses using information on registered ICD-codes from the local care register in Region Jönköping. These data exactly identify the online-consultations with physicians at DCT providers in Jönköping, but have the limitation that they do not include any consultations with DCT providers located elsewhere. Each care contact may have up to five registered diagnoses classified according to the ICD-10. The data include the complete ICD-code, but we define the most common diagnoses categories on a three digit level.

To define the most common diagnoses, we first generate a list of all complete ICD-codes and the corresponding number of recorded registrations at online consultations from 2016 to 2018 (by each gender). We then sum the number of registered diagnoses within each three digit ICD-code, and rank these three digit codes by the number of registrations (again by gender). A three digit ICD-code is defined as being among the most common diagnoses if it is included among the top 80 % of all registered diagnoses for either male or female individuals. For non-administrative ICD-codes (i.e., any ICD-code not in Z00-Z99), we also include three-digit ICD-codes that belongs to the same ICD-block (subchapter) as any of the most common ICD-codes and had been registered at least 10 times during the study period.

We classify the set of common diagnoses into four subsets: upper respiratory infections, skin conditions, diagnoses related to genital and reproductive organs, and a residual category (other). Table B.1, B.2, B.3, and B.4 list the three digit level ICD-codes in each subset. In 2018, 90 % of the online consultations with a physician had at least one registered ICD-code that is included in the definition of the most common diagnoses. The four categories respectively covered 19% 25%, 22%, and 28% of all online consultations in the same year. Because each consultation may have more than one diagnosis, the contacts in these subsets are not completely mutually exclusive.

Table B.1: Upper respiratory infections

icd	icd-block	Description
B27	B25 -B34	Infectious mononucleosis
B30	B25 -B34	Viral conjunctivitis
B34	B25 -B34	Viral infection of unspecified site
J00	J00 -J06	Acute nasopharyngitis [common cold]
J01	J00 -J06	Acute sinusitis
J02	J00 -J06	Acute pharyngitis
J03	J00 -J06	Acute tonsillitis
J06	J00 -J06	Acute upper respiratory infections of multiple and unspecified sites
J30	J30 -J39	Vasomotor and allergic rhinitis
J31	J30 -J39	Chronic rhinitis, nasopharyngitis and pharyngitis
J35	J30 -J39	Chronic diseases of tonsils and adenoids
J45	J40 -J47	Asthma
R05	R00 -R09	Cough

Table B.2: Skin related conditions

icd	icd block	Description
B00	B00 -B09	Herpesviral [herpes simplex] infections
B02	B00 -B09	Zoster [herpes zoster]
B07	B00 -B09	Viral warts
B08	B00 -B09	Other viral infections characterized by skin and mucous membrane lesions, not elsewhere classified
B35	B35 -B49	Dermatophytosis
B36	B35 -B49	Other superficial mycoses
B37	B35 -B49	Candidiasis (excluding B373, B373P, or B374)
L01	L00 -L08	Impetigo
L02	L00 -L08	Cutaneous abscess, furuncle and carbuncle
L03	L00 -L08	Cellulitis
L08	L00 -L08	Other local infections of skin and subcutaneous tissue
L20	L20 -L30	Atopic dermatitis
L21	L20 -L30	Seborrhoeic dermatitis
L23	L20 -L30	Allergic contact dermatitis
L29	L20 -L30	Pruritus
L30	L20 -L30	Other dermatitis
L50	L50 -L54	Urticaria
L60	L60 -L75	Nail disorders
L63	L60 -L75	Alopecia areata
L64	L60 -L75	Androgenic alopecia
L65	L60 -L75	Other nonscarring hair loss
L70	L60 -L75	Acne
L71	L60 -L75	Rosacea
L73	L60 -L75	Other follicular disorders
R21	R20 -R23	Rash and other nonspecific skin eruption
R22	R20 -R23	Localized swelling, mass and lump of skin and subcutaneous tissue
R23	R20 -R23	Other skin changes

Table B.3: Genital & reproductive organs

icd	icd block	Description
B37	B35 -B49	Candidiasis (only B373, B373P, or B374)
F52	F50 -F59	Sexual dysfunction, not caused by organic disorder or disease
N30	N30 -N39	Cystitis
N39	N30 -N39	Other disorders of urinary system
N76	N70 -N77	Other inflammation of vagina and vulva
N77	N70 -N77	Vulvovaginal ulceration and inflammation in diseases classified elsewhere
N92	N80 -N98	Excessive, frequent and irregular menstruation
N94	N80 -N98	Pain and other conditions associated with female genital organs and menstrual cycle
R10	R10 -R19	Abdominal and pelvic pain
Y42	Y40 -Y59	Hormones and their synthetic substitutes and antagonists, not elsewhere classified
Z30	Z	Contraceptive management
Z92	Z	Personal history of medical treatment

Table B.4: Other common diagnoses

icd	icd-block	Description
A08	A00 -A09	Viral and other specified intestinal infections
A09	A00 -A09	Other gastroenteritis and colitis of infectious and unspecified origin
A69	A65 -A69	Other spirochaetal infections
B80	B65 -B83	Enterobiasis
F32	F30 -F39	Depressive episode
F33	F30 -F39	Recurrent depressive disorder
F40	F40 -F48	Phobic anxiety disorders
F41	F40 -F48	Other anxiety disorders
F42	F40 -F48	Obsessive-compulsive disorder
F43	F40 -F48	Reaction to severe stress, and adjustment disorders
F45	F40 -F48	Somatoform disorders
F51	F50 -F59	Nonorganic sleep disorders
G43	G40 -G47	Migraine
G44	G40 -G47	Other headache syndromes
G47	G40 -G47	Sleep disorders
H10	H10 -H13	Conjunctivitis
K12	K00 -K14	Stomatitis and related lesions
K13	K00 -K14	Other diseases of lip and oral mucosa
K14	K00 -K14	Diseases of tongue
K21	K20 -K31	Gastro-oesophageal reflux disease
K29	K20 -K31	Gastritis and duodenitis
K30	K20 -K31	Functional dyspepsia
M54	M50 -M54	Dorsalgia
M79	M70 -M79	Other soft tissue disorders, not elsewhere classified
R00	R00 -R09	Abnormalities of heart beat
R06	R00 -R09	Abnormalities of breathing
R07	R00 -R09	Pain in throat and chest
R11	R10 -R19	Nausea and vomiting
R19	R10 -R19	Other symptoms and signs involving the digestive system and abdomen
R50	R50 -R69	Fever of other and unknown origin
R51	R50 -R69	Headache
R52	R50 -R69	Pain, not elsewhere classified
R53	R50 -R69	Malaise and fatigue
R61	R50 -R69	Hyperhidrosis
T14	T08 -T14	Injury of unspecified body region
T38	T36 -T50	Poisoning by hormones and their synthetic substitutes and antagonists, not elsewhere classified
T78	T66 -T78	Adverse effects, not elsewhere classified
W57	W55 -W65	Bitten or stung by nonvenomous insect and other nonvenomous arthropods
W64	W55 -W65	Exposure to other and unspecified animate mechanical forces
X58	X58 -X59	Exposure to other specified factors
Z00	Z	General examination and investigation of persons without complaint and reported diagnosis
Z02	Z	Examination and encounter for administrative purposes
Z03	Z	Medical observation and evaluation for suspected diseases and conditions
Z71	Z	Persons encountering health services for other counselling and medical advice, not elsewhere classified
Z76	Z	Persons encountering health services in other circumstances

Appendix C Descriptive statistics

Table C.1 shows descriptive statistics for three groups: the subset of individuals who were 20 ($\pm$ 365 days) years old at any point in 2018 and had at least one DCT consultation in that year (*DCT users 2018*); individuals in the same cohort who did not have a DCT consultation in 2018 (*Non-users 2018)*; and individuals who were 20 ($\pm$ 365 days) years old at any point during the four year period before the emergence of the DCT market, i.e., from July 2012 to June 2016 (*cohorts 2012-2016*).

The first row shows the average annual number of in-person physician visits (our main outcome variable). DCT users visited a physician more often than non-users, which may reflect a generally greater propensity to seek care or just a worse health status.[2] Further below in the table, we also note that the proportion of individuals who did not visit a physician during their 19th life year (0 Phys vis 18) was lower among DCT users than in other groups. Women are very much over-represented among DCT users, who also tend to have parents with higher socioeconomic background than non-users in the same as well as the earlier cohorts. A larger fraction of DCT users live in Region Stockholm (vs. Region Västra Götaland), and a larger fraction of DCT users lives in a large city (Stockholm or Gothenburg) rather than a town or a rural area.

[2] A similar pattern was also demonstrated for the study population in Ellegård and Kjellsson (2019).

Table C.1: Descriptives across cohorts and DCT users

	COHORT 2018 (DCT USERS)		COHORT 2018 (NON-USERS)		COHORTS 2012-2016 (NON-USERS)	
	Mean	SD	Mean	SD	Mean	SD
In-person phys visits	1.462	23.160	0.889	18.063	1.053	19.659
0 Phys vis 18	0.381	0.486	0.521	0.500	0.475	0.499
1 Phys vis 18	0.264	0.441	0.246	0.431	0.251	0.434
$\geq$ 2 Phys vis 18	0.356	0.479	0.233	0.423	0.274	0.446
Female	0.727	0.445	0.455	0.498	0.484	0.500
Share Sthlm	0.631	0.483	0.547	0.498	0.542	0.498
Resides Rural	0.071	0.256	0.103	0.303	0.109	0.311
Resides Town	0.171	0.377	0.231	0.422	0.238	0.426
Resides City	0.758	0.428	0.666	0.472	0.654	0.476
Dad uni	0.408	0.492	0.377	0.485	0.352	0.477
Mum uni	0.488	0.500	0.464	0.499	0.428	0.495
Mum inc > median	0.496	0.500	0.426	0.494	0.404	0.491
Dad inc > median	0.628	0.483	0.570	0.495	0.566	0.496
Parent non-nordic	0.242	0.429	0.263	0.440	0.228	0.419
Observations	2,366,695		20,666,406		110,131,616	
Unique Indvs	8,058		86,844		236,601	

Note: Descriptive statistics are presented here with sample split by DCT users 2018 and non-DCT users. Non-users are also presented by cohort; those who were 20 (± 365 days) years old at any point in the period before DCT being introduced (June 2012- July 2016), and those who were 20 (± 365 days) years old at any point in 2018. (These include individuals who potentially consumed DCT 2016 or 2017). Mean and standard deviation of each variable are based on the number of observations (individual and day) within the period. An individual is observed at most 730 days around the 20th birthday. 'In-person phys visits' is our main outcome variable; any in-person visit to a physician which has been scaled to a per capita annual basis. The 0/1/$\geq$2 Phys vis 18 variables are dichotomous indicators of the number of physician visits during the individual's 19th life year.

Appendix D External validity

To obtain an idea of the generalisability of our results, we use care register data for 2015 to compare one of the pre-DCT cohorts in our study population with other age groups with respect to a set of health measures. Specifically, we compare the study cohort in period -1 – who turned 20 in the second half of 2015 or the first half of 2016 – to individuals who resided in the study regions throughout 2015 (and did not move between the two regions). The reason why we use pre-DCT data for this exercise is that DCT, to the extent that it affected care utilisation, may have had different impact on the care utilisation of our post-DCT study cohort and on other age groups.

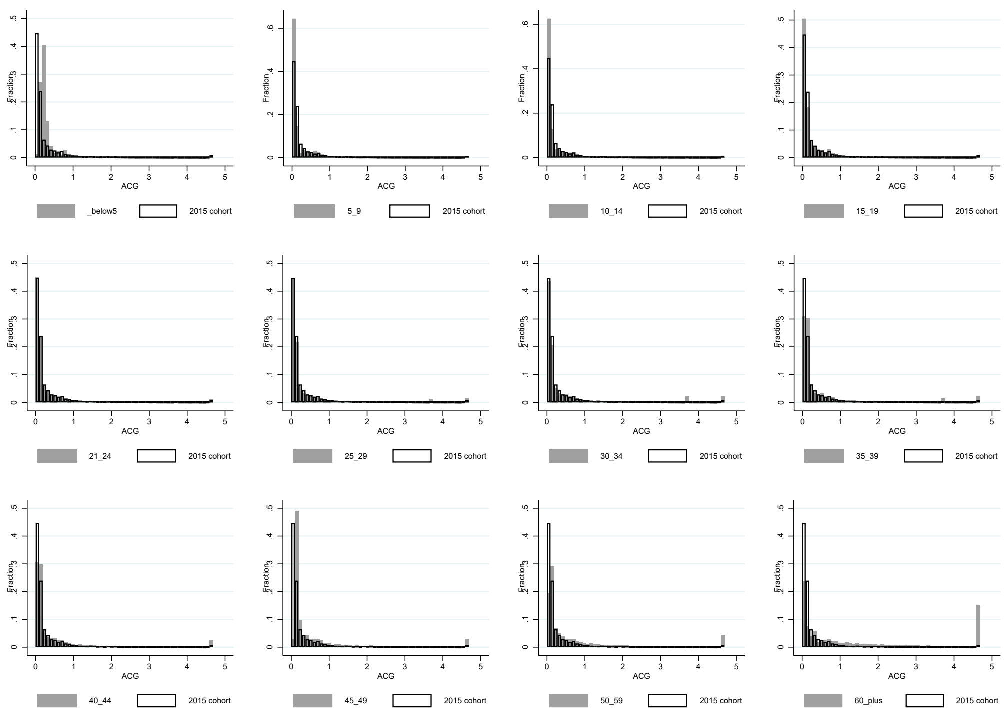
Our first health measure is the individual's predicted health care costs according to the Johns Hopkins ACG(R) System (v 11). This software, which is used for risk-adjustment in many settings (including Swedish primary care), uses diagnoses recorded in care registers to group individuals into *Adjusted Clinical Group* (ACG) with similar expected costs (similar to DRGs). The software produces an value for each individual showing the expected costs relative to the average costs in the region. Due to the presence of outliers, we recode the ACG values of individuals above the 95th percentile or below so that they get the ACG of the 95th percentile (using the *winsor2* package in Stata).

Figure D.1 compares the ACG values of the 2015 cohort (empty bars) to the other residents in 2015 in various age groups. As seen from the figure, the study cohort is very similar to the 21-34 age groups in terms of expected health care costs, and quite similar to the 15-19 age group. The study cohort is less similar to children <15 (in particular 0-4 year-olds) and to individuals above 35.

Secondly, we compare the groups with respect to the share of individuals who had been diagnosed with at least one of the most common DCT-related diagnoses (the "All" category in Table J.5) in primary care in 2015. Notably, this group only includes individuals who visited a (non-DCT) primary care provider in 2015. Figure D.2 shows that roughly 40% of the study cohort received such a diagnosis in 2015, which is similar to most age groups except the very youngest and oldest age groups.

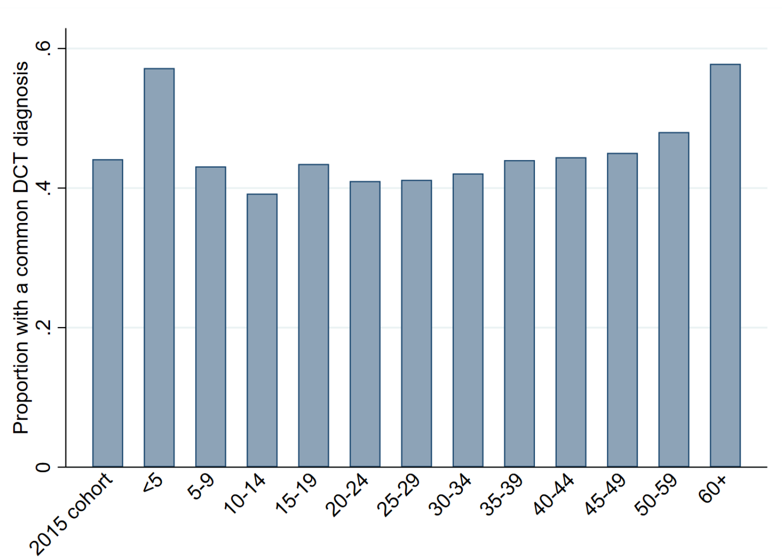
We then look at each of the four categories of common DCT diagnoses (*resp*, *gen/repr*, *skin*, *other*). The proportions with a diagnosis are generally similar in the study cohort and the 15-29 age group, and often also for other age groups.

Figure D.1: ACG distribution



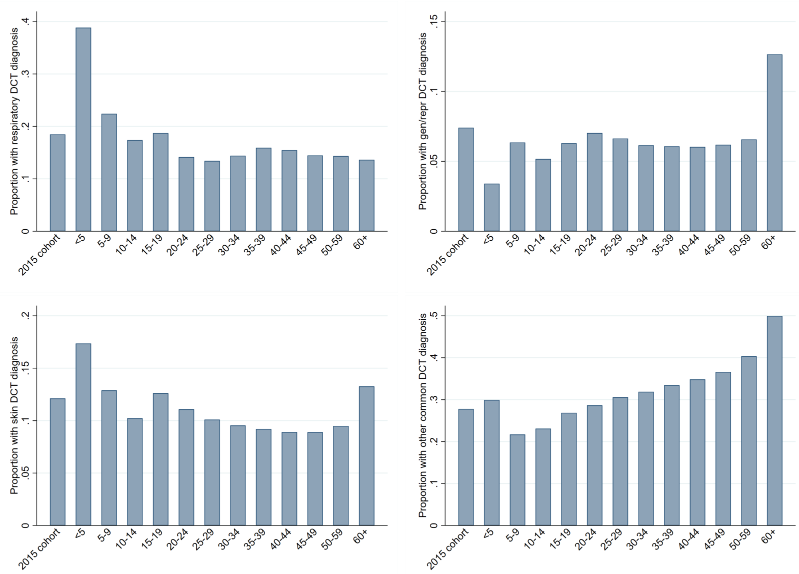
Note: The figure shows the distribution of ACG values in a pre-DCT cohort ("2015 cohort") and the general population in 2015, by age group. ACG captures expected health care costs and is based on diagnoses set during the year. Due to extreme values, the ACG variable is winsorised at the 95th percentile.

Figure D.2: Share with a DCT-relevant diagnosis



Note: The figure shows the proportion of individuals who were diagnosed with one of the most common diagnoses in DCT in traditional primary care in 2015. The leftmost bar ("2015 cohort") shows the proportion for the youngest pre-DCT cohort.

Figure D.3: Share with a DCT-relevant diagnosis; by category



Note: The figure shows the proportion of individuals who were diagnosed with a diagnoses in our *resp*, *gen/rep*, *skin* and *other* diagnosis categories in traditional primary care in 2015. The leftmost bar ("2015 cohort") shows the proportion for the youngest pre-DCT cohort.

Appendix E Results for 2017 cohorts

E.1 Graphical evidence of 2017 cohort

Figure E.1 plots annual in-person physician visits per capita, by the day relative to the 20th birthday (i.e., the DCT user fee threshold). The first panel shows the graph for the whole sample and the next two by gender. The middle and rightmost sub-graphs illustrate the two years after the DCT market emerged (2017 and 2018), and the leftmost sub-graphs show the pooled pre-period (Figure H.5 shows that the pattern is similar for each of the years in the pre-period).

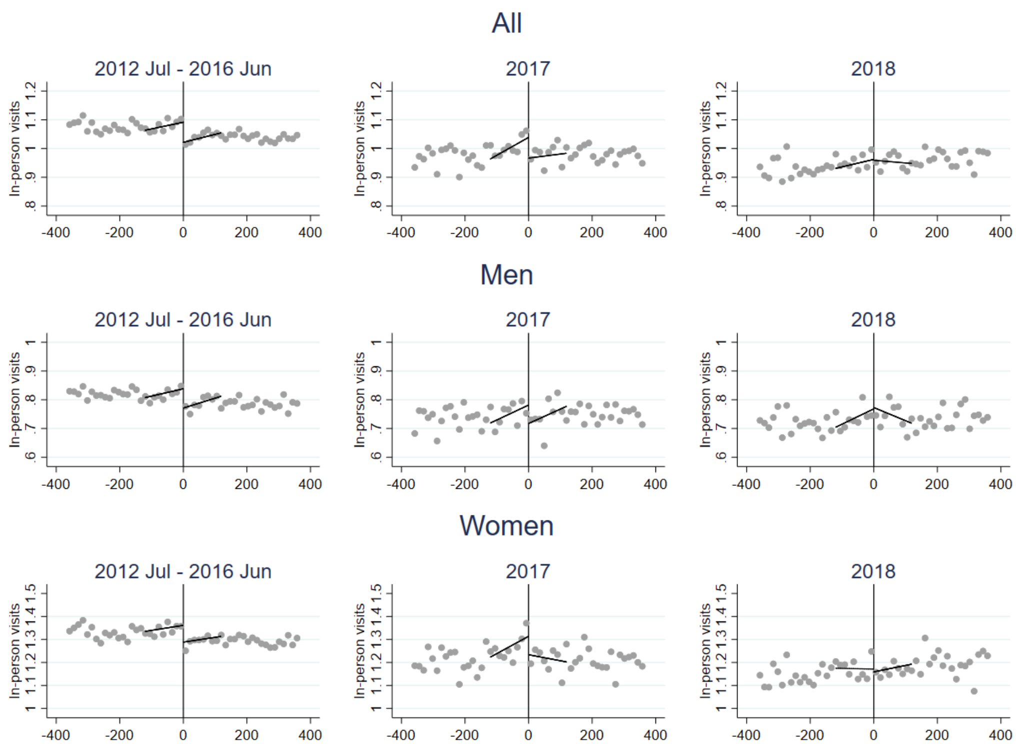


Figure E.1: In-person physician visits

Note: The graphs display the average number of in-person physician visits by age. The age variable is measured as days in relation to the 20th birthday. The number of in person visits is rescaled to reflect annual averages, and is plotted for 730 days grouped in 26 bins.

E.2 Main results for 2017 cohort

Table E.1 presents diff-in-disc estimates equivalent to the main results in Table 2 but compares individuals turning 20 in 2017 (instead of 2018) to the pre-DCT periods. The estimates confirm the results from the graphical analysis in section 4.1: The first stage estimates are generally smaller and weaker compared to 2018, although the negative direction is consistent with our expectation. The sharp diff-in-disc for the full sample provides no indication that the onset of the user fee of DCT increase consumption of in-person visits. The IV estimates are small and statistically insignificant, although for the most precisely estimated model, for men, the coefficient is negative as is consistent with substitution.

Figure E.2 shows estimates for 2017 corresponding to the results fort 2018 in Figure 4 in the main document. The first stage is small but robust to varying the size of both the donut and bandwidths. Comparable to the results from the main specification, the degree of substitution is estimated to be close to zero for different bandwidths, although with some increased precision for longer bandwidths.

Table E.1: Main result for 2017 cohort as post-period

	All	Men	Women
A. First stage (sharp diff-in-disc)			
Online consultations	-0.0499***	-0.0195***	-0.0824***
	(0.00706)	(0.00701)	(0.0114)
B. RF (sharp diff-in-disc)			
In-person visits	-0.00191	0.00279	-0.00679
	(0.0352)	(0.0397)	(0.0563)
C. IV (fuzzy diff-in-disc)			
In-person visits	0.0383	-0.143	0.0825
	(0.704)	(2.032)	(0.681)
Avg. ind/day pre	151525	78130	73395
Avg. ind/day post	31929	16467	15462

Note: Variable names in left column represent the dependent variable. Each model uses a fixed bandwidth of 120 days on either side of the threshold (20th birthday). Avg. ind/day = average number of individuals per day in cells, shown separately for the pre cohorts (which include several birth cohorts) and the post cohort (which only includes individuals turning 20 in 2017). Data are collapsed by region, gender, birth year and day relative to 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

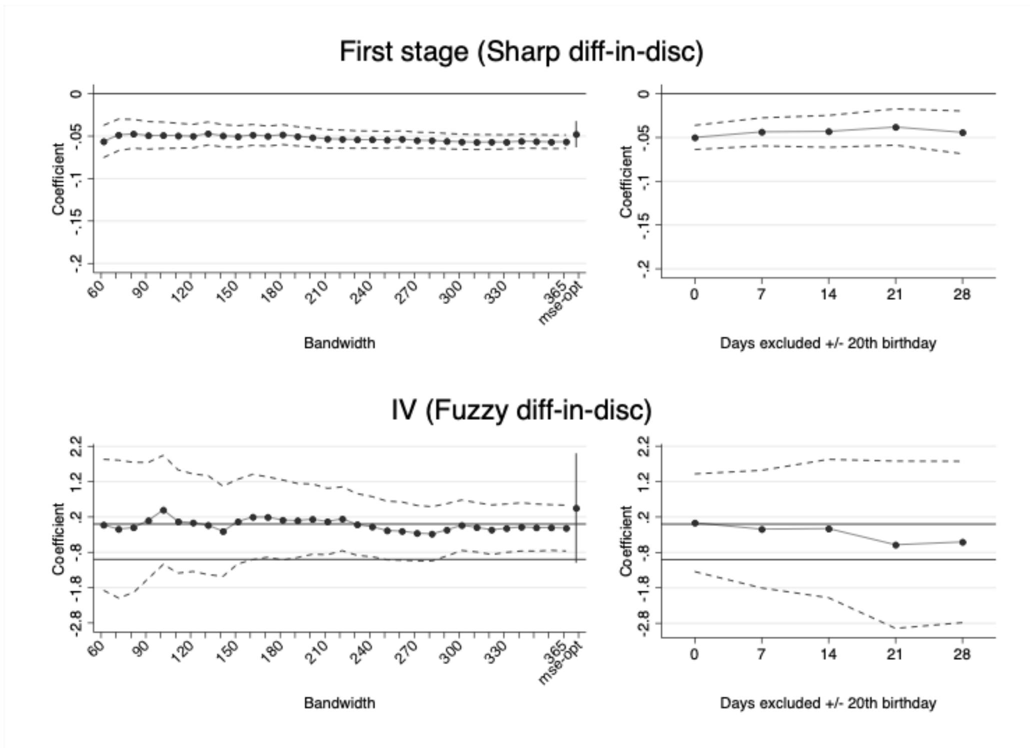


Figure E.2: Robustness 2017 - bandwidth and donut

Note: The graphs show estimates for 2017 corresponding to Fig 4. The upper panel shows coefficients from sharp diff-in-disc models of the effect of the user fee on DCT consultations (the first stage). The lower panel shows coefficients from fuzzy diff-in-disc models of the relationship between DCT consultations and in-person visits (IV). The graphs on the left show how the estimates vary over different bandwidhts (60–365 days + MSE-optimal bandwidth (Cattaneo, Idrobo and Titiunik, 2019)). The graphs on the right shows how the estimates vary when successively removing more observations around the 20th birthday (" donut" estimations), in all cases using a bandwidth of +/- 120 days relative to the 20th birthday.

Appendix F First stage heterogeneity and complier profiling

To provide information on the type of individuals responding to the change in the user fee, we examine the heterogeneity of the treatment effect of the user fee on DCT use (i.e., the first stage) and characterise the compliers of the change in the user fee using the logic from Angrist and Pischke (2008). In our setting a complier is an individual that – on a given day – makes a DCT consultation under the absence of a DCT user fee but chooses not to make a DCT consultation when facing a user fee. We define always-takers (i.e., individuals that would use DCT during the given period independent of the presence of a fee) as the proportion of individuals using DCT before the 20th birthday (when no fee applies) and never-takers (i.e., individuals that would not use DCT during the given period independent of the fee) as the proportion of individuals without any DCT contact in the period after the 20th birthday (when the fee applies). We further assume that there are no defiers (i.e. individuals who abstain from using DCT unless there is a fee).

Under the assumption of no defiers, the proportion of the compliers are equal to the first stage estimate if the outcome is binary (Angrist and Pischke, 2008). While the outcome variable in our application is adjusted to represent the yearly number of DCT contacts, the relative size of the coefficients is still informative. The proportion of compliers among the DCT users before the 20th birthday may therefore be computed as the ratio between the effect of the user fee and the level of consumption in the period with the higher user fee (before the 20th birthday). That is, $\beta_2/(\beta_4+\beta_1)$ in equation 3. That implies that the proportion of the users that are compliers and always-takers are equal to 49.5% and 50.5%.[3] Angrist and Pischke (2008) further show that the ratio between the first stage coefficient for a subpopulation based on a given characteristic and the total population is informative of the relative likelihood that a complier has that characteristic. The left subgraph of Figure F.1 displays sharp diff-in-discs coefficients estimating the effect of the user fee on DCT consultations for various subpopulations compared to the main estimate (depicted as the plotted line). These estimates correspond to β_3 in equation 3 in the main document. The center subgraph displays the corresponding baseline utilisation per subpopulation in 2018 for individuals below the 20th birthday. These estimates correspond to $\beta_1+\beta_4$ in equation 3. The left

[3] To get an estimate of the absolute proportion of compliers and always-takers, these coefficients should be scaled back to the daily level. In a given day, there are 0.000412 always-takers, and 0.000408 compliers using DCT.

subgraph displays the relative reduction (i.e., $\beta_3/\beta_1+\beta_4$).

The graphs shows that women, individuals residing in urban areas, and individuals with previous physician contacts have both larger first stage and a larger baseline utilisation. These results suggest that these characteristics are relatively more common among compliers than in the overall sample. However, the results also suggest that the size of the effect is very much related to the baseline utilisation (and that these characteristics are relatively more common also among always-takers). There are also minor indications that both first stage and baseline utilisation are larger among individuals with high SES parents (fathers with income below the median and no university education) and individuals having two parents being born outside of the Nordic countries. Thus, the first stage is generally larger for subpopulations with a larger baseline utilisation. Notably, the left subgraph show that the relative effect of the onset of the user fee in relation to the utilisation before the 20th birthday is very similar between subpopulations, indicating that price sensitivity is similar across subgroups.

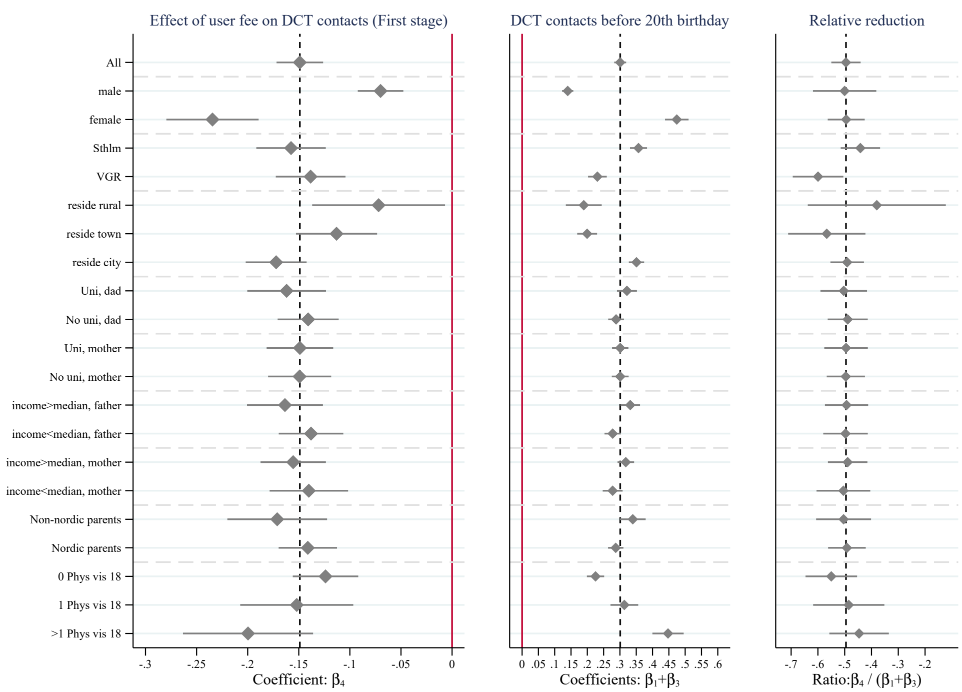


Figure F.1: Effect of user fee on DCT consultations, by subpopulations

Figures F.2 and F.3 display graphically the number of DCT and in-person consultations around the age threshold, by gender.

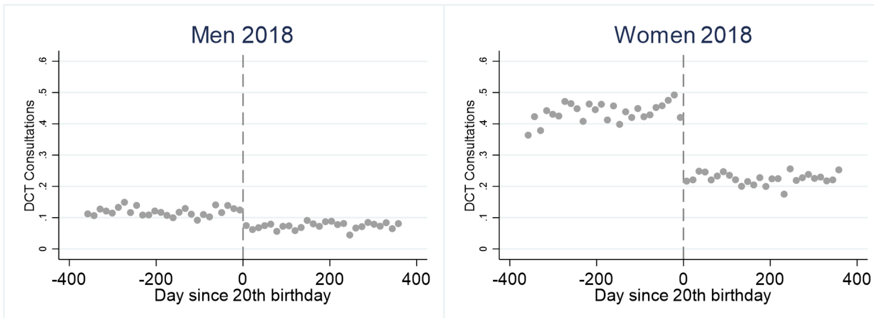


Figure F.2: DCT consultations in the postperiod by gender

Note: The graphs display the average number of DCT consultations by age for each gender in 2018. The age variable is measured as days in relation to the 20th birthday. The number of DCT consultation is rescaled to reflect annual averages, and is plotted for 730 days grouped in 26 bins.

Figure F.3: In-person physician visits

Note: The graphs display the average number of in-person physician visits by age for men and women separately. The age variable is measured as days in relation to the 20th birthday. The number of in person visits is rescaled to reflect annual averages, and is plotted for 730 days grouped in 26 bins.

Appendix G Heterogeneity

Urban and previous care consumption We explore heterogeneity in the degree of substitution in dimensions that are pointed out as important for the first stage heterogeneity in Appendix F. The share of compliers are larger among residents in urban areas and among individual that have consumed care between their 18th and 19th birthday. Table G.1 displays the urban-rural dimension (columns 1-2), as well as by the individual's prior care consumption (columns 3-4). We find, as expected from when splitting the sample by region, that the substitution is mainly driven by the urban sample. In contrast, both the first stage and the substitution effect found are largely comparable for individual with different levels of previous care consumption (although the IV estimates are statistically insignificant).

Table G.1: Complier heterogeneity - Rurality and care consumption

	Rural	Urban	0 visits	$\geq$ 1 visit
FS	-0.101***	-0.172***	-0.124***	-0.176***
	(0.0178)	(0.0152)	(0.0164)	(0.0198)
IV	-0.0490	-0.561***	-0.495	-0.389
	(0.503)	(0.212)	(0.307)	(0.257)

Note: The table shows estimates with sample split by first urban-rural dimension and then by whether individual has consumed care by 18 years old. Each model uses a 120 bandwidth with 2018 as the postperiod. Standard errors in parentheses clustered by relative age (running variable), with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Regional heterogeneity Our main analysis merges data from two independent administrative regions – Region Stockholm and Region Västra Götaland. Given the scope for institutional differences, it is of interest to examine heterogeneity over these regions.

Table G.2 presents our main specification from models where regions are estimated separately for each region. As indicated by the graphical evidence, the size of the Stockholm coefficient (-.77) is considerably larger than the coefficient for Västra Götaland (.01), and it is only in Stockholm that we can reject the null hypothesis of no substitution. The first stage estimates in the first panel shows that the response to the DCT user fee is similar in both regions. Thus, the difference in the degree of substitution is driven by differences in the reduced form estimates.

Table G.2: Fuzzy diff-in-disc results (IV), by region

	All	Men	Women
	A. First stage (Sharp diff-in-disc)		
Sthlm	-0.158***	-0.0783***	-0.243***
	(0.0173)	(0.0165)	(0.0328)
F-stat	82.49	22.38	55.04
VGR	-0.138***	-0.0596***	-0.224***
	(0.0174)	(0.0163)	(0.0316)
F-stat	63.23	13.41	50.35
	B. Reduced form (Sharp diff-in-disc)		
Sthlm	0.121***	0.108**	0.132*
	(0.0441)	(0.0518)	(0.0677)
VGR	-0.00189	0.0263	-0.0325
	(0.0475)	(0.0576)	(0.0700)
	C. IV (Fuzzy diff-in-disc)		
Sthlm	-0.767***	-1.384**	-0.543*
	(0.288)	(0.706)	(0.287)
VGR	0.0137	-0.440	0.145
	(0.343)	(0.974)	(0.313)
Sthlm: Avg. ind/day pre	82264	42380	39884
Sthlm: Avg. ind/day post	17502	9009	8492
VGR: Avg. ind/day pre	69283	35763	33520
VGR: Avg. ind/day post	13985	7291	6695

Note: The table shows estimates with the sample split by regions and sex. Each model uses bandwidth of 120 days on either side of the 20th birthday. 'Average ind/day pre (post)' refers to average individuals per cell in the pre (post) cohort. Standard errors in parentheses clustered by the running variable, with separate clusters for pre and post periods. * P<0.1, ** P<0.05, *** P<0.01.

We continue by further breaking down the results by region and rurality, which is further motivated by urban residents being a large complier group. The estimates in Table G.3 show that the estimated degree of substitution is more similar when comparing the urban areas of the two regions (Stockholm -.631; Västra Götaland -.391), although the estimate for Västra Götaland is not statistically significant. This in turn implies that the differences in the degree of substitution between the two regions is largely explained by differences in the share of urban residents.

Table G.3: IV estimate by urban/rural heterogeneity

IV ESTIMATES		
	Rural	Urban
RS	-2.920	-0.631**
	(2.291)	(0.280)
RVG	0.495	-0.391
	(0.558)	(0.400)
AVG INDIVIDUALS/DAY (CELL)		
Sthlm Pre:	40369	40369
Sthlm Post:	1869	1563
VGR Pre:	36815	36815
VGR Post:	8330	5654

Note: The table shows estimates with sample split by regions and urban/rural dimension, with a fixed 120-days bandwidth. 'Avg. ind/day pre (post)' refers to average individuals per cell in the pre (post) cohort. Standard errors in parentheses clustered by relative age (separate clusters for pre-post periods). * P<0.1, ** P<0.05, *** P<0.01.

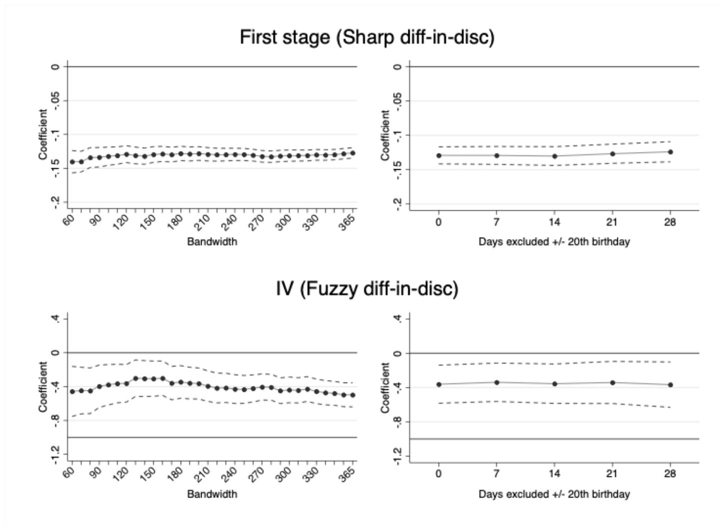


Figure H.1: Zero degree polynomial - bandwidth and donut

Note: The graphs show how removing 0 to 70 days affects the estimates using a bandwidth of +/- 180 (left) or +/- 365 (right) days relative to the 20th birthday. The upper panel shows coefficients from sharp diff-in-disc models of the effect of the user fee on DCT consultations (the first stage). The lower panel shows coefficients from fuzzy diff-in-disc models of the relationship between DCT consultations and in-person visits (IV).

Appendix H Robustness

H.1 Zero-degree polynomial estimates

The graphs in Figure H.1 show first stage and IV estimates from a model using a zero-degree polynomial for varying bandwidths, and a specification using a fixed bandwidth of 120 days. The first stage estimates of the onset of the user fee on DCT consultations are similar – but more stable than the estimates from the linear specification. The estimates of the degree of substitution hover around the estimate from our preferred specification (Table 2). Overall, these estimates are more stable and generally more precisely estimated than those from the linear specification. Thus, in difference to the linear specification, the zero-degree polynomial provides similar estimates of degree of substitution also for larger bandwidths. The graphs to the right hand side in Figure H.1 presents coefficients from donut estimations removing 0 to 28 days, showing in line with the estimates in the main document that these are stable for both the first stage relationship and the degree of substitution.

H.2 Longer donut specifications

This section shows donut estimation results using longer outer bandwidths than in the main document (for a linear and a zero degree polynomial specification). We estimate these models with longer bandwidths to be able to further increase the size of the donut hole without losing too many observations.

Figure H.2 shows estimates from donut specifications using longer bandwidths of 180 days (left graphs) or 365 days (right graphs) using a linear specification. The upper panel shows first stage estimates of the effect of the fee on the number of DCT consultations and the lower panel shows IV estimates of the degree of substitution. The horizontal axis indicates the number of days dropped on each side of the 20th birthday (the donut hole).

For both bandwidths, the first stage estimates are rather stable. The IV estimates are also stable up until we remove 56 days (8 weeks on each side of the 20th birthday). Removing donuts of this size or larger eliminates the substitution effect (for both bandwidths). Since the first stage estimates are stable or even decreasing, the decrease in the estimated degree of substitution stems from variability in the relationship between age and the number of in-person visits far away from the threshold. Revisiting Figure 3 in the main text, it is easy to see that the slope of a linear regression line may vary depending on which portion of the x-axis one selects to fit the lines, but that the relationship looks close to flat when looking at the period as a whole. Thus, to explore the sensitivity to the donut estimations to the linear specification, we also estimate donut specifications using a zero-degree polynomial.

Figure H.3 shows that for larger bandwidths (of 180 or 365 days) the specification using a zero-degree polynomial yields results similar to the main estimates even with the larger donuts of 56-70 days. The point estimates from the IV model indicate that the degree of substitution is between 0.3 and 0.5. The difference to the corresponding estimates from linear specification of the model suggest that for larger donuts (+/-8 weeks) the estimated degree of substitution is sensitive to the (linear) specification of the model. Notably, the estimates from the zero-degree specification, which are less affected by volatility further away from the threshold, are in line with the main estimates. Thus, our conclusions about the degree of substitution are only affected if we rely on estimates that are clearly affected by fluctuations far away from the threshold.

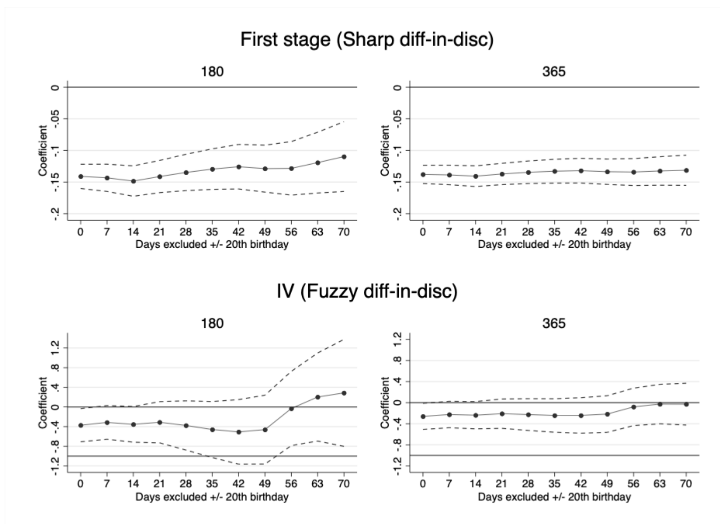


Figure H.2: Longer donut estimations

Note: The graphs show how removing 0 to 70 days affects the estimates using a bandwidth of +/- 180 (left) or +/- 365 (right) days relative to the 20th birthday. The upper panel shows coefficients from sharp diff-in-disc models of the effect of the user fee on DCT consultations (the first stage). The lower panel shows coefficients from fuzzy diff-in-disc models of the relationship between DCT consultations and in-person visits (IV).

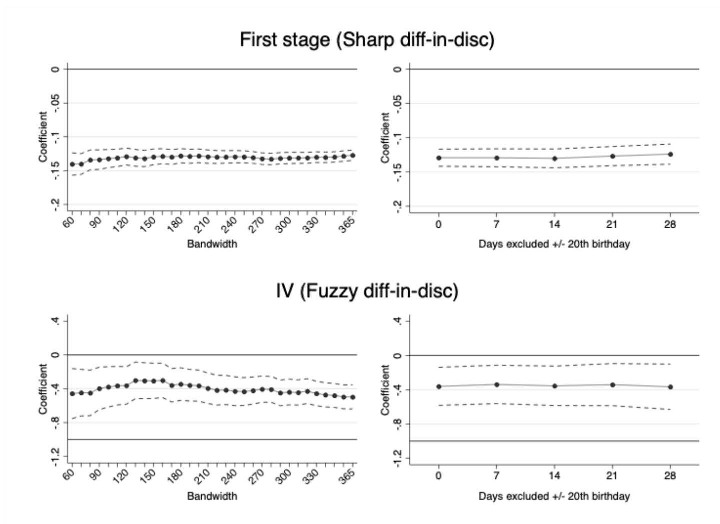


Figure H.3: Zero-degree polynomial - longer donut estimations

Note: The graphs show removing 0 to 70 days affects estimates from a specification using a zero degree polynomial and a bandwidth of +/-180 (left) or +/- 365 (right) days relative to the 20th birthday. The upper panel shows coefficients from sharp diff-in-disc models of the effect of the user fee on DCT consultations (the first stage). The lower panel shows coefficients from fuzzy diff-in-disc models of the relationship between DCT consultations and in-person visits (IV).

H.3 First stage graphs by diagnosis groups

Figure H.4 plots the annual number of DCT consultations per capita in 2018 relative to the individual's 20th birthday, for four types of diagnoses common DCT-diagnoses: upper respiratory infections, skin conditions, diagnoses related to genital and reproductive organs, and other common DCT-diagnoses.

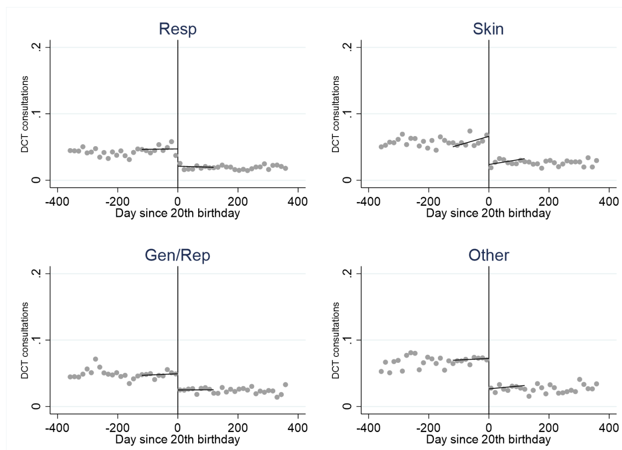


Figure H.4: DCT consultations in 2018 by diagnosis group

Note: The graphs display the average number of DCT consultations by age for four subgroups of diagnoses in 2018. The age variable is measured as days in relation to the 20th birthday. The number of DCT consultation is rescaled to reflect annual averages, and is plotted for 730 days grouped in 26 bins.

H.4 Balance test

Table H.1 shows balance estimation results for the dichotomous background variables displayed in the leftmost column. Each row presents the result of sharp difference-in-discontinuity regressions (Eq. (1)) using the background characteristic in question as the outcome variable. We estimate these models using individual level data and cluster standard errors on the individual.

In regression discontinuity studies *not* using age as the running variable, balance tests of this kind can spot suspected violations of the assumption of random assignment due to, e.g., manipulation. In our case, such manipulation is ruled out by definition, as we are studying a close to balanced panel of individuals observed around their 20th birthday and all the outcomes variables in our regressions are time-invariant (for instance, the variable *Phys visit = 1* equals one if the individual visited a primary care physician exactly once during his/her 19th life year). This implies that all differences reflect sample composition changes (i.e., attrition) and nothing else.

There are some statistically significant estimates, but the magnitudes of these coefficients are negligible in relation to the means of the variables (c.f. Table C.1). For our baseline sample (column All, 2018), the coefficient on the indicator for having made exactly one primary care visit during one's 18th life year is statistically significant at the 10% level. Put in relation to the mean of the outcomes variable, which is 0.251, it is clear that the coefficient of .0017 is not a meaningful difference.

H.5 Including covariates

Table H.2 shows fuzzy diff-in-disc estimates with and without including covariates, i.e., the pre-determined individual characteristics used in our balance test (columns "*No Covs/With covs*"), across a range of bandwidths. We note that, as expected, covariates do not affect the point estimates nor the precision of the estimates.

Table H.1: Balance test on individual data

	ALL		MEN		WOMEN	
	2017	2018	2017	2018	2017	2018
Woman	-0.00020	0.00010				
	(0.00080)	(0.00080)				
Phys visits = 0	0.00195	-0.00187	0.00342**	-0.000786	0.000325	-0.00273
	(0.00119)	(0.00117)	(0.00164)	(0.00161)	(0.00170)	(0.00168)
Phys visits = 1	-0.00219**	0.00170*	-0.00292**	0.000920	-0.00140	0.00251*
	(0.00103)	(0.00102)	(0.00141)	(0.00139)	(0.00151)	(0.00148)
Phys visits ≥ 2	0.000241	0.000162	-0.000501	-0.000134	0.00108	0.000221
	(0.00105)	(0.00101)	(0.00132)	(0.00128)	(0.00162)	(0.00156)
Share Sthlm	-0.00208*	0.00108	-0.00279*	-0.000372	-0.00134	0.00262
	(0.00118)	(0.00116)	(0.00164)	(0.00162)	(0.00171)	(0.00168)
Resides Rural	0.00116	0.0000699	0.000313	0.000536	0.00207**	-0.000417
	(0.000716)	(0.000693)	(0.000992)	(0.000963)	(0.00103)	(0.000997)
Resides Town	-0.000217	-0.000993	0.00190	0.000436	-0.00248*	-0.00253*
	(0.00100)	(0.000983)	(0.00139)	(0.00135)	(0.00145)	(0.00143)
Resides City	-0.000945	0.000923	-0.00221	-0.000973	0.000405	0.00295*
	(0.00112)	(0.00110)	(0.00155)	(0.00152)	(0.00162)	(0.00159)
Dad uni	-0.000909	-0.0000824	-0.00253	0.00129	0.000823	-0.00156
	(0.00115)	(0.00114)	(0.00160)	(0.00158)	(0.00166)	(0.00164)
Mum uni	-0.00203*	-0.000160	-0.00582***	0.00128	0.00202	-0.00172
	(0.00118)	(0.00117)	(0.00164)	(0.00162)	(0.00171)	(0.00168)
Mum inc > median	-0.000910	-0.000283	-0.00252	0.000346	0.000790	-0.000981
	(0.00118)	(0.00116)	(0.00163)	(0.00161)	(0.00170)	(0.00168)
Dad inc > median	0.00110	-0.000767	0.00126	-0.000625	0.000930	-0.000917
	(0.00118)	(0.00116)	(0.00163)	(0.00161)	(0.00170)	(0.00167)
Par non-nordic	0.000163	0.000414	0.00140	-0.000471	-0.00116	0.00135
	(0.00105)	(0.00104)	(0.00145)	(0.00144)	(0.00151)	(0.00149)

Note: Balance tests. Fixed bandwidth 120 days +/- either side of 20th birthday. Each table row represents an outcome variable and each cell shows a sharp diff-in-disc estimate contrasting the changes at age 20 for the cohort turning 20 in 2017 (2018) and the pre-DCT cohort. Robust standard errors clustered by individual in parentheses. * P<0.1, ** P<0.05, *** P<0.01.

Table H.2: IV estimates over different bandwidths

	ALL		MEN		WOMEN	
	No Covs	With Covs	No Covs	With Covs	No Covs	With Covs
BW 30	-0.80	-0.79	-3.69	-3.69	0.04	0.04
	(-2.06, 0.46)	(-2.05, 0.47)	(-7.32, -0.05)	(-7.32, -0.05)	(-1.16, 1.25)	(-1.16, 1.24)
KP F-stat	34.00	34.25	5.55	5.61	19.17	19.40
BW 60	-0.23	-0.22	-0.75	-0.75	-0.05	-0.05
	(-0.83, 0.37)	(-0.82, 0.38)	(-2.22, 0.72)	(-2.22, 0.72)	(-0.65, 0.55)	(-0.64, 0.55)
KP F-stat	93.51	93.86	19.94	20.06	55.21	55.56
BW 90	-0.44	-0.43	-0.96	-0.96	-0.27	-0.26
	(-0.88, 0.00)	(-0.87, 0.02)	(-2.14, 0.21)	(-2.14, 0.22)	(-0.71, 0.18)	(-0.70, 0.19)
KP F-stat	126.15	126.54	29.36	29.47	81.67	82.07
BW 120	-0.45	-0.44	-1.03	-1.03	-0.26	-0.25
	(-0.84, -0.06)	(-0.83, -0.05)	(-2.10, 0.05)	(-2.10, 0.05)	(-0.65, 0.12)	(-0.64, 0.13)
KP F-stat	163.78	164.17	37.76	37.87	103.36	103.75
BW 150	-0.47	-0.46	-1.12	-1.13	-0.27	-0.26
	(-0.85, -0.09)	(-0.84, -0.09)	(-2.20, -0.05)	(-2.21, -0.05)	(-0.63, 0.09)	(-0.62, 0.10)
KP F-stat	176.89	177.27	37.17	37.25	117.75	118.12
BW 180	-0.37	-0.36	-0.84	-0.86	-0.22	-0.20
	(-0.71, -0.03)	(-0.70, -0.02)	(-1.78, 0.10)	(-1.80, 0.08)	(-0.55, 0.12)	(-0.54, 0.13)
KP F-stat	207.79	208.30	47.23	47.31	137.42	137.82
BW 365	-0.26	-0.26	-0.71	-0.74	-0.14	-0.14
	(-0.51, -0.01)	(-0.50, -0.01)	(-1.56, 0.15)	(-1.60, 0.12)	(-0.36, 0.09)	(-0.36, 0.09)
KP F-stat	348.22	349.27	57.53	57.46	279.83	280.72
BW opt	-0.45	-0.40	-0.46	-0.56	-0.18	-0.13
	(-0.86, -0.05)	(-0.82, 0.01)	(-1.63, 0.71)	(-1.75, 0.63)	(-0.59, 0.23)	(-0.56, 0.29)
KP F-stat	154.77	146.39	29.31	28.57	89.28	81.53

Note: The table shows IV estimates for the full sample and for men and women separately. Each panel, other than the last one, is associated with a fixed bandwidth on either side of the the threshold (and for both cohorts). The last panel presents the results using a MSE-optimal bandwidth. 95% confidence intervals shown in parentheses, derived from standard errors clustered by the running variable (days since 20th birthday with separate clusters for pre- and post cohorts).

H.6 Placebo tests using 19th and 21th birthdays

Table H.3 displays a placebo analysis using alternative birthdays as thresholds in sharp diff-in-discs for both DCT and in-person consultations. The thresholds used are the 19th and the 21th birthday. For simplicity, these are coded as 365 days from the 20th birthday.

Table H.3: Placebo:Alternative birthday

	All	Men	Women
A. First stage (19th birthday)			
Online consultations	-0.00547	-0.00547	-0.00547
	(0.00631)	(0.00631)	(0.00631)
B. Reduced form (19th birthday)			
In-person visits	-0.00234	-0.00627	0.00197
	(0.0251)	(0.0286)	(0.0405)
C. First stage (21 birthday)			
In-person visits	-0.00329	-0.00329	-0.00329
	(0.00724)	(0.00724)	(0.00724)
D. Reduced form (21 birthday)			
In-person visits	0.00831	0.0131	0.00281
	(0.0264)	(0.0286)	(0.0420)

Note: Variable names in left column represent the dependent variable. Data are collapsed by region, gender, birth year and day relative to 20th birthday. The first two panels display sharp diff-in-discs using the 365th day before the 20th birthday as the threshold. The last two panels display a sharp diff-in-disc using the 365th day after the 20th birthday as the threshold. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

H.7 Sensitivity to inclusion of pre-periods of varying length

Figure H.5 plots annual in-person physician visits per capita by the day relative to the 20th birthday (i.e., the DCT user fee threshold) for 6 time periods, each corresponding to one year. The first four sub-graphs show each of the four years preceding the introduction of DCT services in mid-2016. The last two sub-graphs present figures for 2017 and 2018 (when the DCT market emerged).

In the pre-period, the 20th birthday was associated with a drop in the number of in-person visits. These graphs show that the drop at the 20th birthday was stable during the pre-period (although somewhat smaller in the the first year, July 2012 – June 2013). The drop in 2017 is similar to the years before the DCT emerged, while in 2018 we no longer observe a drop at the 20th birthday.

The thin black lines on each side of the user fee threshold reflect how the data portrayed in the figure are used to estimate the diff-in-disc estimates. The length of the lines corresponds to the 120 days bandwidth.

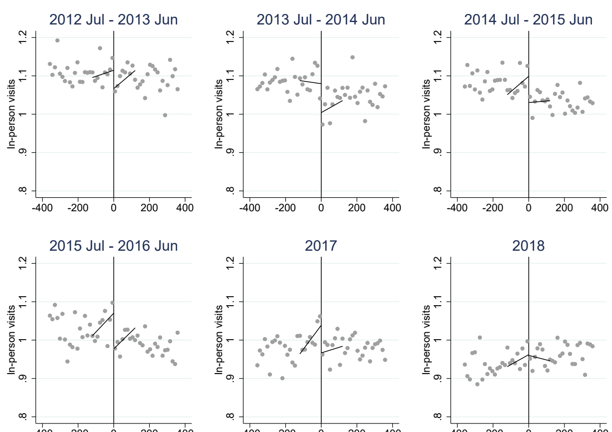


Figure H.5: Physician visits over time, separate graph for each pre-period

Table H.4 presents our main specification (fuzzy diff-in-disc) estimates for different lengths of the pre-period. The first row presents estimates using only the last 12 months before June 31 2016. Each row then adds another 12 months to the pre-period. The last row presents the estimates from the main results in Table 2. For correspondence with the main results, a 120-days bandwidths are used for based on the sample using the full pre-period. Overall, the results are similar across various length of the pre-period. Compared to our main estimate, which includes all pre-periods, we would obtain a slightly higher degree of substitution

Table H.4: Main IV result over different pre-periods

	All	Men	Women
2015 Jul - 2016 Jun	-0.584**	-0.687	-0.544**
	(0.246)	(0.623)	(0.262)
2014 Jul - 2016 Jun	-0.507**	-1.064*	-0.324
	(0.213)	(0.577)	(0.212)
2013 Jul - 2016 Jun	-0.502**	-1.127**	-0.299
	(0.198)	(0.560)	(0.195)
2012 Jul - 2016 Jun	-0.450**	-1.026*	-0.260
	(0.200)	(0.549)	(0.196)
Avg. ind/day pre	151525	78130	73395
Avg. ind/day post	31490	16302	15188

Note: The table present fuzzy diff-in-disc estimates using various lengths of the pre-period. The first row presents estimates using only the last 12 months before June 31 2016. Each row than adds another 12 months to the pre-period. The last row presents the estimates from the main results in Table 2. All estimates are from models using a fixed bandwidth of 120-days. Avg. ind/day = average number of individuals per day in cells, shown separately for the pre cohorts (which include several birth cohorts) and the post cohort (which only includes individuals turning 20 in 2018). Estimates using data collapsed by region, gender, birth year and day relative to 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

if we excluded the first period (for which the observed drop at the 20th birthday is smaller). The estimated degree of substitution is noticeably higher if we remove all but the last year in the pre-period (-.58, with a 95% confidence interval covering -1 but not 0). We conclude that our main specification, which relies on more data points, yields a conservative estimate of the degree of substitution.

H.8 Sensitivity to secular decrease of in-person visits

The graphs in Figure 3 suggest that there is a general decrease in the number of in-person visits between the pre-period and the post period also among individuals above 20 years old. The graphs indicate that this decrease is about 5 to 10%. Not accounting for this overall reduction may bias the diff-in-disc estimates, as it is possible that this secular decrease would have reduced the discontinuity at age 20 even in the absence of DCT user fee.

In section 3.2, we presented the diff-in-disc estimand as:

$$\theta = \frac{\tau_y}{\tau_{DCT}} = \frac{(Y_1^+ - Y_1^-) - (Y_0^+ - Y_0^-)}{(DCT_1^+ - DCT_1^-) - (DCT_0^+ - DCT_0^-)} \tag{1}$$

To identify the degree of substitution this estimand relies on the assumption that any effects of confounding treatments must be time-invariant. (Millán-Quijano, 2020) suggests an estimand that relaxes this assumpion that in our context corresponds to:

$$\theta^M = \frac{\tau_y}{\tau_{DCT}} = \frac{(Y_1^+ - Y_1^-) - (1-\gamma)(Y_0^+ - Y_0^-)}{(DCT_1^+ - DCT_1^-) - (DCT_0^+ - DCT_0^-)} \tag{2}$$

where γ is equal to the overall (proportional) reduction in the in-person consultations between the two periods. Subtracting θ^M from θ (i.e Eq. 1 from Eq. 2) yields an expression of the bias of the standard diff-in-disc estimand under these circumstances:

$$bias = \gamma \frac{(Y_0^+ - Y_0^-)}{\tau_{DCT}}. \tag{3}$$

Thus, in order to get an estimate of the bias, we replace τ_{DCT} and $(Y_0^+ - Y_0^-)$ by estimates from the first stage regression and a standard RD-model in the pre-period. Assuming that γ is equal to the proportional decrease in the in-person consultations among individuals just above 20 (i.e. $(Y_1^+ - Y_0^+)/Y_0^+$), we can use the coefficients from our reduced form equation to compute γ. This exercise yields an estimate of the bias of the main IV estimate amounting to .03 (which corresponds to an overestimation of the degree of substitution of about 7 %, or 42% instead of 45%). The estimated bias for results split by gender is of similar size, or smaller. Thus, we do not consider the general decrease in the number of in-person consultations to be a major concern for our conclusions.

H.9 Fuzzy RD for Stockholm

The diff-in-disc assumes that any confounding policy has time-invariant effects. An alternative estimation strategy that relaxes that assumption would be to focus exclusively on Region Stockholm, where there was no confounding change of in-person visit fees at age 20. Assuming there are no other confounding policies at age 20, we can then use a standard fuzzy RD specification (instead of a diff-in-disc). Table H.5 shows that such a specification yields similar estimates at bandwidths of 90 days or more (including when we we use an MSE-optimal bandwidth for the relevant estimation sample).

Table H.5: Fuzzy Regression Discontinuity, Sthlm 2018

	All	Men	Women
BW 60	0.0844	-0.0211	0.123
	(0.298)	(0.736)	(0.279)
KP F-stat	15.41	13.32	39.86
BW 90	-0.299	-0.528	-0.221
	(0.258)	(0.673)	(0.251)
KP F-stat	19.86	16.65	54.03
BW 120	-0.480**	-0.697	-0.398*
	(0.244)	(0.614)	(0.239)
KP F-stat	22.63	21.59	57.11
BW 150	-0.463**	-0.628	-0.399*
	(0.233)	(0.588)	(0.223)
KP F-stat	27.68	22.92	70.60
BW 180	-0.403*	-0.539	-0.343*
	(0.213)	(0.525)	(0.206)
KP F-stat	31.74	26.73	82.35
BW 365	-0.282*	-0.531	-0.199
	(0.155)	(0.504)	(0.139)
KP F-stat	61.51	29.29	169.6
Opt BW	-0.414*	-0.626	-0.157
	(0.243)	(0.691)	(0.239)
Left bw	115	78	80
Right bw	110	98	105
KP F-stat	22.72	16.62	51.63

Note: The table presents fuzzy regression discontinuity estimates for Stockholm in 2018 (that is, no differencing across cohort as in our main specifications). The bandwidth is fixed for all but the last panel, where the bandwidth is chosen flexibly on either side of the threshold. 'KP F-stat' refers to the Kleibergen-Paap F-statistic. Standard errors clustered by the running variable. * P<0.1, ** P<0.05, *** P<0.01.

H.10 Wild bootstrapping precision adjustment

In this section we examine the sensitivity of our results to the estimation of standard errors. A recent literature discusses methods to obtain bias-corrected estimates and robust confidence intervals for standard RD settings with data-driven bandwidth choices (Calonico, Cattaneo and Titiunik, 2014; He and Bartalotti, 2020). This literature has not yet developed methods adapted to a diff-in-disc setting, so we modify the wild bootstrap procedure developed (and thoroughly described in) He and Bartalotti (2020) to fit such a setting. In short, He and Bartalotti (2020) estimate the bias from choosing an optimal bandwidth, *h*, using a higher order polynomial for a longer bandwidth *b*. Using a given set of *h* and *b* this procedure consists of two algorithms (both *h* and *b* are allowed to vary each side of the threshold). The first algorithm estimates the bias, and the second algorithm estimates the distribution. Both algorithms rely on a higher order polynomial with bandwidth *b* mimicking the data generated process, and the estimation of linear polynomials with bandwidth *h* of a dataset obtained from the data generating process. Using the notation from our study setting, the procedure inHe and Bartalotti (2020) estimates Z_c^+ and Z_c^- for $Z \in (DCT, y)$ and a single cohort, c, in order to obtain the bias and distribution of a fuzzy RD:

$$\theta^{RD} = \frac{(y_c^+ - y_c^-)}{(DCT_c^+ - DCT_c^-)} \tag{4}$$

By contrast, our modified procedure estimates Z_c^+ and Z_c^- for $Z \in (DCT, y)$ and $c \in (0,1)$ in order to obtain the bias and distribution of:

$$\theta = \frac{\tau_y}{\tau_{DCT}} = \frac{(y_1^+ - y_1^-) - (y_0^+ - y_0^-)}{(DCT_1^+ - DCT_1^-) - (DCT_0^+ - DCT_0^-)} \tag{5}$$

The first three rows in each of the three panels of Table H.6 reproduces the coefficients, standard errors and 95% confidence intervals obtained in our main specification using the MSE-optimal bandwidth for the relevant estimation sample (see Table 2). The table further displays the bias corrected estimates and confidence interval obtained from the bootstrap procedure.

The bias-corrected bootstrap results are overall in line with the main results. The bias corrected coefficient for the fuzzy diff-in-disc for the full sample equals -.51 compared to the main estimate of -.45. The bootstrapped confidence interval is only slightly broader, and excludes both zero and one. When splitting the sample by gender, we observe that the bias corrected coefficients are larger for women, but smaller for men, compared to corresponding standard coefficient. While the bootstrapped confidence intervals are broader, they lead to the same conclusions as the standard one.

Notably, both bias corrected coefficients and bootstrapped confidence intervals for the first stage are very similar to the standard coefficients and confidence intervals. Thus, the difference in the fuzzy diff-in-disc comes from the bias-correction of the reduced form.

Table H.6: Bootstrapped standard errors comparison

	All	Men	Women
	FIRST STAGE		
Coefficient	-0.149	-0.0772	-0.238
SE	(0.0119)	(0.0143)	(0.0252)
CI	(-0.172, -0.125)	(-0.105, -0.0493)	(-0.287, -0.189)
Bias Corrected Coeff	-0.151	-0.0828	-0.245
Bootstrapped CI	[-0.176, -0.128]	[-0.119, -0.0622]	[-0.301, -0.201]
F-stat	154.8	29.31	89.28
	REDUCED FORM		
Coefficient	0.0673	0.0356	0.0437
SE	(0.0307)	(0.0464)	(0.0500)
CI	(0.00704, 0.128)	(-0.0553, 0.126)	(-0.0543, 0.142)
Bias Corrected Coeff	0.0767	0.0293	0.0617
Bootstrapped CI	[0.0185, 0.148]	[-0.0731, 0.115]	[-0.0244, 0.180]
	IV		
Coefficient	-0.453	-0.461	-0.184
SE	(0.206)	(0.596)	(0.210)
CI	(-0.857, -0.0496)	(-1.629, 0.708)	(-0.595, 0.227)
Bias Corrected Coeff	-0.510	-0.272	-0.248
Bootstrapped CI	[-0.989, -0.0800]	[-1.599, 1.317]	[-0.718, 0.112]
F-stat	35.01	6.656	24.30

Note: The table shows IV results with sample split by gender and two sets of coefficients, standard errors and confidence inervals. Each model uses the MSE-optimal bandwidth for In-person visits varying by the sex; see bandwidth in first three columns in Table 2. In each panel, the first three rows (Coefficient, SE, CI) represent the results as we have presented so far with the standard way of calculating standard errors. The following two rows in each panel (Bias corrected coeff, Bootstrapped CI) present the same IV coefficient with bootstrapped standard errors corrected for potential bias, and the corrsponding confidence intervals.

H.11 Main specification on individual-level data

Table H.7 shows the main specification estimated on individual-level daily data using a fixed bandwidth of 120 days. In this specification, we cluster standard errors at the individual to account for autocorrelation. The coefficients differ only on the third decimal. Just as for the collapsed data, the individual-level data gives us a 95% confidence interval that excludes both 0 and -1 (CI, -.897, -.027), i.e. we may rule out both zero and complete substitution.

Table H.7: Fuzzy diff-in-disc results; individual-level data

	All	Men	Women
	A. First stage (sharp diff-in-disc)		
DCT consultations	-0.149***	-0.0700***	-0.234***
	(0.0148)	(0.0131)	(0.0273)
	B. RF (sharp diff-in-disc)		
In-person visits	0.0670**	0.0718*	0.0611
	(0.0320)	(0.0394)	(0.0511)
	C. IV (fuzzy diff-in-disc)		
In-person visists	-0.450**	-1.026*	-0.260
	(0.220)	(0.598)	(0.221)
Observations	43923493	22663627	21259866
Individuals	232657	119704	112953

Note: Estimations using individual-level daily data. Variable names in left column represent the dependent variable. Each model uses a bandwidth of 120 days before and after the 20th birthday. Standard errors in parentheses clustered by individual. * P<0.1, ** P<0.05, *** P<0.01.

Appendix I Decomposition by type of visits

Table I.1 displays sharp diff-in-disc estimates of the DCT user fee on DCT and in-person consultations decomposed by whether it is an initial contact in a care episode or a follow-up consultation. An initial consultation is defined as a consultation without either a DCT or an in-person consultation during the preceding 7, 14 or 28 days. (The wash-out period of 14 days is used in the analysis presented in the main text.) The remaining consultations are defined as follow-ups (although these may or may not be an actual part of the same care episode). Overall, the decomposition results are similar across wash-out periods. The onset of the user fee leads to a reduction in initial DCT consultations and DCT follow-ups. These decreases are complemented by an increase in initial in-person visits, but not in in-person follow-ups (in particular for wash-out periods longer than 7 days).

Table I.2 presents a further decomposition for the 14 days wash-out period used in the main document. Initial consultations are decomposed into consultations with and without follow-ups (and by mode of follow-ups: only by the same mode (DCT/in-person) or any follow-up by the other mode (DCT/in-person)). Follow-up consultations are decomposed by the mode of the initial consultation of the episode (DCT/in-person).

Table I.1: Decomposition of intitial and follow-up consultations

	All	Men	Women
	A. 7 DAYS WASH OUT		
Initial DCT consultation	-0.123***	-0.0578***	-0.194***
	(0.0111)	(0.0109)	(0.0218)
DCT follow-ups	-0.0258***	-0.0122***	-0.0405***
	(0.00466)	(0.00385)	(0.00884)
Initial in-person visits	0.0558**	0.0488	0.0627
	(0.0277)	(0.0346)	(0.0442)
Follow up in-person visits	0.0112	0.0230*	-0.00162
	(0.00941)	(0.0118)	(0.0146)
	B. 14 DAYS WASH OUT		
Initial DCT consultation	-0.114***	-0.0563***	-0.176***
	(0.0110)	(0.0110)	(0.0210)
DCT follow up 14 days	-0.0352***	-0.0137***	-0.0584***
	(0.00530)	(0.00479)	(0.00972)
Initial in-person visits	0.0662**	0.0503	0.0826**
	(0.0263)	(0.0338)	(0.0415)
Follow up in-person visits	0.000861	0.0215	-0.0216
	(0.0120)	(0.0148)	(0.0179)
	C. 28 DAYS WASH OUT		
Initial DCT consultation	-0.0969***	-0.0528***	-0.145***
	(0.0105)	(0.0105)	(0.0191)
DCT follow-ups	-0.0520***	-0.0172***	-0.0897***
	(0.00679)	(0.00609)	(0.0125)
Initial in-person visits	0.0692***	0.0536	0.0855**
	(0.0252)	(0.0342)	(0.0392)
In-person follow-ups	-0.00217	0.0182	-0.0244
	(0.0151)	(0.0183)	(0.0234)

Note: Estimations using individual-level daily data. Variable names in left column represent the dependent variable. The analysis decompose DCT and in person visits in initial and follow up visits (with a previous contact within 7/14/28 days). Each model uses a bandwidth of 120 days before and after the 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Table I.2: Further decomposition of consultations

	All	Men	Women
	A. DCT consultations		
Initial DCT (All)	-0.114***	-0.0563***	-0.176***
	(0.0110)	(0.0110)	(0.0210)
-no follow-up	-0.0938***	-0.0503***	-0.141***
	(0.00965)	(0.00981)	(0.0178)
-any follow-up	-0.0205***	-0.00562	-0.0365***
	(0.00472)	(0.00482)	(0.00888)
-only DCT follow-ups	-0.0101***	0.0000182	-0.0211***
	(0.00331)	(0.00286)	(0.00648)
-any in-person follow-ups	-0.0103***	-0.00564	-0.0155**
	(0.00376)	(0.00373)	(0.00689)
DCT follow up (All)	-0.0352***	-0.0137***	-0.0584***
	(0.00530)	(0.00479)	(0.00972)
-initial DCT	-0.0300***	-0.00834*	-0.0535***
	(0.00490)	(0.00425)	(0.00901)
-initial in-person	-0.00512*	-0.00532*	-0.00492
	(0.00262)	(0.00284)	(0.00425)
	B. In-person visits		
Initial in-person (All)	0.0662**	0.0503	0.0826**
	(0.0263)	(0.0338)	(0.0415)
-no follow-up	0.0576**	0.0431	0.0728**
	(0.0242)	(0.0320)	(0.0370)
-any follow-up	0.00955	0.00690	0.0123
	(0.00971)	(0.0122)	(0.0161)
-only in-person	0.0144	0.0125	0.0164
	(0.00964)	(0.0122)	(0.0157)
-any DCT follow-up	-0.00490**	-0.00560**	-0.00415
	(0.00229)	(0.00222)	(0.00429)
In-person follow-ups (All)	0.000861	0.0215	-0.0216
	(0.0120)	(0.0148)	(0.0179)
-initial in-person	0.0130	0.0224	0.00275
	(0.0102)	(0.0138)	(0.0151)
-initial DCT	-0.0122**	-0.000916	-0.0243***
	(0.00515)	(0.00458)	(0.00835)

Note: Estimations using individual-level daily data. Variable names in left column represent the dependent variable. The analysis decompose DCT and in person visits in initial and follow up visits (i.e. a consultation that followed a previous consultation within 14 days). Each model uses a bandwidth of 120 days before and after the 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Appendix J Robustness analysis of diagnosis data

Tables J.1, J.2, J.3, and J.4 present results using the same estimation approach as in the main text but for a range of fixed bandwidths. Table J.5 shows estimations similar to the corresponding table in the main text but with optimal bandwidth for the outcome of interest.

The results from estimations using other bandwidths support the main pattern of the substitution within diagnosis groups from table J.5. The results are overall similar, suggesting an overall degree of substitution of about 30-45%, except for the shortest bandwidth of 60 days. The result that consultations related to upper respiratory infections display the highest degree of substitution holds for all bandwidths.

The overall pattern from Table 3 remains also when splitting the sample by gender. Although there is more variation in size across bandwidths, consultations related to upper respiratory infections display the highest degree of substitution for both genders. The first stage estimates are similar across bandwidths for both genders (and jointly). The variation comes from the reduced form (the sharp diff-in-disc in-person visits).

Table J.1: Decomposition by type of diagnosis, (bw=60)

	A. DCT consultations (sharp diff-in-disc)				
	Common	Resp	Skin	Gen/Rep	Other
All	-0.133***	-0.0218***	-0.0481***	-0.0242***	-0.0478***
	(0.0154)	(0.00808)	(0.00845)	(0.00736)	(0.0112)
Men	-0.0604***	-0.00183	-0.0329***	-0.00553**	-0.0224**
	(0.0141)	(0.00794)	(0.00843)	(0.00271)	(0.00900)
Women	-0.212***	-0.0434***	-0.0646***	-0.0445***	-0.0753***
	(0.0274)	(0.0143)	(0.0139)	(0.0152)	(0.0183)
	B. Substitution (fuzzy diff-in-disc)				
	Common	Resp	Skin	Gen/Rep	Other
All	-0.160	-1.424	-0.0804	0.147	0.385
	(0.275)	(0.956)	(0.320)	(0.474)	(0.522)
Men	-0.312	-13.72	-0.175	-2.393	1.152
	(0.708)	(61.82)	(0.554)	(2.004)	(1.399)
Women	-0.109	-0.860	-0.0269	0.491	0.144
	(0.261)	(0.650)	(0.335)	(0.484)	(0.522)

Note: Table shows results from the first stage equation (sharp diff-in-disc) and the IV-model (fuzzy-diff-in-disc) by diagnosis groups. The first column shows all *Common* diagnoses set by DCT providers. In the second to fourth column, these common diagnoses are decomposed into subgroups: upper respiratory infections (*Resp*), skin related diseases (*Skin*), genital and reproductive organs *Gen/Rep*, and *Other* common diagnoses. Each row presents the diff-in-disc estimates for the diagnosis groups for the given estimation sample (All, Men, Women). Each model uses a fixed bandwidth of 60 days each side of the age cutoff. Estimates using data collapsed by region, gender, year and day relative to 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Table J.2: Decomposition by type of diagnosis, (bw=90)

	A. DCT consultations (sharp diff-in-disc)				
	Common	Resp	Skin	Gen/Rep	Other
All	-0.132***	-0.0272***	-0.0383***	-0.0274***	-0.0480***
	(0.0125)	(0.00648)	(0.00690)	(0.00629)	(0.00844)
Men	-0.0601***	-0.00566	-0.0237***	-0.00466**	-0.0277***
	(0.0114)	(0.00675)	(0.00687)	(0.00224)	(0.00720)
Women	-0.210***	-0.0504***	-0.0541***	-0.0521***	-0.0700***
	(0.0227)	(0.0114)	(0.0114)	(0.0129)	(0.0142)
	B. Substitution (fuzzy diff-in-disc)				
	Common	Resp	Skin	Gen/Rep	Other
All	-0.378*	-1.207**	-0.203	-0.00501	-0.268
	(0.220)	(0.583)	(0.324)	(0.345)	(0.437)
Men	-0.630	-3.224	-0.384	-2.322	-0.136
	(0.593)	(5.282)	(0.650)	(1.916)	(0.892)
Women	-0.295	-0.957**	-0.116	0.223	-0.320
	(0.211)	(0.468)	(0.331)	(0.328)	(0.462)

Note: Table shows results from the first stage equation (sharp diff-in-disc) and the IV-model (fuzzy-diff-in-disc) by diagnosis groups. The first column shows all *Common* diagnoses set by DCT providers. In the second to fourth column, these common diagnoses are decomposed into subgroups: upper respiratory infections (*Resp*), skin related diseases (*Skin*), genital and reproductive organs *Gen/Rep*, and *Other* common diagnoses. Each row presents the diff-in-disc estimates for the diagnosis groups for the given estimation sample (All, Men, Women). Each model uses a fixed bandwidth of 90 days each side of the age cutoff. Estimates using data collapsed by region, gender, year and day relative to 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Table J.3: Decomposition by type of diagnosis, (bw=150)

	A. DCT consultations (sharp diff-in-disc)				
	Common	Resp	Skin	Gen/Rep	Other
All	-0.129***	-0.0263***	-0.0376***	-0.0246***	-0.0457***
	(0.00933)	(0.00486)	(0.00514)	(0.00510)	(0.00590)
Men	-0.0584***	-0.00798	-0.0197***	-0.00178	-0.0308***
	(0.00887)	(0.00539)	(0.00526)	(0.00179)	(0.00555)
Women	-0.204***	-0.0460***	-0.0570***	-0.0493***	-0.0619***
	(0.0175)	(0.00860)	(0.00850)	(0.0104)	(0.0102)
	B. Substitution (fuzzy diff-in-disc)				
	Common	Resp	Skin	Gen/Rep	Other
All	-0.422**	-1.589***	0.0381	0.167	-0.545
	(0.178)	(0.488)	(0.262)	(0.296)	(0.365)
Men	-0.773	-2.460	-0.257	-1.764	-0.797
	(0.489)	(2.518)	(0.636)	(3.694)	(0.669)
Women	-0.309*	-1.420***	0.149	0.247	-0.403
	(0.171)	(0.434)	(0.254)	(0.275)	(0.404)

Note: Table shows results from the first stage equation (sharp diff-in-disc) and the IV-model (fuzzy-diff-in-disc) by diagnosis groups. The first column shows all *Common* diagnoses set by DCT providers. In the second to fourth column, these common diagnoses are decomposed into subgroups: upper respiratory infections (*Resp*), skin related diseases (*Skin*), genital and reproductive organs *Gen/Rep*, and *Other* common diagnoses. Each row presents the diff-in-disc estimates for the diagnosis groups for the given estimation sample (All, Men, Women). Each model uses a fixed bandwidth of 150 days each side of the age cutoff. Estimates using data collapsed by region, gender, year and day relative to 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Table J.4: Decomposition by type of diagnosis, (bw=180)

	A. DCT consultations (sharp diff-in-disc)				
	Common	Resp	Skin	Gen/Rep	Other
All	-0.129***	-0.0297***	-0.0339***	-0.0257***	-0.0459***
	(0.00834)	(0.00432)	(0.00483)	(0.00454)	(0.00542)
Men	-0.0568***	-0.0109**	-0.0168***	-0.000292	-0.0306***
	(0.00808)	(0.00494)	(0.00490)	(0.00169)	(0.00499)
Women	-0.208***	-0.0500***	-0.0526***	-0.0534***	-0.0626***
	(0.0157)	(0.00761)	(0.00798)	(0.00931)	(0.00952)
	B. Substitution (fuzzy diff-in-disc)				
	Common	Resp	Skin	Gen/Rep	Other
All	-0.308*	-0.882**	-0.0191	0.0997	-0.397
	(0.157)	(0.349)	(0.263)	(0.263)	(0.322)
Men	-0.739	-1.193	-0.348	-24.77	-0.741
	(0.454)	(1.322)	(0.685)	(143.7)	(0.607)
Women	-0.174	-0.800**	0.0951	0.252	-0.205
	(0.153)	(0.322)	(0.250)	(0.238)	(0.362)

Note: Table shows results from the first stage equation (sharp diff-in-disc) and the IV-model (fuzzy-diff-in-disc) by diagnosis groups. The first column shows all *Common* diagnoses set by DCT providers. In the second to fourth column, these common diagnoses are decomposed into subgroups: upper respiratory infections (*Resp*), skin related diseases (*Skin*), genital and reproductive organs *Gen/Rep*, and *Other* common diagnoses. Each row presents the diff-in-disc estimates for the diagnosis groups for the given estimation sample (All, Men, Women). Each model uses a fixed bandwidth of 180 days each side of the age cutoff. Estimates using data collapsed by region, gender, year and day relative to 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Table J.5: Decomposition by type of diagnosis (optimal bandwidth)

	A:DCT consultations (First stage: sharp diff-in-disc)				
	Common	Resp	Skin	Gen/Rep	Other
All	-0.129***	-0.0256***	-0.0379***	-0.0259***	-0.0468***
	(0.0110)	(0.00572)	(0.00609)	(0.00584)	(0.00749)
Men	-0.0600***	-0.00581	-0.0206***	-0.00484**	-0.0298***
	(0.0105)	(0.00618)	(0.00632)	(0.00207)	(0.00679)
Women	-0.204***	-0.0468***	-0.0566***	-0.0487***	-0.0651***
	(0.0208)	(0.0104)	(0.0103)	(0.0119)	(0.0129)
	B: Substitution (fuzzy diff-in-disc)				
	Common	Resp	Skin	Gen/Rep	Other
All	-0.440**	-1.447**	-0.105	-0.0393	-0.472
	(0.205)	(0.575)	(0.301)	(0.346)	(0.417)
Men	-0.624	-2.253	-0.391	-1.169	-0.478
	(0.546)	(3.886)	(0.695)	(1.501)	(0.771)
Women	-0.378*	-1.334***	0.00721	0.0863	-0.464
	(0.202)	(0.504)	(0.294)	(0.329)	(0.460)

Note: Table shows results from the first stage equation (sharp diff-in-disc) and the IV-model (fuzzy-diff-in-disc) by diagnosis groups. The first column shows all *Common* diagnoses set by DCT providers. In the second to fourth column, these common diagnoses are decomposed into subgroups: upper respiratory infections (*Resp*), skin related diseases (*Skin*), genital and reproductive organs *Gen/Rep*, and *Other* common diagnoses. Each row presents the diff-in-disc estimates for the diagnosis groups for the given estimation sample (All, Men, Women). All models apply the bandwidths used for the main results in Table 2 (i.e., MSE-optimal bandwidths for In-person visits for both regions and both genders). Estimates using data collapsed by region, gender, year and day relative to 20th birthday. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Appendix K Prescriptions of antibiotics

Table K.1 shows sharp diff-in-disc estimates of the effects of the onset of the DCT user fee on various outcomes related to antibiotic prescriptions.[4] The positive and significant coefficients in the first column indicate that the user fee if anything has a small positive effect on the total number of antibiotic prescriptions. That is, the larger use of DCT consultations among 19–year-olds relative to 20–year-olds does not increase their antibiotics consumption, but rather decreases it.

The results by antibiotic type (columns 2-4) further suggest that the increase is primarily driven by prescriptions of antibiotics related to respiratory infections (rather than antibiotics related to skin conditions or cystitis). Together with the results in section 4.2.2, which suggest that close to all DCT consultations for respiratory infection replace in-person visits, we interpret these results as indications of physicians being more (or at least not less) restrictive during online consultations in terms of prescribing antibiotics for respiratory infections.

As for the main analysis, the diff-in-disc estimates are sensitive to general trends that would affect the size of the drop at the 20th birthday even without the onset of the DCT user fee. Indeed, antibiotic use has declined over the last decade, and a proportional decrease in the number of prescriptions for individuals each side of the cut-off would generate positive diff-in-disc estimates as observed in Table K.1. We therefore study the components of the diff-in-disc estimates - the pre and post RD estimates. We also specifically estimate a standard RD in Stockholm, where there is no confounding user fee for in-person visits.

Table K.2 shows results from sharp RD before (panel A includes the complete pre-period) and after (panel B includes 2018) the introduction of the DCT-services. The 20th birthday is associated with a decrease in the number of prescriptions and defined daily doses (DDD) before the DCT, likely driven by the onset of the user fee in the region of Västra Götaland. In 2018, the same discontinuity is associated with an insignificant increase in the number of prescriptions and DDDs. Thus, these results implies that the estimates in table K.1 is not only driven by a general decrease in the number of prescriptions but an actual change in the sign of the effect at the discontinuity.

Table K.3 presents results for the same outcomes as in the previous tables from a sharp diff-in-disc (panel A) and a RD in 2018 (panel B) for the region

[4] Antibiotics is defined as all at-codes within J01 except Metenamin J01XX05. We follow the Public Health Agency of Sweden defining three categories of antibiotics relating to respiratory infections (J01AA02, J01CE02, J01CA04, J01CR02, J01DB, J01DC, J01DE, J01FA) cystitis (J01CA08, J01EA01, J01MA02, J01MA06, J01XE01) Skin and soft tissues (J01FF01, J01CF05).

Table K.1: Antibiotic prescription, sharp diff-in-disc

	A. Sharp diff-in-disc				
	All	Resp	Skin	Cystit	DDD
All	0.0368***	0.0161	0.00724	0.00865	0.557
	(0.0141)	(0.0108)	(0.00609)	(0.00746)	(0.442)
Men	0.0223	0.00680	0.00270	0.000913	0.733
	(0.0173)	(0.0130)	(0.00904)	(0.00381)	(0.578)
Women	0.0519**	0.0259	0.0121	0.0167	0.365
	(0.0230)	(0.0172)	(0.00867)	(0.0149)	(0.557)

Note: Table shows results from a sharp diff-in-disc (reduced form) on antibiotic prescriptions. The first column presents diff-in-disc estimates for the total number of antibiotic prescriptions, columns 2 to 4 present estimates for the number of prescriptions of various types: respiratory infections, skin conditions, and cystit. Column 5 presents estimates for the total number of Defined Daily Doses (DDD). Standard errors clustered by the running variable, with separate clusters for pre-post periods. Each model uses a fixed bandwidth of 120 days on each side of the threshold (20th birthday). * P<0.1, ** P<0.05, *** P<0.01.

of Stockholm. Although the RD estimates are all insignificant and tend to be smaller than the diff-in-disc estimates, the conclusion is still that physicians are not less restrictive in terms of prescribing antibiotics during DCT consultations – if anything they are more restrictive.

Table K.2: Sharp RD of antibiotic prescriptions, pre/post

	A: Sharp RD of antibiotic prescriptions, pre				
	All	Resp	Skin	Cystit	DDD
All	-0.0177**	-0.0141***	-0.000852	-0.00348	-0.317
	(0.00699)	(0.00469)	(0.00268)	(0.00356)	(0.206)
Men	-0.0112	-0.0119**	0.00000120	-0.000440	-0.178
	(0.00825)	(0.00587)	(0.00351)	(0.00160)	(0.280)
Women	-0.0245**	-0.0164**	-0.00176	-0.00668	-0.464
	(0.0120)	(0.00740)	(0.00442)	(0.00744)	(0.295)
	B: Sharp RD of antibiotic prescriptions, 2018				
	All	Resp	Skin	Cystit	DDD
All	0.0187	0.00426	0.00683	0.00473	0.211
	(0.0133)	(0.0104)	(0.00592)	(0.00721)	(0.403)
Men	0.0111	-0.00509	0.00270	0.000473	0.554
	(0.0153)	(0.0116)	(0.00836)	(0.00346)	(0.507)
Women	0.0274	0.00951	0.0104	0.0100	-0.0989
	(0.0197)	(0.0156)	(0.00748)	(0.0130)	(0.474)

Note: Table shows results from a sharp regression discontinuity before (panel A) and after (paned B) the introduction of DCT-services. In panel B, we show results for 2018. The first row presents the RD-estimates for any type of prescriptions, columns 2 to 4 present estimates for antibiotic types related to respiratory infections, skin conditions, and cystit. Column 5 presents the same estimate for the total number of Defined Daily Doses (DDD). Each model uses a fixed bandwidth of 120 days on each side of the threshold (20th birthday). Data is collapsed by gender, year and day relative to 20th birthday. Standard errors clustered by the running variable. * P<0.1, ** P<0.05, *** P<0.01.

Table K.3: Sharp RD of antibiotic prescriptions (Stockholm 2018)

	A: DIFF-IN-DISC				
	All	Resp	Skin	Cystit	DDD
All	0.0428**	0.0238	0.00292	0.0124	0.826
	(0.0196)	(0.0150)	(0.00877)	(0.0101)	(0.545)
Men	0.0154	0.0131	-0.00601	0.00251	0.732
	(0.0239)	(0.0181)	(0.0127)	(0.00523)	(0.673)
Women	0.0711**	0.0348	0.0124	0.0224	0.919
	(0.0321)	(0.0239)	(0.0114)	(0.0202)	(0.811)
	B: REGRESSION DISCONTINUITY				
	All	Resp	Skin	Cystit	DDD
All	0.0279	0.0106	0.00117	0.00719	0.717
	(0.0173)	(0.0136)	(0.00811)	(0.00863)	(0.465)
Men	0.0111	0.00649	-0.00673	0.00130	0.805
	(0.0214)	(0.0163)	(0.0120)	(0.00479)	(0.562)
Women	0.0449	0.0147	0.00959	0.0130	0.617
	(0.0281)	(0.0220)	(0.00994)	(0.0171)	(0.681)

Note: The table shows estimates from sharp diff-ind-disc in Stockholm in panel A and a sharp regression discontinuity for the Stockholm region using data from 2018 in panel B. The first column presents estimates for any type of prescriptions, columns 2 to 4 present estimates for antibiotic types related to respiratory infections, skin conditions, and cystit. Column 5 presents the same estimate for the total number of Defined Daily Doses. Each model uses a fixed bandwidth of 120 days on each side of the threshold (20th birthday). Data is collapsed by gender, year and day relative to 20th birthday. Standard errors clustered by the running variable. * P<0.1, ** P<0.05, *** P<0.01.

Appendix L Other primary care consultations

Table L.1 displays fuzzy diff-in-disc results for primary care consultations with other health care professionals than physicians (e.g., nurses) using other bandwidths than Table 4 in the main text. (Table L.2 displays the estimates for all consultations (i.e., the sum of consultations with other health care professionals and physicians) for other bandwidths than in the main text.) Table L.3 shows the same estimations using a MSE-optimal bandwidth. The patterns are consistent across tables and bandwidths. The results suggest that the degree of substitution is slightly larger when including consultations with other health care professionals than physicians. The estimate suggests that there is partial substitution, but the estimates are noisier than when we include physician consultations only and the confidence intervals no longer exclude full substitution.

For women, there is a consistent negative (but insignificant) coefficient on consultations at midwife/youth/STD clinics. To further explore this pattern, we retain the same outcome variable but restrict the first stage to only include online consultations with diagnoses related to the genital and reproductive organs. Table L.4 presents the fuzzy diff-in-disc estimates. For women, the results hover around -1, which suggests that women substitute online consultations (with physicians) for in-person midwife visits related to contraceptive management. In other words, the observed the lack of substitution between in-person *physician* visits and online consultations for these diagnoses (Table J.5) is explained by another type of substitution. Note that the large positive coefficients among men is primarily due to a very weak first stage (and likely relate to visits at a STD or youth clinic).

Table L.1: Visits to other health care professionals (fixed bandwidth)

	Fuzzy Diff-in-disc		
	All	Men	Women
	Bandwidth=60		
Nurse visits at a PCC	-0.217	-0.474	-0.130
	(0.166)	(0.416)	(0.176)
Visits at midwife/youth/STD clinic	0.00387	0.457	-0.136
	(0.260)	(0.279)	(0.328)
Nurse+midwife/youth/STD	-0.213	-0.0166	-0.266
	(0.311)	(0.458)	(0.391)
	Bandwidth=90		
Nurse visits at a PCC	-0.0792	-0.721**	0.129
	(0.132)	(0.351)	(0.140)
Visits at a midwife/youth/STD clinic	-0.0717	0.361	-0.201
	(0.205)	(0.228)	(0.258)
Nurse+midwife/youth/STD	-0.151	-0.360	-0.0719
	(0.243)	(0.389)	(0.295)
	Bandwidth=150		
Nurse visits at a PCC	-0.0431	-0.632**	0.135
	(0.110)	(0.317)	(0.109)
Visits at a midwife/youth/STD clinic	-0.0734	0.477**	-0.227
	(0.163)	(0.217)	(0.199)
Nurse+midwife/youth/STD	-0.117	-0.155	-0.0918
	(0.197)	(0.346)	(0.226)
	Bandwidth=180		
Nurse visits at a PCC	-0.0594	-0.608**	0.113
	(0.104)	(0.284)	(0.104)
Visits at a midwife/youth/STD clinic	-0.151	0.459**	-0.323*
	(0.150)	(0.199)	(0.184)
Nurse+midwife/youth/STD	-0.211	-0.149	-0.209
	(0.184)	(0.321)	(0.210)

Note: Table shows fuzzy diff-in-discs estimates of the effect of online consultations on in-person consultations with other health care professionals than physicians at primary care centers: consultations with a nurse at a primary care center; consultations with a midwife/nurse/physician at a midwife/youth/STD clinic, and the sum of these two types of consultations. ; and the sum of all consultations, , including physician consultations at a primary care center. Each model uses a fixed bandwidth of 60/90/150/180 days each side of the threshold. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Table L.2: Visits to other health care professionals, fixed bandwidth

	FUZZY DIFF-IN-DISC All	Men	Women
	BANDWIDTH=60		
All consultations	-0.441	-0.769	-0.316
	(0.497)	(0.929)	(0.552)
	BANDWIDTH=90		
All consultations	-0.591	-1.325*	-0.338
	(0.376)	(0.755)	(0.412)
	BANDWIDTH=150		
All consultations	-0.588**	-1.277*	-0.363
	(0.299)	(0.686)	(0.308)
	BANDWIDTH=180		
All consultations	-0.582**	-0.991	-0.426
	(0.273)	(0.613)	(0.282)

Note: Table shows fuzzy diff-in-discs estimates of the effect of DCT consultations on in-person consultations with physicians or other health care professionals at primary care centers. (consultations with a nurse at a primary care center; consultations with a midwife/nurse/physician at a midwife/youth/STD clinic, and physician consultations at a primary care center). Each model uses a fixed bandwidth of 60/90/150/180 days each side of the threshold. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

Appendix M Cost analysis

M.1 Overview and sources

We lack information to perform a full cost-benefit analysis,[5] but we are able to calculate a rough estimate of the relative costs of providing a consultation in the DCT and PCC settings, respectively.

[5] For instance, we cannot value the utility of patients' decreased worry for minor conditions that DCT but not PCCs would accept to take care of.

Table L.3: Visits to other health care professionals

	FUZZY DIFF-IN-DISC		
	All	Men	Women
Nurse visits at a PCC	-0.0268	-0.519	0.137
	(0.123)	(0.312)	(0.127)
Visits at midwife/youth/STD clinic	-0.0611	0.376	-0.197
	(0.183)	(0.212)	(0.230)
Nurse+midwife/youth/STD	-0.0879	-0.143	-0.0602
	(0.220)	(0.347)	(0.265)
All consultations	-0.541	-1.032	-0.366
	(0.331)	(0.691)	(0.360)

Note: Table shows fuzzy diff-in-discs estimates of the effect of online consultations on in-person consultations with other health care professionals than physicians at primary care centers: consultations with a nurse at a primary care center; consultations with a midwife/nurse/physician at a midwife/youth/STD clinic, the sum of these two types of consultations; and the sum of all consultations, including physician consultations at a primary care center. Each model uses the MSE-optimal bandwidth for main outcome variable (in-person consultations) for both regions and both genders (to remove heterogeneity due to changes in bandwidth). Standard errors clustered by the running variable, with separate clusters for pre-post periods. P<0.1, ** P<0.05, *** P<0.01.

Dahlstrand (2022) provides information about the labour cost per hour and consultation times of the largest DCT provider in Sweden. DCT physicians on average spend 5 minutes on the direct consultation and 11 minutes on administration.

To estimate the consultation time at PCCs, we use the standard consultation length in the Swedish system for certain private-practising physicians ("Taxeläkare"), i.e., 25 minutes including patient administration time (see §5 of SFS (1991:1121)). We decompose the total time into the time on the actual consultation vs administration time using estimates from a recent time use study in Swedish PCCs (See Table 4.1 (p. 85) in Ivarsson Westerberg et al. (2021). Notably, the cited regulated consultation length for private practitioners aims to capture a reasonable average for a much more heterogeneous group of patients than those seeking DCT care (including, e.g., elderly patients with multiple illnesses). It is therefore likely that it provides an overestimate of the time that PCC physicians would spend on

Table L.4: Consultations at a midwife/youth/STD clinic

	Fuzzy Diff-in-disc Instrument: gen/rep DCT		
	All	Men	Women
	Optimal bw, main outcome		
Visits at a midwife/youth/STD clinic	-0.350	5.522	-0.939
	(1.048)	(3.881)	(1.114)
	Bandwidth=60		
Visits at a midwife/youth/STD clinic	0.0231	5.730	-0.691
	(1.549)	(4.424)	(1.673)
	Bandwidth=90		
Visits at a midwife/youth/STD clinic	-0.395	5.509	-0.917
	(1.129)	(4.316)	(1.192)
	Bandwidth=120		
Visits at a midwife/youth/STD clinic	-0.785	7.115	-1.366
	(1.085)	(5.721)	(1.135)
	Bandwidth=150		
Visits at a midwife/youth/STD clinic	-0.426	17.00	-1.052
	(0.947)	(18.52)	(0.943)
	Bandwidth=180		
Visits at a midwife/youth/STD clinic	-0.829	101.6	-1.356*
	(0.829)	(589.4)	(0.798)

Note: Table shows fuzzy diff-in-discs estimates of the effect of online consultations (with a registered diagnosis related to genital and reproductive health) on in-person consultations with a midwife/nurse/physician at a midwife/youth/STD clinic. That is, the excluded instrument is DCT consultations with a registered diagnosis related to genital and reproductive health. See section 4.2.2. Each model uses a fixed bandwidth of 60/12/180 days each side of the threshold. Standard errors clustered by the running variable, with separate clusters for pre-post periods. * P<0.1, ** P<0.05, *** P<0.01.

DCT-equivalent complaints. In some alternative scenarios, we therefore use the same consultation lengths for PCCs and DCT. The main reason why we use the 25 minute length in the baseline scenario is that it is possible that the in-person setting by itself prompts the physician to perform a physical examination, and to accompany the patient from/to the waiting room.

Dahlstrand (2022) also provides information about the company's labour costs (75-90 USD per hour), and the shares of physicians that are specialists (0.31) versus still in specialty training (0.36) or residents only (0.33). We assume that the lower wage goes for the latter two categories and calculate an average wage cost per hour as a weighted sum (i.e., 0.69*75+0.31*90), which gives us SEK734 per hour.

We set PCC physicians' wages to 1,000 SEK per hour, an estimate sourced from an administrator working with the costing model in one of the Swedish regions.[6] We also calculate costs under the assumption of similar physician wages in DCT and PCC in some scenarios.[7]

Before the actual DCT appointment, the patient has to spend some time finding the site or application and filling out a triage form. Similarly, the patient that contacts a PCC first has to speak to a triage nurse. We lack good sources here so we just assume that the amount of time is similar (10 minutes) in both cases. Notably, it is only in the PCC setting that this tiem entails a cost for the health care system, namely the cost for the triage nurse. We use an estimate of an hourly wage using the official average wage statistics for nurses (296 SEK per hour[8] including income and payroll taxes) with basic or specialist education.

The in-person setting also includes some other activities that only involve the patient (i.e., do not affect the costs of the health care system). The patient will normally have to spend time in a telephone queue before talking to the nurse, and then in the waiting room. We set this time to 30 minutes. We further assume that patients spend 22 minutes travelling to/from the PCC. We derive this estimate using the figures provided in a report by the Swedish Competition Authority, which described the shares of the population in the regions that can reach their PCC within 5, 15, 30 or above 30 minutes) (Swedish Competition Authority, 2010).[9]

[6] András Borsodi, Östergötland

[7] The wage differential is not particularly interesting *per se*: The ability to hire junior physician is not the interesting innovation in DCT and is not feasible in a general equilibrium.

[8] We divide the average annual wage times by the statutory working time of 2,080 hours/year.

[9] Specifically, we multiply the population share in each bracket with the upper bound for the time limit in the bracket (e.g., 5 minutes), sum over all population shares,

In our study regions, the vast majority have a short distance from their home to their closest PCC. We also assume that around half of the study population lives in Stockholm. The direct cost for a bus ticket is estimated at 80 SEK (round trip).

Patient's opportunity cost of time is set to 100 SEK in our baseline scenario. Assuming an opportunity cost of 100 SEK is equivalent to assuming that compliers are evenly spaced in the 0-200 SEK range. Our assumption about patients' opportunity cost of time thus reflects the idea that the sensitivity to the user fee should tell us something about the patients' valuation of the visit. However, 100 SEK is likely an understatement of the value of time for many adults. In an alternative scenario, we therefore use average wage per hour (official wage statistics, including income and payroll tax) instead.

Finally, we impose a cost for office space of 36 SEK/visit in DCT and 200 SEK/visit in PCC. The estimate for DCT comes from Ekman (2018) and is probably an overstatement given the subsequent growth in volume (which should led to less than proportional increases in the need for space, especially since the medical workforce works from home). To estimate the office space costs per visit in the PCC setting, we use data from a costing model in a Swedish region (Östergötland), which have a comparably sophisticated such model for primary care (cost data is generally scant in Swedish primary care). To account for the higher rents in the our study regions, we add 70 SEK to the base 130 SEK from the costing model.

All other costs are disregarded. In reality, one may imagine that DCT providers have lower overhead costs (since they manage a less complex service, they might need fewer managers and administrative personnel per clinician, and they do not need a receptionist). However, and notably, to realize such cost savings at the provider level, some PCCs would need to close. Closures are often politically infeasible, and where they are not, they by themselves induce counterveiling cost increases in terms of travel time for some patients.

One notable cost from which we abstract is the development and maintenance of the virtual consultation software. Notably, though, PCCs too nowadays use similar systems for remote communication, so the omission is more a theoretical than a practical problem.

---

weigh by the population shares living in each region (assumed 50-50 in our study populations), and multiply by 2 to get a round trip. 22 minutes = 0.36 of an hour = =(2*(0.5*(0.72*5+0.24*15+0.04*30)+0.5*(0.41*5+0.43*15+0.16*30)))/60

Table M.1: Cost items and values in scenarios 1-6

Cost component	**Mode**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Physician wage (SEK)	DCT	734	734	734	734	734	734
	PCC	1000	1000	734	734	734	1000
Consultation (share of hour)	DCT	0.08	0.08	0.08	0.08	0.08	0.08
	PCC	0.21	0.08	0.21	0.08	0.08	0.21
Admin (share of hour)	DCT	0.18	0.18	0.18	0.18	0.18	0.18
	PCC	0.2	0.18	0.2	0.18	0.18	0.2
Triage nurse wage (SEK)	DCT	N/A	N/A	N/A	N/A	N/A	N/A
	PCC	296	296	296	296	296	296
Booking and triage (share of hour)	DCT	0.17	0.17	0.17	0.17	0.17	0.17
	PCC	0.17	0.17	0.17	0.17	0.17	0.17
Office space (SEK)	DCT	36	36	36	36	36	0
	PCC	200	200	200	200	200	0
Patient's opportunity cost (SEK)	DCT	100	100	100	100	273	100
	PCC	100	100	100	100	273	100
Patient's travel time (share of hour)	DCT	0	0	0	0	0	0
	PCC	0.36	0.36	0.36	0.36	0.36	0.36
Patient's travel cost (SEK)	DCT	0	0	0	0	0	0
	PCC	80	80	80	80	80	80
Patient's waiting time (share of hour)	DCT	0	0	0	0	0	0
	PCC	0.5	0.5	0.5	0.5	0.5	0.5

The table shows the values used to compute relative unit costs for providing a physician consultation in the two modes of care (DCT and PCC). (1) is our baseline scenario.

M.2 Cost scenarios

We estimate costs in six scenarios. Table M.1 gives an overview of the parameters used in each scenario, where (1) is the baseline scenario.

Baseline: In our baseline scenario, we estimate that direct health care production costs of a DCT consultation is 35% of an those for in-person visit. Hence, each DCT consultation is resource-saving even from a health care budget perspective. Adding the value of saved time (indirect costs) of course increases the cost savings even further, although the estimate is still in the same ball park (30% of in-person costs).

Scenario 2: *Like baseline, except assume same physician consultation times as in DCT.* When using the same physician time in the two modes of care, the (pure health care) cost ratio increases to 45% (36% when including the value of patients' time savings). The nurse-led triage step, the office space and higher wages in traditional care thus still makes DCT a cheaper alternative.

Scenario 3: *Like baseline, except assume same labour costs as in DCT.* Compared to the baseline scenario, the cost ratio increases to 42% (34% when including patients' time costs). In this scenario, it is the nurse-led triage step, the office space and the longer consultation times in traditional care that makes DCT a cheaper alternative.

Scenario 4: *Like baseline, except assume same labour costs AND consultation times as in DCT).* This scenario combines scenarios 2 and 3. In scenario, 4, physicians spend the same time on patients in DCT and PCCs, and the wage levels are the same. Compared to baseline, the (pure health care) cost ratio increases to 52% (40% when including patients' time costs). In this case, it is only the nurse-led triage step and the office space that makes DCT a cheaper alternative. Although this estimate, together with our estimated substitution rate, indicate an increase in the health care budget, it is almost budget neutral (and still clearly resource saving when including patients' time savings).

Scenario 5: *Like baseline, except assume same labour costs AND consultation times as in DCT AND assume patient opportunity cost using mean wage in Sweden.* In scenario 5, we use the mean wage (including income and payroll taxes) – around 270 SEK – to see how much the estimate of the opportunity cost of time matters. We use the same wage and time estimates as in scenario 4 and therefore compare these two scenarios. As we only change parameters related to the patient, this scenario only affects the cost ratio that includes the value of patients' time, which decreases from 40% in scenario 4 to 36% in scenario 5, implying that the estimated opportunity cost of time is not crucial for the cost estimates.

Scenario 6: *Like baseline, except assume no office costs.* This scenario abstracts from office space. It is equivalent to assuming that the existence or absence of DCT does not affect PCCs' office space decisions. Otherwise, everything else is as in the baseline scenario, i.e., lower wages and shorter consultation times in DCT. Compared to the baseline scenario, the cost ratio increases from 35% to 42% (30% to 33% when including patients' time costs). Thus, office space plays a role, but does not change the conclusion that the degree of substitution is high enough to establish budget neutrality from a health care perspective and even save resources in a broader perspective.

Comparison with previous estimates: An early cost analysis comparing one of the DCT start-ups with standard primary care in 2016 (Ekman, 2018) arrived at a substantially higher estimate, suggesting that the cost per visit in DCT was 58% of that in the in-person setting (even though the value of patients' time saved was included). One reason why Ekman arrived at a higher costs ratio was that he aimed to include more costs, e.g., admin and support, management, equipment, operations and – importantly – technical development, and write-offs. Since his figures were from a very early phase of the DCT provider's operations, it is likely that these fixed costs are overestimated per visit in a setting with higher volumes and (presumably) lower costs for developing the technical platform (which, as already noted, nowadays exist for PCCs as well). If we subtract the costs for equipment, operations and technical development from Ekman's calculations, the cost ratio is 42%, i.e., similar to our estimates.

To sum up, our cost calculations indicates that the availability of DCT does not increase health care costs, i.e., the unit costs are low enough relative to those in PCCs (35-52%) to make up for the additional demand. Adding the valuation of patients' saved time makes it even more likely that the availability of DCT is resource-saving at the societal level.

References

Angrist, Joshua D, and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics.* Princeton university press.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica*, 82(6): 2295–2326.

Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2019. *A Practical Introduction to Regression Discontinuity Designs: Foundations.* Cambridge:Cambridge University Press.

Dahlstrand, Amanda. 2022. "Defying Distance? The Provision of Services in the Digital Age." Unpublished manuscript.

Ekman, Björn. 2018. "Cost Analysis of a Digital Health Care Model in Sweden." *PharmacoEconomics - Open*, 2(3): 347–354.

Ellegård, Lina Maria, and Gustav Kjellsson. 2019. "Nätvårdsanvändare i Skåne kontaktar oftare vårdcentral och gör inte färre akutbesök." *Läkartidningen*, 116.

He, Yang, and Otavio Bartalotti. 2020. "Wild Bootstrap for Fuzzy Regression Discontinuity Designs: Obtaining Robust Bias-Corrected Confidence Intervals." *Econometrics Journal*, 23(2): 211–231.

Ivarsson Westerberg, Anders, Agneta Andersson, Eva Anskär, Daniel Castillo, Magnus Falk, and Anders Forssell. 2021. *Papper, pengar och patienter : primärvården i administrationssamhället.* Lund:Studentlitteratur.

Millán-Quijano, Jaime. 2020. "Fuzzy difference in discontinuities." *Applied Economics Letters*, 27(19): 1552–1555.

SFS. 1991:1121. "Förordning om läkarvårdsersättning [Act on remuneration of physician care]."

Swedish Competition Authority. 2010. "Uppföljning av vårdval i primärvården – Valfrihet, mångfald och etableringsförutsättningar." Konkurrensverkets rapportserie 2010.