

# Effects of Nurse Visit Copayment on Primary Care Use: Do Low-Income Households Pay the Price?

Tapio Haaga, Petri Böckerman, Mika Kortelainen, and Janne Tukiainen\*

March 2023

## Abstract

Nurses increasingly examine and treat primary care patients, but the evidence on the impacts of cost-sharing for nurse visits is lacking. We employ a staggered difference-in-differences design to examine the effects of adopting a 10-euro copayment for nurse visits on the curative primary care use of Finnish adults, building on our pre-analysis plan. The results show that the moderate copayment reduced nurse visits by 9–12% during a one-year follow-up. We find heterogeneity by income in absolute terms, but not in relative terms. In terms of visits, the estimated decrease is more than twice as large for the bottom 40% of the income distribution compared to the top 40%. The spillover effects on general practitioner (GP) use were negative but small in magnitude, with statistical significance varying depending on the specification. Overall, our results suggest that moderate copayments may create a greater barrier to accessing care for low-income individuals, but their effects on health are inconclusive.

**Keywords:** Cost-sharing, copayments, out-of-pocket costs, utilization, primary care, nurse, general practitioner, difference-in-differences, pre-analysis plan, blind analysis

**JEL codes:** H42, I11, I13, I14, I18

---

\***Haaga:** University of Turku, and Finnish Institute for Health and Welfare (THL) (email: [tapio.haaga@utu.fi](mailto:tapio.haaga@utu.fi)). **Böckerman:** University of Jyväskylä, Labour Institute for Economic Research LABORE, and IZA Institute of Labor Economics (email: [petri.boeckerman@labore.fi](mailto:petri.boeckerman@labore.fi)). **Kortelainen:** University of Turku, InFLAMES Research Flagship Center, VATT Institute for Economic Research, and Helsinki Graduate School of Economics (email: [mika.kortelainen@utu.fi](mailto:mika.kortelainen@utu.fi)). **Tukiainen:** University of Turku, and VATT Institute for Economic Research (email: [janne.tukiainen@utu.fi](mailto:janne.tukiainen@utu.fi)). **Acknowledgements:** We thank Mikko Peltola for support, and Heather Royer, Henri Salokangas, Markku Siikanen, Lauri Sääksvuori, and Jussi Tervola for comments and suggestions. We also thank all seminar participants who have provided comments to this study and our other related projects. This work is supported by THL, Finnish Ministry of Social Affairs and Health, and Yrjö Jahnsson Foundation (research grant No. 20197209). **Replication codes:** <https://github.com/tapiohaa/ASMA3>. **Pre-analysis plan and earlier versions:** <https://osf.io/skuv9/>.

# 1 Introduction

Aging population puts pressure on primary care systems that provide services from health promotion and disease prevention to treatment and rehabilitation as early and as close to the patient as possible. For instance, the U.S. is projected to have a shortage of 18,000–48,000 primary care physicians by 2034 (AAMC, 2021). Task shifting and out-of-pocket costs are two popular policy tools to tackle this challenge. Nurses are increasingly substituted for physicians in examining, diagnosing, and treating patients (McMichael & Markowitz, 2022), playing a critical role in primary care. Out-of-pocket costs generate revenue and reduce the demand for health care services (Einav & Finkelstein, 2018), potentially including nurse visits, and are thus a key tool in the rationing of health care.

Adding a small price to a previously free health care service may have disproportionately large utilization effects (Iizuka & Shigeoka, 2022). The implications are potentially important for accessible entry-level services that serve a gatekeeping purpose, such as primary care nurse visits. Primary care is the first point of contact for the patient in the healthcare system. Any needs-based prioritization by professionals is conditional on patients having contacted the healthcare system in the first place. Cost-sharing ideally reduces wasteful spending if rational and well-informed patients do not seek low-value care. In reality, the patients’ choices can be far from optimal (Chandra et al., 2021). Moreover, out-of-pocket costs arguably constitute a financial barrier to care that disproportionately affects low-income patients, potentially contributing to inequality. Low-income individuals are in a vulnerable position, because they have worse health than the population on average, and the marginal benefit of care may be much larger for them.

We analyze whether adopting a 10-euro copayment for curative primary care nurse visits affects the primary care use of Finnish adults. Our focus is on the heterogeneous effects by income level, as specified in our pre-analysis plan. To this end, we employ a staggered difference-in-differences design and exploit the staggered adoption of the copayment in primary care areas in 2014–2019, as well as comprehensive administrative data. Most Finnish municipalities adopted the copayment at some point in 2014–2019 to collect more revenue, and

the exact timing of the adoption in the treated areas is arguably arbitrary. While there is a lot of previous research on the effects of cost-sharing, mostly focusing on healthcare coverage or copayments for visits to general practitioners (GP), the impact of nurse visit copayments on primary care use has not been studied before.

Regarding the nature of visits under study, nurses carry out triage and book appointments to nurses and GPs in Finland. Acute infectious diseases and chronic conditions, such as diabetes, asthma, or dementia, are common reasons for nurse visits. Nurses work closely with GPs even when providing independent appointments. They consult GPs or book GP appointments if needed. GPs write the vast majority of prescriptions and authorize access to specialized healthcare.

We find that the copayment adoption reduced the number of curative nurse visits by 9–12% during a one-year follow-up. There is statistically significant heterogeneity by income in absolute terms: the estimated decrease in the number of visits is more than two times larger at the bottom 40% of the income distribution than at the top 40%. The effect size increases as income decreases. However, heterogeneity by income level is less clear-cut and statistically insignificant in relative terms (percentage changes). The relevant dimension of heterogeneity is arguably context-specific. Moreover, we examine whether nurse visit copayments have spillover effects on GP use, which is rationed based on gatekeeping. We estimate a 2–5% reduction in GP visits, but our preferred estimates are closer to zero and often insignificant.<sup>1</sup>

Overall, the results suggest that copayments may create a greater barrier to accessing care for low-income individuals. A related question is whether the inequality in accessing care increases inequality in health. We argue that the potential spillover effects on drug prescriptions and referrals to specialists, that proxy doctor-assessed need for diagnosis and treatment (“severity”), were likely close to zero given that we find only small if any effects on GP use.<sup>2</sup> We also conducted exploratory analyses, not included in the pre-analysis plan, to shed light on the question. First, those with a prescription for diabetes or hypertension, whose baseline nurse use is much higher, responded

---

<sup>1</sup>For GP visits, we prefer specifications that allow for a linear pre-trend difference or that require parallel trends only from the last pre-treatment period onwards. See Section 4 for details.

<sup>2</sup>These outcomes are either not included in our data or have serious quality issues.

more strongly in absolute terms to the copayment abolition than those without, but we found no difference in relative terms. Second, we observe statistically insignificant and imprecisely estimated increases in emergency department visits and unplanned hospitalizations for ambulatory care sensitive conditions. We conclude that adopting a 10-euro copayment is likely not large enough to detect meaningful health effects.

The nurse visit copayment was later abolished nationally in July 2021. For this reason, we also examine the impacts of this reform using a DD design, as stated in the pre-analysis plan. We do not obtain any causal conclusions using this reform, because the parallel trends assumption is not credible based on pre-trend patterns during the COVID-19 pandemic. These analyses are reported in the Online Appendix.

Our study relates to several strands of literature. First, it is connected to the large literature on the effects of cost-sharing on healthcare use and health. A seminal contribution to this literature is the RAND Health Insurance Experiment (Newhouse & the Insurance Experiment Group, 1993). Since 2012, when the first results from the Oregon Health Insurance Experiment were published (Finkelstein et al., 2012), studies have for instance examined age-based changes in cost-sharing among the elderly (Shigeoka, 2014) and in insurance status among young adults (Anderson et al., 2014) and the elderly (Chatterji et al., 2022), a switch to a high-deductible insurance plan in a large firm (Brot-Goldberg et al., 2017), major health insurance reforms, as in Massachusetts (Miller, 2012) and Medicaid expansions (Simon et al., 2017), dynamic incentives (Aron-Dine et al., 2015; Cabral, 2017; Einav et al., 2015), and the framing of cost-sharing policies (Iizuka & Shigeoka, 2021). Studies have also examined the impacts of health insurance on mortality (Goldin et al., 2020; Miller et al., 2021) and of childhood Medicaid coverage on later-life mortality (Goodman-Bacon, 2021b) and healthcare use (Wherry et al., 2018).

A considerable share of this literature examines the effects of healthcare coverage (i.e., adjustment at the extensive margin). Our analysis relates more closely to studies that analyze the impacts of copayments in public health insurance systems covering all citizens (i.e., adjustment

at the intensive margin). These studies have evaluated the effects of copayments for GP visits in the Nordic countries among children and adolescents (Haaga et al., 2023; Johansson et al., 2019; Landsem & Magnussen, 2018; Nilsson & Paul, 2018; Olsen & Melberg, 2018), and among Irish children (Walsh et al., 2019) and the elderly (Ma & Nolan, 2017). Han et al. (2020) and Iizuka and Shigeoka (2022) examine the impacts of copayments among small children and adolescents in Taiwan and Japan, while Kruse et al. (2022) study the effects of abolishing a copayment for psychological treatment among Danish adolescents.

We fill several gaps in this literature. Despite the crucial role nurses nowadays play in primary care, we are the first to examine the impacts of copayments for primary care nurse visits. We do so for the adult population by analyzing direct effects on nurse visits and, importantly, also indirect effects on GP use. Moreover, we use a staggered DD design, which we analyze with state-of-the-art methods that are robust to the staggered treatment and heterogeneity of treatment effects. In contrast, most of the studies cited in the paragraph above are either based on regression discontinuity designs or single-event DD designs and focus on children or adolescents. We are also the first to use a comprehensive pre-analysis plan in an observational study on cost-sharing and to conduct pre-specified heterogeneity analysis by income level. Several earlier studies examine treatment effect heterogeneity by income but do so focusing on children or adolescents and not on the adult population (Han et al., 2020; Johansson et al., 2019; Kruse et al., 2022; Nilsson & Paul, 2018).

Second, our study relates to the literature on task shifting from GPs to nurses. By 2022, nurse practitioners (NP) have full practice authority in 31 states in the U.S., allowing them to examine, diagnose, treat, and prescribe medications to the full extent of their training and experience without the supervision of physicians (McMichael & Markowitz, 2022). Studies have examined the impacts of this task shifting on healthcare delivery outcomes and costs (Alexander & Schnell, 2019; Kleiner et al., 2016; Laurant et al., 2018; Traczynski & Udalova, 2018). However, the literature does not consider demand-side factors affecting the use of nurse visits or their role in the rationing of health care services. We contribute by investigating the direct and indirect effects

of nurse visit copayments on primary care use.

Third, we also demonstrate the use of a detailed pre-analysis plan (PAPs) and analysis blinding in a non-experimental economics study that analyzes historical events, with the aim of limiting the concerns for the “garden of forking paths” (Gelman & Loken, 2014; Olken, 2015). There are only a few observational economics studies based on a detailed PAP (Bohm & Lind, 1993; Clemens & Strain, 2021; Korkeamäki & Uusitalo, 2008; Neumark, 2001; Neumark & Yen, 2022) despite replication concerns having been more pronounced in observational studies than in randomized controlled trials (Brodeur et al., 2020; Vivalt, 2019). We randomly assigned areas into placebo policies at the start (observed adoption dates), making the placebo treatment indicator independent of outcomes. Then, we wrote computer codes specifying in detail how we planned to clean, construct, and analyze the data and report results *before* estimating any of the results based on the real treatment assignment. The codes and a corresponding placebo report were registered as our PAP. In the spirit of Banerjee et al. (2020), we took a flexible approach to *ex post* changes and additions, prioritizing transparency and ability to make reasonable changes over puristically following the PAP. We document and discuss the changes and additions we made *ex post* in Section A.4 and denote figures and tables added after the PAP registration by “post-blind”.

Section 2 introduces the institutional background and Section 3 the data. Section 4 presents our empirical approach. Section 5 reports the results for the copayment adoption. Section 6 concludes. Our Online Appendix contains additional (post-blind) supplementary analyses for the copayment adoption, the analyses for the copayment abolition, a description on data construction and a documentation of changes and additions to the PAP.

## 2 Institutional Background

**Publicly-funded primary care in Finland.** Primary care is provided for adults by three sectors: publicly-funded primary care, occupational healthcare, and private clinics. These sectors target different patient populations and differ with respect to gatekeeping, out-of-pocket costs, and

waiting times. Publicly-funded primary care is the main provider for those not entitled to occupational care or who cannot afford the fees for private physicians. These groups include pensioners, the unemployed, and low-income individuals. The employed and more affluent adults usually prefer private clinics or occupational care due to faster access and less gatekeeping. Occupational care is also free of charge at the point of use. In contrast, nurses in publicly-funded primary care do triage on the phone or at health stations and book appointments to both nurses and GPs. A referral is needed to consult a specialist. Copayments are moderate: at maximum 21 euros per GP visit and approximately 10 euros per curative nurse visit. Waiting times vary and may be long for non-urgent care.

Municipalities form publicly-funded primary care areas (health centers) on their own or in cooperation with others. Every citizen has their designated health station determined by the location of residence. In some primary care areas, all health stations may be available on a visit-by-visit basis. Since 2014, citizens have been able to choose their health station once a year, but active choices have been uncommon. Municipal services are financed through state transfers, municipal taxes, copayments, and borrowing. The state guides copayment policies by setting which groups or services are exempted (based on Act on Social and Health Care Client Fees) and maximum copayment levels (in the corresponding Government Decree). Within these constraints, primary care areas set their own policies.

The public system is characterized by limited supply and labor shortages. Cohort sizes in medical schools are fixed, and the public and the private sector compete for doctors. Primary care areas face challenges in hiring nurses at the prevailing wage level determined by collective bargaining. The challenges are reinforced by the fact that central and local governments have been running fiscal deficits for years, which is expected to continue.

**Copayments for nurse visits.** Nurses have a central role in Finnish primary care. Besides their gatekeeping role, they often treat patients with acute infectious diseases or chronic conditions, such as diabetes, asthma, or dementia, working closely with GPs. Finland adopted restricted prescription rights for nurses in 2010. Related to that, the law on copayments was

changed to allow primary care areas to charge a copayment for curative nurse visits, no longer specifying professions (e.g., physicians) whose visits can be subject to copayments. However, the decree continued to mention explicitly only doctor visits. This likely explains why no areas immediately adopted the nurse visit copayment, first introduced in 2014. Many other areas adopted it to collect more revenue once they became aware of the possibility.

The staggered adoption is illustrated in Figure 1. At the end of 2019, half of the population lived in areas charging the copayment with the vast majority of municipalities charging it. Our Online Appendix includes a complementary map (Appendix Figure A1). In 2021, 80% of the municipalities with the copayment charged it for three visits annually, and by far the most common per-visit copayment was 11 euros, the population-weighted mean being 12 euros (Appendix Figure A2). No major area-specific changes have been made to the levels following the adoption except for slight inflation adjustments.

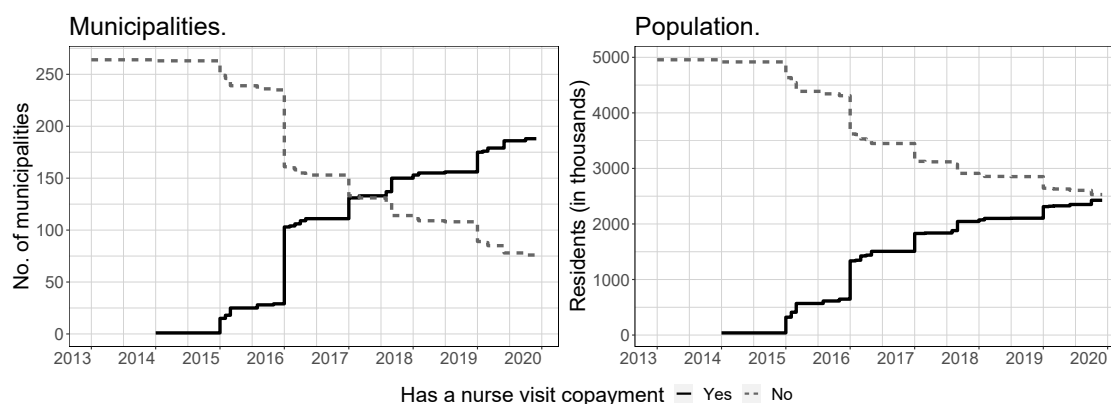


Figure 1: Staggered Adoption of the Nurse Visit Copayment.

*Notes:* We take municipalities in mainland Finland (293 in 2022) and use the 2022 municipal boundaries and population sizes from the end of 2019. The sample contains those municipalities whose policies on copayments for curative nurse visits we observe in our data collection.

Several policies protect financially vulnerable low-income patients from healthcare costs. There are annual out-of-pocket caps for public healthcare services and prescription drugs of 692 euros and 592 euros (in 2022). Households with lowest incomes and only little wealth can apply for social assistance, a means-tested last-resort benefit, which can also cover out-of-pocket costs



for public health care and prescription drugs. The law on copayments requires that, for some public services, financially vulnerable patients can apply for an exemption or a lowered copayment. This right does not apply to nurse visit copayments, but some areas may still exempt individuals based on applications. A few primary care areas provide general exemptions to specific low-income groups, such as those with the lowest national pension or those receiving social assistance.

### 3 Data

We combine several Finnish national administrative registers using person IDs. The data contain contacts in publicly-funded primary care and in hospitals, social assistance recipients, drug prescriptions since 2018, and socioeconomic characteristics of all individuals who have a permanent residence at year's end.<sup>3</sup> We observe age and the municipality of residence which are used to link visits to copayment policies. We construct a variable for equivalized family disposable income and calculate population sizes for each municipality. All data permits are authorized by Findata and Statistics Finland.

We also use publicly available data on each municipality's primary care area in 2021 (from the Association of Finnish Municipalities). Two publicly available registers listing social and healthcare organizations are linked to primary care contacts (THL). We create three tables mapping areas to copayment policies. The first reports whether a given municipality had adopted the nurse visit copayment by the end of 2019 and the possible adoption date. These data were collected from municipal documents, websites, and news in local media. The search was based on the publicly available dataset on nurse and GP visits copayments (from THL), which we also use for GP visit copayments in 2013–2018. The third table reports the copayments in Summer 2021, collected from the websites of primary care areas.<sup>4</sup>

---

<sup>3</sup>The data on primary care use (Register of Primary Health Care Visits), specialized healthcare use (Care Register for Health Care), and social assistance recipients (Register of Social Assistance) are all administered by the Finnish Institute for Health and Welfare. The socioeconomic data comprise Statistics Finland's FOLK modules "basic", "family", and "income". Prescriptions (Kanta Prescription Center) are administered by the Social Insurance Institution of Finland.

<sup>4</sup>We thank Katja Ilmarinen, who had gathered the same information independently, for allowing us to

Our empirical analyses use data from 2013–2019, restricting to pre-pandemic (COVID-19) years. We include those individual-by-year observations in which the person is 25 years or older. The aim is to exclude minors, who are exempted from the copayment, and students, who have access to student healthcare. The primary outcomes are the annualized number of curative nurse and GP visits per capita in publicly-funded primary care, constructed by multiplying monthly visits per capita by 12. The secondary outcomes include the share of individuals receiving social assistance and the annual sum of received basic social assistance, both defined at the family level.<sup>5</sup> Our pre-specified choice is to estimate the effects separately for the bottom 40% and the top 40% of the equivalized family disposable income distribution in all analyses.<sup>6</sup>

We discuss in detail how we clean and construct our analysis data in Section A.3 in the Online Appendix. Ultimately, we have an unbalanced panel at the municipality-by-time-period-by-income-decile-by-outcome level. Time period is a month except for the annually-measured sum of social assistance. The panel is unbalanced because we exclude some observations due to quality problems, mainly for primary care outcomes. When the national data collection started in 2011, not all areas were able to transfer primary care data from their electronic health record (EHR) systems to the national register. Later changes in the EHR systems may also be visible in the data as a sudden but short drop to a near zero value in aggregate contacts. The details of how we detect and exclude observations with data quality concerns are provided in Section A.3.

---

cross-validate our information.

<sup>5</sup>Social assistance is a means-tested last-resort benefit for households.

<sup>6</sup>We focus on two groups for parsimony. Using smaller groups than the bottom 40% and the top 40% has two disadvantages: smaller samples and larger variation, and the fact that the share of social assistance recipients is larger at the bottom of the income distribution, potentially attenuating estimates as the benefit can cover copayments.

## 4 Empirical Approach

**Research design.** We use a staggered difference-in-differences (DD) design with an irreversible treatment. For each event, we have both never-treated and later-treated municipalities as controls.<sup>7</sup> The never-treated areas include the six largest cities and differ from the treated areas. We do not view the decision to adopt the nurse visit copayment as quasi-random. However, the decision on when to adopt conditional on adopting seems much more arbitrary. Our interpretation of municipal decision-making is that there is potential randomness in the timing of when public servants became aware of the possibility to charge the copayment and, consequently, in the treatment timing.<sup>8</sup> We rely on a parallel trends assumption (PTA) for identification instead of assuming (quasi-)random treatment timing among the treated areas. We assume that the outcomes for the treated cohort and for the comparisons (not-yet-treated or never-treated) would have followed parallel trends in the absence of treatment. The PTA may hold in both levels and logs, or in levels but not logs, or *vice versa* (Roth & Sant’Anna, 2023). As in our PAP, we report the main results on primary care use in both levels and logs.<sup>9</sup> Based on pre-treatment trends in event-study plots reported in Section 5, the PTA is plausible in both levels and logs.

We do not have to account for utilization spillovers across areas due to Finland’s publicly-funded primary care system where non-urgent care is provided by a designated health station determined by the location of residence. Neither do we expect noticeable anticipation effects. The implementation time from the political decision is often a month or less, and only a few citizens likely pay much attention to the minutes of the municipal committees.

We model our setting as a single-treatment design in line with our PAP, not controlling for GP visit copayments. With this choice, we likely trade off some bias for lower variance and more external validity compared to using the estimator proposed by de Chaisemartin and D’Haultfoeuille

---

<sup>7</sup>To be specific, copayment policies are set at the primary care area level. However, we analyze the data at the municipality level for practical purposes.

<sup>8</sup>The earlier-treated and later-treated municipalities, defined by median event date, are rather similar in some key features (Table A1, post-blind).

<sup>9</sup>The results in levels are shown first while most of the results in logs are documented in the Online Appendix.

(2022).<sup>10</sup> The state increased the maximum GP visit copayment from 16.10 euros in 2015 to 20.90 euros in 2016. Municipalities responded differently to the increase: many made it instantly in 1/2016, some made it later, and some have not. We find that the nurse visit copayment adoption is correlated with only a 1-euro increase in GP visit copayments based on the methods of our main analysis (Appendix Figure A3, post-blind). Theoretically, the GP visit copayment increases arguably had only small impacts and thus small bias given that the changes were moderate both in euros and percentages to existing (and not new) copayments. Our effect estimates on GP use, that are negative but close to zero, indirectly suggest that the bias (and also the spillover effect) is small. Our underlying assumption is that copayments for nurse and GP visits decrease both nurse and GP use given the gatekeeping system.

We note two potential threats to identification. First, once the copayment is adopted, preventive and curative visits must be distinguished and counted for charging purposes. If this affects how contacts are recorded to the EHR system, the number of recorded visits may change even if the underlying use does not. However, we find no evidence of preventive-labeled visits crowding out curative-labeled visits: adopting the copayment for curative nurse visits is not correlated with the number of preventive nurse visits based on the methods of our main analysis (Appendix Figure A4, post-blind).

The second possible threat is supply reductions to improve fiscal balance, as with the copayments. If a health station is closed or its opening hours are cut, we would expect that GP use decreases as well, potentially as much as nurse visits. This is not what we observe: our effect estimates on GP visits are negative but close to zero and often insignificant. It could be that the cuts are entirely on nurses and nurse visits, but we find this hypothesis unlikely. We are not aware of any anecdotal evidence in support of these kinds of cuts.

**Econometric methods.** We use stacked regressions (Cengiz et al., 2019; Gormley & Matsa, 2011) as our baseline. For robustness checks, we use the Callaway and Sant’Anna (2021)

---

<sup>10</sup>For many-treatment settings, de Chaisemartin and D’Haultfoeuille (2022) propose a DD estimator whose coefficients are not contaminated by the effect of other treatments, which is a potential problem in conventional TWFE regressions. Their estimator controls for other treatments by restricting to a subset of switches in the treatment of interest and a subset of controls whose policy paths fulfill specific requirements.

(CS) estimator. Both estimators are robust to biases in conventional DD two-way fixed effects (TWFE) regression models caused by staggered treatments and treatment effect heterogeneity (Baker et al., 2022). If misspecified, the conventional models project heterogeneous treatment effects onto group and time fixed effects, rather than treatment status (Gardner, 2021). Both the static (Goodman-Bacon, 2021a) and event-study (Sun & Abraham, 2021) TWFE specifications suffer from these biases.

In practice, stacking ensures that earlier-treated units are not used as controls for later-treated units. It transforms the staggered setting into event-specific datasets that are ultimately stacked (or pooled) together before conventional TWFE regressions are fitted. We first create a separate dataset for each event, including the treatment cohort and all clean controls that are unexposed (not-yet-treated) in the window of 24 months before and 12 months after the copayment adoption. We only use data from the 36-month window and include events with at least 12 post-treatment months. Depending on the outcome, we exclude several municipality-year observations due to data quality concerns (see Section A.3). We require balanced panels in event (or relative) time as the baseline. These event-specific datasets are stacked for estimation. Our static TWFE specification includes event-specific unit and time fixed effects:

$$y_{mte} = \alpha_{me} + \gamma_e + \delta^{DD} D_{mt} + \varepsilon_{mte}. \quad (1)$$

Here,  $m$ ,  $t$ , and  $e$  denote municipality, month, and event-specific dataset, and  $\alpha_{me}$  and  $\gamma_e$  represent event-specific municipality and month fixed effects.  $D_{mt}$  is a dummy for post-treatment periods in the treated municipalities. We weight by population due to heterogeneity in municipality size.<sup>11</sup> Standard errors are clustered by municipality.<sup>12</sup>

We use two other stacked specifications modified from Model 1. A dynamic event-study version of Model 1 replaces the term  $\delta^{DD} D_{mt}$  with  $\sum_{l=-24, l \neq -1}^{11} \mu_l D_{mte}^l$  (all leads and lags with  $t = -1$  omitted as a reference), where  $D_{mte}^l$  is a dummy for the treated areas for observations  $l$

<sup>11</sup>We discuss weighting in Section A.3 and present uniformly-weighted robustness checks in Section A.1.

<sup>12</sup>Our stacked data contain multiple copies of the same observation as a municipality can belong to several event-specific datasets. Thus, clustering at the municipality-by-event level is not appropriate.

months from the copayment adoption in the event-specific dataset  $e$  where the area  $m$  is treated. A third version of Model 1 allows for a linear pre-trend difference between the treated and the comparisons, replacing the term  $\delta^{DD}D_{mt}$  with  $\sum_{l=0}^{11}\mu_l D_{mt}^l$  (all lags) and  $\theta d_{me}t_e$  (event-specific differential linear pre-trends). In the latter term,  $d_{me}$  is a dummy for the municipality being treated in that event-specific dataset and  $t_e$  denotes time relative to that event. Here, the PTA concerns deviations from a linear pre-trend difference.

The stacking estimator is efficient as it uses OLS to derive weights on the event-specific DD estimates, trading off bias for efficiency. However, the use of variance weighting may lead to inconsistency for the *sample-average* ATT (Baker et al., 2022). Indeed, Gardner (2021) shows that the estimator identifies an average of event-specific ATTs, weighted by event-specific treatment variance and sample size. For the PAP, we valued the simplicity and implementability of stacking and its ability to accommodate triple difference models for testing treatment effect heterogeneity.

As an alternative to stacking, we use the CS estimator (Callaway & Sant’Anna, 2021). The aim is to identify a group-time average treatment effect, allowing for treatment effect heterogeneity over cohorts and time. The group-time ATTs can be aggregated to construct measures of overall treatment effects. We provide both event-study-type estimates and a static estimate that is the average of all group-time ATTs, weighted by group size.<sup>13</sup> The authors propose several two-step plug-in estimators for group-time ATTs: first estimate nuisance functions and then plug their fitted values into the sample analogue of the group-time ATT. When the never-treated units are the comparisons, the PTA is assumed only from the last pre-treatment period on.<sup>14</sup>

For the CS estimator, we use outcome regression, weight by population, and cluster standard errors by municipality. Events with at least 12 follow-up months are included. The dataset is balanced in calendar time, excluding municipalities with data quality concerns in the study window. Our baseline is to exclude the years 2013 and 2019 when analyzing primary care

<sup>13</sup>Our stacking analyses use a 12-month follow-up. Here, follow-up varies by treatment group. This also implies putting more weight on the earlier-treated cohorts.

<sup>14</sup>Thus, the assumption does not restrict pre-treatment trends. However, the PTA is different and restricts pre-trends when the not-yet-treated are used as comparisons (Callaway & Sant’Anna, 2021).

use to increase the number of sample municipalities.<sup>15</sup> Regarding social assistance use, we use all data from 2013–2019. The data are aggregated to the municipality-by-time-period level for estimation.

For testing treatment effect heterogeneity, we use a triple difference (DDD) model with the stacked data. We compare the evolution of outcomes at the bottom 40% of the income distribution to that at the top 40% in both treatment and comparison areas. The PTA is now assumed in ratios (Olden & Møen, 2022), concerning the relative outcomes of the income groups. We use the following specification:

$$\begin{aligned}
y_{mgte} = & \alpha + \beta_{1e}Treat_{me} + \beta_{2e}Affected_{ge} + \beta_{3e}Post_{te} + \beta_{4e}Treat_{me} \times Affected_{ge} \\
& + \beta_{5e}Treat_{me} \times Post_{te} + \beta_{6e}Affected_{ge} \times Post_{te} \\
& + \gamma Treat_{me} \times Affected_{ge} \times Post_{te} + \epsilon_{mgte}.
\end{aligned} \tag{2}$$

Here,  $m$ ,  $g$ ,  $t$ , and  $e$  denote municipality, socioeconomic group, month, and event-specific dataset.  $Treat_{me}$  and  $Post_{te}$  are dummies for treated municipalities and post-treatment periods, both defined within event-specific datasets.  $Affected$  is a dummy for the bottom 40%, and  $\gamma$  is the coefficient of interest. Other coefficients are event-specific. We again weight by population size and cluster standard errors by municipality.

## 5 Results: Staggered Adoption

**Pre-trend plots.** Figure 2 plots the trends in curative nurse visits for the bottom 40% and the top 40% of the income distribution in treatment and comparison municipalities based on the stacked dataset. We find that nurse visits decreased in the treated municipalities after the adoption of the copayment compared to the comparison municipalities. The decrease was 0.10–0.15 annualized visits at the bottom 40% of the income distribution and approximately 0.05 visits at the top 40%.

---

<sup>15</sup>The exclusion of 2013 trades off one event and 12 months of data for a greater number of municipalities. In many cases, the primary care data quality concerns reported in Section A.3 occurred early in the panel. The exclusion of 2019 leads us to keep one large never-treated municipality that changed its EHR system in Spring 2019 and consequently had data quality concerns.

Nurse visits were increasing in both policy groups before the copayment adoption, and the PTA is arguably plausible. After the adoption, the growth continued in the comparison municipalities, but nurse use decreased in the treated municipalities. The effects on GP visits, in contrast, are small or zero, and there may be a small decreasing pre-trend in GP visits in the treated areas relative to the comparisons (Appendix Figure A5).<sup>16</sup> Neither do we observe any clear effects on receiving social assistance (Appendix Figure A6).

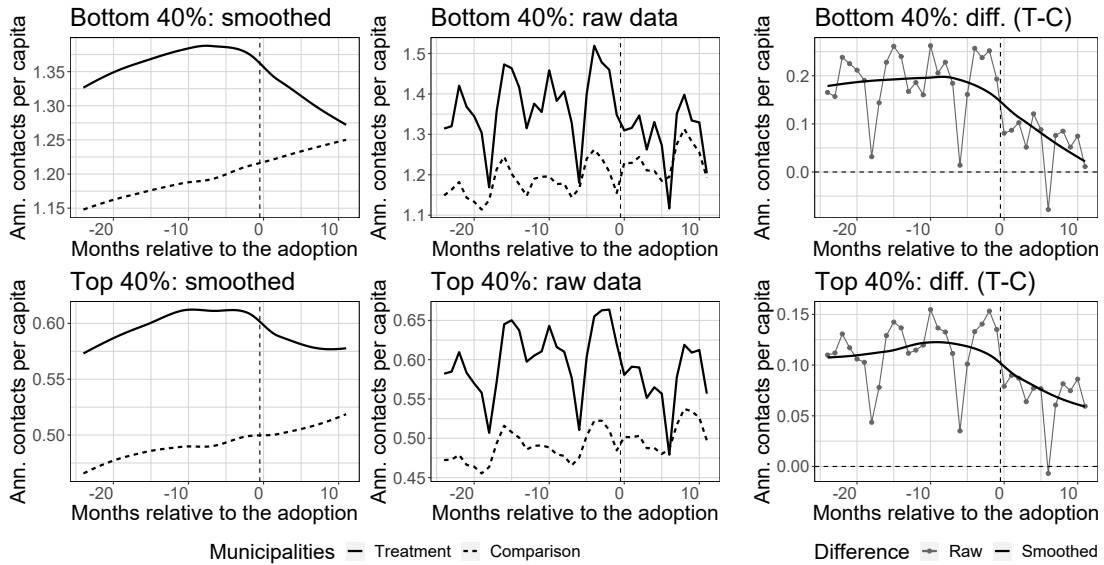


Figure 2: Adoption: Evolution in Nurse Visits.

*Notes:* The dataset is stacked, and event-specific datasets balanced. The outcome is the number of annualized curative nurse visits per capita. Treatment municipalities adopted the nurse visits copayment at time 0 in relative time. The left column contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle column, while the difference between treatment and comparison areas is depicted in the right column. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. The observed reductions in nurse use occurring every twelve months are likely explained by summer holidays and reduced supply in July, which appears to disproportionately affect the treated municipalities that are smaller on average. Figure 1 shows that January was a common adoption month.

We also estimate dynamic event-study regressions using the stacked data (see Section 4), comparing the evolution of outcomes between the treated and unexposed municipalities. The

<sup>16</sup>For GP visits, we consequently prefer the stacked specification that allows for a linear pre-trend difference or the CS estimator with the never-treated units as comparisons that assumes parallel trends only from the last pre-treatment period on. These estimates should be closer to zero than the estimates from our baseline stacked specification, assuming parallel trends in every period.



event-study plots on nurse and GP visits are reported in Figure 3. Consistent with the pre-trend plots, nurse use decreased in the treated municipalities after the copayment adoption compared to the comparison municipalities. The potential effects on GP visits are negative but close to zero. No clear effects on receiving social assistance are observed (Appendix Figure A7). The reduction in nurse visits is larger at the bottom 40% of the income distribution in absolute terms. However, we do not find such a pattern in relative terms using logarithmized outcomes (Appendix Figure A8, post-blind).

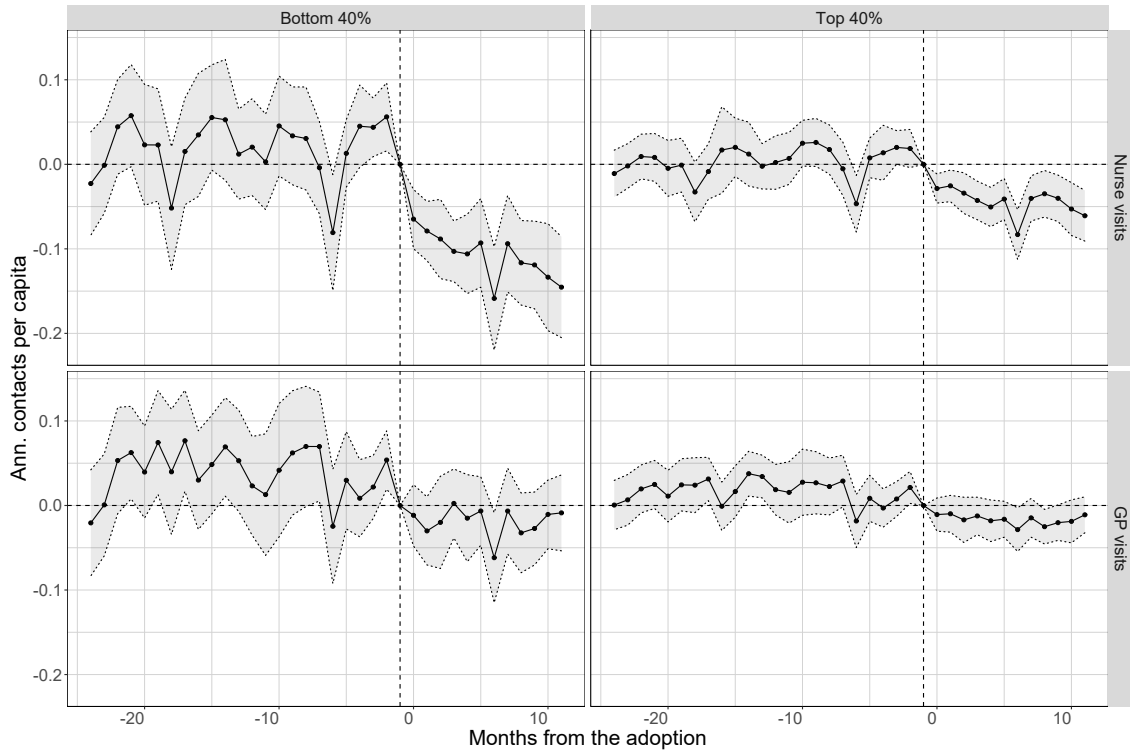


Figure 3: Adoption: Event-Study Plot on Primary Care Use with Stacked Data.

*Notes:* The point estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The dataset is stacked, and event-specific datasets balanced. We use the dynamic event-study variant of Model 1, comparing the evolution of annualized contacts per capita between treated and unexposed municipalities. Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income.

**Main results.** We construct static treatment effect estimates by fitting Model 1 to the stacked data. The results on annualized primary care contacts per capita are reported in Table 1.

Both nurse visits and GP visits decrease at both the bottom 40% and the top 40% of the income distribution, and these reductions are statistically significant. Annualized nurse visits decrease by  $-0.13$  visits ( $-9.3\%$ ) at the bottom 40% and by  $-0.05$  visits ( $-8.0\%$ ) at the top 40%. The reductions in GP visits are closer to zero:  $-0.06$  visits ( $-3.9\%$ ) at the bottom 40% and  $-0.03$  ( $-4.7\%$ ) at the top 40%. The estimates are clearly heterogeneous in absolute terms, and the lower end of the income distribution responds more. However, such a pattern is not found in relative terms. The estimates on social assistance outcomes are insignificant: the estimate on the share of the population receiving social assistance is close to zero, while the estimate on the annual sum of received social assistance is positive and marginally insignificant (Appendix Table A2). However, the inclusion of a linear pre-trend difference attenuates both estimates.

Table 1: Adoption: DD Comparisons, Primary Care Use.

Metric	Nurse visits		GP visits	
	Bottom 40%	Top 40%	Bottom 40%	Top 40%
Level	1.373	0.603	1.450	0.705
Estimate	$-0.127$	$-0.048$	$-0.056$	$-0.033$
Std. error	0.032	0.018	0.021	0.010
P-value	0.000	0.008	0.007	0.001
Change (%)	$-9.252$	$-8.020$	$-3.879$	$-4.669$
Estimate (trends)	$-0.131$	$-0.057$	$-0.048$	$-0.026$
Change (%) (trends)	$-9.565$	$-9.490$	$-3.298$	$-3.706$
Events	17	17	17	17
Treated areas	152	152	152	152
All areas	245	245	245	245

*Notes:* The dataset is stacked and balanced. We use 1) Model 1 and 2) its modified version allowing for differential linear pre-trend (“trends”). Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively.

**Robustness checks.** The stacking estimates for nurse visits are not sensitive but attenuate for GP visits and social assistance outcomes when using a modified specification that allows for

linear pre-trend differences (Table 1 and Appendix Table A2).<sup>17</sup> The stacking results on primary care use are robust to logarithmized outcomes and to unbalanced event-specific datasets that have more municipalities and observations than the balanced datasets in the main analysis (Appendix Table A3).

Next, we report the results constructed using the CS estimator (see Section 4). The findings on nurse visits are similar in both stacked and CS event-study plots. Regarding GP visits, the CS estimates show virtually no effect, while the stacked estimates suggest a small decrease (Appendix Figure A9 and Appendix Figure A10, post-blind). The CS event-study plots point to a decrease in the share of individuals receiving social assistance (Appendix Figure A11), but this finding is not supported by the stacked results.

Our static CS effect estimate averages over all group-time ATEs, weighted by group size. We estimate the effects on primary care use in eight cases, using either never-treated or not-yet-treated municipalities as the comparison group and exclude data from either 2013 or 2019 or both. The CS estimates differ from the stacked estimates in two ways (Appendix Figure A12 and Appendix Figure A13). First, the effects on nurse visits are larger. Annualized nurse use decreases by  $-0.18$  to  $-0.23$  visits ( $-13\%$  to  $-17\%$ ) at the bottom 40% and by  $-0.08$  to  $-0.10$  visits ( $-13\%$  to  $-16\%$ ) at the top 40%. Different estimands and accumulating effects plausibly explain this. The stacked results are based on a one-year follow-up, while the CS estimand averages over all group-time ATEs, having for most units a longer follow-up. Second, the estimates on GP visits are all negative but close to zero and insignificant. The results on the social assistance outcomes (insignificant) are in Appendix Figure A14.

**Effect heterogeneity by income level.** We apply a triple difference (DDD) model (Model 2) using the stacked data to test treatment effect heterogeneity by income, comparing the evolution of outcomes at the bottom 40% of the income distribution to that at the top 40% both in the treatment and comparison areas. The results on primary care use are presented in Table 2. In absolute terms, nurse use decreases by  $-0.07$  to  $-0.08$  annualized visits ( $-5.3\%$  to  $-5.6\%$ )

---

<sup>17</sup>Considering all our results, the number of general practitioner (GP) visits is reduced by 2–5%. The estimates accounting for a linear pre-trend difference are closer to zero in that interval.

at the bottom 40% relative to the top 40%, and the estimates are significant. GP visits appear to decrease more at the bottom 40%, but the estimates are insignificant. In relative terms (logs), all the estimates are insignificant.

Table 2: Adoption: DDD Comparisons, Primary Care Use.

Metric	Balanced datasets		Unbalanced datasets	
	Nurse Visits	GP Visits	Nurse Visits	GP Visits
<b>A. Annualized contacts per capita</b>				
Level	1.373	1.450	1.386	1.444
Estimate	−0.073	−0.025	−0.078	−0.021
Std. error	0.017	0.013	0.017	0.013
P-value	0.000	0.058	0.000	0.115
Change (%)	−5.304	−1.718	−5.635	−1.462
Events	17	17	19	19
Treated areas	152	152	175	175
All areas	245	245	264	264
<b>B. Logarithmized annualized contacts per capita</b>				
Estimate	−1.544	0.711	−1.575	0.818
Std. error	1.559	0.903	1.523	0.925
P-value	0.322	0.431	0.301	0.377
Events	17	17	19	19
Treated areas	126	135	175	175
All areas	209	225	264	264

*Notes:* The dataset is stacked. We use Model 2. Estimates and standard errors are multiplied by 100 if the outcome is the logarithm of annualized contacts per capita. Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality.

Complementarily, we plot the effects on primary care use by income decile in Figure 4 (post-blind). We use stacking with balanced event-specific datasets to obtain the results. The pattern is clear in absolute terms: the estimate attenuates as income increases. The bottom 10% is an exception. A plausible partial explanation is that the share of social assistance recipients is

decreasing in income. About 37% (80%) of social assistance recipients are at the bottom 10% (30%) of the income distribution (Appendix Figure A15, post-blind). Social assistance can cover copayments for these individuals, potentially attenuating the effects. Apart from the first decile, there may be heterogeneity also in relative terms: the lower end of the income distribution is somewhat more sensitive than the top.

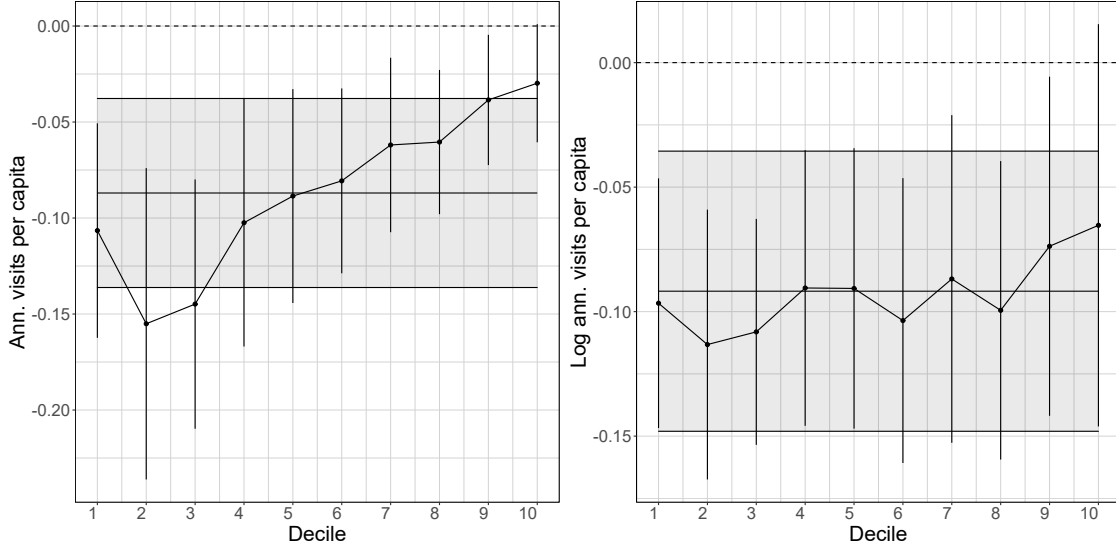


Figure 4: Adoption: Estimates on Nurse Visits by Income Decile.

*Notes:* This figure was not pre-registered and is post-blind. The dataset is stacked and balanced. We use Model 1. Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. We use the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse visits and its logarithm. The gray block, centered at the black horizontal line, shows the ATT estimate for the whole population and its confidence interval.

**Estimates for all individuals.** We provide the results for the entire sample population in Online Appendix Section A.1 (post-blind). Annualized nurse use decreases by  $-0.09$  to  $-0.10$  visits ( $-9\%$  to  $-10\%$ ) using stacking (one-year follow-up) and by  $-0.13$  to  $-0.16$  visits ( $-13\%$  to  $-17\%$ ) using the CS estimator (a longer follow-up). Stacked estimates on GP visits show a decrease ( $-3\%$  to  $-5\%$ ), but the corresponding CS estimates are closer to zero and insignificant.

**Potential health effects.** We first highlight that the spillover effects on drug prescriptions and referrals to specialists, that proxy the doctor-assessed need for diagnosis and

treatment (“severity”), are likely close to zero given that we find only small if any effects on GP use. These outcomes are either not included in our data or have serious quality issues. We also examine the heterogeneity of the effects with respect to having received a drug prescription in 2018–2019 with an ATC code referring to diabetes or hypertension (A10, C02–C03, and C07–C09), proxying a diagnosis of these conditions. Those with a prescription for diabetes or hypertension responded more strongly in absolute terms to the copayment abolition, but we found no difference in relative terms (Appendix Table A4, post-blind). Online Appendix Section A.1 presents the results on emergency department (ED) visits and unplanned hospitalizations for ambulatory care sensitive conditions (ACSC), both outcomes being post-blind. The effects – more exploratory than confirmatory in nature – are positive but imprecisely estimated and statistically insignificant for both outcomes. We conclude that the change in copayment is likely not large enough to detect statistically significant and quantitatively meaningful health effects.

**Additional Supplementary Analyses.** Online Appendix Section A.1 adds and discusses more (post-blind) analyses. The effect estimates for nurse visits are larger when a longer follow-up is used, and the main findings are qualitatively robust to weighting municipalities uniformly. We also assume quasi-random treatment timing among the later-treated municipalities, excluding the never-treated, and use the estimator proposed by Roth and Sant’Anna (2022). On average, the nurse visit copayment reduces nurse visits but has no effect on GP use, mostly in line with our main results. However, there is a caveat to our main heterogeneity findings: the differences in the effects in absolute terms are not that large nor clear using the RS estimator. Thus, the relative effects appear even larger for the top 40%.

## 6 Conclusion

We analyze the effects of a staggered adoption of a nurse visit copayment (approximately 10 euros) on public primary care use of Finnish adults. The copayment adoption reduced curative nurse visits by 9–12% during a one-year follow-up. There is statistically significant heterogeneity by income

in absolute terms: the decrease in the number of visits is more than two times larger at the bottom 40% of the income distribution than at the top 40%. However, such heterogeneity is much less clear-cut and statistically insignificant in relative terms (percentage changes).<sup>18</sup> The estimates for GP visits are negative but close to zero and often insignificant (from  $-2\%$  to  $-5\%$ ).

For comparison, we convert our estimates for all individuals to the semi-arc elasticity of the number of nurse visits with respect to price, as in Brot-Goldberg et al. (2017). These elasticities represent changes in quantities, normalized by the baseline, divided by the price change:  $\frac{(q_1 - q_0)/(q_1 + q_0)}{(p_1 - p_0)/2}$ . Here,  $q$  denotes the number of nurse visits and  $p$  is their “price”. Following Nilsson and Paul (2018), we define the price as the share of out-of-pocket costs of the total cost of the visit. The elasticity is sensitive to the chosen parameters, so we provide two estimates: a baseline and a large estimate. The parameter values are listed in Section A.5. The baseline is  $-0.41$  and the large estimate  $-1.24$ . Nilsson and Paul (2018) report a semi-arc elasticity of  $-0.88$  at the 20th birthday and  $-0.55$  at the 7th birthday when individuals face copayments of 10–15 euros for outpatient doctor visits in Sweden. Also in Sweden, the effect of a 10-euro copayment for GP visits at the 20th birthday map to an elasticity of  $-1.11$  (Johansson et al., 2019).<sup>19</sup> All these estimates are lower than the elasticities of  $-2.11$  and  $-2.26$  that Brot-Goldberg et al. (2017) calculate based on the RAND Health Insurance Experiment.

Regarding the fiscal impacts, our back-of-the-envelope calculations compare estimated copayment revenue to estimated savings from lower primary care use. Assuming 4 million sample individuals, a 10-euro copayment, that only 70% of the copayment revenue are collected<sup>20</sup>, a pre-treatment mean of 1.000 annualized nurse visits, an effect of  $-0.089$  annualized visits<sup>21</sup>, and an average cost of 35 euros per nurse visit (Mäklin & Kokko, 2020), annual copayment revenue

---

<sup>18</sup>However, there is a caveat to these heterogeneity results. In a robustness check using the Roth and Sant’Anna (2022) estimator and later-treated areas as controls, added after registering our pre-analysis plan, the differences in effects in absolute terms are not clear. Thus, the relative effects appear even larger for the top 40%. However, the results from other robustness checks are in line with the finding of treatment effect heterogeneity by income in absolute terms.

<sup>19</sup>We use the estimates for all individuals from their Table 1 and use a copayment of SEK 100 and the total cost of SEK 1500 per visit.

<sup>20</sup>This share is an *ad hoc* choice to account for collection costs, increased social assistance use, and the fact that the copayment was often charged for the first three visits annually.

<sup>21</sup>These are an average over two specifications of Table A5.

would be 26 million compared to saving 12 million due to lower nurse use. If the utilization effect is larger, say  $-15\%$  (Figure A19), the copayment revenue would be 24 million and the savings 21 million. If both the collection costs and production costs were higher than the values used above, then the fiscal importance of the short-term savings from reduced service use is prominent relative to the collected revenue. Notably, this conclusion does not account for the potentially very important health effects.

The reduction in nurse use in terms of visits was largest at the lower end of the income distribution in our pre-specified main analyses. This finding is consistent with Johansson et al. (2019) and Nilsson and Paul (2018). In contrast to these studies, we observe heterogeneity only in levels but not in relative terms (effects compared to baseline use). The relevant dimension of heterogeneity is arguably context-specific. For instance, it may be appropriate to use equal elasticities (defined in relative terms) for both low-income and high-income individuals in microsimulation models used in *ex ante* evaluations of cost-sharing changes. At the same time, the observed heterogeneity in terms of visits is arguably an important finding for a politician placing a large weight on equality and viewing cost-sharing as an instrument to affect the allocation of public resources. Income-related heterogeneity can be especially important as low-income individuals have on average much worse health and may thus have a higher marginal benefit of health care. The exact severity of the inequality as well as the potential health effects warrant further research.



**CRedit author statement:** **Haaga:** Conceptualization, Formal analysis, Writing - Original Draft, Writing - Review & Editing. **Böckerman:** Conceptualization, Writing - Review & Editing, Supervision. **Kortelainen:** Conceptualization, Writing - Review & Editing, Supervision. **Tukiainen:** Conceptualization, Writing - Review & Editing, Supervision.

## References

- AAMC. (2021). The complexities of physician supply and demand: Projections from 2019 to 2034. Prepared for the AAMC by IHS Markit Ltd. <https://www.aamc.org/news-insights/press-releases/aamc-report-reinforces-mounting-physician-shortage>
- Alexander, D., & Schnell, M. (2019). Just what the nurse practitioner ordered: Independent prescriptive authority and population mental health. *Journal of Health Economics*, 66, 145–162. <https://doi.org/10.1016/j.jhealeco.2019.04.004>
- Anderson, M. L., Dobkin, C., & Gross, T. (2014). The effect of health insurance on emergency department visits: Evidence from an age-based eligibility threshold. *The Review of Economics and Statistics*, 96, 189–195. [https://doi.org/10.1162/REST\\_a\\_00378](https://doi.org/10.1162/REST_a_00378)
- Aron-Dine, A., Einav, L., Finkelstein, A., & Cullen, M. (2015). Moral hazard in health insurance: Do dynamic incentives matter? *The Review of Economics and Statistics*, 97, 725–741. [https://doi.org/10.1162/REST\\_a\\_00518](https://doi.org/10.1162/REST_a_00518)
- Baker, A. C., Larcker, D. F., & Wang, C. C. Y. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144, 370–395. <https://doi.org/10.1016/j.jfineco.2022.01.004>
- Banerjee, A., Duflo, E., Finkelstein, A., Katz, L., Olken, B., & Sautmann, A. (2020). In praise of moderation: Suggestions for the scope and use of pre-analysis plans for RCTs in economics. NBER Working Paper No. 26993. <https://doi.org/10.3386/w26993>
- Bohm, P., & Lind, H. (1993). Policy evaluation quality: A quasi-experimental study of regional employment subsidies in Sweden. *Regional Science and Urban Economics*, 23, 51–65. [https://doi.org/10.1016/0166-0462\(93\)90028-D](https://doi.org/10.1016/0166-0462(93)90028-D)
- Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, 110, 3634–3660. <https://doi.org/10.1257/aer.20190687>

- Brot-Goldberg, Z. C., Chandra, A., Handel, B. R., & Kolstad, J. T. (2017). What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics. *The Quarterly Journal of Economics*, 132, 1261–1318. <https://doi.org/10.1093/qje/qjx013>
- Cabral, M. (2017). Claim timing and ex post adverse selection. *The Review of Economic Studies*, 84, 1–44. <https://doi.org/10.1093/restud/rdw022>
- Callaway, B., & Sant’Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225, 200–230. <https://doi.org/10.1016/j.jeconom.2020.12.001>
- Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134, 1405–1454. <https://doi.org/10.1093/qje/qjz014>
- Chandra, A., Flack, E., & Obermeyer, Z. (2021). The health costs of cost-sharing. NBER Working Paper No. 28439. <https://doi.org/10.3386/w28439>
- Chatterji, P., Nguyen, T., & Yörük, B. K. (2022). The effects of Medicare on health-care utilization and spending among the elderly. *American Journal of Health Economics*, 8, 151–180. <https://doi.org/10.1086/716544>
- Clemens, J., & Strain, M. (2021). The heterogeneous effects of large and small minimum wage changes: Evidence over the short and medium run using a pre-analysis plan. NBER Working Paper No. 29264. <https://doi.org/10.3386/w29264>
- de Chaisemartin, C., & D’Haultfoeulle, X. (2022). Two-way fixed effects and differences-in-differences estimators with several treatments. NBER Working Paper No. 30564. <https://doi.org/10.3386/w30564>
- Einav, L., & Finkelstein, A. (2018). Moral hazard in health insurance: What we know and how we know it. *Journal of the European Economic Association*, 16, 957–982. <https://doi.org/10.1093/JEEA/JVY017>
- Einav, L., Finkelstein, A., & Schrimpf, P. (2015). The response of drug expenditure to nonlinear contract design: Evidence from Medicare Part D. *The Quarterly Journal of Economics*, 130, 841–899. <https://doi.org/10.1093/qje/qjv005>

- Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., & Baicker, K. (2012). The Oregon Health Insurance Experiment: Evidence from the first year. *The Quarterly Journal of Economics*, 127, 1057–1106. <https://doi.org/10.1093/qje/qjs020>
- Gardner, J. (2021). Two-stage differences in differences.
- Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, 102, 460–465. <https://doi.org/10.1511/2014.111.460>
- Goldin, J., Lurie, I. Z., & McCubbin, J. (2020). Health insurance and mortality: Experimental evidence from taxpayer outreach. *The Quarterly Journal of Economics*, 136, 1–49. <https://doi.org/10.1093/qje/qjaa029>
- Goodman-Bacon, A. (2021a). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*. <https://doi.org/10.1016/j.jeconom.2021.03.014>
- Goodman-Bacon, A. (2021b). The long-run effects of childhood insurance coverage: Medicaid implementation, adult health, and labor market outcomes. *American Economic Review*, 111, 2550–2593. <https://doi.org/10.1257/aer.20171671>
- Gormley, T. A., & Matsa, D. A. (2011). Growing out of trouble? Corporate responses to liability risk. *The Review of Financial Studies*, 24, 2781–2821. <https://doi.org/10.1093/rfs/hhr011>
- Haaga, T., Böckerman, P., Kortelainen, M., & Tukiainen, J. (2023). Do adolescents from low-income families respond more to cost-sharing in primary care? Version 2. <https://osf.io/vmuzf/>
- Han, H.-W., Lien, H.-M., & Yang, T.-T. (2020). Patient cost-sharing and healthcare utilization in early childhood: Evidence from a regression discontinuity design. *American Economic Journal: Economic Policy*, 12, 238–278. <https://doi.org/10.1257/pol.20170009>
- Iizuka, T., & Shigeoka, H. (2021). Asymmetric demand response when prices increase and decrease: The case of child healthcare. *The Review of Economics and Statistics*. [https://doi.org/10.1162/rest\\_a\\_01110](https://doi.org/10.1162/rest_a_01110)

- Iizuka, T., & Shigeoka, H. (2022). Is zero a special price? Evidence from child health care. *American Economic Journal: Applied Economics*, 14, 381–410. <https://doi.org/10.1257/app.20210184>
- Johansson, N., Jakobsson, N., & Svensson, M. (2019). Effects of primary care cost-sharing among young adults: Varying impact across income groups and gender. *The European Journal of Health Economics*, 20, 1271–1280. <https://doi.org/10.1007/s10198-019-01095-6>
- Kleiner, M. M., Marier, A., Park, K. W., & Wing, C. (2016). Relaxing occupational licensing requirements: Analyzing wages and prices for a medical service. *The Journal of Law and Economics*, 59, 261–291. <https://doi.org/10.1086/688093>
- Korkeamäki, O., & Uusitalo, R. (2008). Employment and wage effects of a payroll-tax cut—evidence from a regional experiment. *International Tax and Public Finance*, 16, 753. <https://doi.org/10.1007/s10797-008-9088-6>
- Kruse, M., Olsen, K. R., & Skovsgaard, C. V. (2022). Co-payment and adolescents’ use of psychologist treatment: Spill over effects on mental health care and on suicide attempts. *Health Economics*, 31, 92–114. <https://doi.org/10.1002/hec.4582>
- Landsem, M. M., & Magnussen, J. (2018). The effect of copayments on the utilization of the GP service in Norway. *Social Science & Medicine*, 205, 99–106. <https://doi.org/10.1016/j.socscimed.2018.03.034>
- Laurant, M., van der Biezen, M., Wijers, N., Watananirun, K., Kontopantelis, E., & Vught, A. V. (2018). Nurses as substitutes for doctors in primary care. *Cochrane Database of Systematic Reviews*. <https://doi.org/10.1002/14651858.CD001271.pub3>
- Ma, Y., & Nolan, A. (2017). Public healthcare entitlements and healthcare utilisation among the older population in Ireland. *Health Economics*, 26, 1412–1428. <https://doi.org/10.1002/hec.3429>
- Mäklin, S., & Kokko, P. (2020). Terveysten- ja sosiaalihuollon yksikkökustannukset Suomessa vuonna 2017. <https://urn.fi/URN:ISBN:978-952-343-493-6>

- McMichael, B. J., & Markowitz, S. (2022). Toward a uniform classification of nurse practitioner scope of practice laws. NBER Working Paper No. 28192. <https://doi.org/10.3386/w28192>
- Miller, S. (2012). The impact of the Massachusetts health care reform on health care use among children. *American Economic Review*, 102, 502–507. <https://doi.org/10.1257/aer.102.3.502>
- Miller, S., Johnson, N., & Wherry, L. R. (2021). Medicaid and mortality: New evidence from linked survey and administrative data. *The Quarterly Journal of Economics*, 136, 1783–1829. <https://doi.org/10.1093/qje/qjab004>
- Moynihan, R., Sanders, S., Michaleff, Z. A., Scott, A. M., Clark, J., To, E. J., Jones, M., Kitchener, E., Fox, M., Johansson, M., Lang, E., Duggan, A., Scott, I., & Albarqouni, L. (2021). Impact of COVID-19 pandemic on utilisation of healthcare services: A systematic review. *BMJ Open*, 11, e045343. <https://doi.org/10.1136/bmjopen-2020-045343>
- Neumark, D. (2001). The employment effects of minimum wages: Evidence from a prespecified research design the employment effects of minimum wages. *Industrial Relations: A Journal of Economy and Society*, 40, 121–144. <https://doi.org/10.1111/0019-8676.00199>
- Neumark, D., & Yen, M. (2022). Effects of recent minimum wage policies in California and nationwide: Results from a pre-specified analysis plan. *Industrial Relations: A Journal of Economy and Society*, 61, 228–255. <https://doi.org/10.1111/irel.12297>
- Newhouse, J. P., & the Insurance Experiment Group. (1993). Free for all? Lessons from the RAND Health Insurance Experiment. <https://doi.org/10.7249/CB199>
- Nilsson, A., & Paul, A. (2018). Patient cost-sharing, socioeconomic status, and children’s health care utilization. *Journal of Health Economics*, 59, 109–124. <https://doi.org/10.1016/j.jhealeco.2018.03.006>
- Olden, A., & Møen, J. (2022). The triple difference estimator. *The Econometrics Journal*. <https://doi.org/10.1093/ectj/utac010>
- Olken, B. A. (2015). Promises and perils of pre-analysis plans. *Journal of Economic Perspectives*, 29, 61–80. <https://doi.org/10.1257/jep.29.3.61>

- Olsen, C. B., & Melberg, H. O. (2018). Did adolescents in Norway respond to the elimination of copayments for general practitioner services? *Health Economics*, 27, 1120–1130. <https://doi.org/10.1002/hec.3660>
- Roth, J., & Sant’Anna, P. H. C. (2022). Efficient estimation for staggered rollout designs. <https://doi.org/10.48550/ARXIV.2102.01291>
- Roth, J., & Sant’Anna, P. H. (2023). When is parallel trends sensitive to functional form? *Econometrica* (forthcoming).
- Shigeoka, H. (2014). The effect of patient cost sharing on utilization, health, and risk protection. *American Economic Review*, 104, 2152–2184. <https://doi.org/10.1257/aer.104.7.2152>
- Simon, K., Soni, A., & Cawley, J. (2017). The impact of health insurance on preventive care and health behaviors: Evidence from the first two years of the ACA Medicaid expansions. *Journal of Policy Analysis and Management*, 36, 390–417. <https://doi.org/10.1002/pam.21972>
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225, 175–199. <https://doi.org/10.1016/j.jeconom.2020.09.006>
- Traczynski, J., & Udalova, V. (2018). Nurse practitioner independence, health care utilization, and health outcomes. *Journal of Health Economics*, 58, 90–109. <https://doi.org/10.1016/j.jhealeco.2018.01.001>
- Vivalt, E. (2019). Specification searching and significance inflation across time, methods and disciplines. *Oxford Bulletin of Economics and Statistics*, 81, 797–816. <https://doi.org/10.1111/obes.12289>
- Walsh, B., Nolan, A., Brick, A., & Keegan, C. (2019). Did the expansion of free GP care impact demand for emergency department attendances? A difference-in-differences analysis. *Social Science & Medicine*, 222, 101–111. <https://doi.org/10.1016/j.socscimed.2018.12.029>

Wherry, L. R., Miller, S., Kaestner, R., & Meyer, B. D. (2018). Childhood Medicaid coverage and later-life health care utilization. *The Review of Economics and Statistics*, 100, 287–302.  
[https://doi.org/10.1162/REST\\_a\\_00677](https://doi.org/10.1162/REST_a_00677)



## A Online Appendix

### A.1 Supplementary Results: Staggered Adoption

**Estimates for all individuals.** We provide the results also for the entire sample population: pre-trend plots (Figure A16, post-blind), dynamic event-study plots using stacking (Figure A17, post-blind) and the CS estimator (Figure A18, post-blind), and aggregated static stacking (Table A5, post-blind) and CS (Figure A19, post-blind) estimates. Annualized nurse use decreases by  $-0.09$  to  $-0.10$  visits ( $-9\%$  to  $-10\%$ ) using stacking (one-year follow-up) and by  $-0.13$  to  $-0.16$  visits ( $-13\%$  to  $-17\%$ ) using the CS estimator (a longer follow-up). Stacked estimates on GP visits show a decrease ( $-3\%$  to  $-5\%$ ), but the corresponding CS estimates are closer to zero and insignificant.

**Follow-up length, and cohort-specific effects.** To examine how the estimates depend on follow-up length, Figure A20 (post-blind) shows the CS event-study plots on the number of annualized primary care visits per capita for all observable durations. The effect size for nurse visits appears to grow steadily for two years before attenuating.<sup>22</sup> Motivated by this finding, Table A6 (post-blind) shows the stacked results using a 24-month follow-up. Compared to main results (Table 1), the effect sizes increase slightly for nurse visits and decrease for GP visits, and more than that with the specification allowing for a linear pre-trend difference. Finally, Figure A21 (post-blind) shows CS estimates by treatment cohort.

**Weighting municipalities uniformly.** We repeated the adoption analyses, but instead of population weighting, we uniformly weighted municipalities when using the CS estimator and municipality-by-income-decile observations when using stacked TWFE regressions. The pre-treatment healthcare use is now higher as small municipalities have a larger weight. Qualitatively, the main findings are robust to the weighting scheme: the copayment adoption is associated with a reduction in nurse use in the treated areas. The estimates on GP visits are negative but small, and the estimates on social assistance outcomes are inconclusive. The

---

<sup>22</sup>Note that the composition of areas changes in relative time as the sample size decreases and uncertainty increases the further we are from the event.

uniformly-weighted stacking results on annualized primary care contacts per capita are reported in Table A7 (post-blind): nurse visits decrease by  $-0.17$  visits ( $-11.2\%$ ) in the bottom 40% and by  $-0.07$  visits ( $-9.4\%$ ) in the top 40%. In contrast, the population-weighted estimates in Table 1 show smaller decreases:  $-0.13$  visits ( $-9.3\%$ ) in the bottom 40% and  $-0.05$  visits ( $-8.0\%$ ) in the top 40%. The rest of the figures and tables are provided in the replication codes folder.

**The Roth and Sant’Anna (2022) estimator (RS), and the later-treated as comparisons.** We use not-yet-treated municipalities as comparisons in the main stacked analysis, and causal inference relies on the parallel trends assumption. However, if only later-treated municipalities are the comparisons, we consider as plausible a stronger assumption of quasi-random treatment timing. This assumption would allow us to use the estimator proposed by Roth and Sant’Anna (2022). The key benefit of the RS estimator is that it can produce considerably more precise estimates in a staggered setting than the earlier methods (e.g., the CS DD estimator).

As a post-blind complementary analysis, we use later-treated municipalities as comparisons and estimate the effects using the RS estimator with the *staggered* R package. We continue to include events that have at least 12 follow-up months. We use a balanced panel in calendar time and drop treatment cohorts that have only a single treated municipality, as required by the software. Following the assumption of quasi-random treatment timing, municipalities are weighted equally and not by population. We use the refined variance estimator proposed by Roth and Sant’Anna (2022) for standard errors.<sup>23</sup> As with the CS estimator, we use 2013 or 2014 as the panel start year and 2018 or 2019 as the end year. Given that we restrict to balanced panels (in calendar time), shorter study windows tend to have more treatment cohorts to study but less monthly observations for a given included municipality.

Three estimands proposed by Roth and Sant’Anna (2022) are considered. The key building block is  $ATE(t, g)$ , defined as the average treatment effect on the outcome in period  $t$

---

<sup>23</sup>In 12 cases out of 96, the software produces a warning that a more conservative Neyman-style variance estimate is less than an estimated adjustment factor. In these cases, we consequently use the unrefined Neyman-style variance estimator discussed in Roth and Sant’Anna (2022).

of being first-treated in period  $g$  relative to not being treated at all. Our first estimand, “simple”, is a simple average of the  $ATE(t, g)$  weighted by cohort size. Our second estimate, “cohort”, first averages the  $ATE(t, g)$  for each cohort  $g$  before averaging these cohort effects weighting by cohort size. Finally, “calendar” first averages the  $ATE(t, g)$  for each time period  $t$  weighting by cohort size before taking a mean of the calendar effects.

Table A8 (post-blind) contains the RS results on annualized primary care visits per capita. Our main finding of the nurse visit copayment reducing nurse use is qualitatively robust, but the pattern of heterogeneity by income level differs from the pre-specified analyses. The estimates show statistically significant reductions of  $-0.12$  to  $-0.17$  annualized nurse visits ( $-8\%$  to  $-11\%$ ) for the bottom 40% of the income distribution and of  $-0.09$  to  $-0.14$  visits ( $-12\%$  to  $-18\%$ ) for the top 40%. There is no longer clear heterogeneity in absolute terms (levels) by income level. Consequently, the effects in relative terms appear even larger for the higher end of the income distribution than for low-income individuals. With respect to GP visits, the estimates vary around zero and are mostly insignificant. The effects show changes of  $+0.07$  to  $-0.06$  annualized GP visits ( $+5\%$  to  $-4\%$ ) for the bottom 40% of the income distribution and of  $+0.02$  to  $-0.03$  visits ( $+2\%$  to  $-4\%$ ) for the top 40%. Table A9 (post-blind) reports the same results but using a logarithmized outcome.

**ED visits and unplanned hospitalizations for ACSCs.** Our focus in the main analyses is on the effects of nurse visit copayments on primary care use and the potentially heterogeneous impacts by income level. A natural follow-up question is, did the reduction in nurse use have undesirable offset effects on emergency department (ED) use or health. First, we believe that the potential effects on drug prescriptions and fills are close to zero given that the effects on GP visits, where nearly all prescriptions are written, are close to zero and insignificant. However, the reduced nurse use may still have spillovers outside primary care. Here, we present results on two additional post-blind outcomes: ED visits and unplanned hospitalizations for ambulatory care sensitive conditions (ACSC). Two main changes to the pre-specified main analyses are that we 1) estimate the impacts on all individuals to increase sample size and 2) use the CS estimator because

it has a longer follow-up than our baseline stacking analyses with a one-year follow-up. The longer follow-up is reasonable because the negative health impacts likely accumulate over time.

The CS event-study plots are in Figure A22 (post-blind). There appears to be no systematic pre-trend differences in either of the outcomes, but the estimates on ACSC hospitalizations are very noisy (these events are rare). In a one-year follow-up, the copayment adoption is associated with some statistically insignificant increases in ED visits and decreases in ACSC hospitalizations. The estimates are positive for both outcomes after two years, continuing to increase. However, we caution against strong interpretations far from the treatment, because the composition of areas changes and the effective sample size decreases the further we move from the adoption.

The static CS estimates are in Figure A23 (post-blind). For both outcomes, the estimates are positive but statistically insignificant, varying from +2% to +6% for ED visits and from +4% to +10% for ACSC hospitalizations. Overall, we view the nature of these statistically insignificant estimates more exploratory than confirmatory given that the outcomes and the chosen methods were post-blind.

## A.2 Results: Later Simultaneous Abolition

**Background.** In July 2021, the government conducted a broader reform to the act on copayments to reduce barriers to access and health inequality. A key change was to exempt nurse visits from copayments. More than 200 municipalities and almost three million people were affected by the policy change (see Figure A24). In contrast, there were no major changes to GP visit copayments at the time of the nurse visit copayment abolition. Transfers to municipalities were increased to compensate for reduced copayment revenue. Figure A25 shows the pre-abolition policies by municipality, also presenting the population size by using bubble size. The treatment group consists of many municipalities that are on average small and rural, while the comparison group contains the largest six cities by population size.

**Choices for Data Construction and Analysis.** We analyze the effects of the nationwide nurse visit copayment abolition of July 2021 using 12 pre-treatment and 11 post-treatment months (from 7/2020 to 5/2022), requiring a balanced panel. This limits the abolition analysis to pandemic times, excluding the early 2020 when the supply and demand shocks of the pandemic were largest. We assume that the effects of the earlier adoption accumulate fully within one year. For this reason, we exclude those municipalities that adopted the copayment less than 12 months before the start of the study window.

Besides our primary outcomes, we also examine the effects on the annualized number of prescriptions per capita written by public-sector organizations, extracted from the Kanta Prescription Center which is administered by the Social Insurance Institution. For year 2020, we use socioeconomic data from the end of that year, but for years 2021 to 2022 we use data from the end of 2020 due to a data release lag.

**Pre-analysis plan.** We did not have data from 2021 or 2022 at the time of writing our PAP, which can be verified from Statistics Finland who pseudonymize any batch of data before they are available for our use. Instead of the actual post-treatment data, we wrote the statistical programs as if the abolition occurred three years earlier on July 1st, 2018.

**Key Findings.** Regarding the nationwide abolition, we do not obtain causal conclusions

since imposing a parallel trends assumption does not seem to be reasonable based on pre-trend patterns.

**The COVID-19 pandemic.** Primary care utilization in 2020–2022 was to a large extent affected by the global COVID-19 pandemic. At the onset of the pandemic, major supply and demand shocks reduced the utilization of a broad range of services all around the world Moynihan et al. (2021). Interestingly, the number of curative nurse visits had not recovered to pre-pandemic levels by May 2022 in Finland (Figure A26). This finding is likely explained by supply-side factors. Public discussion about the shortage of nurses in publicly funded Finnish healthcare has intensified during the COVID-19 pandemic. The fact that nurses have been allocated to testing, tracing, and vaccination tasks has reduced the capacity of the system to treat non-COVID patients. Covid patients have also been a substantial additional burden on already strained public healthcare.

The observed reduction in nurse visits affects our ability to make valid inferences from the copayment abolition. First, it is likely that primary care utilization is even more supply-driven than in normal times. Assuming a shortage of appointment slots, the copayment abolition may not increase aggregate nurse visits even if demand for them increases. Patients who get appointments may also have more urgent conditions on average than before the pandemic. Second, the parallel trends assumption (PTA) required for causal inference may be less plausible in a context where the epidemiological situation and thus COVID-related burden to the system have varied during the pandemic. Moreover, the pandemic may have affected urban and rural areas differently. Urban areas are overrepresented in our comparison group for the abolition analyses.

**Pre-trend plots.** Figure A27 (post-blind) plots the trends in curative nurse and GP visits for all sample individuals. The differences in levels between the treatment areas are rather stable in the nine months prior to the abolition. After the abolition, nurse use appears to increase in the treated areas compared to our comparison municipalities. This trend break is not observable in GP visits.

It seems possible that the abolition increased nurse use. However, we are not willing to make a much stronger conclusion that the observed increase in nurse visits is the causal effect of

the copayment abolition. First, even if the pre-treatment differences in outcomes are stable for the nine months before the abolition, the same cannot be stated of the whole lead-up period. The nurse use was 0.03 to 0.07 annualized visits lower in the treated areas in 10/2020–6/2021 (months  $-9$  to  $-1$ ). In contrast, the nurse use was even higher in the treated areas than in our comparison areas in 8/2020–9/2020 (months  $-11$  to  $-10$ ). The quantitative magnitude of this relative decrease in pre-treatment periods is approximately similar to the observed relative increase in post-treatment periods. The observed relative changes in pre-treatment periods make it more plausible that there could be changes of similar magnitude in the post-treatment periods even in the absence of the policy change. Second, the relative increase in nurse visits in the treated areas essentially occurs in 12/2021–5/2022 (months  $+5$  to  $+10$ ), six months after the policy change. Nurse use does not increase notably in the treated areas relative to the comparison areas in the first five post-treatment months.

In the Online Appendix, Figure A28 plots the trends in curative nurse visits for the bottom 40% and the top 40% of the distribution of equivalized family disposable income in treatment and comparison municipalities. The same graphs, but for curative GP visits and prescriptions, are in Figure A29 and Figure A30. Figure A31 (post-blind) plots the pre-trends in nurse and GP visits for all individuals but this time weighting municipality-by-income-decile observations uniformly instead of population weights.

**Regression estimates.** The above analysis illustrates visually that the conventional PTA is not plausible in this application. Consequently, few causal conclusions can be made. We still provide highly suggestive point estimates for the magnitude of the observed changes in nurse use, either caused causally by the copayment abolition or driven by unobserved confounders correlated with the treatment assignment. In contrast to the PAP, we do not report standard errors in the table to highlight the suggestive nature of the estimates. At least it seems possible that the abolition increased nurse visits.

We use the following three methods that differ by the PTA required, weighted by population size. We view none of the three assumptions superior to the rest. First, the TWFE

DD specification without a pre-trend difference contains a static indicator for post-treatment periods for treated municipalities, and municipality and time fixed effects. It assumes parallel trends throughout the study window. Second, the TWFE DD specification with a pre-trend difference replaces the static treat x post indicator by lags of every post-treatment period for treated municipalities and by linear pre-trend differences in relative time between the treated and comparison areas. The average of the lags is reported. It assumes parallel trends in deviations of the outcome from a linear time trend. Third, the CS estimator (see Section 4) assumes parallel trends but only from the last pre-treatment period on.

The estimates for all sample individuals are reported in Table A10 (post-blind). With respect to nurse visits, extrapolating a linear pre-trend difference (TWFE with trends) leads to the largest point estimate while the TWFE model without trends produces the smallest estimates. The abolition is associated with an increase of +2.6% to +10.6% in the treated municipalities. The corresponding estimates for GP visits are in every case smaller and closer to zero, varying from −1.7% to +8.0%. The estimates on prescriptions vary from −1.7% to +1.8%.

Table A11 (post-blind) contains the corresponding estimates but after weighting municipalities uniformly instead of using population weights. The TWFE model without trends and the CS estimator produce estimates on nurse visits that vary between +1.3% and +10.9%. The TWFE model extrapolating a linear pre-trend difference, in contrast, produces a large 21 to 22 percent increase. The estimates on GP visits vary between −4.4% and +5.7% and the estimates on prescriptions between −1.7% and +0.2%.

The estimates for the bottom 40% and the top 40% of the income distribution are reported in Table A12. Regardless of the method, the estimates on nurse visits are always higher in relative terms for the top 40%. In two cases, they are also larger in absolute terms. Thus, there are no signs that the correlation between the copayment abolition and nurse use is larger at the lower end of the income distribution.

**Time-placebo analyses.** Our pre-analysis plan provides placebo estimates from experiments in which we fix the treatment and comparison municipalities and the treatment date



(July 1st) but proceed as if the treatment occurred in 2018 or 2019. We exclude municipalities that adopted the copayment in the study period or less than 12 months before the start of the window. The implicit assumption is that the effects of copayment adoption have accumulated fully in a year, so that earlier adoptions do not confound our placebo analysis. These placebo results show that the estimates can be sensitive to the specific version of the PTA. However, in these time-placebo runs it is much easier to select a preferred PTA out of the three alternatives than it is with the data from 2020–2022.

### A.3 Constructing our Analysis Data

**Copayment policies:** In analyses, we do not use every municipality in mainland Finland. The policy is not observed for some municipalities. Some municipalities are excluded because they participated in such municipal mergers where some of the municipalities had a different copayment policy than others before the merger. Regarding the abolition, we exclude those areas that introduced the copayment between 8/2019 and 6/2020, less than 12 months before the start of the study window in 7/2020. Basically, we assume that the effects of the copayment adoption have accumulated fully within one year so that the areas are unaffected by the earlier adoption in the study period for the analysis of the abolition. Two municipalities are excluded because they abolished the copayment already some months before the national reform. Two municipalities were excluded because their nurse visit copayment covered only a very small set of nurse visits.

After the above restrictions, we have 264 out of 293 municipalities in mainland Finland for a study on staggered adoption. In total, they had 5.0 million residents compared to the Finnish population of 5.5 million. Regarding the simultaneous abolition, we have 249 municipalities with 5.0 million residents. Figure 1 illustrates the staggered adoption in terms of treatment areas and treated population, and Figure A24 does the same for the abolition. Figure A1 (the staggered adoption) and Figure A25 (the abolition) show the municipal policies graphically. Figure A2 shows how the copayment level varied in Summer 2021.

Regarding the data on GP visit copayments using the 2013 municipal borders, we take municipal mergers into account and have to make choices about uncertainty in some municipality-month observations due to the observed documents not always being explicit and clear. These choices are observable in the replication codes.

**Socioeconomic data:** We exclude those ID-year observations where equivalized family disposable income is exactly zero (less than 1% of the rows) and only include those observations where an individual is aged 25 years or more as we want to exclude minors, who are exempted from the copayment, and students, who have access to student healthcare. Without any other restrictions on the data, this leaves us with approximately four million individuals out of the population of

5.5 million. With the population remaining after the above two restrictions, we compute the distribution of the equivalized family disposable income and sort individuals into income deciles.

**Primary care contacts:** We extract curative primary care outpatient visits fulfilling the following conditions: 1) person ID and visit date are observed, 2) variables related to cancellations are missing, and 3) the healthcare professional was either a nurse or a doctor. The distinction between curative and preventive contacts is important, as the nurse visit copayment was charged for curative visits. In contrast, preventive nurse visits appear to include vaccinations as well (e.g., seasonal flu, or COVID-19) for which no copayments are charged. Since 2013, the coding rate with respect to the curative/preventive indicator has been close to one, but in 2012 the information was missing for approximately 7% of the rows. An exemption are the four municipalities who adopted the Apotti electronic health record system (EHR) system: Vantaa in Spring 2019, and Helsinki, Kerava, and Kauniainen in Spring 2021. The share of missing values has been large after the adoption of Apotti in these municipalities. Consequently, we start our analysis from 1/2013 and exclude the four Apotti municipalities from the abolition analyses and 2019 for Vantaa from the adoption analyses. The coding rate for profession containing both nurses and doctors has been rather steady, varying between 4–7% in 1/2013–12/2019 and between 5–10% in 7/2020–5/2022.

Weekend visits are excluded from the analysis to reduce the potential bias resulting from changes in the way emergency department visits are coded in the registers. During the study period, some primary care areas and hospitals have formed joint emergency departments, and these contacts may be coded either to the primary care register or to the specialized healthcare register. Duplicate contacts are dropped. That is, an individual cannot have more than one curative visit on the same date and time with the same profession. Visits are linked to municipalities and, thus, to copayment policies via clients' municipality of residence.

Since 2019, the register also contains outpatient contacts in private clinics that we want to exclude. We do this by linking each visit in 2019–2022 to TOPI and SOTE organization registers that contain information on the visit provider. Both registers are continuously updated. We have an annual cross-sectional dataset on TOPI and two cross-sections on SOTE from early 2020 and

early 2022. We use SOTE from early 2020 for years 2019 and 2020 and SOTE from early 2022 for years 2021 and 2022. In 2019–2022, the linking of TOPI does not work for 4% of the extracted visits while the same figure for SOTE is 1%. Then, we include those visits whose provider a) has a TOPI service area code that refers to health centers (120, 121, or 122) or b) is a public sector organization in SOTE.<sup>24</sup>

After having aggregated primary care contacts to municipality-month observations, we exclude several observations due to quality issues. Not all areas were able to transfer data from their EHR to the national register when the national primary care data collection started in 2011. Changes in EHR systems can also be seen in the data as a sudden drop to a near-zero value in aggregate contacts. Even if these issues, mostly missing contacts, were unrelated to copayment policies, we want to mitigate the potential bias from missing visits by excluding the corresponding observations.

To identify suspiciously low or high values of service use, we first sum up curative nurse and GP visits at the municipality-month level and compute a distribution of means by permutationally excluding every combination of four consecutive months. The largest mean is our reference value with which we define an observation as suspiciously low if its value is less than  $X\%$  of the largest mean. The threshold  $X$  depends on the outcome. July is not considered because many people, both professionals and patients, are on vacation. Next, we again compute a distribution of means but this time after excluding the suspiciously low observations. We define an observation to be suspiciously high if its value is greater than  $120 + X\%$  of the largest mean. We mark all municipality-year pairs to be excluded if they contain suspicious months. We conduct the algorithm separately before (1/2013–12/2019) and after (7/2020–5/2022) the onset of the COVID-19 pandemic.

90 municipalities (out of 293) have suspiciously low or high values of primary care use in the pre-pandemic study period using a threshold of  $X = 0.4$ . The evolution of the sum of

---

<sup>24</sup>There are also private sector organizations providing publicly-funded primary care if the primary care area has outsourced services. In 2020, there were 21 such contracts between primary care areas (mostly small municipalities) and private sector organizations according to the Association of Finnish Municipalities. However, all primary care contacts in health centers should have a TOPI service area code that refers to health centers.

curative nurse and GP visits in these municipalities is illustrated in Figure A32, Figure A33, and Figure A34. Gray segments highlight municipality-year pairs with suspiciously low values. Based on visual inspection, the algorithm appears to be good at detecting irregularities. Regarding the abolition analyses, we set a higher threshold of 0.55 to make the algorithm more sensitive. 8 municipalities have suspicious values (Figure A35). In the replication codes folder, we illustrate primary care use in the remaining municipalities for which we find no suspicious observations.

Besides the effects of the copayment adoption on primary care use, we would have wanted to examine the impacts on waiting times as well. However, the share of visits that contain the required variables to construct waiting times vary greatly across areas and time. Thus, it is not clear to us whether the waiting times we could construct would be comparable across areas or over time. For this reason, we do not use waiting times as an outcome.

**Social assistance recipients:** The raw data contain ID-year observations for social assistance recipients, including monthly indicators for having received social assistance and annual sums of social assistance. One person is an applicant in the data even if the benefit is applied at the household level. Using data on family relations, we construct for each individual two variables: a monthly indicator for belonging to a family where someone received social assistance in a given month, and an annual sum of received basic social assistance. When aggregating outcomes to municipality-by-time-period level, the share of recipients is in percentages, and the sum of euros is per capita. We look for potentially missing values with the same algorithm as for primary care contacts with a threshold of 0.4. The algorithm comes up with 24 municipalities with suspectible municipality-year observations, but after a closer inspection we decide to exclude only one municipality-year observation, namely municipality no. 892 in 2016.

**Prescriptions:** We extract prescriptions written at public-sector units, containing both initializations of drug therapies and renewed prescriptions. Ideally, we would have wanted to consider only drug therapy initializations as patients can apply for renewals online or at pharmacies without a GP visit. However, we had some doubts about the quality of the variable and decided to include all prescriptions. Excluding private sector prescriptions may be a problem in a small set of

mostly small municipalities that have outsourced their primary care services to a private provider. We search for potentially missing values with the same algorithm as with primary care contacts. The algorithm does not find any irregularities even if we use a higher threshold of 0.6, making it more sensitive to outliers.

**ED visits and unplanned hospitalizations for ACSCs.** These data for 2013–2019 come from the Care Register for Health Care. In 2019, there was a change in variables with which these outcomes are recorded to the register. The key question in defining unplanned hospitalizations for ACSCs is how to separate different inpatient episodes. We first restrict to episodes lasting 0–100 days, excluding 0.6% of the rows. We aggregate the individual’s episodes starting on the same day into one by using the most recent discharge date. Then, we drop overlapping sub-episodes. Two episodes are defined to be separate if there is more than one week between them. Finally, we include only unplanned hospitalizations for ACSCs. The ACSCs are defined based on main diagnoses: if there is one or more ACSC ICD-10 diagnoses for the episode, we treat it as an ACSC episode. Our list for ACSC diagnosis codes follows in spirit the classifications used by the THL in Finland that are based on the classifications used by the NHS in the UK.

We search for potentially missing values with the same algorithm as with primary care contacts. Regarding ED visits, we exclude 57 weird municipality-year observations after setting a threshold of 0.3 for the algorithm. The same threshold leads to detecting 182 irregular observations for ACSC hospitalizations. Our conclusion is that the ACSC data are too noisy for our algorithm to work well. For this reason, we do not exclude ACSC observations due to data quality concerns. Instead, we assume that the potential data quality concerns are not related to the treatment status.

**Weighting by population size.** In the main analysis, we weight all regressions by population size to increase the precision of our estimates. Finnish municipalities are heterogeneous by population size. In 2020, the smallest municipality in mainland Finland had 700 residents, while the largest had 657,000. The median population size was 6,000. Our effective municipal sample sizes are smaller, as we restrict to individuals aged 25 or older and focus on the bottom 40% or the top 40% of the income distribution. Therefore, it is obvious that the outcomes of small

municipalities can be much noisier than the outcomes of large municipalities in our data.

The plausibly increased precision due to population weighting comes at a cost. First, the ATT estimates may not generalize to the whole country if the variation essentially comes from the largest municipalities. The ATTs may not need to be homogeneous by municipality size. Second, an institutional change (say, a change in the EHR system) in one large municipality can more easily bias the estimates than in a case where the variation comes from a large pool of municipalities, each municipality receiving the same weight.

As a robustness check, we estimate the results also without population weighting. This means that we uniformly weight municipalities when using the CS estimator and municipality-by-income-decile observations when using the TWFE regression (stacked or not). We made the following commitments in the PAP. If the results are reasonably insensitive to the form of weighting, we only report the main population-weighted estimates in the report and its appendices, and the uniformly-weighted results are provided in the replication folder. If, however, the results are sensitive to weighting, we include some of these uniformly-weighted result tables to the report as well to provide balance.

## A.4 Changes and Additions to the Pre-Analysis Plan

**Data: detecting quality issues in the data (change).** In the PAP, the algorithm which we use to detect quality issues in the primary care data was designed to find periods with abnormally low health care use. Here, we modify it so that it can detect periods with abnormally high values as well, which we think is a reasonable change. The implications can be seen for municipality no. 10 in Figure A32: now the spike in late 2019 is detected as an abnormally high value. We also use a higher threshold of 0.55 for the abolition analyses, making the algorithm more sensitive to outliers. Using a higher threshold of 0.6 for prescriptions does not change the number of excluded observations.

**Data: the Apotti EHR system (change).** We anticipated in the PAP that the adoption of the Apotti EHR system in three municipalities in Spring 2021 might cause data quality issues in the register data, which it did. The algorithm described above detects these three municipalities, but we also manually exclude Vantaa from the abolition analyses. Vantaa adopted the Apotti EHR already in 2019. After that, the share of missing values in the curative/preventive categorical has been large.

**Data: socioeconomic data (change).** The latest statistical year available is currently 2020. We expect that the socioeconomic data from the end of 2021 will be available by June 2023. In the final research report, we plan use values from the end of 2021 for year 2022.

**Data: social assistance recipients (change).** We added the year 2019 to the analysis. These data were not yet extracted to us at the time of writing the PAP.

**Data: does data quality improve when the copayment is adopted (addition)?** Once the copayment is adopted, an area needs to distinguish between curative and preventive nurse visits and be able to count the number of curative nurse visits. This may affect data quality in a way in which the number of recorded curative nurse visits changes even if the underlying utilization does not. We checked in Figure A4 (post-blind) that the number of preventive nurse visits does not increase in the treated municipalities after the copayment adoption, even if the number of curative nurse visits decreases considerably.



**Results: estimates on all individuals (addition).** In the PAP, we separately estimated the effects at the bottom 40% and the top 40% of the income distribution. None of the PAP results were estimated using the whole sample. This was motivated in two ways. First, we expected that the aggregate effects would be small due to supply rigidity in the public primary care system. If copayments reduce the use of primary care services among low-income individuals, the waiting time may be reduced, potentially attracting more patients who previously used occupational or private healthcare. Second, we have a major focus on the potential heterogeneity of effects. Still, aggregate estimates are relevant to the policymaker as well. Thus, we do now provide the main results (pre-trend plots and stacking and CS estimates) using all sample individuals (Table A5, Figure A16, Figure A17, Figure A18, and Figure A19, all post-blind).

**Results: estimates by income decile (addition).** We committed in the PAP to show the main estimates for primary care outcomes by income quintile and decile to allow for a more flexible analysis on treatment effect heterogeneity by income. Specifically, we now estimate and show in Figure 4 (post-blind) the adoption results on the number of nurse visits by income decile using stacking with balanced event-specific datasets as in main analysis. However, we do not estimate the abolition results by income deciles, as the abolition results are highly sensitive to a specific version of the parallel trends assumption.

**Adoption: summary statistics (addition).** We added Table A1 to examine whether the earlier-treated and later-treated municipalities are similar in their observable characteristics.

**Adoption: weighting municipalities uniformly (addition).** We re-estimated the results weighting our municipality-by-income-decile observations uniformly instead of population-weights as a robustness check. In the PAP, the idea was generally speaking to report the population-weighted estimates in the report and its appendixes and provide the uniformly-weighted results in the replication folder with the exception of picking some uniformly-weighted results, which we view important to show to the reader, to the online appendix. Regarding the adoption analyses, we added Table A7 (post-blind; uniformly-weighted stacked baseline results) to the appendix. Two figures from the PAP (uniformly-weighted CS event-study plots and static CS

estimates) are no longer in the appendix, but they are provided in the replication folder as are the rest of the uniformly-weighted results. In the PAP, the uniformly-weighted results were discussed in the main text, but we moved this discussion to the Online Appendix to shorten the body text.

**Adoption: changes to GP visit copayments (addition).** In the PAP, we committed to adding a plot showing when the GP visit copayment increases were made relative to the adoption of the nurse visit copayment. Details were fixed after the PAP registration, and the plot in question is in Figure A3 (post-blind).

**Adoption: potentially diverging trends.** The PAP discussed our plans to adjust the analysis of the copayment adoption in a hypothetical case where pre-trend plots and event-study plots hint of diverging pre-trends. We observe no such pre-trend differences in nurse visits, but there may be a small trend difference in GP visits. Overall, we view the planned analyses as appropriate and sufficient. For GP visits, we recommend putting more weight to the specifications that allow for a linear pre-trend difference or assume parallel trends only from the last pre-treatment period on.

**Adoption: logarithmized results (addition).** In the PAP, we showed the key results (static stacked baseline results, static CS results, and the DDD results) both in levels and logs. The logarithmized results were (mostly) in the appendix. After registering the PAP, we also added event-study plots (stacked in Figure A8 and CS in Figure A10, both post-blind) using logarithmized outcomes. The purpose was to visually examine whether the pre-treatment trends in outcomes seemed parallel in both levels and logs.

**Adoption: the follow-up length and cohort-specific effects (addition).** The observed differences in estimates from our pre-registered stacking and CS specifications raised the possibility that the effects are increasing in follow-up length. To examine this, we added a CS event-study plot on the annualized number of primary care visits in Figure A20 (post-blind) and static stacked results using a 24-month follow-up (Table A6, post-blind). Additionally, Figure A21 (post-blind) shows the CS results for each treatment cohort.

**Adoption: are social assistance recipients mainly in the first income decile**

**(addition)?** We added Figure A15 (post-blind) to answer the question. The question arose when we discovered from Figure 4 (post-blind) that the effects tend to grow the lower the income decile is except for the first decile. The hypothesis is that receiving social assistance can attenuate the effect *ceteris paribus* as the benefit can be applied for to reimburse copayments.

**Adoption: the Roth and Sant’Anna (2022) estimator and later-treated comparisons (addition).** We use not-yet-treated municipalities as comparisons in the main stacked analyses. As a robustness check, we exclude the never-treated from analyses and thus use the later-treated as comparisons. After this change, it is a plausible assumption that treatment timing is quasi-random. Making this assumption, we estimate the effects on primary care use with the estimator proposed by Roth and Sant’Anna (2022) that can be much more precise than the earlier DD based alternatives (e.g., the CS estimator). The results are in Table A8 (post-blind) in levels and in Table A9 (post-blind) in logs.

**Adoption: ED visits and unplanned hospitalizations for ACSCs (addition).** We also estimated the effects on emergency department (ED) visits and unplanned hospitalizations for ambulatory care sensitive conditions (ACSC), both post-blind outcomes. We estimate the impacts on all individuals to increase sample size and use the CS estimator because it has a longer follow-up than our baseline stacking analyses with a one-year follow-up. The longer follow-up is reasonable because the negative health impacts likely accumulate over time. The CS event-study plots are in Figure A22 (post-blind) and the static estimates in Figure A23 (post-blind).

**Adoption: effects for diabetes and hypertension patients (addition).** We also examine treatment effect heterogeneity with respect to having received a drug prescription in 2018–2019 with an ATC code referring to diabetes or hypertension (A10, C02–C03, and C07–C09), proxying a diagnosis of these conditions. The stacking results are in Table A4 (post-blind).

**Abolition: follow-up length (change).** The pre-registered follow-up for the abolition analyses was 12 months. In this version, we use an 11-month follow-up since we currently have data only until the end of 5/2022. This change is not related to the nurses’ strike in April which we listed in the PAP as a potential reason to shorten the follow-up time.

**Abolition: referrals as an outcome (change).** Our initial plan was to use referrals to specialist care written by public primary care as an outcome that proxies GP-assessed need for diagnosis. However, we noticed already when writing the PAP that there were too many referrals missing in the early 2010s so that referrals could not be used as an outcome for the adoption analyses. Consequently, the PAP used referrals as an outcome for the abolition analyses only. In this version of the study, we do not use referrals as an outcome at all. In contrast to our plans, we only observe referrals that have led to a specialist visit or a procedure by the time the data were extracted (June 2022). As waiting times can be long for nonurgent care and our follow-up ends in May 2022, we observe only a subset of referrals we would want to observe.

**Abolition analyses to the appendix (change).** This is the most obvious change from the PAP. Originally, we planned to present both the adoption and the abolition results in the main text. Ultimately, we were not willing to make a PTA for the abolition analyses. Thus, we did not reach any causal conclusions about the abolition. Moving the abolition analyses to the appendix makes the main text simpler and more readable and frees some space to extend and improve the adoption analyses. In the PAP, we wrote down that we may need to adjust our analysis based on pre-trend plots when analyzing the final data. The pre-trend differences do not behave as well as expected when writing the PAP.

**Abolition: results presented in the appendix (change).** The set of abolition tables and figures we present to the reader has also changed. Regarding figures, we added the pre-trend plots for all individuals with respect to both nurse and GP visits in Figure A27 (post-blind) before showing similar figures in our two income-based subgroups. We also added uniformly-weighted pre-trend plots for all individuals in Figure A31 (post-blind).

Regarding tables, one change is that we do not report standard errors in tables to highlight the suggestive nature of the abolition estimates. Instead of statistical uncertainty, we think that design-based uncertainty is more important here. We added two tables containing regression estimates for all individuals, one population-weighted (Table A10, post-blind) and the other uniformly-weighted (Table A11, post-blind). These two tables replace two tables from the

PAP that had otherwise similar content but were estimated for our income-based subgroups. The motivation for this change was that as the parallel trends assumption does not seem plausible for all individuals nor our income-based subgroups, it may be simpler to just show the problems and report the estimates using all individuals. Due to the identification problems, we saw little value in reporting the DDD results for the abolition that were described in the PAP but ultimately dropped.

**Fixing bugs (change).** The raw data on social assistance contain more than one row for some person-year pairs. Our codes did not previously account for that. Version 1 and Version 2 of this article used a smaller sample for the abolition analyses than originally intended, dropping 11 municipalities due to missing a merge. The scripts have been correct, but we most likely made a human error when running *0\_copayment\_policies.R* manually. The abolition results are robust to the smaller sample.

## A.5 Parameter Values for Semi-Arc Elasticities

To obtain the price of care after the copayment adoption, we multiply a per-visit copayment of 10 euros by 0.8 that is selected to present the probability that the patient pays the copayment. The copayment is not always charged. As Figure A2 shows, by far the most common policy is to charge it for the first three visits annually. Finally, we divide the multiplication by the total cost of the visit. At baseline, we use the estimated average cost of 35 euros in 2017 in public healthcare (Mäklin & Kokko, 2020). To get the upper bound, we use a total cost of 61 euros. This is the total cost of a 20-minute nurse appointment in a major private provider in July 2022 (66 euros), adjusted to the 2017 prices.

Regarding the quantities, we use Table A5 (post-blind) as the baseline. These estimates present the stacked results for all individuals in a one-year follow-up. Specifically, we average over the two specifications in Panel A (both balanced and unbalanced event-specific datasets): the pre-treatment mean is thus 1.000 annualized nurse visits, and the effect estimate is  $-0.089$ . To get the upper bound, we fix the pre-treatment mean of 1.000 annualized visits, but use a  $-15\%$  reduction as the effect estimate, which is based on Figure A19 (post-blind) showing the CS results for all individuals. In this case, the follow-up time is on average longer and depends on the municipality.

These parameter values lead us to the following semi-arc elasticities: the baseline is  $-0.41$  and the upper bound is  $-1.24$ .

## A.6 Additional Figures and Tables

Table A1: Adoption: Characteristics of the Never-Treated, Earlier-Treated, and Later-Treated.

Feature	Earlier-treated	Later-treated	Never-treated
Population size	13,212	13,003	31,791
Population change, %	+0.069	+0.079	+0.897
Degree of urbanization	78.479	77.805	92.228
Upper secondary education, %	67.280	68.827	71.318
Employment rate, %	69.365	66.749	68.764
Pensioners, %	26.149	26.984	21.969
Nurse visit per capita	0.836	0.923	0.719

*Notes:* This table was not pre-registered and is post-blind. The never-treated had not adopted the nurse visit copayment by the end of 2019. The later-treated adopted it after the median event time (January 1st, 2016). Values represent population-weighted means from 2013. Data are publicly available statistics from Statistics Finland and Sotkanet.

Table A2: Adoption: Social Assistance Use.

Metric	Share receiving	Euros received
Level	2.657	105.674
Estimate	−0.040	6.279
Std. error	0.039	3.279
P-value	0.311	0.057
Change (%)	−1.487	5.942
Estimate (trends)	−0.004	3.771
Change (%) (trends)	−0.133	3.569
Events	19	6
Treated areas	174	131
All areas	264	264

*Notes:* The dataset is stacked and balanced. We use 1) Model 1 and 2) its modified version allowing for differential linear pre-trend (“trends”). Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Outcomes are the share of individuals in a family receiving social assistance (in percentages) and the annual sum of of the family’s received basic social assistance (in euros) per capita. With the latter outcome, we only include events that occurred on January 1st.



Table A3: Adoption: DD Comparisons, Primary Care Use, Robustness Checks.

Metric	Nurse visits		GP visits	
	Bottom 40%	Top 40%	Bottom 40%	Top 40%
A. Unbalanced data				
Level	1.386	0.605	1.444	0.698
Estimate	−0.133	−0.050	−0.053	−0.031
Std. error	0.030	0.017	0.020	0.009
P-value	0.000	0.004	0.008	0.001
Change (%)	−9.627	−8.219	−3.649	−4.386
Estimate (trends)	−0.129	−0.054	−0.032	−0.019
Change (%) (trends)	−9.316	−9.000	−2.228	−2.794
Events	19	19	19	19
Treated areas	175	175	175	175
All areas	264	264	264	264
B. Balanced data, logarithmized outcome				
Estimate	−10.198	−8.215	−5.113	−6.059
Std. error	2.464	3.326	2.001	2.098
P-value	0.000	0.014	0.011	0.004
Estimate (trends)	−10.280	−8.593	−4.425	−4.555
Events	17	17	17	17
Treated areas	126	126	135	135
All areas	209	209	225	225

*Notes:* The dataset is stacked. We use 1) Model 1 and 2) its modified version allowing for differential linear pre-trend (“trends”). Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively.

Table A4: Adoption: DD Comparisons, Primary Care Use, Diabetes or Hypertension.

Metric	Nurse visits		GP visits	
	Has prescription	The rest	Has prescription	The rest
A. Visits per capita				
Level	1.533	0.589	1.547	0.747
Estimate	−0.141	−0.058	−0.058	−0.042
Std. error	0.039	0.016	0.021	0.011
P-value	0.000	0.000	0.008	0.000
Change (%)	−9.173	−9.831	−3.731	−5.648
Estimate (trends)	−0.161	−0.050	−0.065	−0.021
Change (%) (trends)	−10.469	−8.492	−4.195	−2.857
Events	17	17	17	17
Treated areas	152	152	152	152
All areas	245	245	245	245
B. Logarithmized visits per capita				
Estimate	−9.584	−11.438	−4.900	−7.157
Std. error	2.530	3.208	1.862	2.149
P-value	0.000	0.000	0.009	0.001
Estimate (trends)	−10.355	−9.196	−5.120	−3.704
Events	17	17	17	17
Treated areas	126	126	135	135
All areas	209	209	225	225

*Notes:* The dataset is stacked. We use 1) Model 1 and 2) its modified version allowing for differential linear pre-trend (“trends”). Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. The table examines treatment effect heterogeneity with respect to having received a drug prescription (“has prescription”) in 2018–2019 with an ATC code referring to diabetes or hypertension (A10, C02–C03, and C07–C09), proxying a diagnosis of these conditions. Outcomes are the annualized number of curative nurse and GP visits, respectively.

Table A5: Adoption: DD Comparisons, Primary Care Use, All Individuals.

Metric	Contacts per capita		Log. contacts per capita	
	Balanced	Unbalanced	Balanced	Unbalanced
A. Nurse visits				
Level	0.995	1.004		
Estimate	−0.087	−0.091	−9.180	−9.669
Std. error	0.025	0.023	2.853	2.623
P-value	0.001	0.000	0.002	0.000
Change (%)	−8.737	−9.114		
Estimate (trends)	−0.098	−0.096	−9.756	−9.424
Change (%) (trends)	−9.797	−9.587		
Events	17	19	17	19
Treated areas	152	175	126	175
All areas	245	264	209	264
B. GP visits				
Level	1.091	1.085		
Estimate	−0.044	−0.041	−5.355	−5.172
Std. error	0.015	0.014	1.991	1.851
P-value	0.003	0.003	0.008	0.006
Change (%)	−4.048	−3.818		
Estimate (trends)	−0.039	−0.028	−4.589	−3.847
Change (%) (trends)	−3.607	−2.610		
Events	17	19	17	19
Treated areas	152	175	135	175
All areas	245	264	225	264

*Notes:* This table was not pre-registered and is post-blind. The dataset is stacked. We use 1) Model 1 and 2) its modified version allowing for differential linear pre-trend (“trends”). Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Outcomes are the annualized number of curative nurse and GP visits (or their logarithm). Depending on the column, event-specific datasets are either balanced or unbalanced.

Table A6: Adoption: DD Comparisons, Primary Care Use, 24-Month Follow-Up.

Metric	Nurse visits		GP visits	
	Bottom 40%	Top 40%	Bottom 40%	Top 40%
Level	1.361	0.593	1.442	0.700
Estimate	−0.143	−0.051	−0.047	−0.029
Std. error	0.042	0.022	0.026	0.013
P-value	0.001	0.024	0.075	0.023
Change (%)	−10.541	−8.555	−3.234	−4.181
Estimate (trends)	−0.169	−0.071	−0.036	−0.016
Change (%) (trends)	−12.394	−12.049	−2.480	−2.306
Events	14	14	14	14
Treated areas	130	130	130	130
All areas	235	235	235	235

*Notes:* This table was not pre-registered and is post-blind. The dataset is stacked and balanced. We use 1) Model 1 and 2) its modified version allowing for differential linear pre-trend (“trends”). In contrast to the main analysis, we use a 24-month follow-up. Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively.

Table A7: Adoption: DD Comparisons, Primary Care Use, Uniform Weighting.

Metric	Nurse visits		GP visits	
	Bottom 40%	Top 40%	Bottom 40%	Top 40%
Level	1.512	0.754	1.558	0.857
Estimate	−0.170	−0.071	−0.046	−0.036
Std. error	0.024	0.014	0.017	0.010
P-value	0.000	0.000	0.006	0.001
Change (%)	−11.224	−9.425	−2.934	−4.203
Estimate (trends)	−0.144	−0.093	−0.073	−0.054
Change (%) (trends)	−9.519	−12.392	−4.698	−6.314
Events	17	17	17	17
Treated areas	152	152	152	152
All areas	245	245	245	245

*Notes:* This table was not pre-registered and is post-blind. The dataset is stacked and balanced. We use 1) Model 1 and 2) its modified version allowing for differential linear pre-trend (“trends”). We weight our municipality-by-income-decile observations uniformly. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively.

Table A8: Adoption: the RS estimator, Primary Care Use.

Estimand	Window	Nurse visits		GP visits	
		Bottom 40%	Top 40%	Bottom 40%	Top 40%
simple	2013-18	-0.128 (0.054) [-8.6%]	-0.124 (0.024) [-16.5%]	+0.052 (0.031) [+3.3%]	+0.013 (0.022) [+1.6%]
simple	2013-19	-0.145 (0.031) [-9.7%]	-0.103 (0.015) [-13.8%]	-0.024 (0.027) [-1.5%]	-0.014 (0.018) [-1.6%]
simple	2014-18	-0.137 (0.053) [-9.2%]	-0.113 (0.024) [-14.8%]	+0.014 (0.031) [+0.9%]	+0.007 (0.020) [+0.8%]
simple	2014-19	-0.155 (0.040) [-10.4%]	-0.092 (0.021) [-12.1%]	-0.045 (0.008) [-2.9%]	-0.018 (0.012) [-2.1%]
cohort	2013-18	-0.149 (0.050) [-10.0%]	-0.135 (0.028) [-17.9%]	+0.074 (0.030) [+4.8%]	+0.017 (0.023) [+1.9%]
cohort	2013-19	-0.159 (0.028) [-10.7%]	-0.104 (0.014) [-14.0%]	-0.025 (0.026) [-1.6%]	-0.011 (0.017) [-1.3%]
cohort	2014-18	-0.151 (0.048) [-10.2%]	-0.123 (0.026) [-16.1%]	+0.034 (0.031) [+2.2%]	+0.009 (0.022) [+1.0%]
cohort	2014-19	-0.165 (0.038) [-11.1%]	-0.094 (0.021) [-12.4%]	-0.044 (0.030) [-2.8%]	-0.015 (0.009) [-1.8%]
calendar	2013-18	-0.116 (0.042) [-7.8%]	-0.109 (0.021) [-14.5%]	+0.007 (0.027) [+0.5%]	-0.015 (0.018) [-1.8%]
calendar	2013-19	-0.129 (0.030) [-8.6%]	-0.097 (0.016) [-13.0%]	-0.037 (0.025) [-2.4%]	-0.027 (0.016) [-3.2%]
calendar	2014-18	-0.127 (0.040) [-8.6%]	-0.102 (0.020) [-13.2%]	-0.020 (0.024) [-1.3%]	-0.022 (0.016) [-2.5%]
calendar	2014-19	-0.140 (0.034) [-9.4%]	-0.089 (0.019) [-11.7%]	-0.060 (0.018) [-3.9%]	-0.034 (0.011) [-3.9%]

*Notes:* This table was not pre-registered and is post-blind. We use the Roth and Sant’Anna (2022) estimator (RS) and estimate the effects on annualized number of curative nurse and GP visits per capita. We report point estimates, standard errors (in parenthesis), and the effect as a percentage change (in square brackets). Pre-treatment mean for the treated is averaged over 12 pre-treatment months. Later-treated municipalities are used as comparisons. Events that have at least 12 follow-up months are included. We use a balanced panel in calendar time and drop treatment cohorts that have only a single treated municipality. Municipalities are weighted equally. We use the refined variance estimator proposed by Roth and Sant’Anna (2022) for standard errors unless the software produces a warning (12 cases out of 96) that a more conservative Neyman-style variance estimate is less than an estimated adjustment factor. In these cases, we consequently use the unrefined Neyman-style variance estimator discussed in Roth and Sant’Anna (2022). We show the results for four study windows. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Three aggregated estimands proposed by Roth and Sant’Anna (2022) are reported. “Simple” is a simple average of the  $ATE(t, g)$  weighted by cohort size. “Cohort” first averages the  $ATE(t, g)$  for each cohort  $g$  before averaging these cohort effects weighting by cohort size. “Calendar” first averages the  $ATE(t, g)$  for each time period  $t$  weighting by cohort size before taking a mean of the calendar effects.

Table A9: Adoption: the RS estimator, Logarithmized Primary Care Use.

Estimand	Window	Nurse visits		GP visits	
		Bottom 40%	Top 40%	Bottom 40%	Top 40%
simple	2013-18	-0.084 (0.039)	-0.165 (0.034)	+0.017 (0.023)	-0.012 (0.031)
simple	2013-19	-0.095 (0.025)	-0.134 (0.024)	-0.032 (0.019)	-0.045 (0.023)
simple	2014-18	-0.096 (0.039)	-0.156 (0.033)	-0.003 (0.023)	-0.017 (0.028)
simple	2014-19	-0.106 (0.030)	-0.126 (0.027)	-0.044 (0.018)	-0.048 (0.020)
cohort	2013-18	-0.099 (0.035)	-0.180 (0.035)	+0.034 (0.022)	-0.008 (0.032)
cohort	2013-19	-0.106 (0.022)	-0.137 (0.021)	-0.032 (0.018)	-0.043 (0.022)
cohort	2014-18	-0.107 (0.035)	-0.169 (0.034)	+0.011 (0.023)	-0.014 (0.029)
cohort	2014-19	-0.114 (0.028)	-0.130 (0.026)	-0.044 (0.013)	-0.045 (0.017)
calendar	2013-18	-0.080 (0.034)	-0.147 (0.033)	-0.012 (0.021)	-0.040 (0.026)
calendar	2013-19	-0.086 (0.026)	-0.124 (0.026)	-0.040 (0.018)	-0.056 (0.021)
calendar	2014-18	-0.092 (0.032)	-0.145 (0.031)	-0.025 (0.019)	-0.042 (0.023)
calendar	2014-19	-0.098 (0.028)	-0.120 (0.009)	-0.053 (0.018)	-0.060 (0.018)

*Notes:* This table was not pre-registered and is post-blind. We use the Roth and Sant’Anna (2022) estimator (RS) and estimate the effects on logarithmized annualized number of curative nurse and GP visits per capita. We report point estimates and standard errors (in parenthesis). Later-treated municipalities are used as comparisons. Events that have at least 12 follow-up months are included. We use a balanced panel in calendar time and drop treatment cohorts that have only a single treated municipality. Municipalities are weighted equally. We use the refined variance estimator proposed by Roth and Sant’Anna (2022) for standard errors unless the software produces a warning (12 cases out of 96) that a more conservative Neyman-style variance estimate is less than an estimated adjustment factor. In these cases, we consequently use the unrefined Neyman-style variance estimator discussed in Roth and Sant’Anna (2022). We show the results for four study windows. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Three aggregated estimands proposed by Roth and Sant’Anna (2022) are reported. “Simple” is a simple average of the  $ATE(t, g)$  weighted by cohort size. “Cohort” first averages the  $ATE(t, g)$  for each cohort  $g$  before averaging these cohort effects weighting by cohort size. “Calendar” first averages the  $ATE(t, g)$  for each time period  $t$  weighting by cohort size before taking a mean of the calendar effects.

Table A10: Abolition: DD Comparisons, All Individuals.

Metric	Nurse Visits	GP Visits	Prescriptions
A. Annualized contacts per capita			
Level	0.808	0.795	4.313
Municipalities	241	241	249
Estimate (w/o trends)	0.021	−0.005	−0.007
Change (%)	2.562	−0.575	−0.165
Estimate (with trends)	0.086	0.038	−0.072
Change (%)	10.609	4.834	−1.672
Estimate (CS)	0.046	−0.010	−0.040
Change (%)	5.727	−1.269	−0.930
B. Logarithmized annualized contacts per capita			
Municipalities	240	241	249
Estimate (w/o trends)	3.907	0.760	0.107
Estimate (trends)	8.298	7.978	1.776
Estimate (CS)	5.611	−1.664	−0.494

*Notes:* This table was not pre-registered and is post-blind. The following methods are used: 1) a TWFE DID model without a pre-trend difference that includes an indicator for post-treatment periods in treated municipalities and municipality and time fixed effects, 2) a TWFE DID model with a pre-trend difference that replaces the static treat x post indicator by lags of every post-treatment period for treated municipalities and by a linear pre-trend difference in relative time. The mean of the estimated dynamic effects is reported. 3) The CS estimator with outcome regression (Callaway & Sant’Anna, 2021). Due to heterogeneity in municipality size, we weight by population size.



Table A11: Abolition: DD Comparisons, All Individuals, Uniform Weighting.

Metric	Nurse Visits	GP Visits	Prescriptions
A. Annualized contacts per capita			
Level	0.969	0.947	4.572
Municipalities	241	241	249
Estimate (w/o trends)	0.012	−0.037	−0.075
Change (%)	1.283	−3.881	−1.630
Estimate (with trends)	0.210	0.036	−0.046
Change (%)	21.661	3.795	−1.001
Estimate (CS)	0.110	0.056	−0.003
Change (%)	10.890	5.666	−0.057
B. Logarithmized annualized contacts per capita			
Municipalities	240	241	249
Estimate (w/o trends)	3.510	−4.355	−1.733
Estimate (with trends)	20.666	2.227	−0.333
Estimate (CS)	10.616	2.303	0.175

*Notes:* This table was not pre-registered and is post-blind. The following methods are used: 1) a TWFE DID model without a pre-trend difference that includes an indicator for post-treatment periods in treated municipalities and municipality and time fixed effects, 2) a TWFE DID model with a pre-trend difference that replaces the static treat x post indicator by lags of every post-treatment period for treated municipalities and by a linear pre-trend difference in relative time. The mean of the estimated dynamic effects is reported. 3) The CS estimator with outcome regression (Callaway & Sant’Anna, 2021). Due to heterogeneity in municipality size, we weight by population size.

Table A12: Abolition: Estimates by Income.

Metric	Nurse Visits	GP Visits	Prescriptions
<b>A. Bottom 40%</b>			
Level	1.075	1.057	5.929
Municipalities	241	241	249
Estimate (w/o trends)	−0.011	−0.011	−0.029
Change (%)	−0.978	−1.067	−0.490
Estimate (with trends)	0.085	0.053	−0.090
Change (%)	7.862	5.033	−1.511
Estimate (CS)	0.035	−0.013	−0.081
Change (%)	3.272	−1.218	−1.371
<b>B. Top 40%</b>			
Level	0.526	0.513	2.718
Municipalities	241	241	249
Estimate (w/o trends)	0.041	0.005	0.018
Change (%)	7.788	0.943	0.666
Estimate (with trends)	0.064	0.026	−0.044
Change (%)	12.180	5.045	−1.624
Estimate (CS)	0.041	−0.002	0.001
Change (%)	7.809	−0.367	0.019

*Notes:* The following methods are used: 1) a TWFE DID model without a pre-trend difference that includes an indicator for post-treatment periods in treated municipalities and municipality and time fixed effects, 2) a TWFE DID model with a pre-trend difference that replaces the static treat x post indicator by lags of every post-treatment period for treated municipalities and by a linear pre-trend difference in relative time. The mean of the estimated dynamic effects is reported. 3) The CS estimator with outcome regression (Callaway & Sant’Anna, 2021). Due to heterogeneity in municipality size, we weight by population size. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number contacts per capita.

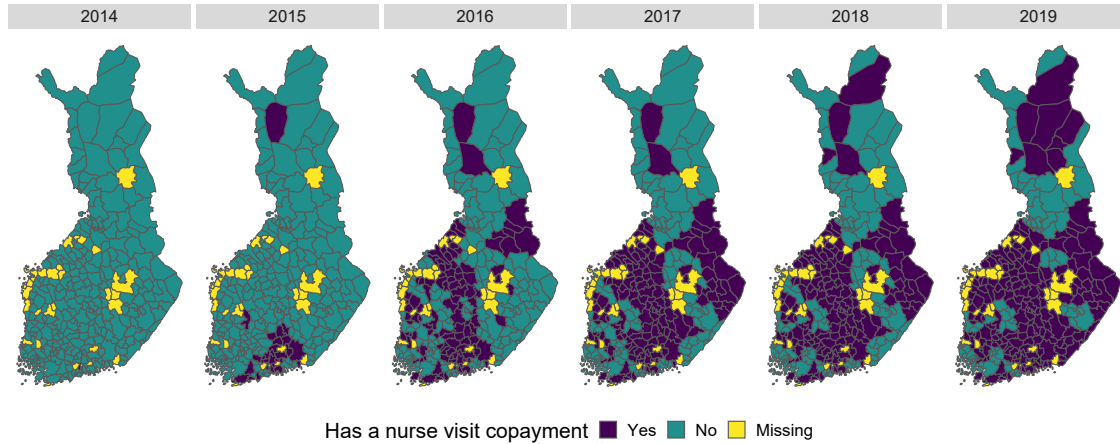


Figure A1: Staggered Adoption on Map.

*Notes:* The plot shows copayment policies by municipality at the end of a given year using the 2021 municipal boundaries (294 municipalities in mainland Finland). The sample contains those municipalities whose policies on copayments for curative nurse visits we observe in our data collection.

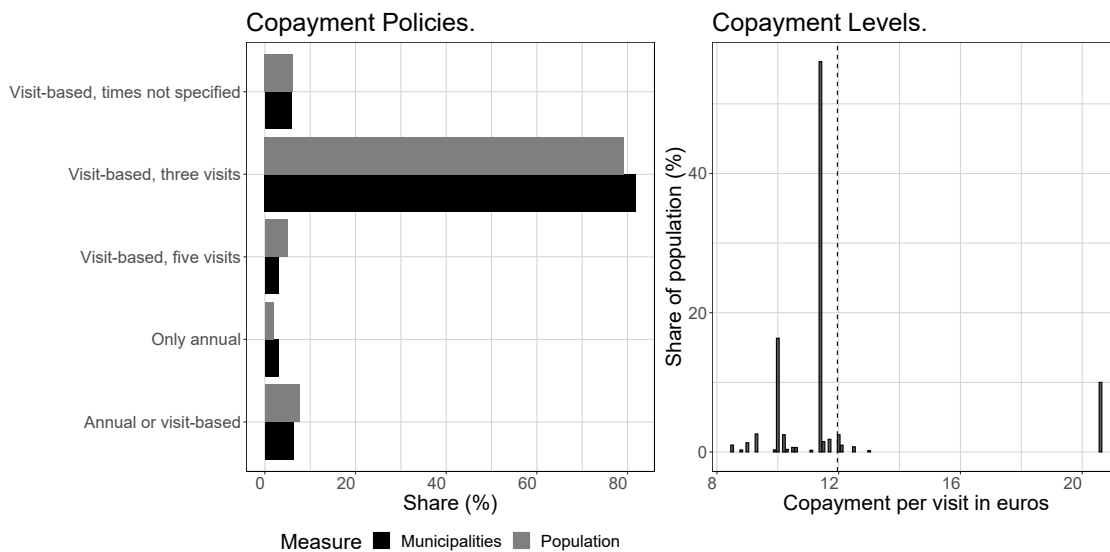


Figure A2: Copayment Levels and Policies in Summer 2021.

*Notes:* The plot shows how common different copayment options and levels were in Summer 2021 among the copayment municipalities of Section A.2. That is, municipalities who adopted the copayment less than 12 months before the start of the study window (7/2020) are excluded - see Section A.3 for details. The population sizes are from the end of 2019. The annual copayment is often twice the amount of the per-visit copayment. Some municipalities allow the patient to choose between the annual and the per-visit copayment. In most municipalities, a per-visits copayment was charged for the first three visits annually.

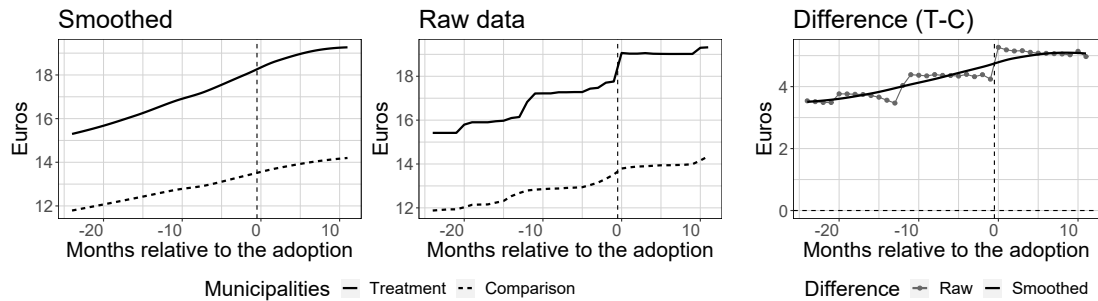


Figure A3: Adoption: Evolution in GP Visit Copayments.

*Notes:* This figure was not pre-registered and is post-blind. The dataset is stacked and balanced. The outcome is the GP visit copayment, paid for the first three visits annually. Treatment municipalities adopted the nurse visits copayment at time 0 in relative time. The left column contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle column, while the difference between treatment and comparison areas is depicted in the right column.

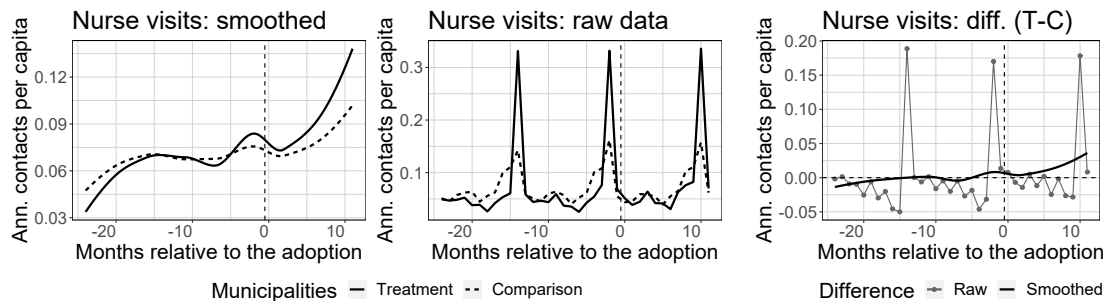


Figure A4: Adoption: Evolution in Preventive Nurse Visits, All Individuals.

*Notes:* This figure was not pre-registered and is post-blind. The dataset is stacked and balanced. The outcome is the number of annualized preventive nurse visits per capita. Treatment municipalities adopted the nurse visits copayment at time 0 in relative time. The left column contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle column, while the difference between treatment and comparison areas is depicted in the right column. The spikes occurring every 12 months likely represent seasonal influenza vaccinations, starting often in November.

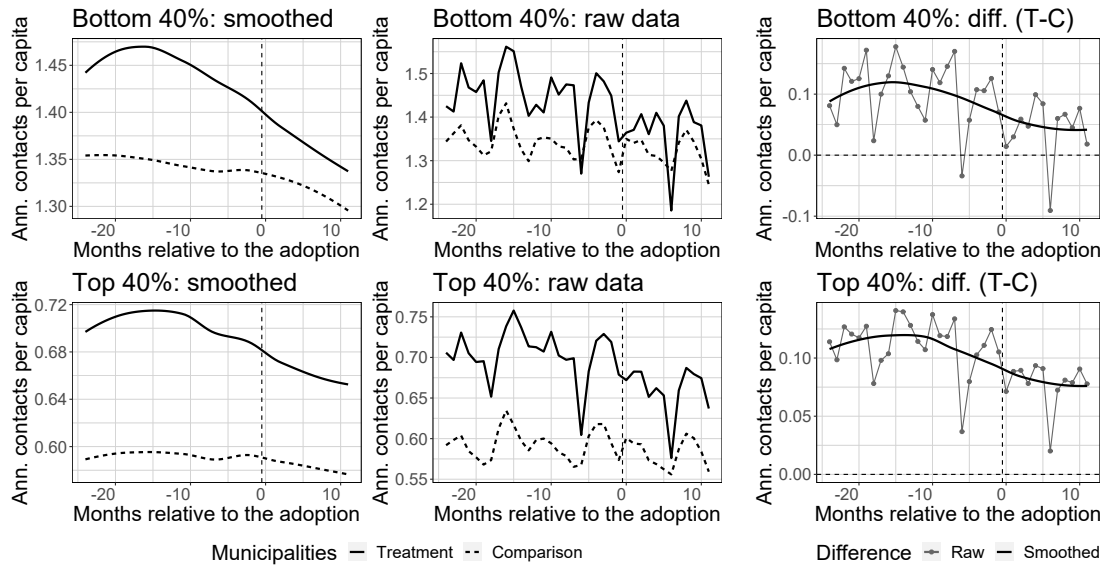


Figure A5: Adoption: Evolution in GP Visits.

*Notes:* The dataset is stacked and balanced. The outcome is the number of annualized curative GP visits per capita. Treatment municipalities adopted the nurse visits copayment at time 0 in relative time. The left column contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle column, while the difference between treatment and comparison areas is depicted in the right column. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income.

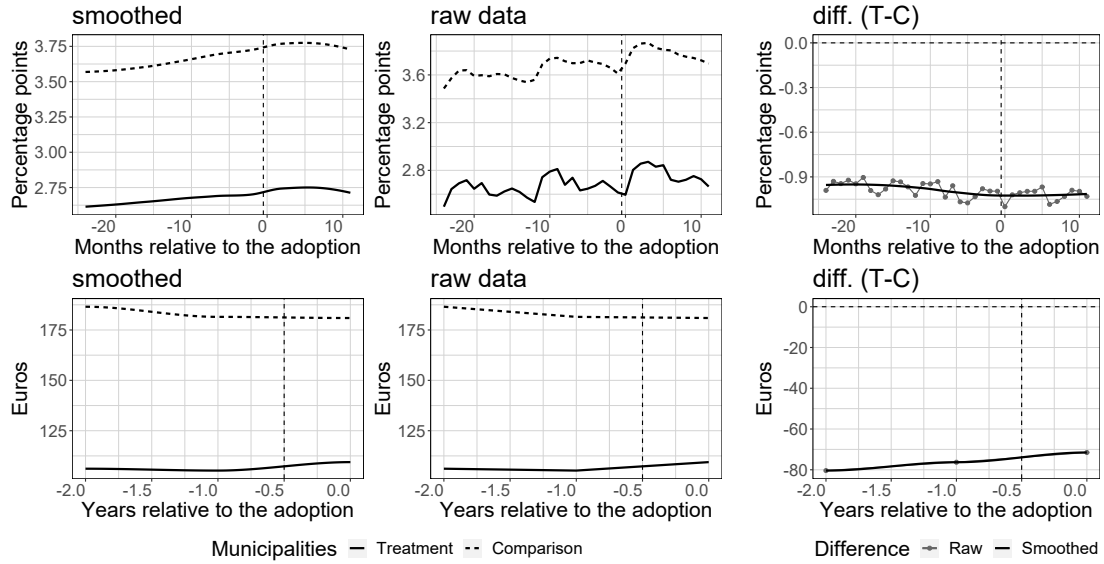


Figure A6: Adoption: Evolution in Social Assistance Recipients and Sums.

*Notes:* The dataset is stacked and balanced. In the top row, the outcome is the share of individuals in a family receiving social assistance (in percentages). In the bottom row, the outcome is the annual amount of social assistance received. As the latter is measured only annually, we include only those events that occurred on January 1st (and not in the middle of a year). Treatment municipalities adopted the nurse visits copayment at time 0 in relative time. The left column contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle column, while the difference between treatment and comparison areas is depicted in the right column.

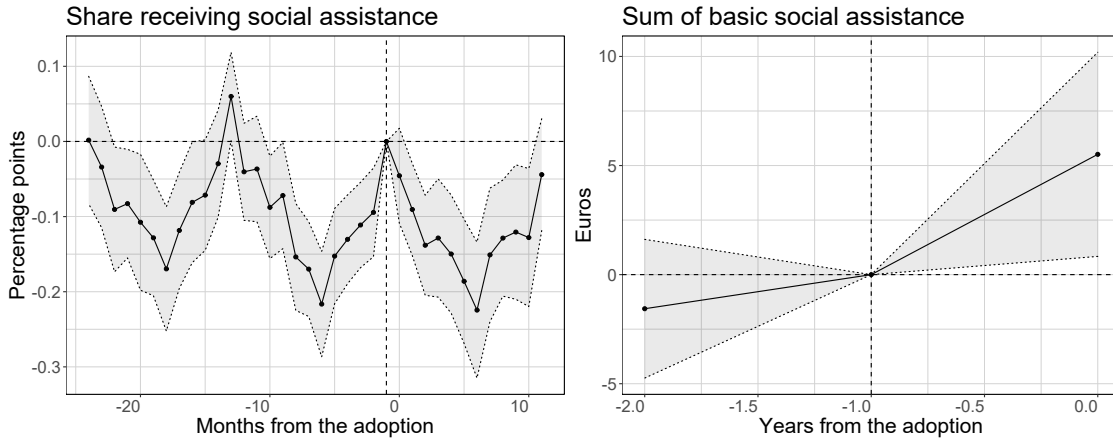


Figure A7: Adoption: Event-Study Plots on Social Assistance Use with Stacked Data.

*Notes:* The point estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The dataset is stacked, and event-specific datasets balanced. We use the dynamic event-study variant of Model 1, comparing the evolution of outcomes between treated and unexposed municipalities. With respect to the annual data on the sum of received social assistance, we include only events that occurred on January 1st. Due to heterogeneity in municipality size, we weight by population size. The standard errors are clustered by municipality.

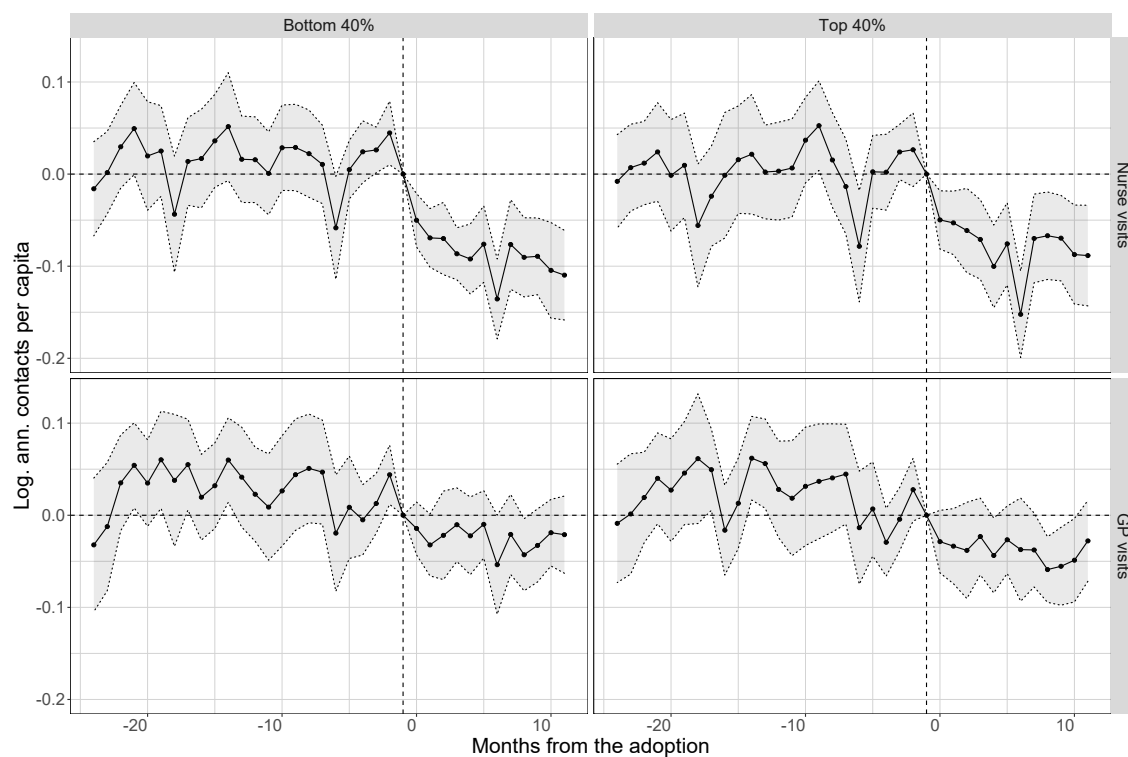


Figure A8: Adoption: Event-Study Plot on Logarithmized Primary Care Use with Stacked Data.

*Notes:* This figure was not pre-registered and is post-blind. The point estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The dataset is stacked, outcome logarithmized, and event-specific datasets balanced. We use the dynamic event-study variant of Model 1, comparing the evolution of annualized contacts per capita between treated and unexposed municipalities. Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income.

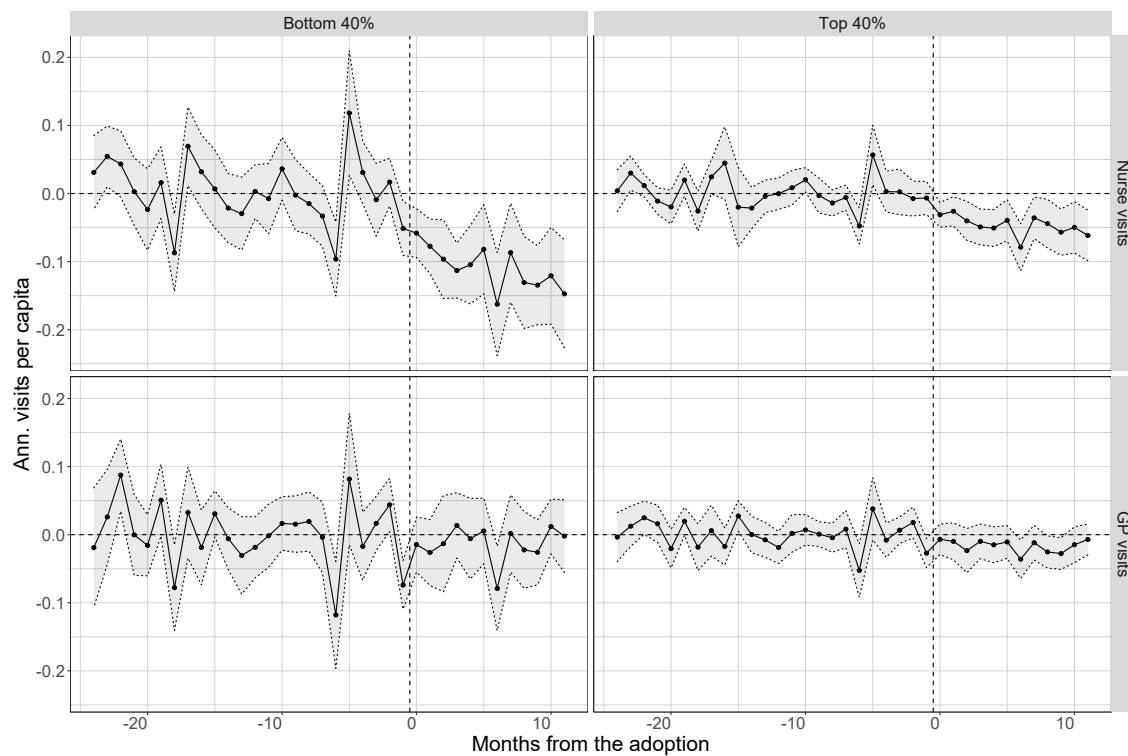


Figure A9: Adoption: Event-Study Plots Using the CS Estimator, Primary Care Use.

*Notes:* The post-treatment estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The pre-treatment estimates represent pseudo-ATTs from period  $t - 1$  to period  $t$  and thus differ from event-study estimates. We use the CS estimator (Callaway & Sant'Anna, 2021) with outcome regression, weight by population size, and cluster standard errors by municipality. Units that are not yet treated are used as a comparison. The dataset from 2014 to 2018 is balanced. The estimates compare the evolution of annualized contacts per capita between treated and unexposed municipalities. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively.



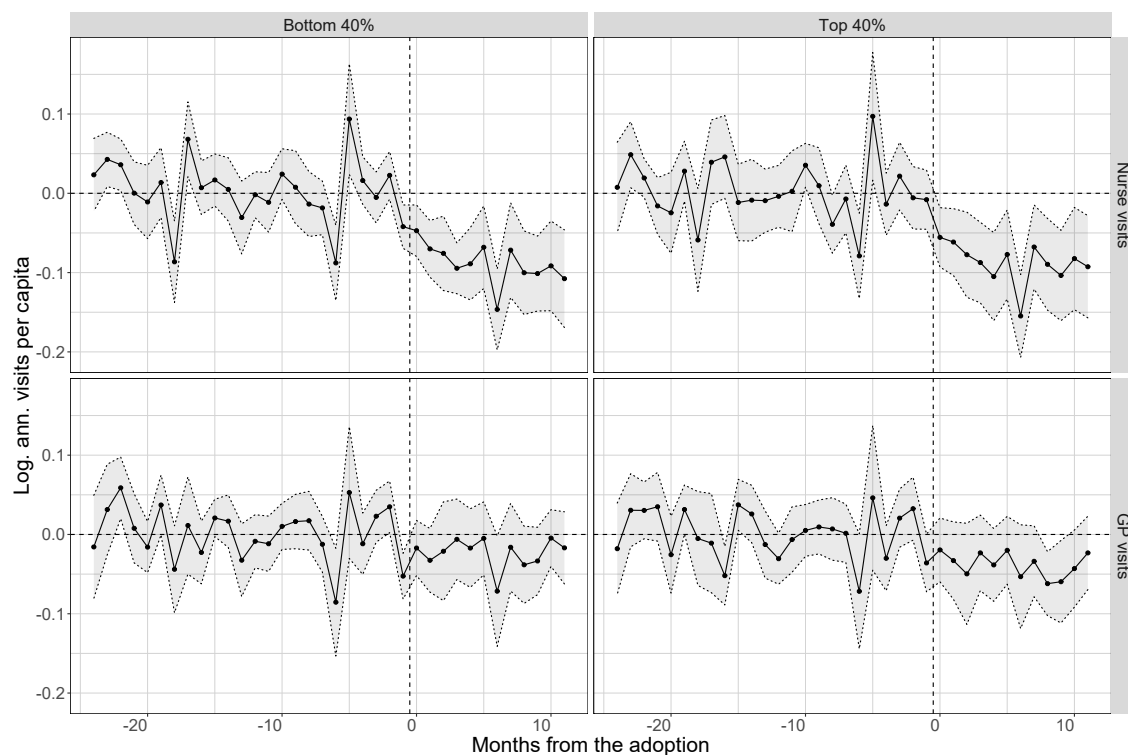


Figure A10: Adoption: Event-Study Plots Using the CS Estimator, Logarithmized Primary Care Use.

*Notes:* This figure was not pre-registered and is post-blind. The post-treatment estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The pre-treatment estimates represent pseudo-ATTs from period  $t - 1$  to period  $t$  and thus differ from event-study estimates. We use the CS estimator (Callaway & Sant'Anna, 2021) with outcome regression, weight by population size, and cluster standard errors by municipality. Units that are not yet treated are used as a comparison. The dataset from 2014 to 2018 is balanced. The estimates compare the evolution of logarithmized annualized contacts per capita between treated and unexposed municipalities. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively.

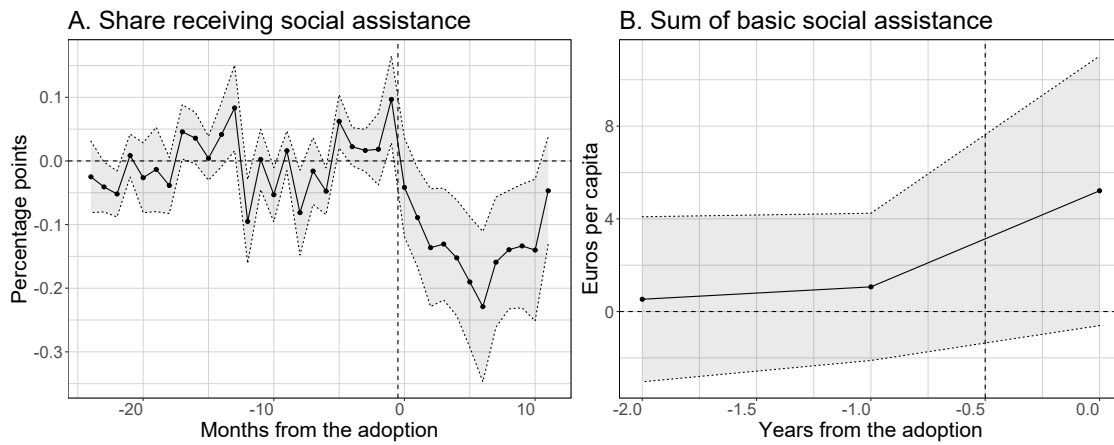


Figure A11: Adoption: Event-Study Plots Using the CS Estimator, Social Assistance Use.

*Notes:* The post-treatment estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The pre-treatment estimates represent pseudo-ATTs from period  $t - 1$  to period  $t$  and thus differ from event-study estimates. We use the CS estimator (Callaway & Sant'Anna, 2021) with outcome regression, weight by population size, and cluster standard errors by municipality. Units that are not yet treated are used as a comparison. The dataset is balanced. The estimates compare the evolution of outcomes between treated and unexposed municipalities. With respect to the annual data on the sum of received social assistance, we only include events that occurred on January 1st.

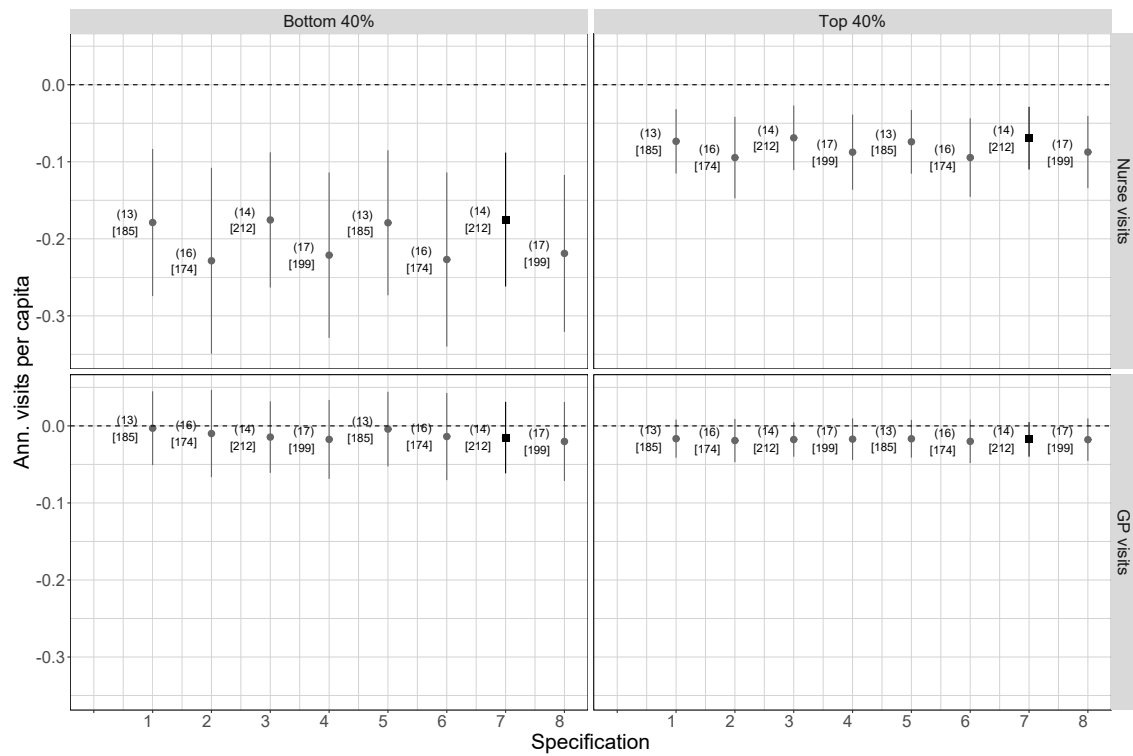


Figure A12: Adoption: the CS Estimator, Primary Care Use.

*Notes:* The point estimates represent static effect estimates for the treatment group. We use the CS estimator (Callaway & Sant'Anna, 2021) with outcome regression, weight by population size, and cluster standard errors by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively. Comparison group consists of the never-treated areas in specifications 1–4 and the not-yet-treated areas in specifications 5–8. Study period start year is 2013 in specifications 1, 2, 5, and 6, and 2014 in specifications 3, 4, 7, and 8. Study period end year is 2018 in specifications 1, 3, 5, and 7, and 2019 in specifications 2, 4, 6, and 8. The baseline is highlighted by black. Sample sizes are reported in the number of events (in parentheses) and municipalities (in square brackets).

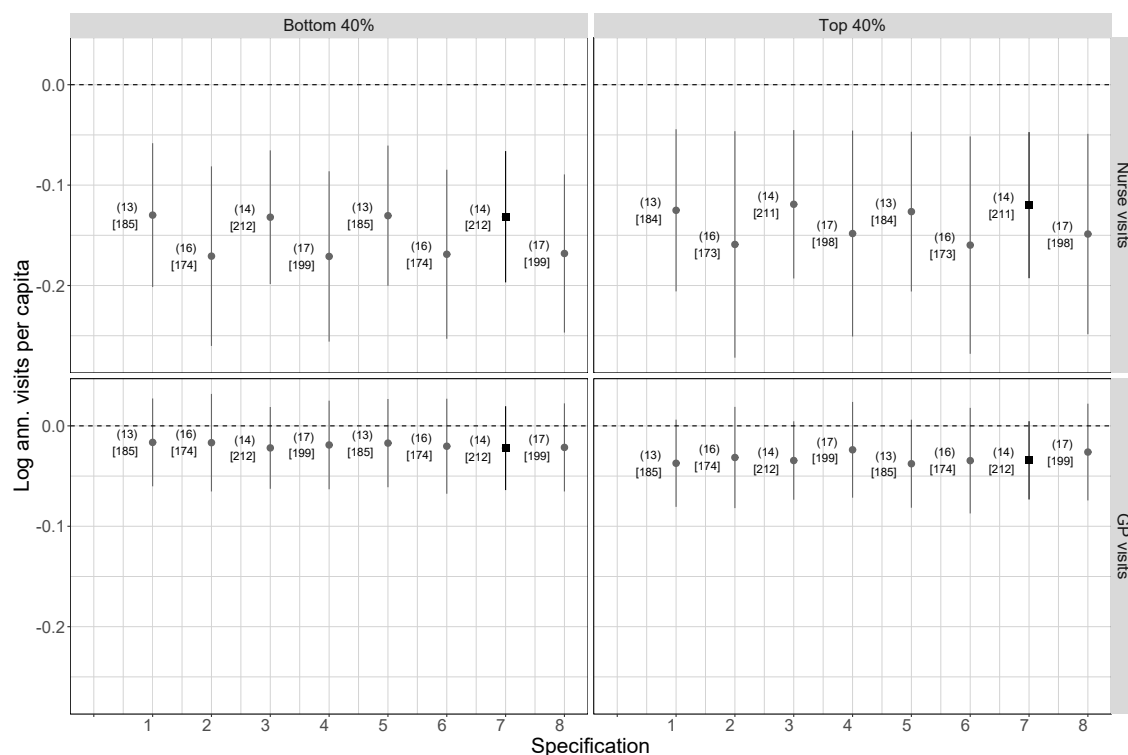


Figure A13: Adoption: the CS Estimator, Logarithmized Primary Care Use.

*Notes:* The point estimates represent static effect estimates for the treatment group. We use the CS estimator (Callaway & Sant'Anna, 2021) with outcome regression, weight by population size, and cluster standard errors by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the logarithmized annualized number of curative nurse and GP visits, respectively. Comparison group consists of the never-treated areas in specifications 1–4 and the not-yet-treated areas in specifications 5–8. Study period start year is 2013 in specifications 1, 2, 5, and 6, and 2014 in specifications 3, 4, 7, and 8. Study period end year is 2018 in specifications 1, 3, 5, and 7, and 2019 in specifications 2, 4, 6, and 8. The baseline is highlighted by black. Sample sizes are reported in the number of events (in parentheses) and municipalities (in square brackets).

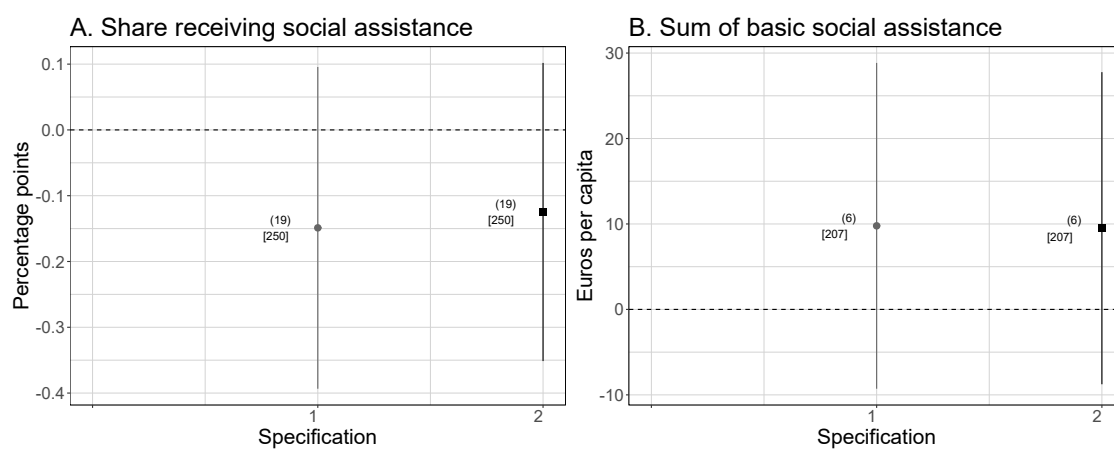


Figure A14: Adoption: the CS Estimator, Social Assistance Use.

*Notes:* The point estimates represent static effect estimates for the treatment group. We use the CS estimator (Callaway & Sant'Anna, 2021) with outcome regression, weight by population size, and cluster standard errors by municipality. Outcomes are the share of individuals in a family receiving social assistance (in percentages) and the annual sum of received social assistance. Specifications (comparison units): 1) the never-treated, and 2) the not-yet-treated. The baseline is highlighted by black. Sample sizes are reported in the number of events (in parentheses) and municipalities (in square brackets).

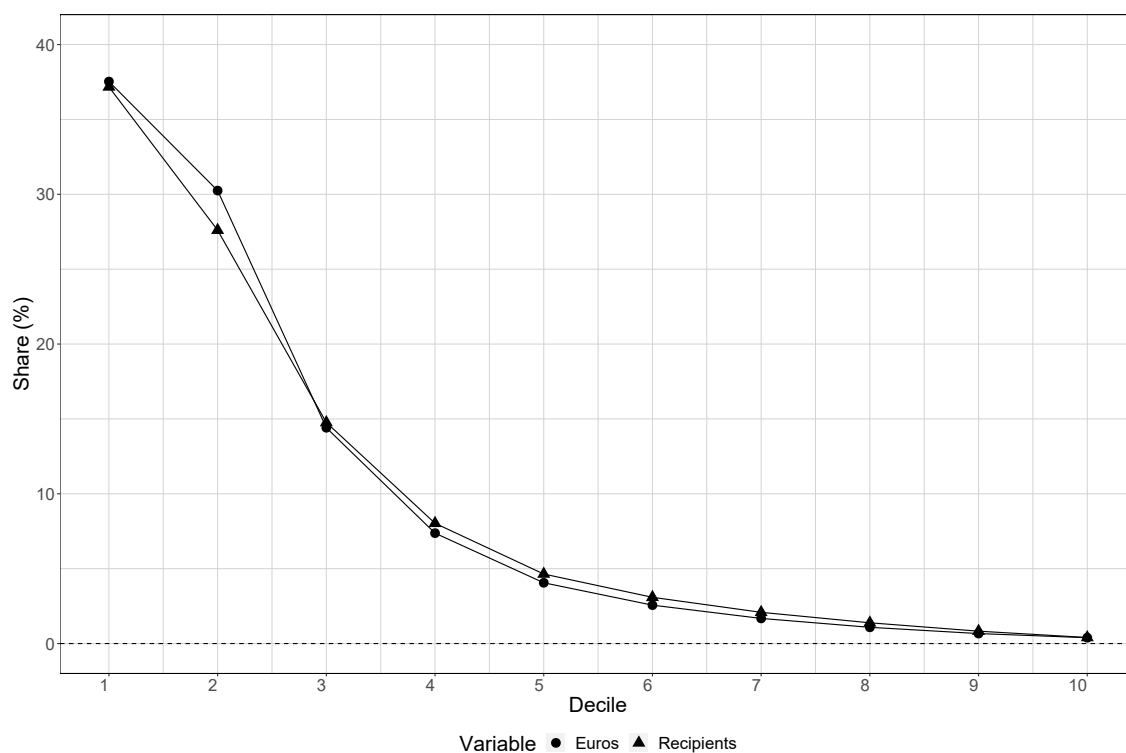


Figure A15: How Social Assistance Is Allocated between Income Deciles.

*Notes:* This figure was not pre-registered and is post-blind. The figure shows for each income decile the share of social assistance of the total sum of basic social assistance recipients and euros in 2013–2019. We restrict to those aged 25 or more, as in main analyses. Social assistance is a last-resort means-tested benefit for those with low income and little wealth, and it can also be applied to cover copayments in public healthcare. Based on the figure, there are much more people at the bottom 20% of the income distribution (equivalized disposable family income) that may potentially apply for social assistance to cover health expenses than in the remaining 80%.

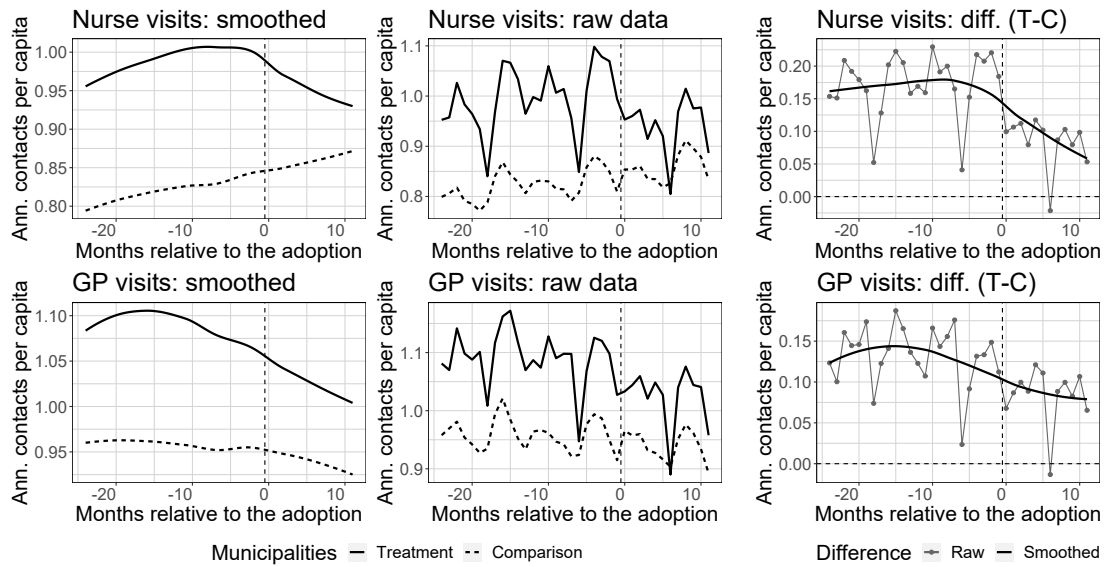


Figure A16: Adoption: Evolution in Outcomes, All Individuals.

*Notes:* This figure was not pre-registered and is post-blind. The dataset is stacked, and event-specific datasets balanced. The outcomes are the number of annualized curative nurse visits and GP visits per capita. Treatment municipalities adopted the nurse visits copayment at time 0 in relative time. The left column contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle column, while the difference between treatment and comparison areas is depicted in the right column.

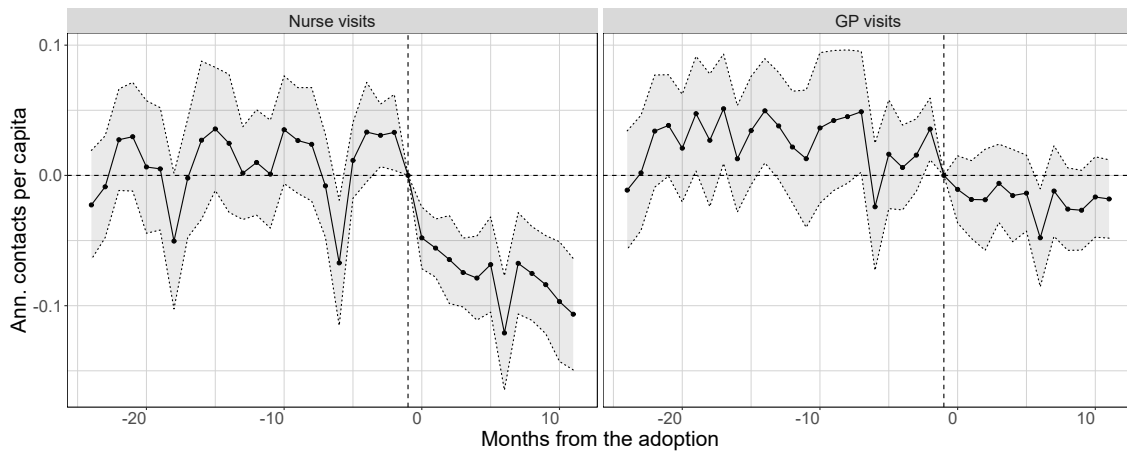


Figure A17: Adoption: Stacked Event-Study Plot on Primary Care Use, All Individuals.

*Notes:* This figure was not pre-registered and is post-blind. The point estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The dataset is stacked, and event-specific datasets balanced. We use the dynamic event-study variant of Model 1, comparing the evolution of annualized contacts per capita between treated and unexposed municipalities. Due to heterogeneity in municipality size, we weight by population size. The standard errors are clustered by municipality.

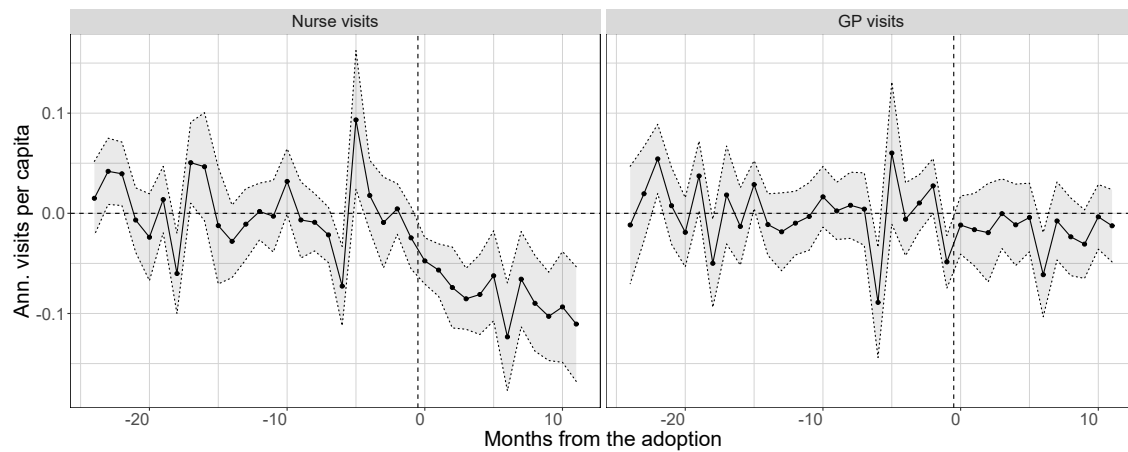


Figure A18: Adoption: the CS Event-Study Plots, Primary Care Use, All Individuals.

*Notes:* This figure was not pre-registered and is post-blind. The post-treatment estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The pre-treatment estimates represent pseudo-ATTs from period  $t - 1$  to period  $t$  and thus differ from event-study estimates. We use the CS estimator (Callaway & Sant'Anna, 2021) with outcome regression, weight by population size, and cluster standard errors by municipality. Units that are not yet treated are used as a comparison. The dataset from 2014 to 2018 is balanced. The estimates compare the evolution of annualized contacts per capita between treated and unexposed municipalities. Outcomes are the annualized number of curative nurse and GP visits.



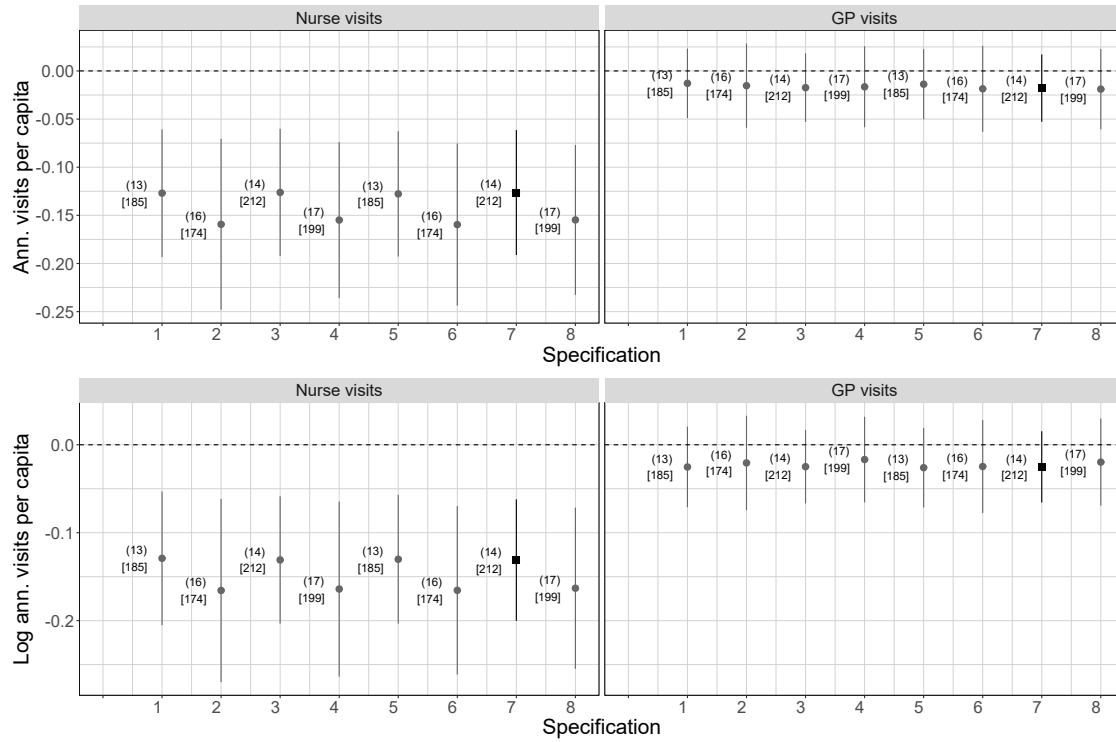


Figure A19: Adoption: the CS Estimator, Primary Care Use, All Individuals.

*Notes:* This figure was not pre-registered and is post-blind. The point estimates represent static effect estimates for the treatment group. We use the CS estimator (Callaway & Sant’Anna, 2021) with outcome regression, weight by population size, and cluster standard errors by municipality. Outcomes are the annualized number of curative nurse and GP visits, respectively. Comparison group consists of the never-treated areas in specifications 1–4 and the not-yet-treated areas in specifications 5–8. Study period start year is 2013 in specifications 1, 2, 5, and 6, and 2014 in specifications 3, 4, 7, and 8. Study period end year is 2018 in specifications 1, 3, 5, and 7, and 2019 in specifications 2, 4, 6, and 8. The baseline is highlighted by black. Sample sizes are reported in the number of events (in parentheses) and municipalities (in square brackets).

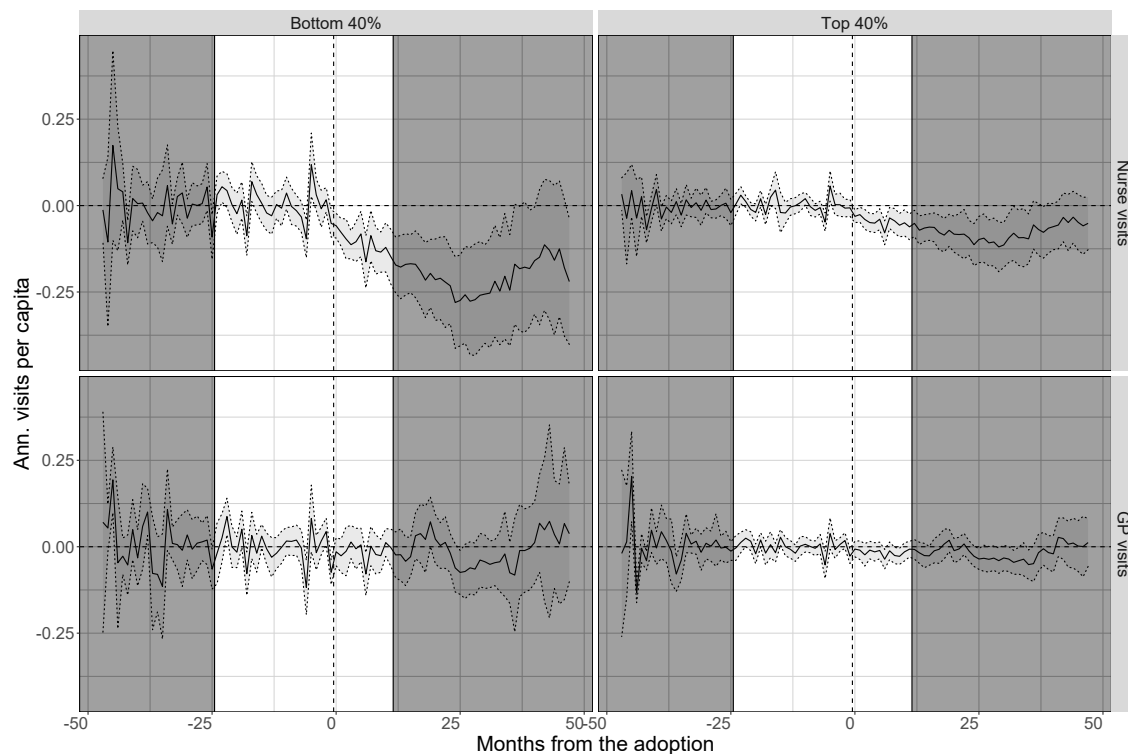


Figure A20: Adoption: Event-Study Plots Using the CS Estimator, Primary Care Use, All Leads and Lags.

*Notes:* This figure was not pre-registered and is post-blind. The post-treatment estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The pre-treatment estimates represent pseudo-ATTs from period  $t - 1$  to period  $t$  and thus differ from event-study estimates. We use the CS estimator (Callaway & Sant'Anna, 2021) with outcome regression, weight by population size, and cluster standard errors by municipality. Units that are not yet treated are used as a comparison. The dataset from 2014 to 2018 is balanced. The estimates compare the evolution of annualized contacts per capita between treated and unexposed municipalities. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively. Our PAP choice was to show the effects in a 36-month window (24+12), visible as the white segment around the event.

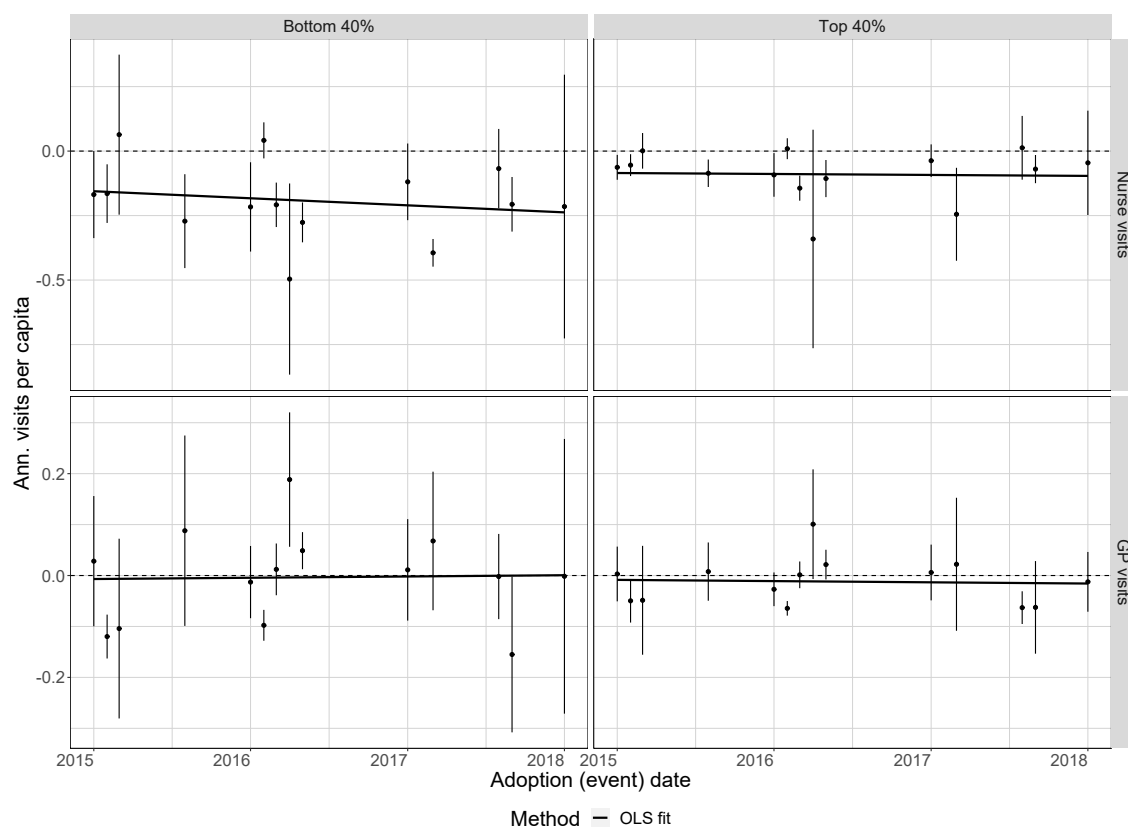


Figure A21: Adoption: Cohort-Specific CS Results, Primary Care Use.

*Notes:* This figure was not pre-registered and is post-blind. The point estimates represent static effect estimates for each treatment cohort. We also regress treatment month on the corresponding point estimate with OLS. We use the CS estimator (Callaway & Sant'Anna, 2021) with outcome regression, weight by population size, and cluster standard errors by municipality. Units that are not yet treated are used as a comparison. The dataset from 2014 to 2018 is balanced. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively. Our PAP choice was to show the effects in a 36-month window (24+12), visible as the white segment around the event.

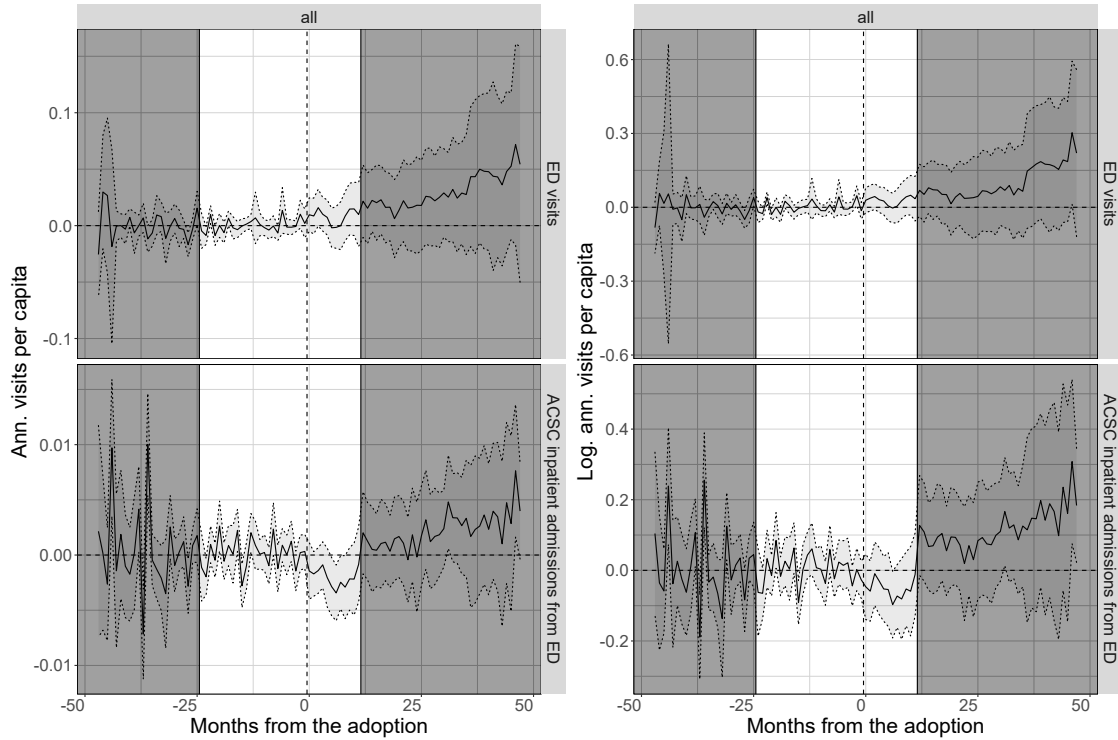


Figure A22: Adoption: Event-Study Plots Using the CS Estimator, ED Visits and Unplanned Hospitalizations for ACSCs.

*Notes:* This figure was not pre-registered and is post-blind. The post-treatment estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The pre-treatment estimates represent pseudo-ATTs from period  $t - 1$  to period  $t$  and thus differ from event-study estimates. We use the CS estimator (Callaway & Sant'Anna, 2021) with outcome regression, weight by population size, and cluster standard errors by municipality. Units that are not yet treated are used as a comparison. The dataset from 2014 to 2018 is balanced. The estimates compare the evolution of annualized contacts per capita (or its logarithm) between treated and unexposed municipalities. Outcomes are the annualized number of emergency department (ED) visits and unplanned hospitalizations for ambulatory care sensitive conditions (ACSC). Our PAP choice for the main analyses was to show the effects in a 36-month window (24+12), visible as the white segment around the event.

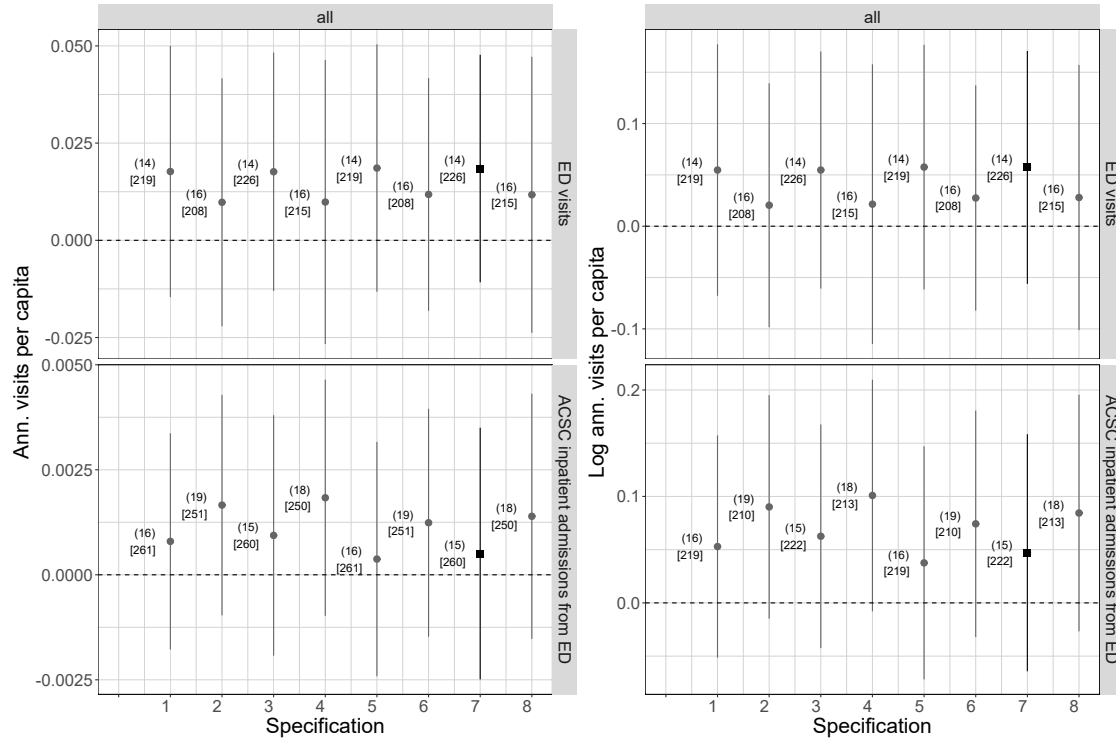


Figure A23: Adoption: the CS Estimator, ED Visits and Unplanned Hospitalizations for ACSCs.

*Notes:* This figure was not pre-registered and is post-blind. The point estimates represent static effect estimates for the treatment group. We use the CS estimator (Callaway & Sant'Anna, 2021) with outcome regression, weight by population size, and cluster standard errors by municipality. Outcomes are the annualized number (or its logarithm) of emergency department (ED) visits and unplanned hospitalizations for ambulatory care sensitive conditions (ACSC). Comparison group consists of the never-treated areas in specifications 1–4 and the not-yet-treated areas in specifications 5–8. Study period start year is 2013 in specifications 1, 2, 5, and 6, and 2014 in specifications 3, 4, 7, and 8. Study period end year is 2018 in specifications 1, 3, 5, and 7, and 2019 in specifications 2, 4, 6, and 8. The baseline is highlighted by black. Sample sizes are reported in the number of events (in parentheses) and municipalities (in square brackets).

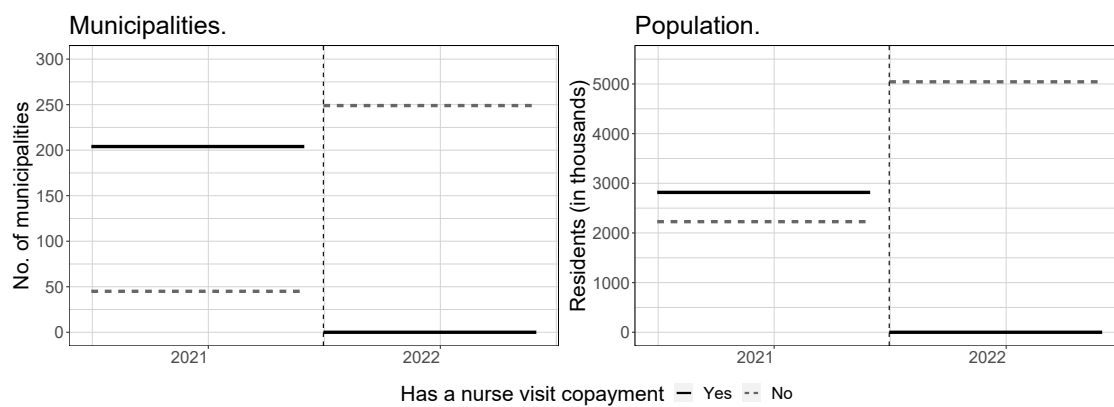


Figure A24: Simultaneous Abolition of the Nurse Visit Copayment.

*Notes:* We take municipalities in mainland Finland (293 in 2022) and use the 2022 municipal boundaries and population sizes from the end of 2019. The sample on contains those municipalities whose policies on copayments for curative nurse visits we observe in our data collection. We also require the adoption to have occurred at least 12 months before the start of the study window (7/2020). We assume that the effects have fully accumulated within 12 months after the adoption.

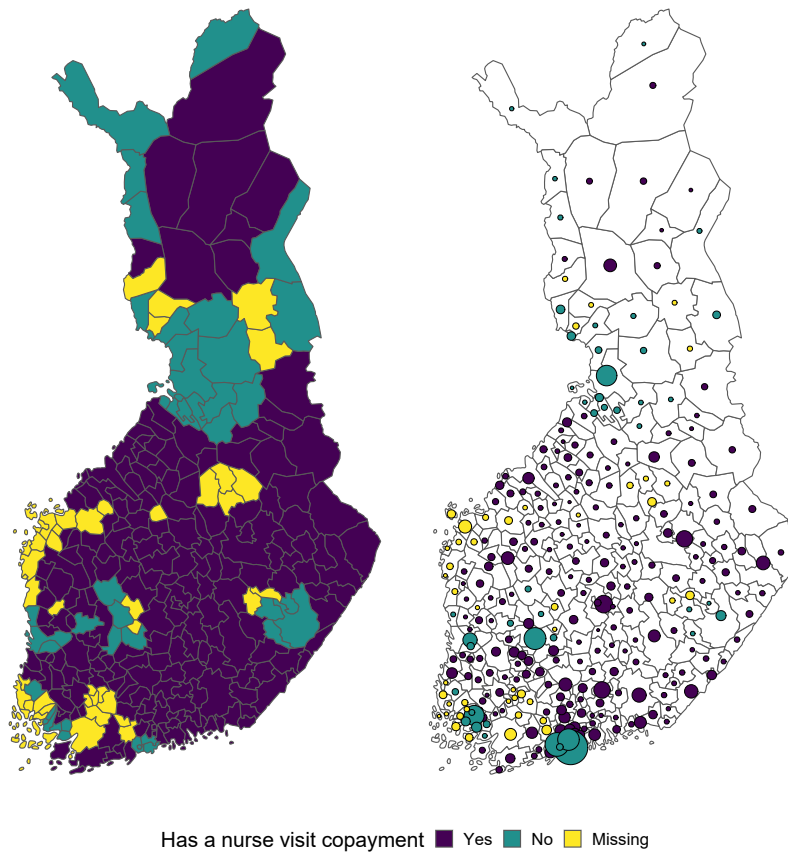


Figure A25: The Abolition on Map.

*Notes:* The plot shows copayment policies by municipality before the law change that abolished the nurse visit copayment in 7/2021, using the 2021 municipal boundaries (294 municipalities in mainland Finland). The group of missing municipalities contain municipalities for whom the policy is unobserved and municipalities who adopted the copayment less than 12 months before the start of the study window (7/2020), see Section A.3 for details. We assume that the effects have fully accumulated within 12 months after the adoption, which motivates the latter restriction. In the bubble plot, the size of the bubble is proportional to the 2018 population size.

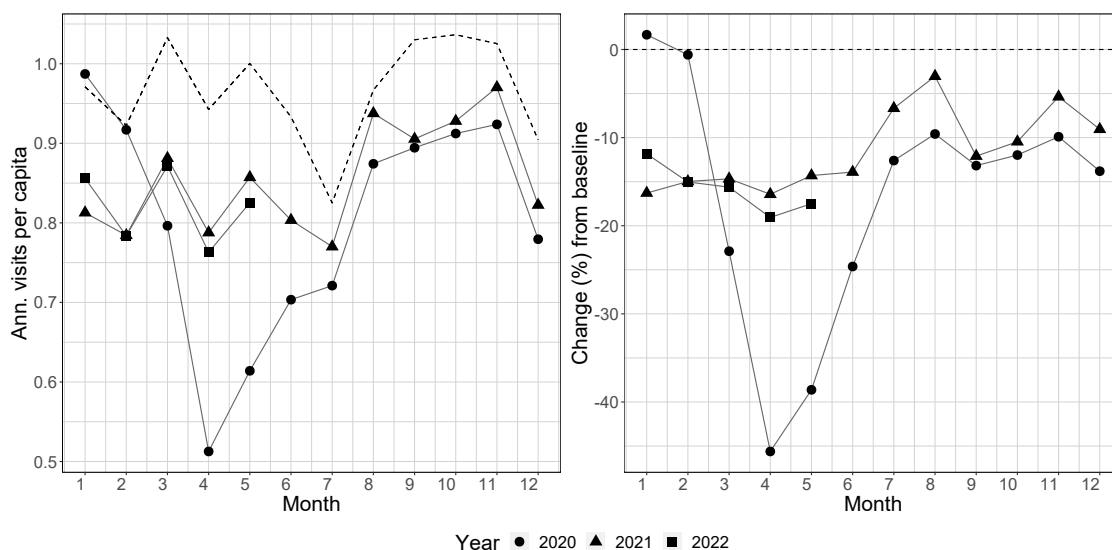


Figure A26: The Evolution of Nurse Visits during the COVID-19 Pandemic.

*Notes:* The figure shows the evolution of curative nurse visits in primary care relative to a baseline (monthly means from 2018–2019). On the left, the baseline is depicted by the dashed line. On the right, we show a change relative to the baseline. The data are from individuals aged 25 or more residing in Mainland Finland. Four municipalities that adopted the Apotti EHR system either in 2019 or 2021 are excluded due to missing values in the curative/preventive categorical. To filter out supply-side noise due to the changing number of workdays in a given month, we divide our measure for healthcare use by the number of workdays in a given month and then multiply it by the mean number of monthly workdays over the years.



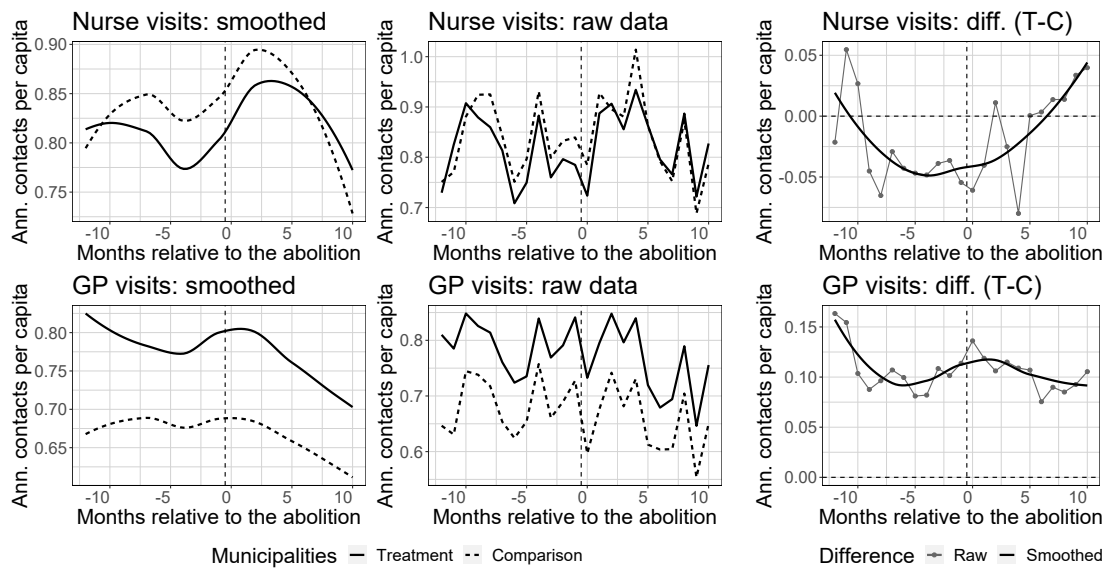


Figure A27: Abolition: Evolution in Outcomes, All Individuals.

*Notes:* This figure was not pre-registered and is post-blind. The outcomes are the number of annualized curative nurse visits and GP visits per capita. The left column contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle column, while the difference between treatment and comparison areas is depicted in the right column.

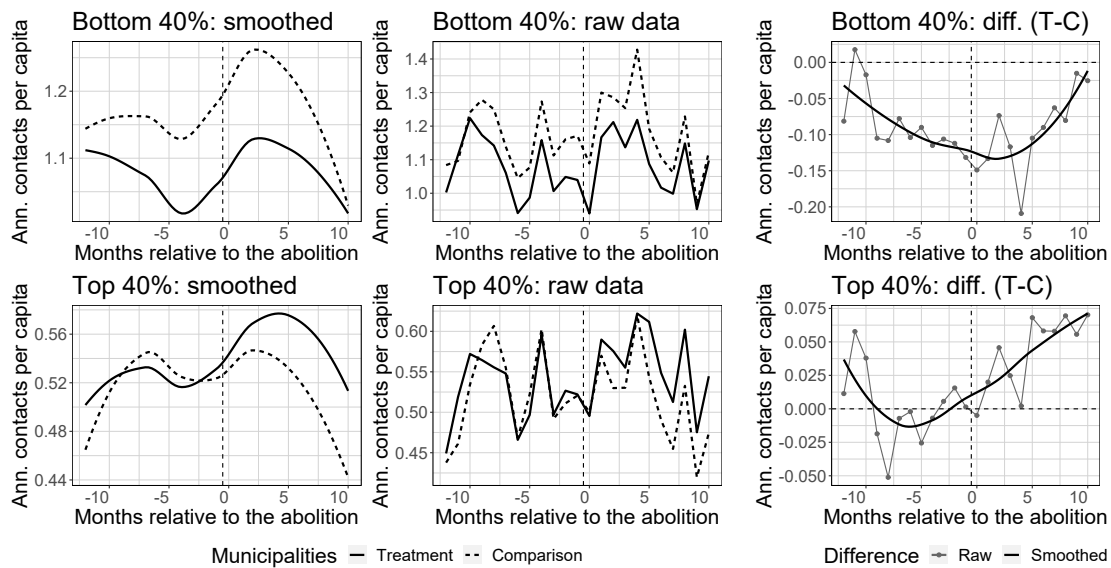


Figure A28: Abolition: Evolution in Nurse Visits.

*Notes:* The outcome is the number of annualized curative nurse visits per capita. The left column contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle column, while the difference between treatment and comparison areas is depicted in the right column. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income.

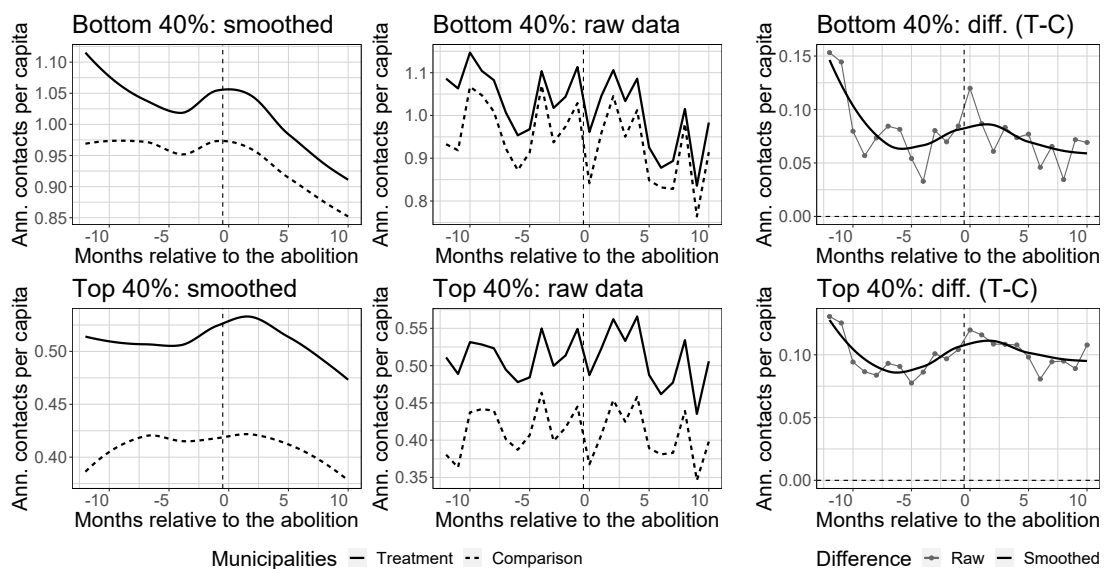


Figure A29: Abolition: Evolution in GP Visits.

*Notes:* The outcome is the number of annualized curative GP visits per capita. The left column contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle column, while the difference between treatment and comparison areas is depicted in the right column. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income.

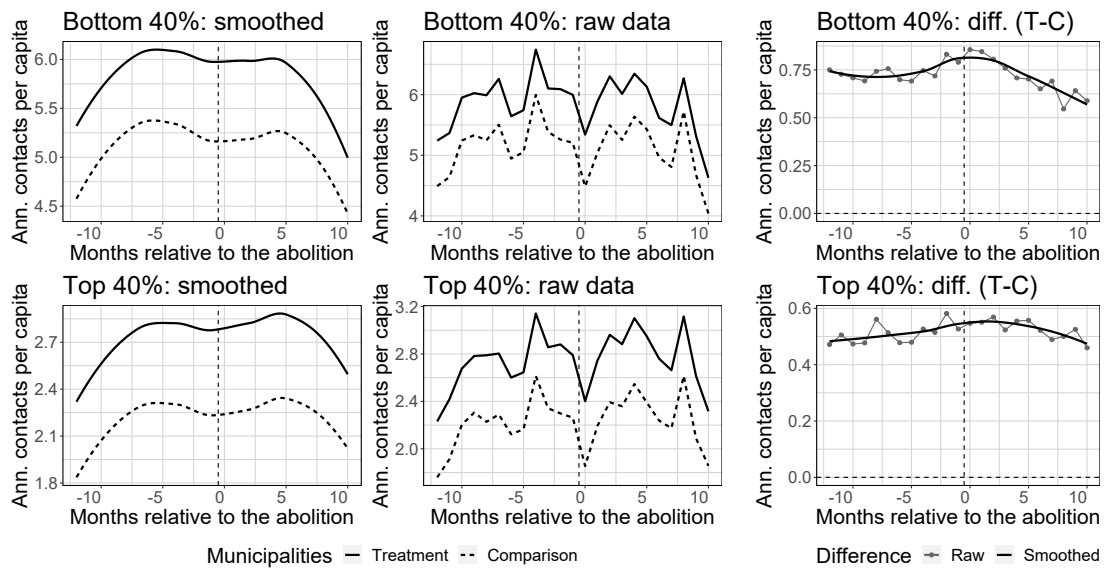


Figure A30: Abolition: Evolution Prescriptions.

*Notes:* The outcome is the number of annualized prescriptions. The left column contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle column, while the difference between treatment and comparison areas is depicted in the right column. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income.

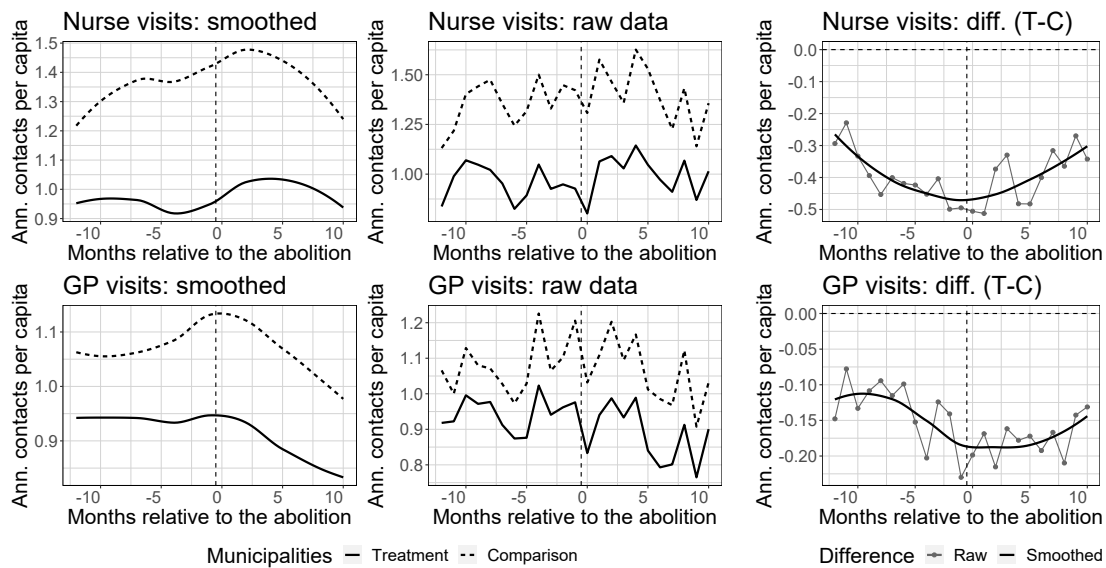


Figure A31: Abolition: Evolution in Outcomes, All Individuals, Uniform Weighting.

*Notes:* This figure was not pre-registered and is post-blind. Municipalities are weighted uniformly instead of population weights. The outcomes are the number of annualized curative nurse visits and GP visits per capita. The left column contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle column, while the difference between treatment and comparison areas is depicted in the right column.

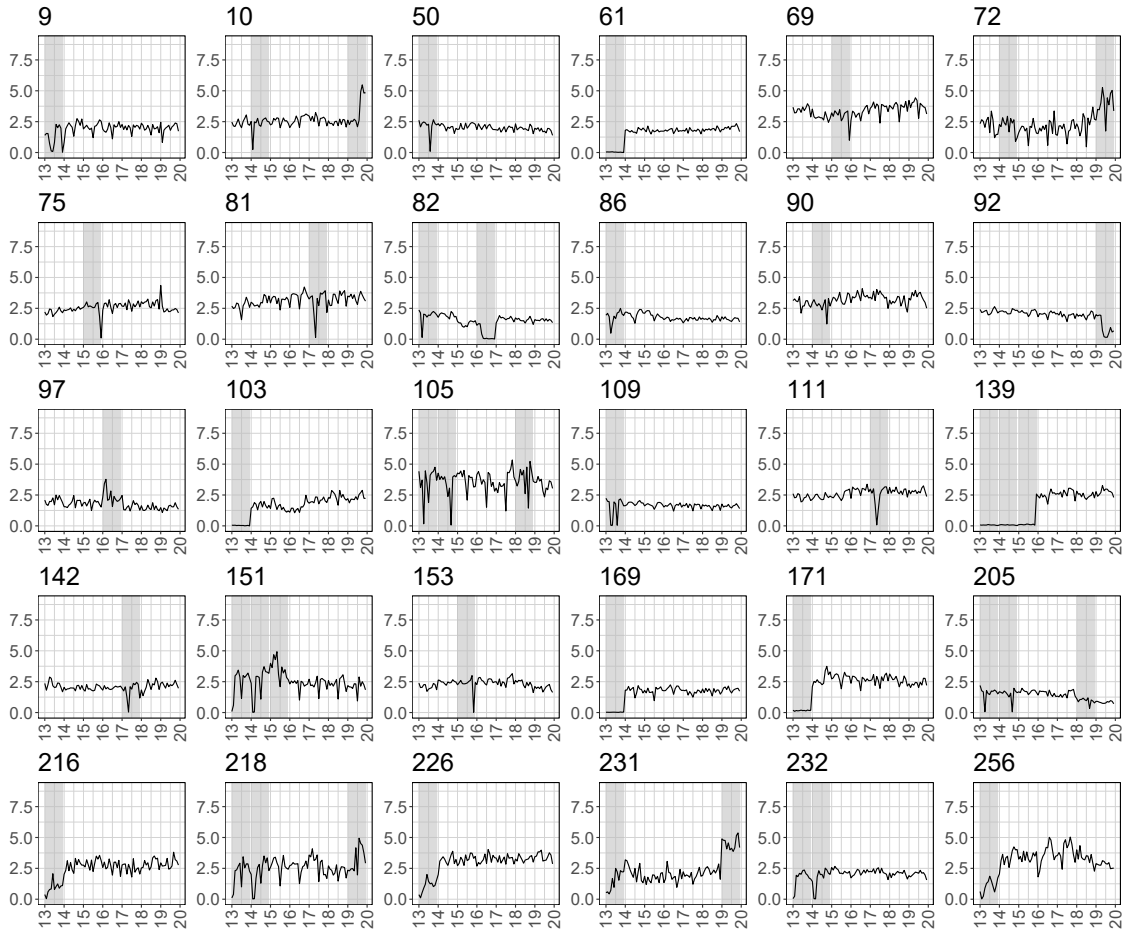


Figure A32: Adoption: Municipalities with Issues in the Primary Care Data, 1.

*Notes:* Using data from 1/2013 to 12/2019, we show the evolution in the annualized number of curative primary care visits (both nurse and GP visits; y axis) over time. Municipality-year observations highlighted by gray show values that are detected by our algorithm (see Section A.3) and that we view as suspiciously low or high and exclude from analyses.

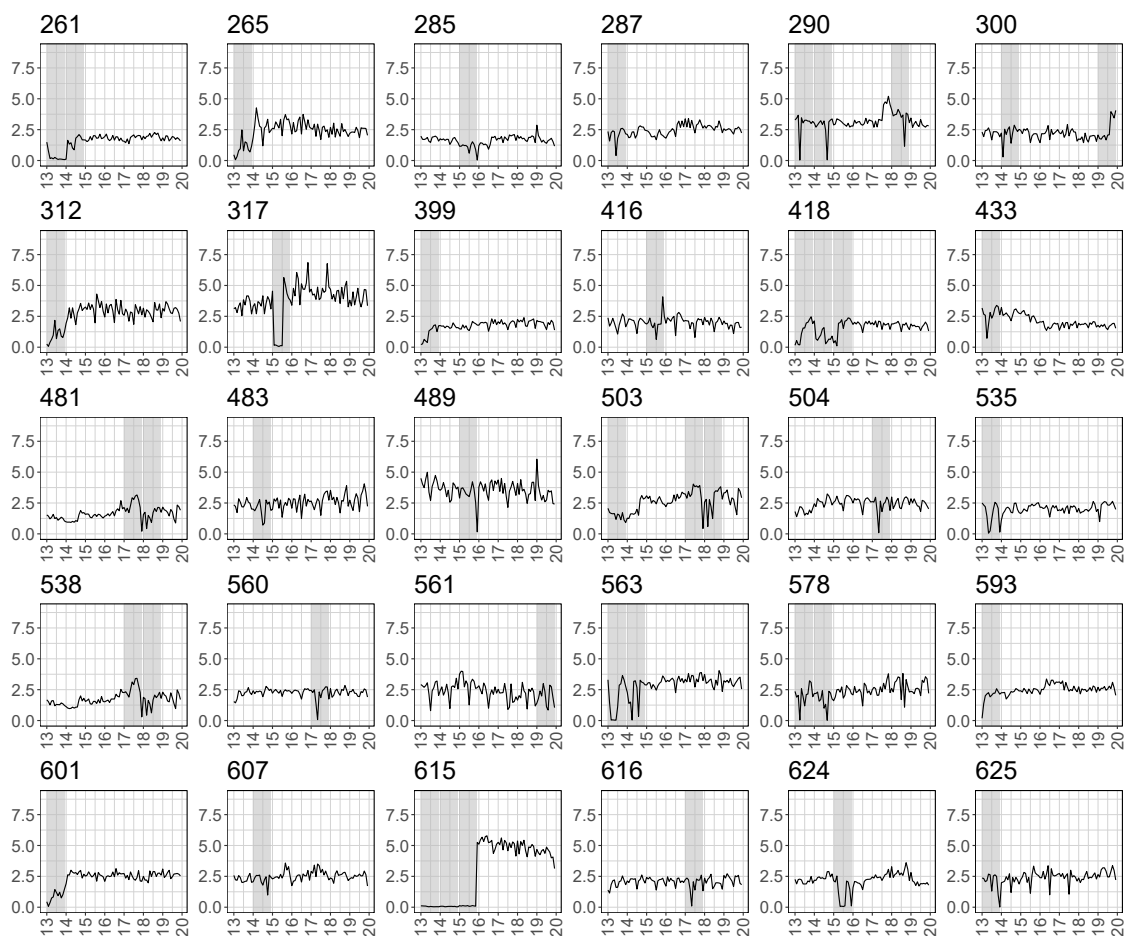


Figure A33: Adoption: Municipalities with Issues in the Primary Care Data, 2.

*Notes:* Using data from 1/2013 to 12/2019, we show the evolution in the annualized number of curative primary care visits (both nurse and GP visits; y axis) over time. Municipality-year observations highlighted by gray show values that are detected by our algorithm (see Section A.3) and that we view as suspiciously low or high and exclude from analyses.

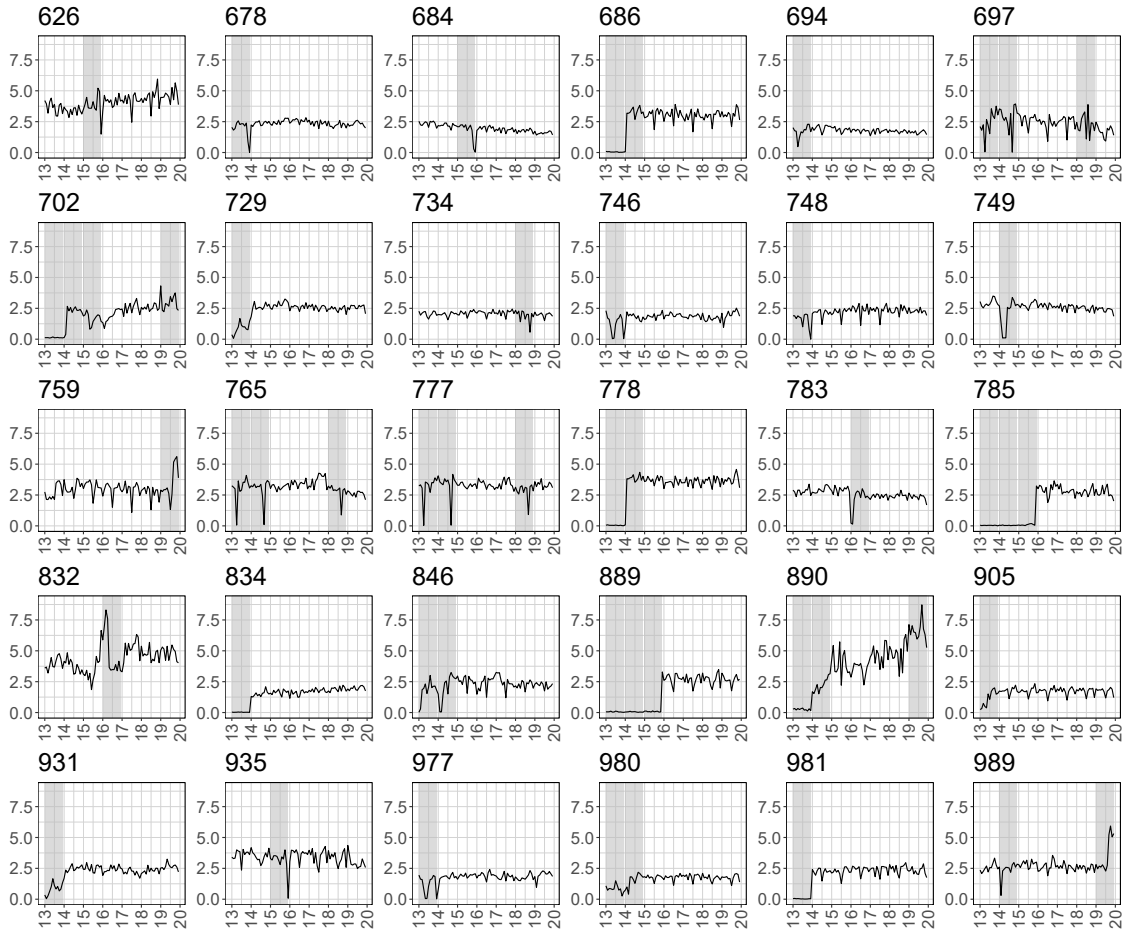


Figure A34: Adoption: Municipalities with Issues in the Primary Care Data, 3.

*Notes:* Using data from 1/2013 to 12/2019, we show the evolution in the annualized number of curative primary care visits (both nurse and GP visits; y axis) over time. Municipality-year observations highlighted by gray show values that are detected by our algorithm (see Section A.3) and that we view as suspiciously low or high and exclude from analyses.



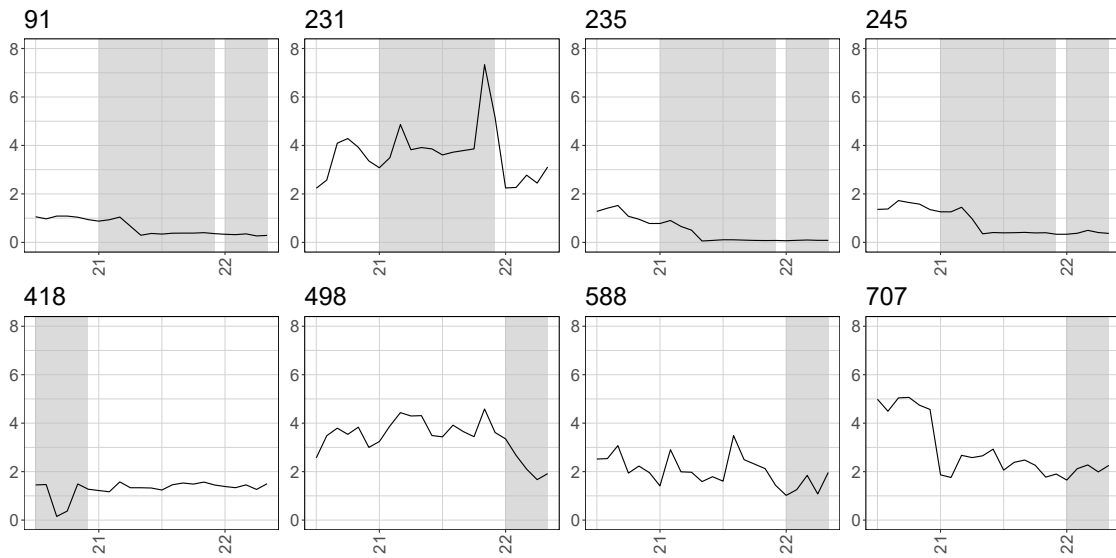


Figure A35: Abolition: Municipalities with Issues in the Primary Care Data.

*Notes:* Using data from 7/2020 to 5/2022, we show the evolution in the annualized number of curative primary care visits (both nurse and GP visits; y axis) over time. Municipality-year observations highlighted by gray show values that are detected by our algorithm (see Section A.3) and that we view as suspiciously low or high and exclude from analyses.