Ostrobothnia Digital Clinic Experiment: Pre-Analysis Plan for a Randomized Controlled Trial

Tapio Haaga, Mika Kortelainen, Oskari Nokso-Koivisto, Tanja Saxell, Meeri Seppä, Lauri Sääksvuori*

March 2025†

Abstract

This randomized controlled trial (RCT) examines the impacts of a publicly provided digital clinic that offers digital primary care services to consumers. Our intervention grants access to a public digital clinic that provides chat-based primary care consultations via a mobile phone application and website, including care needs assessment, diagnoses, follow-up care recommendations, and prescriptions. The digital clinic supplements traditional public primary care services, including in-person visits and phone consultations. The trial takes place in Ostrobothnia, Finland, a healthcare district serving a population of 178,000 residents. We randomize access to the digital clinic at the household level, providing access to 50% of the households. By doing so, we aim to evaluate whether digital services can substitute for, complement, or increase the utilization of traditional primary care, particularly in-person visits or calls to traditional clinics. At the end of the nine-month trial, access to the digital clinic will be expanded to the entire population.

Keywords: digital healthcare, telemedicine, digital clinic, primary care, healthcare utilization, randomized controlled trial

*__Haaga:__ Finnish Institute for Health and Welfare (THL) and University of Turku (UTU). __Kortelainen:__ UTU and THL. __Nokso-Koivisto:__ Aalto University (Aalto). __Saxell:__ Aalto and VATT Institute for Economic Research. __Seppä:__ Aalto. __Sääksvuori:__ THL.

†__Funder:__ The Finnish Ministry of Social Affairs and Health (STM).
__Ethical assessment:__ Finnish Institute for Health and Welfare Institutional Review Board (THL/5935/6.02.01/2024).
__Replication codes:__ https://github.com/tapiohaa/SoteDataLab/tree/main/ostrobothnia_digital_clinic_experiment.
__Acknowledgements:__ the Wellbeing Services County of Ostrobothnia (and Suvi Einola, Sofi Nyman, Petra Fager, Marina Kinnunen, Pia Haglund, and Satu Hautamäki); SAG Flowmedik (and Michael Rossing and Roope Vuorijärvi) which provided the digital clinic platform for Pohjanmaa; STM (Markku Heinäsenaho, Aleksi Yrttiaho and the STM-led coordination group for the project; Andreas Blanco Sequeiros); major supporters and enablers Mikko Peltola and Sonja Lumme (THL) and Heikki Kauppi (UTU); Alex Kivimäki, Aurora Morén, and all other members of our SoteDataLab research group (sotedatalab.fi); Aleksi Reito who provided comments on the PAP; research partners Tuomas Koskela, Kaisa Kujansivu and Elina Tolvanen; Annaleena Okuloff (THL) for extracting the study population and conducting the randomization, Niko Kivimäki Wilders (THL) for data extraction; Reetta Salokannel (Statistics Finland) for data pseudonymisation; Kirk Bansak, Lina Maria Ellegård, Tarja Heponiemi, Mikko Herzig, Suvi Hämäläinen, Gustav Kjellsson, Carol Propper, Katja Rääpysjärvi, Markku Satokangas, Juha Tolvanen, Tuulikki Vehko, Dan Zeltzer, and all wellbeing services counties, private providers, and collegues who have kindly discussed with us and provided ideas and comments.

Contents

1 Introduction

This document outlines our analysis plan for a randomized controlled trial (RCT) entitled *Ostrobothnia Digital Clinic Experiment*. Our complete analysis plan and experimental protocol consist of three separate documents written and made public in the following order:

1. A general-level pre-analysis plan (PAP) describing the main idea, study design, data sources, primary and secondary outcomes, and our statistical approach. This is the document you are currently reading. This PAP was registered *before* initiating the trial.

2. A statistical analysis plan (SAP), which maps the idea into code based on blinded/placebo data. Statistical codes are included at this step of the registration, illustrating in detail how we plan to construct our variables and analysis data. Our current objective is to register the SAP *after* initiating the trial and observing aggregate statistics on the utilization of digital services, but *before* linking the treatment assignment and outcome data. This process will be verifiable by a third party. We may include additional analyses in the SAP compared to the PAP.

3. A populated SAP, which replaces the tables and figures in the SAP with real data after the trial has ended. At this stage, no deviations from the SAP will be made.

Finally, we will produce a research paper, or a series of papers, intended for publication in scholarly journals. The final research papers are distinct from the populated SAP as they may contain not only confirmatory analyses registered in the PAP and SAP but also post-blind and exploratory analyses conducted after linking the treatment indicator and outcome data. In our research papers, we will clearly differentiate between pre-registered analyses and post-blind analyses.

The structure of this PAP is as follows: Section 1 discusses the key motivation for our experiment and articulates our key research questions. Section 2 describes the institutional context. Section 3 describes the intervention. Section 4 provides details on the extraction of the study population, our inclusion and exclusion criteria, and the randomization process. Section 5 presents our data sources and specifies our outcomes of interest. Section 6 presents our main

statistical approach and planned key tables and figures. Section 7 outlines potential extensions to the pre-registered analyses.

This document is accompanied by four appendices, which present the statistical analyses that have informed the design of our experiment and discuss the ethical aspects of the study. Appendix A documents our simulations used to assess how different randomization schemes and sets of covariates affect the precision of suggested estimations. Appendix B documents our power analyses. Appendix C discusses the likelihood of interference in the context of our experiment and presents a back-of-the-envelope model to assess the Stable Unit Treatment Value Assumption (SUTVA) in our experimental context. Appendix D discusses research ethics and documents the administrative steps preceding the implementation of the experiment.

1.1 Study motivation

Digital services leverage technological solutions to deliver online services to consumers through digital platforms, such as mobile applications, websites, and marketplaces. As in other digital service industries, digital platforms in healthcare offer consumers fast and easy access to services, potentially shifting demand from in-person visits and phone hold queues to online platforms. In healthcare, this trend is largely driven by increased convenience and lower costs of online services (Dorsey & Topol, 2020). While there is a growing body of literature examining the effects of health information technologies (Goldfarb & Tucker, 2019), the impacts of digital services remains much less well understood. Importantly, digital healthcare services may not only substitute for traditional health care but also induce new demand and utilization that would not have occurred without the digital channel (Dahlgren et al., 2023; Ellegård et al., 2022). The convenience of digital care (no time spent traveling, shorter waiting times, extended opening hours, and access from any geographical location) can lower the barriers to access, potentially reducing underdiagnosis and undertreatment, or, conversely, exacerbating the utilization of medically low-value services.

Through random assignment of patients to a digital clinic, we aim to contribute to a better understanding of the impacts of digital healthcare services. We conduct a large, region-wide RCT

in Finland, where we randomly grant access to digital primary care services. Our intervention is implemented at the household level and provides digital medical services during a nine-month period to 50% of households in an administrative region with approximately 170,000 individuals.

The digital clinic provides chat-based primary care consultations through a mobile app and website, offering consumers fast access to healthcare professionals. Services include care needs assessments, diagnoses, follow-up recommendations, and prescriptions. Thus, the digital clinic supplements traditional primary care services, including in-person visits and phone consultations. Consequently, a key question of scientific and policy relevance is whether and how digital clinics affect the utilization of primary care services.

1.2 Research questions

Our main research questions are the following:

A. What is the impact of having *access* to the digital clinic on the use of digital clinic services (*i.e.*, take-up)?

B. What is the impact of having *access* to the digital clinic on the utilization of traditional primary care services (intent-to-treat effect, ITT)?

C. **What is the impact of *using* the digital clinic on the utilization of traditional primary care services (average causal response, ACR)?**

D. What is the impact of having *access* to the public digital clinic on the overall utilization of public primary care services, including both the digital clinic and traditional primary care services (ITT)?

In our research, we prioritize the importance of research question C. Beyond its policy relevance, the ACR estimate accounts for expected non-compliance in using the digital clinic.[1] However, we list our research questions here in a sequential order from A to D, as we are unlikely to detect the

[1] See Angrist and Hull (2023) for an illustration for the importance of accounting for non-compliance in pragmatic randomized trials.

impact of using the digital clinic on any downstream outcomes if the take-up of the digital clinic during the trial period is not large enough.

Beyond studying the use of primary health care services (this PAP), we are naturally interested to extend our analyses to health-related outcomes, even though the power to detect differences in some of these outcomes is expected be lower than for our pre-registred outcomes. The objective of these additional analyses is to assess whether improved access to digital primary care can, for example, reduce reliance on more expensive specialized healthcare, such as emergency department visits, hospitalizations, referrals to specialized care, or specialist consultations (see Section 7). These broader effects will be informative about the digital clinic's role in optimizing the delivery of healthcare services.

2 Institutional background

2.1 Primary care in Finland

Finland has a decentralized, universal healthcare system that primarily relies on the public provision of health services. By law, the wellbeing service counties (21 in total) are responsible for organizing public health and social care services, including public primary care (PPC), for their residents. PPC services are characterized by gatekeeping, varying and sometimes long waiting times, and moderate copayments.[2] Primary care services provided by employer-sponsored occupational healthcare and the private sector complement the services provided by PPC, which is disproportionally important for low-income individuals, unemployed, and pensioners. Fast access and no copayments make occupational healthcare an attractive alternative to PPC for those who have access to it. There are also private clinics with fast access and no gatekeeping (even for specialists), albeit with fees much higher than in PPC.

[2]There are no co-payments for nurse visits. The maximum co-payment for general practitioner visits is 23€ for the first three visits annually.

2.2 Traditional public primary care

Traditionally, patients first contact a nurse by phone (hold queue or a call-back service) or by visiting a traditional PPC clinic. The nurse then conducts a care needs assessment, provides potential self-care guidance, and, acting as a gatekeeper, books a phone consultation or an in-person appointment for a physician or other professionals if needed. Team-based models, based on collaboration and consultative interactions between physicians and nurses, are common in PPC, with the goal of optimizing healthcare delivery and addressing issues even during the first contact, reducing the need for follow-ups.

2.3 Public digital clinic

We study the launch of a public digital clinic by a wellbeing services county. Following the widespread adoption and use of digital clinics in occupational healthcare and the private sector, several wellbeing service counties (PPC providers) are launching their own digital clinics as a remote, chat-based access channel to complement their traditional clinic-based service provision.[3] Digital clinics aim to provide patients fast access to healthcare professionals (here: nurses and physicians) through chat, available through a mobile application or website, with extended opening hours and waiting times measured in minutes. Compared with traditional clinics, digital clinics can reduce barriers to healthcare access through extended opening hours, shorter waiting times, and reduced travel time to a health clinic for an in-person visit. While digital clinic services are not suitable for all patients and conditions, these services offer many patients fast, easy, and user-friendly access to primary care. The most common medical issues treated in digital clinics are cold symptoms, stomach problems (e.g., diarrhea, vomiting), gynecological problems, skin problems, allergies, eye infections, and mental health problems.

Figure 1 illustrates the use of a public digital clinic from the patient's perspective and the potential care paths after the digital clinic contact. Logging into the digital clinic via a mobile

[3]Patients can choose to contact a public or private clinic (traditional or digital). In the private sector, patients can access digital clinic services through occupational healthcare (employed working-age population), voluntary private health insurance, or by paying the full out-of-pocket cost.

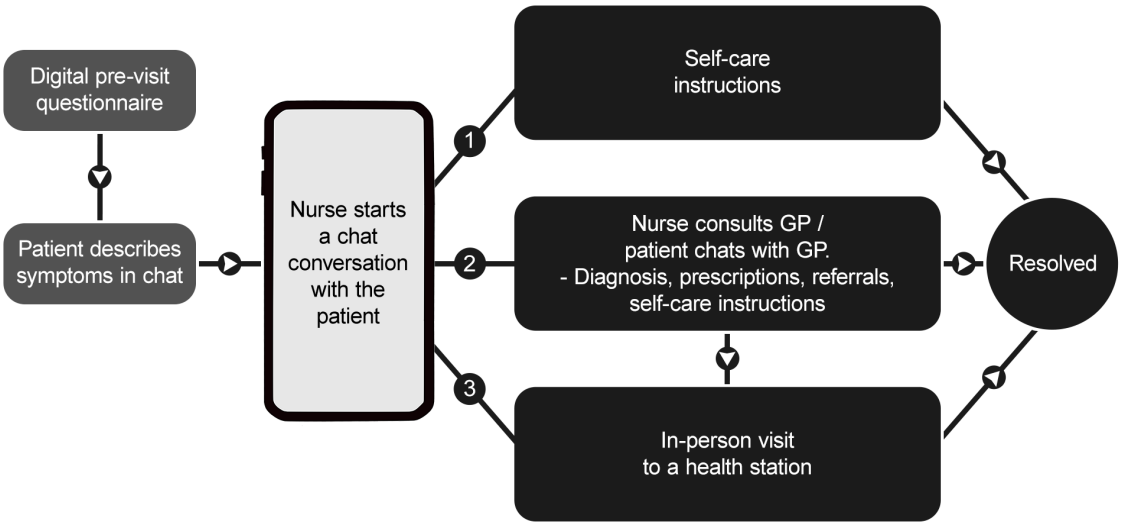

Figure 1: Potential care paths for a patient who experiences a care need and chooses to contact the public digital clinic instead of traditional public primary care by phone or by visiting a traditional PPC clinic in person.

application or website requires verifying a person's identity through online banking credentials or a national authentication service for public services. Thereafter, the patient fills out a digital pre-visit questionnaire before a healthcare professional, often a nurse, performs a care needs assessment, asking follow-up questions via chat. In broad terms, the patient may i) receive instructions from the nurse for self-care and health monitoring, ii) be directed to a physician at the digital clinic (nurse-physician consultation or patient-physician chat), for diagnoses, prescriptions, or referrals (for lab tests or specialist visits), or iii) be directed to an in-person visit to an appropriate professional at the traditional PPC clinic.

From the provider's perspective, digital clinics allow healthcare professionals to manage several patients at the same time via chat, unlike in traditional clinics, where they manage one patient at a time in person or by phone. Consequently, compared with traditional clinics, digital clinics can save professionals' time as no time is wasted waiting for the next patient. The model also allows professionals to specialize in telemedicine and in the chat-based user interface. The task of asking routine questions is automated in the digital clinic via a pre-visit questionnaire.

Having a large customer base with enough contacts outside typical office hours makes extended opening hours possible.

3 Intervention

The Wellbeing Services County of Ostrobothnia, a mid-sized administrative region in Western Finland (Figure 2), will launch its *digital clinic platform*, a website and app for its digital services, on April 15, 2025. Over time, the digital clinic platform will include several different chat channels for various services. The main channel, the *digital clinic*, will be a chat channel to contact primary care professionals. The initial contact will be with a nurse, after which the nurse has the opportunity to consult with a physician. Primary care patients with new health issues who choose to contact the digital clinic are expected to log in with strong identification for a care needs assessment and treatment. At the time of writing, the plan is to keep the digital clinic open from 8 AM to 3 PM Monday through Thursday and from 8 AM to 2 PM on Fridays. Other chat channels that Ostrobothnia plans to launch using the platform, such as a chat for social services, a chat for rehabilitation, and a chat for customer service, do not require strong identification, and are not intended to serve as a substitute for the digital clinic.

3.1 Access to the digital clinic

Our intervention is to randomize access to the digital clinic for a nine-month period, starting on April 15. The randomization assigns households to two groups that either have access (*the treatment group*) or do not have access (*the control group*) to the newly launched digital clinic. Individuals in the treatment group will have access to the digital clinic immediately after its opening. The entire population residing in the region will have access to the digital clinic after the nine-month trial period. The trial does not affect access to other available alternatives for contacting primary care, such as traditional PPC, occupational healthcare, or private clinics.

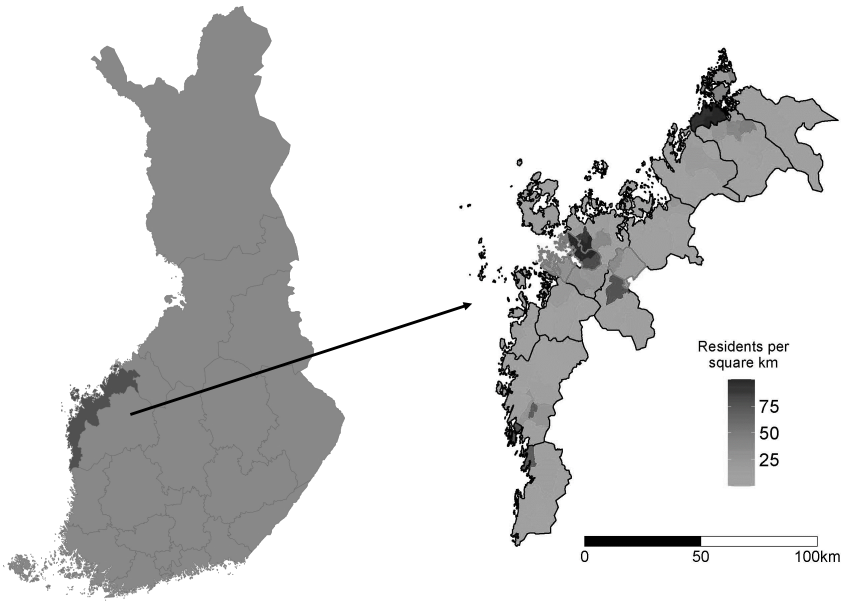

Figure 2: The Wellbeing Services County of Ostrobothnia is located in Western Finland.

Notes: By law, the Wellbeing Services County of Ostrobothnia is responsible for organizing public health, social, and rescue services, including public primary care (PPC), for its residents. Ostrobothnia has the highest share of Swedish speakers in Finland, with a Swedish-speaking majority in many municipalities. Unlike most of Finland, which is Finnish-dominated, Ostrobothnia has a strong bilingual culture. Unlike the heavily urbanized capital region, Ostrobothnia has a balanced mix of urban centers and strong rural communities. The region also has strong religious traditions, which have contributed to socially conservative values in some areas.

Individuals in the treatment group can access the digital clinic by logging into the clinic via a mobile application or website, using strong identification and personal identity number. We expect that some individuals belonging to the control group may try to log in to the digital clinic. In this case, their access will be automatically denied based on their identity number. Moreover, the digital clinic platform will display a short message notifying them that their access to the platform is currently blocked because the digital clinic is being tested with a subset of the population, but that they will ultimately get access to the digital clinic after the test period.

3.2 Communication with the treatment group

The intervention (access to the digital clinic) is accompanied by an information campaign targeting all households in the treatment group. The primary communication channel with the treatment group is through mailed letters. These letters will inform recipients about their option to use the digital clinic during the trial period and provide instructions on what the digital clinic is, and how to use it, as well as the rationale for granting access initially only to a randomly-selected subgroup of the population.

We will send these information letters to all households belonging to the treatment group. We will send one letter per household and randomize the recipient within the household so that all household members over the age of eighteen have the same probability of receiving the letter. In a small sample of households consisting only of minors, all individuals aged 15 to 18 have an equal probability of receiving the letter. There will be no information letters addressed to individuals under the age of fifteen. The letters will be mailed shortly before the launch of the digital clinic.

3.3 Population-wide information during the study period

The scale of the digital clinic launch and the significant changes in available healthcare services are expected to generate public discussion and interest in the reform. In response, the Wellbeing Services County of Ostrobothnia will issue a press release about the digital clinic launch, its staggered implementation, and the associated informational letters to the treatment group at the end of March 2025, a couple of weeks before the launch of the digital clinic and the distribution of informational letters. We expect that local newspapers, and possibly some national newspapers, will write about the digital clinic and the trial. We expect that there will be content (*e.g.*, news, reminders, and a mention of the digital clinic as a contact method for those in the treatment group) on the webpage of the Wellbeing Services County of Ostrobothnia about the digital clinic and the trial throughout the study period. We will document trial-related media coverage and population-wide information released by the Wellbeing Services County of Ostrobothnia during the study in the SAP, in the populated SAP, and in the final research paper(s).

4 Study design

4.1 Target and study population

We extracted individuals whose municipality of residence was within the Ostrobothnia region on March 14, 2025 (target population). Our inclusion criteria required individuals to be alive at the time of extraction and to have a registered permanent address.[4] We additionally excluded individuals residing in the city of Kristinestad, as PPC services in this municipality are outsourced to a private health care provider. Finally, we aimed to exclude individuals residing in institutional care homes. We defined institutional care homes as residences where more than two individuals aged over 80 years lived or where more than four individuals over 60 years lived. We identified no other scientific or ethical reasons to exclude any other individuals who met the inclusion criteria from randomization. See Figure 3 for our target population and sample sizes.

However, in our analyses, we will restrict the sample to individuals aged 0 to 70 to have more statistical power (see Figure 4 in Appendix B1). Moreover, we exclude from the analysis, but not from the randomization, those individuals who are observed in the Finnish Population Information System (study population) but not in the Statistics Finland datasets (background covariates used in analysis) – see Section 5. The number of such individuals is expected to be small. We can report the number only after linking the treatment indicators with the relevant administrative datasets.

4.2 Randomization

We randomized treatment at the household level based on permanent addresses, ensuring that all members of a household were assigned to the same treatment group. Households were stratified by size to maintain balance across different household compositions. Within each stratum, we

[4]The permanent address was missing for 1% of the population. Age, gender, or language did not appear to be correlated with the address being missing. The permanent address is not recorded for individuals with a protection order (approximately 0.2% of the population nationally). The protection order is a legal measure in Finland that restricts the disclosure of an individual's address and other personal information in official registries to protect their safety and privacy.

randomly assigned 50% of the households to the treatment group (a 1:1 ratio). Specifically, for each household ID cluster, we generated a random floating-point number and sorted the clusters by this value within each household size group. Households in the top 50% of these sorted values were assigned to the treatment group.[5]

Moreover, we randomized one recipient of the information letter (see Section 3.2) per treated household as follows: All household members over the age of eighteen had the same probability of receiving the letter. In a small sample of households consisting only of minors, all individuals aged 15 to 18 had an equal probability of receiving the letter. The randomization code, like all other code written for this PAP, is available in the Github repository of this project.[6]

5 Data and outcomes

5.1 Data sources

This study uses multiple Finnish administrative data sources containing individual-level data. The datasets are merged via pseudonymized person identifiers (IDs). Figure 3 summarizes different steps in our research design, including target population construction, randomization, and the construction of the analysis data and study population. We use the following data sources:

- Finnish Population Information System maintained by the Digital and Population Data Services Agency. This dataset was used to extract the target population on March 14, 2025.
- Register of Primary Care Visits maintained by the Finnish Institute for Health and Welfare. This dataset contains contacts with public primary care, private outpatient care, and occupational healthcare. We use data from 4/2024–1/2026.

[5]The actual proportion of treated households and individuals may not be exactly 50%. If the remainder when dividing the stratum size by 2 was not zero, we randomly varied between using the floor and ceiling function within each stratum to select the number of treated units. This approach ensures that approximately half of the units are in the treatment group.

[6]Source: https://github.com/tapiohaa/SoteDataLab/tree/main/ostrobothnia_digital_clinic_experiment

- Full population data on the socioeconomic and sociodemographic characteristics of Finnish residents, maintained by Statistics Finland (FOLK population data and INFRA location data) from 2024. *Note: At the time of writing, the latest data is available for year 2022. Data for the year 2023 is expected to be released in spring 2025 and the data for the year 2024 in spring 2026. Due to this release lag, we plan to use the data for year 2023 for the SAP and the populated SAP, and the data for the year 2024 for the final research paper(s). The choice of the FOLK statistical data year affects the sample sizes of the analysis data. For example, restricting the analysis to the year 2023 data excludes from the analysis sample individuals born between January 1, 2024, and March 14, 2025.*

- Hospitalizations and contacts with specialized health care from the Care Register for Health Care, maintained by the Finnish Institute for Health and Welfare. We use these data from 4/2024 to 1/2026. *Note: this dataset is not required for this PAP, but we may use it for the additional analyses (see Section 7).*

- Drug prescriptions and fills from the Prescription Centre, maintained by the Finnish Social Insurance Institution. We use data from 4/2024 to 1/2026. *Note: this dataset is not required for this PAP, but we may use it for the additional analyses (see Section 7).*

- Entitlements to medicine reimbursements at a special rate from the Finnish Social Insurance Institution from 2024. *Note: At the time of writing, the latest data is available for year 2023.*

- Moreover, our data set includes self-collected data on the location (address) of traditional PPC clinics in September 2023 from the websites of PPC providers.

5.2 Data cleaning and preparation

As stated in Section 1, we will write a statistical analysis plan (SAP), which provides detailed code on the construction of our variables, analysis data, and the study population. Our objective is to register the SAP *after* the digital clinic launch and *after* observing aggregate statistics on

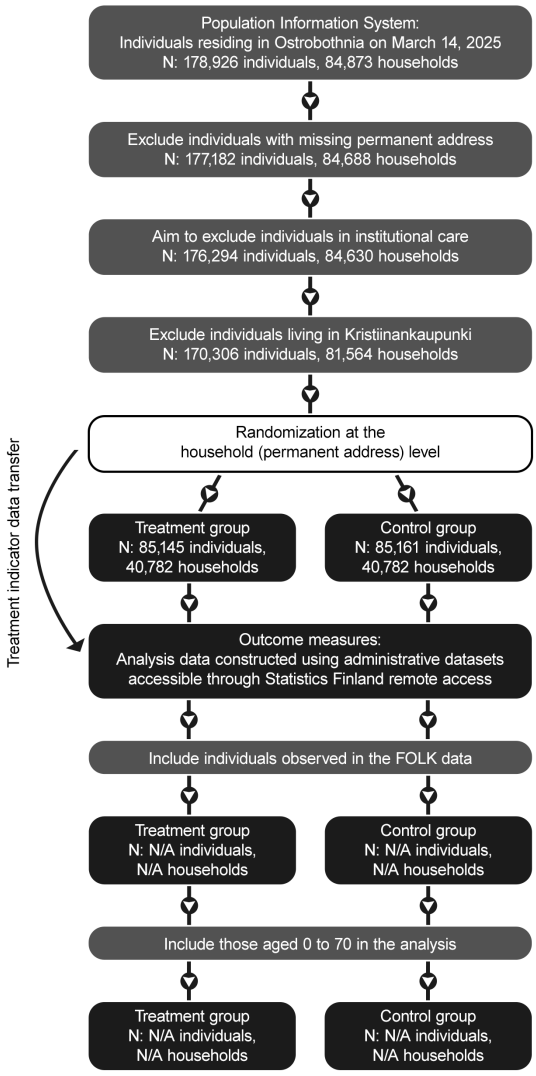

Figure 3: Target Population, Randomization, and Data Construction

Notes: At the time of writing, we cannot report the exact sample sizes of the analysis data, only the sample sizes of the randomization data. This is mainly because the exact sample sizes of the final research paper will be observed only after merging Statistics Finland full population data (FOLK) of 2024 with the treatment indicator data (or the FOLK data for 2023 for the statistical analysis plan, SAP, see Section 5). At the time of writing, the latest statistical year fully available is 2022. The year 2023 is expected to become fully available in spring 2025 and the year 2024 in spring 2026. For example, restricting the analysis to the FOLK statistical year 2023 would exclude from the analysis sample individuals born between January 1, 2024, and March 14, 2025, as they are not observed in the FOLK data for 2023. The second reason is that our plan is to bring the treatment indicator data into the remote access computer (*i.e.*, unblind the data) only after registering our SAP. This is done to create a credible and verifiable firewall between planning the SAP and observing any results.

the utilization of the digital clinic, but *before* linking treatment indicators and outcome data. This process will be verifiable by a third party.[7] The SAP code defines how variables are constructed, missing data are handled, and transformations or aggregations are conducted. However, note already that:

- If the permanent residence is not observed in the Population Information System, the individual is excluded from the target population.

- If an individual from the target population, extracted from the Population Information System, is not observed in Statistics Finland datasets, which is defined as not having a municipality of residence (missing data) at the end of 2024, we will not use that individual in the study population in the analyses because we do not have covariates for the individual. *Note: The Statistics Finland datasets are expected to become fully available for the year 2024 in spring 2026. Due to this release lag, we plan to use the statistical year 2023 for the SAP and the populated SAP, and the statistical year 2024 for the final research paper(s). The choice of the FOLK statistical year affects the sample sizes of the analysis data. For example, restricting the analysis to the statistical year 2023 excludes from the analysis sample individuals born between January 1, 2024, and March 14, 2025.*

- Some values in the Statistics Finland datasets may be missing. We will construct the cleaned covariates used in the analyses in a way that missing values in raw covariates do not lead to the corresponding individual being left out of estimation. These details will be available in the SAP code.

- Some people move out of the Ostrobothnia area or die during the trial. These people remain in the analysis sample. We expect that the events of moving out or dying occur similarly in the treatment and control groups.

[7] This verification can be based on the registration date of our SAP and a confirmation from Statistics Finland regarding the date they transferred the treatment indicator data to the secure cloud server where our research data is maintained.

- Although we preserve the right to set the final data construction choices in the SAP, we note that our replication codes for the PAP (simulations and power analyses) contain early examples of how we currently plan to extract the key outcomes and construct the set of covariates.

- Some analyses in the final research paper will be post-blind, *i.e.*, implemented after unblinding the data by linking treatment indicators and outcomes. In these cases, the data construction choices will also be post-blind.

- There may be outliers, *e.g.*, individuals with suspiciously high health care utilization, leading to suspicion about duplicate values in the data (one underlying contact could be linked to several rows in the data). By defining health care contacts as the number of days with any contact (see Section 5.3) should partially alleviate the risk that outliers pose to the estimates. Our plan is to include potential outliers in the main analysis, but we may conduct a robustness check to examine whether the estimates change as a result when excluding individuals with suspiciously high health care utilization from the analysis.

5.3 Outcomes

The analyses proposed in this PAP are restricted to curative outpatient contacts in the public primary care.[8] The definitions of outcome variables are based on three additional variables: health care provision unit's identifier, contact type (telemedicine or traditional), and the profession (nurse or physician) of the provider.

As stated in Section 1, our main interest is in estimating the impact of *using* the public digital clinic on the use of traditional PPC services. Accordingly, we need to estimate impacts not only on the utilization of traditional PPC (ultimate outcome of interest) but also on the utilization of the digital clinic (take-up). We measure the annualized number of digital clinic contacts in PPC (D. digital clinic utilization; take-up) and the annualized number of contacts to traditional PPC

[8]Restricting to curative contacts should exclude preventive contacts, such as seasonal influenza vaccinations. Restricting service type to outpatient care should exclude visits to other service types, such as school and student healthcare, dental care, or occupational healthcare.

clinics (Y. traditional PPC utilization; reduced form).[9] We count all contacts received during the same day as one contact or visit.

Our primary outcomes include the following types of contacts:

- **in-person visits in PPC (Y1.1)**. Our main research question is whether the use of digital clinics can reduce contacts with traditional PPC, including in-person visits and phone contacts. Of these contacts, we chose in-person visits as our primary outcome. In-person visits are more expensive to provide than phone contacts and require face-to-face interaction. At the same time, we expect that other contacts with traditional PPC, involving telemedicine (mainly phone calls) and professional-to-professional interactions, are a closer substitute for digital clinic contacts than in-person visits. *This outcome includes in-person visits to nurses and physicians in traditional PPC clinics.*

- **the number of public digital clinic contacts (D.1)**. This outcome is required for estimating the impact of *using* the public digital clinic on the utilization of traditional PPC. *This outcome includes care needs assessments, remote appointments to nurses and physicians (via chat and video), and professional-to-professional interactions between nurses and physicians in digital PPC clinics.*

Our secondary outcome can help provide a more nuanced picture of whether the use of the digital clinic can reduce pressure on traditional PPC as compared to focusing strictly on in-person visits, which represent only a minority of contacts in traditional PPC. It includes all traditional PPC contacts except in-person visits:

- **other contacts with traditional PPC (Y1.2)**. We expect that the substitution rate between digital clinics and traditional PPC is higher with respect to this outcome, compared to in-person visits (Y1.1). Other traditional PPC contacts involve remote contacts (*e.g.*, phone

[9] We compute the annualized number of contacts during our 9-month follow-up period by dividing the total number of contacts by 9 and multiplying it by 12.

calls between professionals and patients and professional-to-professional interactions) and are likely to be a closer substitute for digital clinic contacts than in-person visits. *This outcome includes care needs assessments, remote appointments to nurses and physicians, and professional-to-professional interactions between nurses and physicians in traditional PPC clinics.*[10]

Finally, we have two tertiary outcomes:

- **the total number of PPC contacts (Y1.3)**. We expect that the digital clinic access will increase the total number of contacts to PPC, including the digital clinic and the traditional PPC. The question is: by how much? *This outcome includes in-person visits to nurses and physicians, care needs assessments, remote appointments to nurses and physicians, and professional-to-professional interactions between nurses and physicians in digital and traditional PPC clinics.*

- **an indicator for having any public digital clinic contact during the follow-up (D.2)**. The purpose of adding this outcomes is to allow interested readers to construct the Local Average Treatment Effect (LATE) parameter by dividing reduced-form estimates on Y1.1 and Y1.2 by D.2. In other words, this outcome is an alternative approach for estimating the impact of *using* the public digital clinic on the utilization of traditional PPC. *This outcome includes care needs assessments, remote appointments to nurses and physicians (via chat and video), and professional-to-professional interactions between nurses and physicians in digital PPC clinics.*

[10] While our institutional knowledge suggests that nurses' care needs assessments and professional-to-professional interactions in traditional PPC are often done remotely on rather than in-person with the patient being present, our data does not distinguish between these contact types for these outcomes.

6 Statistical analysis

6.1 Estimation and inference

To answer our research questions specified in Section 1.2, we estimate two statistical models. Our first model uses Ordinary Least Squares (OLS) to estimate the impacts of *access* to the public digital clinic. We estimates the impacts of digital clinic *access* on outcomes related to i) the utilization of the public digital clinic (D.1 and D.2; *take-up*), ii) the utilization of traditional PPC services (Y1.1 and Y1.2), and iii) the total utilization of PPC (Y1.3), including traditional and digital services. We estimate the following model for individual i:

$$Y_i = \beta_0 + \beta_1 1(Treatment)_i + \beta_2 X_i + \varepsilon_i, \tag{1}$$

where Y_i is the outcome of interest and $1(Treatment)_i$ is an indicator variable equal to one if the individual belongs to a household randomly assigned to the treatment group and zero if the individual belongs to a household randomized to the control group. X_i is a vector of control variables, defined at the end of this subsection. β_0 is an intercept, and ε_i is the error term. Our parameter of interest is β_1, which measures the causal effect of *access* to the public digital clinic on the corresponding outcome (D.1–D.2, Y1.1–Y1.3).

Our primary objective is to estimate the impact of *using* the public digital clinic on downstream outcomes. For this purpose, we account for non-compliance in the RCT, unlike in the estimation of ITT effects using Model 1. This non-compliance occurs because not all individuals in the treatment group are expected to use the digital clinic.[11] Due to this non-compliance, the impact of *using* the digital clinic can be expected to be different from the impact of *having access to* the clinic (ITT).

[11] Our experiment has one-sided non-compliance. All individuals in the treatment group cannot be expected to use digital health care services during the follow-up period, but none of the individuals in the control group are by construction able to use the digital clinic during the follow-up period.

Our second model estimates the impacts of *using* the public digital clinic (D.1) on the utilization of traditional PPC services (Y1.1 and Y1.2), using Two-Stage Least Squares (2SLS) and the random assignment into the treatment group as an instrumental variable for the utilization of the public digital clinic.[12] Specifically, our parameter of interest is *the average causal response (ACR)*, which is defined as a generalization of the Local Average Treatment Effect (LATE) to settings where the treatment variable is multi-valued rather than binary (Angrist & Imbens, 1995). It represents the expected causal effect of a unit increase in the treatment variable (the number of digital clinic contacts) for individuals whose treatment status (digital clinic use) is influenced by the instrument (randomization).[13]

We estimate for individual i the following model using 2SLS:

$$\begin{aligned} D_i &= \alpha_0 + \alpha_1 1(Treatment)_i + \alpha_2 X_i + \varepsilon_i, \\ Y_i &= \pi_0 + \pi_1 \widehat{D}_i + \pi_2 X_i + \zeta_i, \end{aligned} \tag{2}$$

where $1(Treatment)_i$ is an indicator variable equal to one if the individual belongs to a household randomly assigned to the treatment group and zero if the individual belongs to a household randomized to the control group; D_i is the number of public digital clinic contacts (D.1) and $\widehat{D}$ is its predicted value based on the first equation, and Y_i is the outcome of interest (Y1.1–Y1.2). Moreover, X_i is a vector of control variables, defined at the end of this subsection. α_0 and π_0 are intercepts, and ε_i and ζ_i are the error terms. The parameter of interest, π_1, is the estimated average causal response (ACR).We use outcome D.1 as the take-up outcome in the 2SLS regressions.

Covariates: We include the following fixed effects as covariates (X_i) in all regressions (OLS and 2SLS): the previous number of in-person visits in PPC (Y1.1), an indicator variable for having

[12]See Angrist and Imbens (1995), Angrist et al. (1996), and Imbens and Angrist (1994) for the econometric and statistical background of using 2SLS estimation in randomized controlled trials.

[13]Formally, the ACR is a weighted average, over all values of d (potential intensity of the treatment), of the effect of increasing treatment from $d-1$ to d among switchers whose treatment status goes from strictly below to above d over time (Angrist & Imbens, 1995). In our application, we interpret it to measure the degree of substitution between the public digital clinic and traditional PPC for compliers who consult the digital clinic more only because they were offered access to it.

at least one previous contact with occupational healthcare during 12 months preceding the trial, age (in years), gender, municipality, income percentile, language, distance quartile to the nearest traditional PPC clinic, and indicators for having a common chronic disease and multimorbidity.[14] These covariates are expected to be uncorrelated with the treatment indicator, but can substantially improve the precision of our estimates as illustrated in Appendix A.

Standard errors: Standard errors are clustered at the permanent address level, which corresponds to the level of randomization.

Multiple hypothesis testing in Table 3: In this PAP, we do not pre-specify a plan to adjust for multiple comparisons, but we specify a hierarchy of outcomes (primary, secondary, tertiary) in Section 5.3. Consequently, we report p-values only for the primary outcome (the ACR and ITT effects on the number of in-person visits in PPC, Y1.1) and report 95% confidence intervals without p-values for all secondary and tertiary outcomes. The confidence intervals for secondary and tertiary outcomes will not be adjusted for multiple comparisons, suggesting that inferences drawn from these outcomes may therefore not be reproducible.

6.2 Planned tables

We describe here our key pre-registered analyses to be included in our research paper intended for publication in a scholarly journal.

Table 1: Characteristics of the Residents at Baseline. This table presents the means, standard deviations (SD), percentage differences in means (difference %), and standardized mean differences (SMD) of several baseline covariates for the treated and control individuals, as well as means and SDs for the populations of Ostrobothnia and Finland.

[14]The common diseases covered here include special reimbursement rights for cardiovascular diseases, diabetes, respiratory diseases, rheumatic diseases, cancer, neurological diseases, and severe mental health disorders. Multimorbid individuals are defined as those with special reimbursement rights for at least two out of three: cardiovascular diseases, respiratory diseases, and diabetes.

Table 2: Characteristics of Public Digital Clinic Users vs. Traditional PPC Clinic Users in the Treatment Group in Ostrobothnia. This table presents the means, standard deviations (SD), percentage differences in means (difference %), and standardized mean differences (SMD) of several baseline covariates for the users of public digital clinics and for the users of traditional PPC services among treated individuals, as well as means and SDs for the total population of Ostrobothnia. The hypothesis is that the digital clinic users differ noticeably from the general population and from the population that use traditional PPC services.

Table 3: Effects of Access and Utilization of Public Digital Clinic on the Utilization of Traditional Public Primary Care. This table illustrates how we plan to report the results of our pre-registered confirmatory analyses. In Panel A, we report the impact of having *access* to the public digital clinic on the use of the public digital clinic (take-up). In Panel B, we report the impact of having *access* to the public digital clinic on the use of traditional PPC services and the total use of PPC (intent-to-treat effect, ITT). In Panel C, we report the impact of *using* the public digital clinic on the use of traditional public primary care services (average causal response, ACR). Here, we use outcome D.1 (the total utilization of the digital clinic in PPC) as the take-up outcome in the 2SLS regressions (see Section 5.3).

Subgroup analyses: We do not pre-register any subgroup analyses (*e.g.*, visits provided by nurses versus physicians, different patient subgroups) in this PAP, but we may register subgroup analyses later in the SAP. Alternatively, we may later conduct subgroup analyses to complement the SAP, but we acknowledge that such post-blind subgroup analyses are more hypothesis-generating than hypothesis-confirming in nature.

Table 1: Characteristics and Means Comparisons of Residents at Baseline.

	Treated N:	Control N:	Treated - Control	Ostrobothnia N:	Finland N:
	Mean [SD]	Mean [SD]	Diff. (%) [SMD]	Mean	Mean
A. Prior health care use					
PPC: in-person visits (days)					
PPC: other contacts (days)					
Private HC: in-person visits (days)					
Private HC: other contacts (days)					
Occup. HC: in-person visits (days)					
Occup. HC: other contacts (days)					
B. Sociodemographic covariates					
Age (in years)					
Is female (share)					
Language: Finnish (share)					
Language: Swedish (share)					
Relationship or widowed (share)					
Living in a city (share)					
Dist. to nearest trad. PPC clinic (km)					
Tertiary education (share)					
Pensioner (share)					
Employed (share)					
Income (thousands of euros)					
C. Morbidities					
Common chronic disease (share)					
Has multimorbidity (share)					
Observations					

Notes: The table presents the means, standard deviations (SD), percentage differences in means (difference %), and standardized mean differences (SMD) of several covariates for the treated and control individuals, as well as means and SDs for the total population of Ostrobothnia and Finland. The analysis sample is restricted to those aged 0–70. In Panel A, health care use is measured in 12 months preceding the trial and represent annualized health care utilization. Health care contacts are here defined in terms of contact dates: individuals get value 1 if they have any relevant contact on the given day. For private clinics and occupational health care, we include curative contacts conducted by nurses or doctors. Covariates in Panel B are measured at the end of 2024. Living in a city is defined as classes K1–K3 of the city–countryside classification. Income is defined as the equivalent family disposable income. Distance to the neareast traditional PPC clinic is a straight-line distance. The list of traditional PPC clinics was collected in late 2023. In Panel C, morbidity is defined based on special reimbursement rights in 2024. The common diseases covered here include special reimbursement rights for cardiovascular diseases, diabetes, respiratory diseases, rheumatic diseases, cancer, neurological diseases, and severe mental health disorders. Multimorbid individuals are defined as those with special reimbursement rights for at least two out of three: cardiovascular diseases, respiratory diseases, and diabetes.

Table 2: Characteristics of Public Digital Clinic Users vs. Traditional PPC Clinic Users in the Treatment Group in Ostrobothnia.

Clients of:	digital clinics N:		trad. PPC clinics N:		Digi - trad. clinics		Ostrobothnia N:	
	Mean	[SD]	Mean	[SD]	Diff. (%)	SMD	Mean	[SD]
A. Prior health care use								
PPC: in-person visits (days)								
PPC: other trad. contacts (days)								
Private HC: in-person visits (days)								
Private HC: other contacts (days)								
Occup. HC: in-person visits (days)								
Occup. HC: other contacts (days)								
B. Sociodemographic covariates								
Age (in years)								
Is female (share)								
Language: Finnish (share)								
Language: Swedish (share)								
Relationship or widowed (share)								
Living in a city (share)								
Dist. to nearest trad. PPC clinic (km)								
Tertiary education (share)								
Pensioner (share)								
In labor market (share)								
Income (thousands of euros)								
C. Morbidities								
Common chronic disease (share)								
Has multimorbidity (share)								
Observations								

Notes: The table presents the means, standard deviations (SD), percentage differences in means (difference %), and standardized mean differences (SMD) of several covariates for the users of public digital clinics versus the users of traditional PPC clinics among the treated individuals. The two groups are constructed as follows: 1) those in the treatment group with at least one digital clinic contact in PPC are defined as users of the digital clinic, and 2) those in the treatment group with at least one contact in traditional PPC but zero digital clinic contacts are defined as the users of traditional PPC clinics. The table also contains means and SDs for the total population in Ostrobothnia. The analysis sample is restricted to those aged 0–70. In Panel A, health care use is measured in 12 months preceding the trial, representing annualized utilization. Health care contacts are defined in terms of contact dates: individuals get value 1 if they have any eligible contact on the given day. For private clinics and occupational health care, we include curative contacts conducted by nurses or doctors. Covariates in Panel B are measured at the end of 2024. Living in a city is defined as classes K1–K3 of the city–countryside classification. Income is defined as the equivalized family disposable income. Distance to the neareast traditional PPC clinic is a straigth-line distance. The list of traditional PPC clinics was collected in late 2023. In Panel C, morbidity is defined based on special reimbursement rights in 2024. The common diseases covered here include special reimbursement rights for cardiovascular diseases, diabetes, respiratory diseases, rheumatic diseases, cancer, neurological diseases, and severe mental health disorders. Multimorbid individuals are defined as those with special reimbursement rights for at least two out of three: cardiovascular diseases, respiratory diseases, and diabetes.

Table 3: Effects of Access and Use of Public Digital Clinic on the Use of Traditional Public Primary Care.

	Digital clinic contacts (D.1)	Any digital clinic contact (D.2)	
A. Impact of *access* on the use of the digital clinics			
Effect			
Control group mean			
SE			
CI	[,]	[,]	
	In-person visits (Y1.1)	Other traditional contacts (Y1.2)	Primary care contacts in total (Y1.3)
B. Impact of *access* on the use of traditional primary care and primary care in total			
Effect			
Control group mean			
SE			
CI	[,]	[,]	[,]
Relative effect (%)			
Relative CI (%)	[,]	[,]	[,]
	In-person visits (Y1.1)	Other traditional contacts (Y1.2)	
C. Impact of *using* digital clinic (D.1) on the use of traditional primary care			
Effect			
SE			
CI	[,]	[,]	
N			

Notes: The table contains our baseline intention-to-treat (ITT) results for the impact of access to the digital clinic and our baseline average causal response (ACR) results for the impact of the use of the digital clinic, using a 9-month follow-up. The analysis sample is restricted to those aged 0–70. We estimate the impacts on both the annualized number of digital clinic contacts (DCT utilization; take-up) in Panel A and the annualized number of contacts to traditional PPC (traditional PPC utilization; reduced form) or the annualized number of total contacts to PPC (ITT) in Panel B. Estimators: OLS with Model 1 in Panel A (take-up) and in Panel B (reduced-form), and 2SLS with Model 2 in Panel C (ACR). Covariates: fixed effects as listed in Section 6. Standard errors are clustered at the permanent address level (the level of randomization). Relative effects are calculated by dividing effect estimates by control group means and multiplying by 100. Outcome D.1 (the total utilization of the digital clinic in PPC) is used as the take-up outcome in the 2SLS estimation. The p-values for the primary outcome Y1.1: ACR N/A, ITT N/A.

7 Complementary analyses and next steps

As stated in Section 1, our next step is to write a statistical analysis plan (SAP) that details, in code, how our variables and analysis data are constructed. Our objective is to register the SAP *after* the digital clinic launch and observing aggregate statistics on the utilization of the digital clinic, but *before* linking treatment indicators and outcomes. This process will be verifiable by a third party.

For the SAP, we may add more analyses compared to the PAP. For example, we are considering extending the analyses with the aim of answering some of the following questions, noting that the more outcomes there are, the more crucial it is to appropriately account for multiple hypothesis testing. Note also that adding these analyses does not change our key research questions as stated in Section 1.

- Does the utilization of the public digital clinic reduce urgent and emergency department visits, hospitalizations, referrals to specialized care, or specialist visits (potential savings)?
- Does the utilization of the public digital clinic increase the number of prescription initiations of certain drugs or referrals to medical examinations (a potential benefit/cost of improved access)?
- Does the utilization of the public digital clinic reduce telemedicine contacts with occupational health care or private clinics (spillovers to other sectors)?
- Does the utilization of the public digital clinic affect continuity of care?

Adjustments for multiple hypothesis testing: When later analyzing additional outcomes beyond our main results in Table 3, such as the ones proposed above, we commit to either 1) adjusting inference appropriately to account for multiple hypothesis testing, or 2) not adjusting inference but explicitly stating that the results are more hypothesis-generating (or exploratory) than hypothesis-confirming relative to the main results in Table 3.

What are our initial thoughts on addressing multiple hypotheses testing? Suppose we want to examine the impacts of the use of the public digital clinic on the use costly services

provided by hospitals, such as emergency department visits, specialist visits, and hospitalization. Within this family of outcomes, we would find it natural to control for Family-Wise Error Rate (FWER). Our understanding is that the Holm-Bonferroni correction would often be chosen for this kind of setting and is easily implementable. We also consider computing sharpened q-values, which adjust inference for the probability of false discoveries (Type I errors) when testing multiple hypotheses.

Another potential family of outcomes would focus on the impacts of digital clinic use on the use of occupational healthcare and private clinics. In this domain (spillovers to other health care sectors), we would have four potential outcomes: telemedicine / in-person visits to occupational healthcare / private clinics. As the two families of outcomes mentioned above belong to different domains and test fundamentally different questions, our understanding is that the appropriate approach would be to apply the Holm-Bonferroni correction separately within each family.

Precision of the estimates: Based on the simulations documented in Appendix A, we anticipate that incorporating pre-exposure data with fixed effects estimates will effectively reduce variance in our estimations. However, we also may assess the robustness of our estimation strategy by exploring machine learning-based tools for flexible covariate adjustment in experimental data (see, *e.g.*, List et al., 2024).

Appendix A: Simulations

The purpose of the following simulations was to assess how different randomization schemes and sets of covariates affect precision. These results guided us in designing the study.

A1: Data for simulations

We extract the population residing in the area of the Wellbeing Services County of Ostrobothnia at the end of 2022, using Statistics Finland's FOLK Basic module. We exclude the residents of Kristiinankaupunki due to its outsourced primary care service production, leading to a population of 170,066 in 86,034 families. We also observe age, gender, family ID, municipality of residence, income, and language.

The data on the utilization of traditional primary care from 4/2022–3/2024 comes from the Avohilmo Register of Primary Health Care Visits, administered by the Finnish Institute for Health and Welfare. We proceed as if the trial had started on 4/2023 to allow us to have a 12-month follow-up as well as 12 months of outcome data preceding the trial. We create several outcomes (Y1.1 and Y1.2)[15] and count all contacts received during the same day as one. That is, we measure the number of days with at least one contact. We will consider several follow-ups (9, 10, 11, or 12 months) and annualize the outcomes. For instance, the number of contacts in the 9-month follow-up is divided by 9 and multiplied by 12 to get the annualized figure. To come up with reasonable parameters for the take-up of digital clinics, we use Avohilmo data from the residents of East Uusimaa from 9/2023–3/2024. East Uusimaa launched its digital clinic in 9/2023.

We stratify the population in two ways. First, we stratify by gender and age in years, combining the gender-specific cells where age is at least 90. Second, we stratify by baseline traditional primary care use, that is, by the number of contacts with traditional primary care in the 12 months preceding the trial. Strata with fewer than two observations are merged with adjacent strata from above.

[15](Y1.1): in-person visits in PPC; (Y1.2): other contacts with traditional PPC.

A2: Methods for simulations

Sampling. The study population is a full fixed, finite population, and we treat potential outcomes as fixed but treatment assignments as random. In line with this, the simulations are conducted by randomizing treatment assignments 2,000 times. If the remainder when dividing the stratum size by two is not zero in stratified randomizations, we randomly vary between using the floor and ceiling functions within strata to select the number of treated individuals. This is done to have approximately $\frac{1}{2}$ of the sample individuals treated.

Imputing the unobserved potential outcomes. In our data, we observe traditional primary care use when the digital clinic was not available and use that figure as the potential untreated outcome $Y_i(0)$ for individual i. However, we need to impute the potential digital clinic use under treatment $D_i(1)$ and without exposure $D_i(0)$, as well as the potential outcome under treatment $Y_i(1)$.

We assume that the digital clinic use is zero, $D_i(0) = 0$, if individual i is not offered access. For those offered access, we draw the digital clinic use $D_i(1)$ from the negative binomial distribution under parameters $mean = 0.23$ and $size = 0.0225$ (dispersion parameter). The corresponding density function is strictly decreasing (zero contacts have the highest density). The realized draw has the following characteristics: 1) the share of individuals with at least one digital clinic contact is 5.24%, 2) the number of digital clinic contacts per capita is 0.23, and 3) the number of digital clinic contacts per client is 4.39. These simulated figures match well with the actual figures from the first year (2021) of the digital clinic of Harjun Terveys, which were 0.23 digital clinic contacts per capita, 4.4 contacts per client, and 5.3% of individuals had at least one digital clinic contact.[16] Moreover, the digital clinic in the Wellbeing Services County of East Uusimaa had approximately 0.22 annualized digital clinic contacts per capita (with strong identification) in its first three quarters.[17]

[16] Source: presentation by Joonas Turunen at the EHMA2022 Conference (LINK), 12.6.2024.

[17] Source: publicly available Power BI report of the Wellbeing Services County of East Uusimaa (LINK) and our own calculations, 10.6.2024.

In reality, the use of the digital clinic is likely strongly correlated with at least age and gender (Haaga, Herzig, et al., 2024, Figure 1). Since age and gender are expected to predict digital clinic use well, they are likely important covariates for the precision of our analysis – even more so than our simulations currently suggest.

The unobserved treated potential traditional primary care use $Y_i(1)$ for individual i is imputed using the following model: $Y_i(1) = Y_i(0)$. That is, we assume that the ITT effect is zero (and, consequently, the ACR is zero) and constant for all individuals. This is a rather strict and unrealistic model that rules out heterogeneity in effects.

Metrics for evaluating performance. We use three metrics averaged over 2,000 replications:

A. Mean squared error (MSE) of the point estimator $\hat{\tau}$ compared to the true τ, measuring the precision of the point estimator.
B. The average rejection probability of testing $\hat{\tau} = \tau$. This is the empirical size of the test (Type 1 error) and should ideally be close to the nominal size of 0.05.
C. The average standard error (SE) which is proportional to the confidence interval of the effect.

Estimators. We analyze intent-to treat (ITT) effects by using ordinary least squares (OLS) and average causal responses (ACR) by using two-stage least squares (2SLS). Most specifications include fixed effects for some combination of observables, including age (in years), gender, and baseline outcome. Heteroskedasticity-robust standard errors are used when randomizing at the individual level. All estimations are conducted by using the *fixest* R package.

Specifications. We assess performance of the following specifications, randomizing at the individual level:

A. Completely randomized experiment; no fixed effects included. This is the baseline against which we compare the performance of other methods.
B. Completely randomized experiment; fixed effects for age and gender.
C. Completely randomized experiment; fixed effects for baseline traditional primary care use.

D. Completely randomized experiment; fixed effects for baseline traditional primary care use, age, and gender.

E. Completely randomized experiment; fixed effects for baseline traditional primary care use, age, gender, municipality, income percentile, and language.

F. Stratified (age-by-gender) randomized experiment; fixed effects for age and gender.

G. Stratified (age-by-gender) randomized experiment; fixed effects for baseline traditional primary care use, age, and gender.

H. Stratified (baseline traditional primary care use) randomized experiment; fixed effects for baseline traditional primary care use.

I. Stratified (baseline traditional primary care use) randomized experiment; fixed effects for baseline traditional primary care use, age, and gender.

A3: Simulation results, different specifications

Table 4 reports the mean performance of nine specifications across 2,000 replications in a nine-month follow-up when we randomize at the individual level in each stratum. Several findings emerge. First, the empirical size is close to its nominal value of 5% for all specifications, which is good. Second, specifications that account for baseline traditional primary care use (specifications 3, 4, 5, 7, 8, and 9) lead to significant precision gains relative to those not using any fixed effects at all (Specification 1) or those only accounting for age and gender (Specifications 2 and 6). Specifications that account for the baseline outcome have 35–39% smaller average MSE and 21–23% smaller average SE than Specification 1. Third, there is little benefit from stratifying the randomization by the baseline outcome if fixed effects for the baseline outcome are already included in regression analysis (Specification 3 vs. Specification 8; Specification 4 vs. Specification 9).

Based on these results, we favor the combination of a completely randomized experiment and a rich set of fixed effects in regression analysis. This randomization method is the easiest to implement, and it does not require any additional data beyond the personal IDs (and permanent

addresses when randomizing at the permanent address level) of the target population. At the same time, this specification leads to good performance in the simulations we conducted. Note, however, that we do not know the optimal form of stratification/regression adjustment. With additional effort, covariates, and computational resources, we could likely find specifications that perform better in simulations than Specification 5. Additionally, while Specification 5 has attractive properties *ex ante*, is not necessarily better than the alternatives *ex post*.

Table 4: Comparison of Different Specifications, Simulation Results.

			Reduced form (OLS) $Z \to Y$			ACR (2SLS) $D \to Y$		
	Randomization	Fixed effects	MSE	Size	SE	MSE	Size	SE
1	Complete	None		5.35			5.35	
2	Complete	Age, gender	-7.4%	5.10	-2.9%	-7.5%	5.10	-2.9%
3	Complete	Y_baseline	-36.3%	5.15	-21.7%	-36.3%	5.15	-21.7%
4	Complete	Y_baseline, age, gender	-38.1%	5.25	-22.6%	-38.1%	5.25	-22.5%
5	Complete	Rich FEs	-38.7%	5.30	-22.8%	-38.7%	5.30	-22.7%
6	Strata (age X gender)	Age, gender	-5.7%	4.95	-2.9%	-5.6%	4.95	-2.9%
7	Strata (age X gender)	Y_baseline, age, gender	-37.6%	5.75	-22.6%	-37.5%	5.70	-22.5%
8	Strata (Y_baseline)	Y_baseline	-35.3%	5.30	-21.0%	-35.2%	5.25	-20.9%
9	Strata (Y_baseline)	Y_baseline, age, gender	-37.0%	5.45	-21.8%	-36.9%	5.40	-21.8%

Notes: The table reports the simulated performance of several combinations of randomization and fixed effects in regressions across 2,000 replications, using a nine-month follow-up and randomizing at the individual level in each stratum. The outcome is Y1.1 (in-person visits in PPC). The study population is a full fixed, finite population ($N = 170,066$), and we treat potential outcomes as fixed but treatment assignments as random. The reduced form ITT effect (the impact of access to the digital clinic on the use of traditional primary care; $Z \to Y$) is set to be zero and constant for all individuals. We assume that the digital clinic use (D) is zero if an individual i is not offered access. For those offered access, we draw their digital clinic use from the negative binomial distribution with parameters $mean = 0.23$ and $size = 0.0225$ (dispersion parameter). We analyze ITT effects using ordinary least squares (OLS) and average causal responses (ACR) using two-stage least squares (2SLS). Most specifications include fixed effects for some combination of observables. Specification 5 contains FEs for baseline traditional primary care use, age, gender, municipality, income percentile, and language. Heteroskedasticity-robust standard errors are used. The mean squared error of the point estimator (MSE) and standard errors (SE) are reported as a percentage change relative to the baseline of a completely randomized trial with no covariates in regression analysis. Size is the empirical size of the test (Type 1 error) and should ideally be close to the nominal size of 0.05. 'Strata' refers to stratified randomization, and 'Y' refers to the outcome (the annualized number of contacts with traditional public primary care). For details on the data, see Section A1.

A4: Simulation results, observational data on digital clinic use

In the simulations reported above, the simulated digital clinic use was independent of observables. However, digital clinic use is highly correlated with observables, such as age, in the real world. In

this subsection, we conduct simulations based on data from the residents of East Uusimaa using a 7-month follow-up after East Uusimaa adopted its digital clinic in 9/2023.[18] Although the 7-month follow-up is shorter than we propose for the experiment and although we behave as if the digital clinic use in East Uusimaa had no effect on the traditional PPC use, the clear benefit of these simulations is that they are based on a realistic empirical distribution of digital clinic use. The population of East Uusimaa (98,967) is smaller than the population of Ostrobothnia (170,066). We match the number of families (86,034) by randomly sampling with replacement families from the population of East Uusimaa until we have the same number of families in East Uusimaa than there are in Ostrobothnia.[19] Finally, the analysis sample is restricted to those aged from 0 to 69 to increase power (see Figure 4 in Appendix B).

Table 5 reports the simulated average standard errors as a function of outcome using our preferred specification, Specification 5, after randomizing at the family ID level. Focusing on the ACR analysis, which assesses the impact of the use of the digital clinic on the use of traditional public primary care, the simulated standard errors were 0.089 for Outcome Y1.1 and 0.128 for Outcome Y1.2. Centered around zero, the corresponding confidence intervals would be [–0.174, +0.174] for Outcome Y1.1 and [–0.251, +0.251] for Outcome Y1.2. The definition of the ACR is somewhat technical,[20] but we interpret it to measure the degree of substitution between the digital clinic and traditional public primary care services for compliers who consult the digital clinic more only because they were offered access to it. For reference, Ellegård et al. (2022) estimate a substitution rate of 0.45 ($SE = 0.20$) between direct-to-consumer telemedicine and *in-person* physician visits at the 20th birthday in Sweden in an earlier, observational study. Overall, we are satisfied with this precision that the simulations suggest.

[18] We should change the follow-up to 9 months once we receive more data. This change should lead to smaller mean SEs in the simulations.

[19] Due to differences in family size, the total number of sampled individuals is sligthly smaller in East Uusimaa (167,530) than it is in Ostrobothnia (170,066).

[20] The ACR is a weighted average, over all values of d (potential intensity of the treatment), of the effect of increasing treatment from $d-1$ to d among switchers whose treatment status goes from strictly below to above d over time.

Table 5: Simulated Average Standard Errors Based on Data from East Uusimaa.

		Reduced form (OLS) $Z \to Y$	ACR (2SLS) $D \to Y$
Outcome	Control mean (Y)	SE [$1.96 \times SE$]	SE [$1.96 \times SE$]
In-person visits in PPC (Y1.1)	0.814	0.011 [0.022]	0.089 [0.174]
Other contacts with PPC (Y1.2)	1.545	0.016 [0.031]	0.128 [0.251]

Notes: The table reports the simulated average standard errors (SE) across 2,000 replications as a function of outcome (Y1.1–Y1.2), using randomization at the family level, a seven-month follow-up, and Specification 5 (FEs for baseline traditional primary care use, age, gender, municipality, income percentile, and language). To construct the study population, 1) we take the population of East Uusimaa, 2) randomly sample extra families (duplicates) from that population, treating these as distinct families, to mimic the number of families in Ostrobothnia (East Uusimaa has smaller population). Finally, the analysis sample is restricted to those aged 0 to 69. We treat potential outcomes as fixed but treatment assignments as random. We set digital clinic use (D) to be zero if individual i is not offered access. The reduced-form ITT effect (the impact of access to the digital clinic on the use of traditional primary care; $Z \to Y$) is set to be zero and constant for all individuals. We analyze ITT effects using ordinary least squares (OLS) and average causal responses (ACR) using two-stage least squares (2SLS). Standard errors are clustered by family ID. 'Y' refers to the outcome (the annualized number of contacts to traditional public primary care). For details on the data, see Section A1.

Appendix B: Statistical power

Statistical power is the probability of detecting an effect if it truly exists in the population. High statistical power is desirable because it reduces the chances of failing to detect an effect when it is present (Type 2 error). The desired power level is typically 80% or higher, e.g., 90%. We complement the simulations of Appendix A by conducting a power analysis for intent-to-treat (ITT) effects, that is, the impact of *access to* the digital clinic on the use of traditional public primary care, and for local average treatment effects (LATE), that is, the impact of having *used* the digital clinic at least once on the use of traditional public primary care.

Limitations of ITT and LATE power analyses. The approach for LATE power analysis proposed by Bansak (2020) produces bounds for power. The lower bounds that we will report are thus conservative. Moreover, the approach applies to the LATE effect with binary treatment (whether or not an individual used the digital clinic). However, our ultimate estimand of interest is the average causal response (ACR) with multivalued treatment (how many digital clinic visits an individual has).[21] Our power analysis for the ITT effects does not account for covariates, and our ultimate estimand of interest is the ACR, not ITT. As the target estimand is different, the ITT power analysis can diverge from the results of the appropriate power analysis for ACR. Focusing on the LATE, which would be the impact of *using* the digital clinic *at least once*, Bansak (2020) shows that ITT power analyses can diverge substantially (to both directions) from the results of a LATE power analysis, especially if the compliance rate is low (which is the case in our application). The intuition behind this difference is that the compliance rate is a quantity to be estimated in the LATE analysis (but not in the ITT analysis), and that estimate is generally correlated with the estimate of the ITT (Bansak, 2020).

[21] The ACR is a weighted average, over all values of d (potential intensity of the treatment), of the effect of increasing treatment from $d-1$ to d among switchers whose treatment status goes from strictly below to above d over time. We interpret it to measure the degree of substitution between the digital clinic and traditional public primary care services for compliers who consult the digital clinic more only because they were offered access to it.

Note that our power analyses are based on randomizing at the family level. If we instead randomized at the permanent address level, which we propose in the PAP, it would mean fewer clusters than when randomizing at the family level and plausibly wider CIs.[22]

Parameter choices for LATE and ITT power analyses:

- 5% significance level (Type 1 error probability).
- follow-up of 9, 10, 11, or 12 months.
- treatment group size of 50%.
- substitution rate between digital clinic contacts and traditional primary care contacts of 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, or 0.40 — these are selected, and they represent expected effect size.

B1: Power analysis for local average treatment effect (LATE)

Power analysis for LATE is complicated by the fact that the compliance rate (here: the number of people who use the digital clinic) must be estimated and that estimate is correlated with the estimate of the ITT. Bansak (2020) proposes an approach to LATE power analysis employing the Wald IV estimator, bounding the power of the test of a null hypothesis. As such, the methodology provides a conservative "worst-case scenario" for power. We implemented the analysis using the *powerLATE* R package.

The following parameters were estimated from data:

- the number of randomization units (baseline: families) — based on FOLK data. Overall, we observe 170,066 individuals in 86,034 families.
- the impact of access to the digital clinic on the share of people who used the digital clinic as a function of follow-up (9, 10, 11, or 12 months). We used data from the digital clinic use of the residents of East Uusimaa from 9/2023–3/2024 and extrapolated the number of patients

[22]For instance, cousins living in the same apartment and having the same permanent address do not belong to the same family. For comparison, there were 86,034 families and 79,980 household-dwelling units in Ostrobothnia (excluding Kristiinankaupunki) in 2022 (source: FOLK data and Statistics Finland.)

by a linear fit for longer follow-ups. Finally, we multiplied this parameter by 0.94, 1.00, or 1.06 to get baseline, optimistic, and pessimistic scenarios for take-up.

- the annualized number of digital clinic contacts per digital clinic user. We used data from the digital clinic use of the residents of East Uusimaa from 9/2023–3/2024 and extrapolated the number of contacts by a linear fit for longer follow-ups.
- the standard deviation of (untreated) traditional primary care use as a function of follow-up (9, 10, 11, or 12 months) and outcome (Y1.1–Y1.2). This parameter was estimated with the data from the residents of Ostrobothnia from 4/2023–3/2024 after aggregating (averaging) outcomes to the level of randomization (baseline: family ID).
- the proportion of variation in the digital clinic use D left unexplained by the randomized access Z that is explained by covariates (r2dw). This parameter was estimated (adjusted R2) with the data from the residents of East Uusimaa from 9/2023–3/2024, regressing the indicator for digital clinic use on fixed effects for age, gender, municipality, income percentile, and language.
- the proportion of variation in traditional PPC use Y left unexplained by the randomized access Z that is explained by covariates (r2yw). This parameter was estimated (adjusted R2) with the data from the residents of Ostrobothnia from 4/2023–3/2024 using a follow-up of 9, 10, 11, or 12 months and regressing the outcome (Y1.1–Y1.2; the number of contacts to traditional PPC) on fixed effects for baseline traditional primary care use, age, gender, municipality, income percentile, and language.

The absolute LATE effect on the use of traditional primary care was calculated by multiplying the selected substitution rate and the estimated number of digital clinic contacts per digital clinic user. The LATE effect size was constructed by dividing the absolute LATE effect by the standard deviation of the traditional primary care use which was estimated from the data. The substitution rate can thus be interpreted as the selected MDE because it is proportional to the absolute LATE effect and the LATE effect size. In other words, a reasonable MDE was constructed by assuming (based on an earlier digital clinic launch in East Uusimaa) how much the digital clinic would

actually be used and how many traditional PPC contacts one digital clinic contact replaces (substitution rate; our ultimate parameter of interest).

Table 6 reports the lower bound of power for different values of the substitution rate, the coefficient with which we multiply the projected compliance, and follow-up, restricting the analysis sample to those aged from 0 to 69. Starting with Outcome Y1.1, a 9-month follow-up should lead to conventional power (more than 80%) for MDEs mapping to a substitution rate of 0.40, 72–82% power for MDEs mapping to a substitution rate of 0.35, and low power (47–56%) for MDEs mapping to a substitution rate of 0.25. Longer follow-ups reduce MDEs: a 10-month (11-month) [12-month] follow-up should lead to conventional power for MDEs mapping to a substitution rate of 0.35 (0.30) [0.25]. For reference, Ellegård et al. (2022) estimate a substitution rate of 0.45 ($SE = 0.20$) between direct-to-consumer telemedicine and *in-person* physician visits at the 20th birthday in Sweden in an earlier, observational study.

Second, focus on Outcome Y1.2 for which we expect the effect size to be larger than for Outcome Y1.1.[23] A 9-month (10-month) [12-month] follow-up should lead to conventional power for MDEs mapping to a substitution rate of 0.30 (0.25) [0.20], which is good.

Next, we assess how much the following changes can affect power: i) randomizing at a more granular level than family and ii) making the sample more homogeneous by restricting the analysis sample by age which is highly correlated with the share of digital clinic users. Excluding the elderly, who rarely use the digital clinic, from the analysis sample would reduce sample size but increase compliance and reduce the noise of the outcome. Ultimately, this sample restriction could increase power although it would change the estimand as the full sample is not used. As an alternative randomization scheme, we examine the following. Use person ID, age, and family ID to create clusters as follows: i) each adult living in a family without minors forms their own cluster, and ii) all individuals in a family with minors form a cluster. Randomly assign these clusters to treatment and control groups.[24] This would likely increase power compared to randomizing at

[23] In general, power may be larger for outcomes closer in the causal chain.

[24] Overall N: 118,156 clusters with the alternative randomization scheme vs. 86,934 families.

Table 6: Conservative Lower Bound of Power for the LATE Effect.

Substitution rate	Compliance coefficient	Power			
		9-month follow-up	10-month follow-up	11-month follow-up	12-month follow-up
A. Outcome Y1.1: in-person visits in PPC					
0.40	1.06	90%	96%	99%	100%
0.40	1.00	86%	94%	98%	99%
0.40	0.94	81%	91%	96%	99%
0.35	1.06	82%	91%	96%	99%
0.35	1.00	77%	88%	94%	98%
0.35	0.94	72%	83%	91%	97%
0.30	1.06	70%	82%	91%	96%
0.30	1.00	65%	78%	87%	94%
0.30	0.94	60%	72%	83%	91%
0.25	1.06	56%	69%	79%	89%
0.25	1.00	52%	64%	75%	85%
0.25	0.94	47%	58%	69%	80%
B. Outcome Y1.2: other contacts with traditional PPC					
0.30	1.06	88%	94%	97%	99%
0.30	1.00	84%	91%	96%	98%
0.30	0.94	79%	87%	93%	97%
0.25	1.06	76%	85%	91%	96%
0.25	1.00	71%	80%	88%	94%
0.25	0.94	65%	75%	84%	91%
0.20	1.06	58%	69%	78%	86%
0.20	1.00	54%	64%	73%	82%
0.20	0.94	49%	58%	67%	77%

Notes: The table reports power calculations for the impact of *having used* the digital clinic *at least once during the follow-up* on the use of traditional public primary care (local average treatment effect, LATE). The analysis sample is restricted to those aged from 0 to 69. Treatment is randomized at the family level, sample size refers to the number of families, and standard deviation of the outcome is estimated from family-level outcome means. We used the *powerLATE* R package to construct a conservative lower bound of power for LATE analysis employing the Wald IV estimator (Bansak, 2020) at the 5% significance level with a treatment group size of 50% as a function of outcome, follow-up, substitution rate between digital clinic contacts and traditional primary care contacts, and the coefficient for multiplying projected compliance. The following parameters were selected based on empirical data: 1) sample size, 2) compliance, 3) the annualized number of digital clinic contacts per digital clinic user, 4) the standard deviation of (untreated) traditional primary care use, and 5) r2dw and r2yw (the proportion of variation in the digital clinic use D / traditional PPC use Y left unexplained by the randomized access Z that is explained by covariates). The absolute LATE effect on the use of traditional primary care was calculated by multiplying the selected substitution rate and the selected number of digital clinic contacts per digital clinic user. The LATE effect size was constructed by dividing the absolute LATE effect by the standard deviation of the traditional primary care use. The substitution rate can thus be interpreted as the selected MDE because it is proportional to the absolute LATE effect and the LATE effect size.

the family level, although it would create situations is which treatment assignment varies within all-adult households.

The results in Figure 4 are based on a 9-month follow-up, substitution rate of 0.25, and compliance coefficient of 1.00. The figure shows that the lower bound for power is relatively stable when the upper age limit of the analysis sample is from 40 to 70. Power is lower if the upper age limit is lower than 40 or if the elderly are included in the analysis sample. The power gain from excluding those older than 70 is 5–10 percentage points for Outcome Y1.1 and more than 10 percentage points for Outcome Y1.2. The gain from using the alternative, more granular randomization scheme is slightly smaller.[25] The figure also illustrates that, for a given selected substitution rate, associated power is larger for Outcome Y1.2 than for Y1.1.[26]

Conclusions. We propose randomizing at the permanent address level, and we prefer to restrict the analysis to those aged from 0 to 70. With a 9-month follow-up, we should have conventional power levels to detect effects of reasonable sizes for Outcome Y1.2 (*e.g.,* a substitution rate of 0.30). Regarding Outcome Y1.1, we have conventional power levels for a substitution rate of 0.40, which we think is acceptable but not ideal. If possible, we would recommend a longer follow-up for Outcome Y1.1. For instance, with 12-month (11-month) [10-month] follow-up we should have conventional power levels to detect a substitution rate of 0.25 (0.30) [0.35] for Outcome Y1.1.

[25]Based on this finding, we expect that the reduction in power from randomizing at the permanent address level (vs. at the family level) is small.

[26]We would expect *a priori* that effect size is larger for Outcome Y1.2.

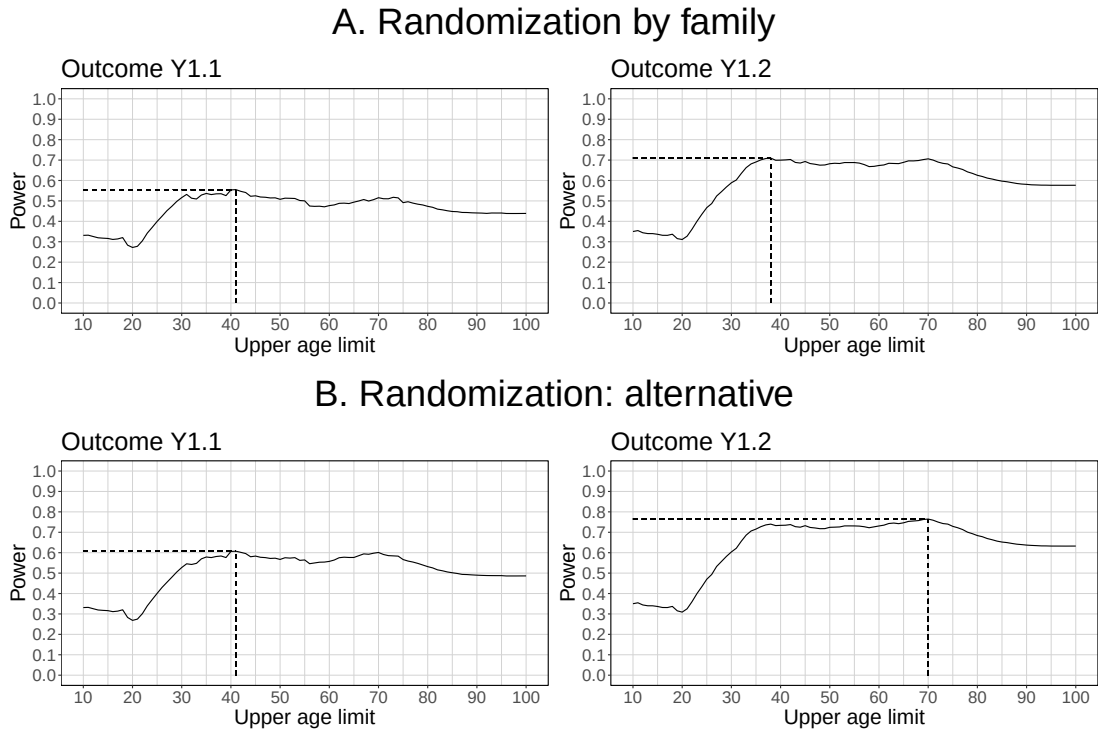

Figure 4: Conservative Lower Bound of Power for the LATE Effect, Age Restriction.

Notes: The figure reports power calculations for the impact of *having used* the digital clinic *at least once during the 9-month follow-up* on the use of traditional public primary care (local average treatment effect, LATE). Outcomes: (Y1.1): in-person visits in PPC; (Y1.2): other contacts with traditional PPC. Treatment is randomized at the family level or at an alternative, more granular level (see the main text), sample size refers to the number of randomization units, and the standard deviation of the outcome is estimated from randomization-unit-level outcome means. We used the *powerLATE* R package to construct a conservative lower bound of power for LATE analysis employing the Wald IV estimator (Bansak, 2020) at the 5% significance level with a treatment group size of 50% for a 9-month follow-up, a compliance coefficient of 1.00, and a substitution rate of 0.25 as a function of outcome, randomization scheme, and the upper age limit for the analysis sample. The following parameters were selected based on empirical data: 1) sample size, 2) compliance, 3) the annualized number of digital clinic contacts per digital clinic user, 4) the standard deviation of (untreated) traditional primary care use, and 5) r2dw and r2yw (the proportion of variation in the digital clinic use D / traditional PPC use Y left unexplained by the randomized access Z that is explained by covariates). The absolute LATE effect on the use of traditional primary care was calculated by multiplying the selected substitution rate and the selected number of digital clinic contacts per digital clinic user. The LATE effect size was constructed by dividing the absolute LATE effect by the standard deviation of the traditional primary care use. The substitution rate can thus be interpreted as the selected MDE because it is proportional to the absolute LATE effect and the LATE effect size.

B2: Power analysis for intent-to-treat effect (ITT)

We used the *pwr* R package to implement scaled ITT power calculations for the two-sample t-test.

The following parameters were estimated from data:

- the number of randomization units (baseline: families) — based on FOLK data. Overall, we observe 170,066 individuals in 86,034 families.
- the impact of access to the digital clinic on the number of annualized digital clinic contacts per capita as a function of follow-up (9, 10, 11, or 12 months). We used data from the digital clinic use of the residents of East Uusimaa from 9/2023–3/2024 and extrapolated the number of contacts by a linear fit for longer follow-ups. Finally, we multiplied this parameter by 0.94, 1.00, or 1.06 (contacts coefficient) to get baseline, optimistic, and pessimistic scenarios for digital clinic use.
- the standard deviation of (untreated) traditional primary care use as a function of follow-up (9, 10, 11, or 12 months) and outcome (Y1.1–Y1.2). This parameter was estimated using data from the residents of Ostrobothnia from 4/2023–3/2024 after aggregating (averaging) outcomes to the level of randomization (baseline: family ID).

The absolute ITT effect on the use of traditional primary care was calculated by multiplying the selected substitution rate and the selected impact on digital clinic use. The ITT effect size was constructed by dividing the absolute ITT effect by the standard deviation of the traditional primary care use. The substitution rate can be interpreted as the minimum detectable effect (MDE) because it is proportional to the absolute ITT effect and the ITT effect size. In other words, a reasonable MDE was constructed by assuming (based on an earlier digital clinic launch in East Uusimaa) how much the digital clinic would actually be used and how many traditional PPC contacts one digital clinic contact replaces (substitution rate; our ultimate parameter of interest).

Table 7 reports statistical power for different values of the substitution rate, the coefficient with which we multiply the projected digital clinic contacts per capita, and follow-up, restricting the analysis sample to those aged 0 to 69 and randomizing at the family level. The difference from the LATE power analyses is that i) power for ITT does not decrease as drastically for shorter follow-ups as it did for LATE, and ii) for a 12-month follow-up, the projected power for ITT is slightly lower than the projected power for LATE. Starting with Outcome Y1.1, a 9-month follow-up should lead to conventional power (more than 80%) for MDEs mapping to a substitution

rate of 0.35 (a sligthly lower MDE than in LATE analyses), 72–81% power for MDEs mapping to a substitution rate of 0.30, and low power (56–66%) for MDEs mapping to a substitution rate of 0.25. Longer follow-ups reduce MDEs: an 11-month follow-up should lead to conventional power for MDEs mapping to a substitution rate of 0.30. Second, focusing on Outcome Y1.2, for which we expect the effect size to be larger than for Outcome Y1.1, a 9-month follow-up should lead to conventional power for MDEs mapping to a substitution rate of 0.25, which is good.

Table 7: Power Calculations for the ITT Effect.

Substitution rate	Contacts coefficient	Power: 9-month follow-up	Power: 10-month follow-up	Power: 11-month follow-up	Power: 12-month follow-up
A. Outcome Y1.1: in-person visits in PPC					
0.35	1.06	91%	93%	94%	96%
0.35	1.00	88%	90%	92%	94%
0.35	0.94	84%	86%	88%	91%
0.30	1.06	81%	84%	86%	89%
0.30	1.00	77%	80%	82%	85%
0.30	0.94	72%	74%	77%	80%
0.25	1.06	66%	69%	72%	75%
0.25	1.00	61%	64%	67%	70%
0.25	0.94	56%	59%	61%	65%
B. Outcome Y1.2: other contacts with traditional PPC					
0.25	1.06	85%	86%	87%	89%
0.25	1.00	81%	82%	83%	85%
0.25	0.94	76%	77%	78%	80%
0.20	1.06	67%	68%	70%	72%
0.20	1.00	62%	63%	64%	67%
0.20	0.94	57%	58%	59%	62%
0.15	1.06	44%	45%	46%	48%
0.15	1.00	40%	41%	42%	44%
0.15	0.94	36%	37%	38%	39%

Notes: The table reports power calculations for the impact of *access to* the digital clinic on the use of traditional public primary care (intent-to-treat, ITT). The analysis sample is restricted to those aged from 0 to 69. Treatment is randomized at the family level, sample size refers to the number of families, and the standard deviation of the outcome is estimated from family-level outcome means. We used the *pwr* R package to implement the power calculations for the two-sample t-test at the 5% significance level with a treatment group size of 50% as a function of outcome, follow-up, substitution rate between digital clinic contacts and traditional primary care contacts, and the coefficient for multiplying projected digital clinic use per capita. The following parameters were selected based on empirical data: 1) sample size, 2) the annualized number of digital clinic contacts per capita, and 3) the standard deviation of (untreated) traditional primary care use. The absolute ITT effect on the use of traditional primary care was calculated by multiplying the selected substitution rate and the selected impact on the digital clinic use. The ITT effect size was constructed by dividing the absolute ITT effect by the standard deviation of the traditional primary care use. The substitution rate can be interpreted as the minimum detectable effect (MDE) because it is proportional to the absolute ITT effect and the ITT effect size. The power calculations can be viewed as conservative as they do not account for covariates that could increase precision and power (see Table 4).

Appendix C: Potential spillovers

C1: Are there spillovers to traditional PPC?

There may be both a direct and an indirect impact of the experiment on access to traditional PPC. The direct effect is negative and caused by moving labor to the digital clinic (nurses and physicians, mostly nurses) from traditional PPC. The positive indirect effect potentially exists because digital clinic contacts (at least partially) substitute for traditional PPC. We expect that the spillover on access is largest for the care needs assessment telephone service and smaller for physician contacts due to institutional factors (gatekeeping by nurses at both digital clinics and traditional PPC).[27] The telephone service for the care needs assessment is centralized in Ostrobothnia (as the digital clinic will be) while other services in traditional PPC, such as in-person visits, are not as they are produced by traditional PPC clinics. The fact that the care needs assessment telephone service is centralized creates challenges for potential random saturation designs.

Consider the potential bias from spillovers via the following back-of-the-envelope model. Before the trial, the utilization of traditional PPC is the same (Y_0) for the treated and the control group due to the randomized treatment assignment. During the trial, suppose that the equilibrium is reached in three steps. In the first step, access to traditional PPC is reduced because some labor is transferred to the digital clinic. This reduces the use of traditional PPC by X units in both the treatment and control groups, with utilization being $Y_0 - X$ in both groups. In the second step, the treatment group has D digital clinic contacts (the control group 0). The substitution rate is α so that the utilization of traditional PPC is $Y_0 - X$ for the controls and becomes $Y_0 - X - \alpha D$ for the treated. In the third step, αD contacts to traditional PPC that became available because the treatment group did not consume them due to digital clinic contacts are allocated between the treatment and the control groups, increasing the use of traditional PPC. Ultimately, the utilization of traditional PPC is $Y_0 - X + \beta\alpha D$ for the controls and $Y_0 - X - \alpha D + (1-\beta)\alpha D$ for the treated.

[27] For reference, the digital clinic of the Wellbeing Services County of Pirkanmaa produced 31,000 contacts during its first two months. During the first month, there were 80,900 calls to the care needs assessment telephone service, which was 8,300 fewer calls than in the previous month. Source: https://www.pirha.fi/w/digiklinikalle-rekisteroitynyt-jo-yli-45-000-kayttajaa-pirkanmaalla, accessed on July 24th 2024.

Thus, the difference is $-2\beta\alpha D \in [-2\alpha D, 0]$ for $\beta \in [0,1]$. From this expression, we can see that it is the coefficient β which can make the estimated effect differ from the true effect ($-\alpha D$).

The back-of-the-envelope model suggests that there are two cases in which the estimated effect is unbiased: either 1) the substitution rate is zero ($\alpha = 0$) or 2) the potential spillover affects the potential outcomes of the treatment and the control group similarly ($\beta = 0.5$). In other words, the αD contacts with traditional PPC that became available because the treatment group did not consume them due to digital clinic contacts should be allocated equally between the treatment and the control groups. If these contacts are disproportionately allocated to the control (treatment) group, we would overestimate (underestimate) the substitution rate.

We believe that $\beta > 0.5$ and that our empirical approach overestimates the substitution rate. The logic is as follows. Suppose that 1) all health shocks can be ordered based on their severity, 2) the gatekeeping system leads to a situation where only the most severe (but not all) health shocks lead to a PPC contact, and 3) the distribution of health shock severity is uniform (unrealistic but assumed for simplicity) and the same for the treated and controls (in other words, no health effects from access to the digital clinic). Then, the difference in PPC use (including both the traditional PPC and digital clinics) between the treated and the controls would be $(1-\alpha)D$ at the second step of the above-described back-of-the-envelope model, before the αD contacts with traditional PPC that became available because the treatment group did not consume them due to digital clinic contacts are reallocated. If $\alpha \in]0, 0.5]$, then $(1-\alpha)D \geq \alpha D$, implying that the control group has αD or more untreated health shocks that are more severe than all the untreated health shocks in the treatment group, leading to $\beta = 1$. If $\alpha = 1$, then the utilization of PPC is the same for the treated and the controls ($Y_0 - X$), implying that $\beta = 0.5$ because the distribution of the untreated health shocks is the same for the treated and the controls. Finally, if $\alpha \in]0.5, 1[$, then $(1-\alpha)D \in]0, 0.5D[$ while $\alpha D \in]0.5D, D[$, implying that the majority of the contacts with traditional PPC that became available because the treatment group did not consume them due to digital clinic contacts are reallocated to the control group, with $\beta \in]0.5, 1[$. In fact, $\beta = \frac{1}{2\alpha}$

decreases for $\alpha \in]0.5, 1[$.[28] Under the assumptions listed above, the previous results suggest that for $\alpha \in]0, 0.5]$ our estimator would overestimate the substitution rate by a factor of 2 while for $\alpha \in]0.5, 1[$ we would always estimate a substitution rate of 1.

The main limitation of the proposed experiment and the related analyses is that they do not take into account the potential spillover effects of the experiment on access to traditional PPC. How serious is this limitation? The answer is ultimately subjective, but we are not overly worried about the potential spillovers. Importantly, it would be unrealistic to assume that the αD contacts with traditional PPC that were not consumed by the treatment group due to the digital clinic would all be reallocated to the treatment and the control group (traditional PPC is not that supply-constrained). A more realistic model would thus contain an additional parameter $\gamma \in (0, 1)$ in the expressions for the utilization of traditional PPC, which would become $Y_0 - X + \beta\gamma\alpha D$ for the controls and $Y_0 - X - \alpha D + (1-\beta)\gamma\alpha D$ for the treated. Thus, the difference is $-2\beta\gamma\alpha D - (1-\gamma)\alpha D$ instead of the earlier $-2\beta\alpha D$. From these expressions, it can easily be verified using inequalities that the bias is smaller once $\gamma \in (0, 1)$ is included in the model.

Second, it would be unrealistic to assume that the newly available traditional PPC contacts are allocated entirely based on health shock severity. In other words, it would be unrealistic to assume that the newly available traditional PPC contacts are allocated entirely to one of the two treatment groups ($\beta = 1$ or $\beta = 0$), limiting the size of the potential bias. For example, $\beta = \frac{2}{3}$ ($\beta = \frac{3}{5}$) [$\beta = \frac{4}{7}$] would lead us to overestimate the substitution rate only by a factor of 1.33 (1.2) [1.14].

One way to measure changes in access to traditional PPC is to track the mean and maximum response time of the call-back service for care needs assessment. The Wellbeing Services County of Ostrobothnia follows these indicators regularly, and they should be also followed regularly throughout the trial. However, a simple before-and-after analysis of response times can be highly sensitive to time effects. Alternatively, we could compare the evolution in the outcomes before and after the experiment between the control group in Ostrobothnia and

[28] $\beta = \frac{(1-\alpha)D + [\alpha D - (1-\alpha)D]/2}{\alpha D} = \frac{1}{2\alpha}$ for $\alpha \in]0.5, 1[$

individuals residing in other wellbeing services counties in a difference-in-differences (DID) framework. However, we believe that the signal would be modest relative to noise in this analysis (lack of power).[29]

[29]This hypothesis is based on the findings of two earlier studies that examine the impacts of copayment changes on the utilization of traditional PPC in Finland, both based on a DID approach and having a much larger population in analysis than we would have for testing spillovers (Haaga, Böckerman, et al., 2024a; Haaga, Böckerman, et al., 2024b).

Appendix D: Research ethics

The design of this study relies on two key factors. First, randomized controlled trials often offer an unparalleled tool for estimating internally valid causal effects. Second, the implementation of our study coincides with a major reform in the provision of health care services that would have happened even in the absence of our study. Consequently, our study leverages a large health care reform and implements it in a staggered fashion in collaboration with the provider of health care services, offering the service provider an opportunity to avoid potential service congestion at the initial phase of the new service and ensure the availability of the workforce needed to operate the services.

D1: Informed consent

This study has been exempted from the requirement for informed consent by the Institutional Review Board (IRB) of the Finnish Institute for Health and Welfare. This waiver is based on two factors. First, our study leverages a health care reform in the public provision of health care services. Thus, the intervention would be available in the absence of our study. Moreover, using the digital clinic in the treatment group is voluntary and the existing channels for accessing health care will remain available. Second, this research uses only de-identified secondary data that is recorded in statutory electronic health records independently of our study. The use of these datasets is governed in Finland by the Act on the Secondary Use of Health and Social Data (Act 552/2019), allowing the use of these datasets for research as specified in the data permit.

D2: Risk assessment

We highlight in the risk assessment for our study three key aspects: 1) offering the digital clinic to the population, 2) assigning the treatment group through active randomization, and 3) communication with the treatment group during the trial. Overall, we assess that the risks associated with the first two aspects are very low. However, the communication with the treatment

group during the trial presents potential minor risks, which we seek to avoid and minimize through careful planning and close collaboration with the public provider of health and social care services in the region, the Wellbeing Services County of Ostrobothnia.

First, we note that the Wellbeing Services County of Ostrobothnia has decided to implement the digital clinic platform independently of this study. Therefore, any potential benefits and drawbacks of offering the digital clinic will occur regardless of the study. Digital clinics have already been widely introduced across different jurisdictions worldwide, and we estimate that solely in Finland, over three million Finns already live in areas where a digital clinic is available alongside traditional health care services.

The use of the digital clinic is always voluntary. However, at least theoretically, it is possible that offering a digital clinic may direct individuals toward digital services where their health issues may not be appropriately addressed, potentially hindering timely access to the best available treatment. In this regard, we emphasize that i) there is no reliable research evidence suggesting that digital services negatively impact health conditions, and ii) as mentioned above, the Wellbeing Services County of Ostrobothnia would implement the digital clinic regardless of this study.

To our knowledge, there are no previous empirical studies using such large-scale randomized trials to assess the benefits and drawbacks of digital clinics. This study aims to fill this knowledge gap. The results of this study may benefit service providers nationally and internationally in making more informed decisions regarding the allocation of resources between physical primary healthcare services and digital services.

Second, during the trial phase, digital clinic services will be offered for a limited time only to a randomly assigned treatment group. This means that the full benefits and drawbacks of the digital clinic will not be realized during the trial phase. However, since there is no credible empirical evidence regarding the advantages and disadvantages of offering a digital clinic, our trial can be seen as a means to mitigate both negative and positive risks related to digital clinic visits. We acknowledge that the study could be potentially ethically challenging if the benefits of

offering a digital clinic were deemed significantly greater than its drawbacks in advance, as this would mean the full benefits would not be realized during the pilot. However, to date, there is no certainty about the benefits and drawbacks of the digital clinic. Moreover, the fact that the public provider of health care services has agreed to conduct this experiment in collaboration with our research team indicates that there are significant open questions about the potential benefits and drawbacks of offering a digital clinic. Finally, we note that assigning 50% of the individuals to the treatment group not only maximizes statistical power but also seems optimal from an ethical point of view, as we do not know whether the benefits of offering a digital clinic outweigh its drawbacks.

Third, we note that our communication strategy during the trial phase serves three main objectives: i) to provide the treatment group with clear and timely information about their opportunity to use the digital clinic, ii) to explain why the digital clinic is being offered only to 50% of the population during the pilot and how this selection was made, and iii) to prevent confusion and potential disappointment for those who are not included in the treatment group. The minor risks associated with communication relate to the latter two objectives and could materialize if we fail to clearly explain the reasons for the trial (ensuring service capacity in physical and digital healthcare and conducting a reliable impact assessment) and the need for randomization (a fair way to create a credible control group) or if the public does not share our views. For example, some individuals in the control group may feel disappointed that they were not included in the treatment group. Moreover, there could be confusion among individuals who are unsure whether they belong to the treatment or control group.

The primary means to mitigate these risks is open communication about the motivation and implementation of our research in collaboration with the key personnel of the Wellbeing Services County of Ostrobothnia. These communication measures include clear and transparent information about the trial using the service provider's website in addition to targeted communication to the treatment group, 2) a public announcement of a targeted letter campaign related to the trial phase, 3) the provision of clear instructions on a website to allow individuals to verify through strong authentication, whether they belong to the treatment or control group.

D3: Data management and governance

Data Management. This research project relies on two types of data: treatment assignment data and administrative data (outcomes). The treatment assignment data includes the personal identification number, the household identification number and an indicator variable documenting whether the individual belongs to the treatment or control group. This treatment assignment data is linked to the administrative register data containing information on healthcare use and the socioeconomic information of individuals drawn from the data sources listed in Section 5.1.

The linking of these data sources is conducted using pseudonymized IDs (national identification numbers) by personnel working at Statistics Finland. Researchers using the dataset do not have access to the encryption key used by Statistics Finland. Thus, all data are processed such that researchers analyzing the data do not have access to directly identifiable information. All administrative health records and demographic data are stored at the Statistics Finland remote access system, a secure computing infrastructure for maintaining and processing research data. The use of the remote access system requires two-factor authentication and prevents the data from being stored outside the server.

Data governance. The research data are governed by their owners listed in Section 5.1. The research team cannot grant rights to use the data. Access to the healthcare data is regulated under the Act on the Secondary Use of Health and Social Data (552/2019) and can be obtained by sending a request to the Finnish Social and Health Data Permit Authority, Findata (https://findata.fi/en/). Access to demographic administrative data can be obtained by sending a request to Statistics Finland (https://www.stat.fi). User licenses for the research data use are granted in accordance with established practices by these organizations. All researchers handling the data must fulfill all necessary conditions for a data permit.

Data Openness. The dataset containing individual-level health data and demographic data is regulated under the Act on the Secondary Use of Health and Social Data (552/2019) and cannot be made readily available for the purpose of direct replication. However, as described above, access to the data can be obtained by sending requests to the Finnish Social and Health Data

Permit Authority, Findata, and to Statistics Finland. We commit to publishing all statistical code and other details of the computations that are sufficient to permit the validation of our empirical work. We commit to storing the treatment assignment code and data for future use, allowing interested researchers to construct the full dataset required for further research utilizing our original treatment assignment.

D4: Study and data permits

This study is conducted by an independent research team in cooperation with the Wellbeing Services County of Ostrobothnia. Before starting the trial, we have obtained the following study and data permits:

- *Research approval for scientific research*: approved by the research approval committee of the Wellbeing Services County of Ostrobothnia. This is a formal authorization for the study by the service provider in charge of providing public healthcare services during our study (5685/13.01/2024).
- *An IRB approval* granted by the Institutional Review Board of the Finnish Institute for Health and Welfare (THL/5935/6.02.01/2024).
- *A user license* to the full population individual-level sociodemographic and socioeconomic data, granted by Statistics Finland (TK/410/07.03.00/2025).
- *A user license* to the Population Information System data and administrative healthcare records from the Finnish Institute for Health and Welfare and the Finnish Social Insurance Institution, granted by the Finnish Institute for Health and Welfare (THL/5259/6.02.00/2023).

References

Angrist, J. D., & Hull, P. (2023). Instrumental variables methods reconcile intention-to-screen effects across pragmatic cancer screening trials. *Proceedings of the National Academy of Sciences*, *120*(51), e2311556120. https://doi.org/10.1073/pnas.2311556120.

Angrist, J. D., & Imbens, G. W. (1995). Two-Stage Least Squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, *90*(430), 431–442. https://doi.org/10.1080/01621459.1995.10476535

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using Instrumental Variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Bansak, K. (2020). A generalized approach to power analysis for Local Average Treatment Effects. *Statistical Science*, *35*(2), 254–271. https://doi.org/10.1214/19-STS732.

Dahlgren, C., Spånberg, E., Sveréus, S., Dackehag, M., Wändell, P., & Rehnberg, C. (2023). Short-and intermediate-term impact of DTC telemedicine consultations on subsequent healthcare consumption. *The European Journal of Health Economics*. https://doi.org/10.1007/s10198-023-01572-z.

Dorsey, E. R., & Topol, E. J. (2020). Telemedicine 2020 and the next decade. *The Lancet*, *395*(10227), 859. https://doi.org/10.1016/S0140-6736(20)30424-4.

Ellegård, L. M., Kjellsson, G., & Mattisson, L. (2022). An app call a day keeps the patient away? Substitution of online and in-person doctor consultations among young adults. http://hdl.handle.net/2077/68545.

Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, *57*(1), 3–43. https://doi.org/10.1257/jel.20171452.

Haaga, T., Böckerman, P., Kortelainen, M., & Tukiainen, J. (2024a). Does abolishing copayment increase doctor visits: A comparative case study? *The B.E. Journal of Economic Analysis & Policy*, *1*(24), 187–204. https://doi.org/10.1515/bejeap-2023-0056.

Haaga, T., Böckerman, P., Kortelainen, M., & Tukiainen, J. (2024b). Effects of nurse visit copayment on primary care use: Do low-income households pay the price? *Journal of Health Economics*, *94*, 102866. https://doi.org/10.1016/j.jhealeco.2024.102866.

Haaga, T., Herzig, M., Kortelainen, M., Nokso-Koivisto, O., Saxell, T., & Sääksvuori, L. (2024). Digitaalisten terveyspalvelujen käyttö, käyttäjät, tuotanto ja vaikuttavuus: Esiselvitys. *Sosiaali- ja terveysministeriön julkaisuja 2024:12*. http://urn.fi/URN:ISBN:978-952-00-5665-0.

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of Local Average Treatment Effects. *Econometrica*, *62*(2), 467–475. https://doi.org/10.2307/2951620

List, J. A., Muir, I., & Sun, G. (2024). Using machine learning for efficient flexible regression adjustment in economic experiments. *Econometric Reviews*, *44*(1), 2–40. https://doi.org/10.1080/07474938.2024.2373446.