

Cost Sharing and Primary Care Use: Evidence from Staggered Copayment Adoption and Later Abolition

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

April 2022

Abstract

This is a pre-analysis plan and a placebo report of a study that will examine the effects of a staggered adoption and a later simultaneous abolition of a nurse visit copayment on primary care use in adult population using difference-in-differences methods and focusing on the potential heterogeneity by income. Starting from 2014, many Finnish primary care areas gradually adopted a copayment of about 10 euros for curative nurse visits to collect revenue, but in July 2021, these copayments were abolished by the government to reduce barriers to access. We examine the effects of these policies on the number of curative nurse visits and GP visits by income-based subgroups. As secondary outcomes, we analyze the effects of the adoption on receiving social assistance, a last-resort benefit, and the effects of the abolition on the number of prescriptions and referrals to specialized healthcare.

Keywords: Cost sharing, copayments, out-of-pocket costs, healthcare use, primary care, general practitioner, difference-in-differences

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

***Haaga:** Finnish Institute for Health and Welfare (THL), and University of Turku (email: tapio.haaga@thl.fi). **Böckerman:** University of Jyväskylä, Labour Institute for Economic Research, and IZA Institute of Labor Economics (email: 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). **Tukiainen:** University of Turku, and VATT Institute for Economic Research (email: janne.tukiainen@utu.fi). **Acknowledgements:** We thank Mikko Peltola for support, and Lauri Sääksvuori and Jussi Tervola for comments and suggestions. 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>.

Contents

1	Introduction	1
1.1	Introduction to the Study	1
1.2	Pre-Analysis Plan	3
2	Institutional Background	4
3	Data	8
4	Staggered Adoption: Placebo Results	11
5	Simultaneous Abolition: Placebo Results	20
A	Online Appendix	A1

1 Introduction

1.1 Introduction to the Study

Primary care improves health, health system efficiency, and equitable access to healthcare (WHO, 2018). Through its gatekeeping function, scarce medical resources can be allocated more efficiently. However, any needs-based prioritization by healthcare professionals is conditional on patients having contacted the system in the first place. Out-of-pocket costs can ideally save resources if rational and informed patients do not seek care that they believe is of low value. However, out-of-pocket costs are potentially a barrier to entry that affects population groups heterogeneously, and thus, may affect inequality through some population groups having a *de facto* worse access to primary care than others.

We examine whether a small copayment of approximately 10 euros for curative primary care nurse visits affects primary care use in the adult population in Finland, focusing on the potentially heterogeneous effects by income. To this end, we exploit the staggered adoption of the copayment for curative primary care nurse visits in 2014-2019 and its simultaneous abolition in July 2021 by using difference-in-differences (DID) methods and administrative data. Most primary care areas adopted the copayment at some point between 2014 and 2019 to collect more revenue. In July 2021, the law on out-of-pocket costs was reformed, and nurse visits were set to be provided free of charge, thereby abolishing the copayment.

In Finnish primary care, nurses conduct triage and book appointments to nurses and GPs, and general practitioners (GP) authorize access to specialized healthcare. These gatekeeping institutions enable our second contribution: we examine whether copayments for curative nurse visits are causally linked to changes in the use of other services to which access is rationed using gatekeeping and which proxy a professional-assessed need for diagnosis and treatment. Specifically, we not only estimate the effects on the number of curative nurse visits, but also on the number of primary care GP visits (our primary outcomes).

When analyzing the abolition, we also examine the effects on the number of prescriptions and referrals to specialized healthcare. To better understand the financial burden on the low-income individuals, we also assess the effects of the copayment adoption on receiving social assistance, a last-resort benefit which can be used to cover out-of-pocket costs.

Our study stands out from the related literature in several ways. First, our exposure is a copayment for curative primary care nurse visits and one of the two primary outcomes is the number of those visits. Curative nurse visits have a large role in any functioning primary care system. Nurses provide treatment and monitoring for the chronically ill, act as care coordinators collaborating with other professionals, and are for many the first contact to the system. Many other studies examining the effects of out-of-pocket costs on healthcare use totally different types of exposures, such as access to health insurance (Card et al., 2008; Kondo and Shigeoka, 2013) or changes to coinsurance rates (Shigeoka, 2014; Fukushima et al., 2016), and do not focus on primary care and primary care nurse visits. The earlier studies using moderate copayments in primary care have focused on copayments for doctor visits (Nilsson and Paul, 2018; Johansson et al., 2019; Ma and Nolan, 2017).

Second, we focus on the potentially heterogeneous effects by income, which is motivated in three ways. The copayment is small, but it constitutes a larger share of disposable income for low-income individuals. The supply of primary care is rather fixed, and we expect aggregate effects to be small. Observing a reduction in aggregate primary care utilization *per se* tells little about whether valuable care was being missed. If, however, the effects were heterogeneous by income, one could argue that the copayment affects access to services in an unequal way. In DID analyses, we estimate all the models - both main results and complementary analyses - separately in the bottom 40% and the top 40% of the distribution of equivalised family disposable income and also use triple differences models to conduct tests on this heterogeneity. Stratifying by income has not been that common in the literature, potentially due to lack of data on income, but some studies have done so (Johansson et al., 2019; Nilsson and Paul, 2018; Han et al., 2020). In these papers, however,

income has been only one of the potentially many stratifying dimensions, while it is a major theme of our paper - both in main analysis and robustness checks. Note also that the three studies use an age-based regression discontinuity (RD) design to estimate treatment effects for small children and adolescents that are local to a specific age cutoff (they use ages 3, 7, and 20).

Third, we estimate treatment effects for the total adult population and use a staggered DID design with irreversible treatment with methods that are robust to the combination of staggered timing and treatment effect heterogeneity. Age-based RD designs are common in the field: of the above mentioned studies, only Kondo and Shigeoka (2013) and Ma and Nolan (2017) use some other design. RD designs have their own benefits, but the continuity-based RD estimand is only informative at a specific cutoff where the policy changes, not for the total population. Two other studies use a single-event DID design (Chandra et al., 2010, 2014). Studies using staggered DID designs and thus exploiting many events (e.g., Iizuka and Shigeoka, 2021) that we found use conventional TWFE modeling while recent advances in DID and event-study methodologies (Goodman-Bacon, 2021; Sun and Abraham, 2021; Baker et al., 2022) show that these kind of estimates may be severely biased in the presence of treatment effect heterogeneity.

Fourth, we use a registered pre-analysis plan (Olken, 2015), accompanied by codes for replication, that specify in detail how we plan to clean the data, construct our analysis data, conduct analyses, and report results *before* estimating any of the actual results.

1.2 Pre-Analysis Plan

Our justification to use a registered pre-analysis plan (PAP) is twofold. First, we want to separate the writing of statistical programs from the actual analysis to avoid receiving feedback from the actual data, which could possibly lead to an unintentional but systematic bias. If the actual estimates are against the priors, the researcher may look more intensively for coding errors than in the case where estimates are unsurprising and expected. Second,

we want to make a clear distinction between confirmatory and exploratory analyses. By defining the confirmatory analyses and how their results are presented beforehand, we can guard against unintentionally changing the study’s focus because of what we observe in analysis. That is, the PAP reduces the possibility of data-mining.

Specifically, we write the statistical programs with placebo data, therefore expecting null results. For the analysis of the staggered adoption, we use the actual adoption dates to construct policy paths, but we randomize the sample municipalities for these policy paths. Consequently, treatment status (whether the area is ever treated or not) and the potential treatment date are, in expectation, unrelated to the outcome. With the randomized policy variable, we write the statistical programs and modify them based on our observations (e.g. on data quality) without knowing how these changes affect the actual results. Randomization is required because we had access to all the relevant data at the time of writing.

Regarding the copayment abolition on July 1st, 2021, we do not have data from 2021 or 2022 at the time of writing, which can be verified from Statistics Finland who pseudonymize any batch of data before they are available for our remote access use. Instead of the actual post-treatment data, we write the statistical programs as if the abolition occurred three years earlier on July 1st, 2018 (before the COVID-19 pandemic).

This PAP defines the set of confirmatory analyses and illustrates how we plan to present the results. As a supplement, we publish the related statistical programs. We can later add to analysis (e.g. conduct more robustness checks or do exploratory analysis on mechanisms), but these additions will be marked as exploratory in the footnotes of figures and tables and listed in an online appendix.

2 Institutional Background

Primary care services are provided for the adult population by three sectors in Finland: publicly-funded primary care, private clinics, and occupational healthcare. These sectors

essentially target different population groups and differ with respect to gatekeeping, out-of-pocket costs, and wait times. Publicly-funded primary care is the main provider of primary care services for the pensioners, the unemployed and the low-income individuals. Nurses do triage on the phone or at health stations and book appointments to 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 (if the nurse visit is charged). Wait times vary and may be long for nonurgent care.

The employed are entitled to preventive occupational healthcare services, but many employers pay for curative services too. Usually, a professional must be contacted before being allowed to book an appointment. These services are free of charge at the point of use, and wait times are short. Private clinics do not use gatekeeping: one can book an appointment directly to a specialist. The state offers a small reimbursement for these services, but the out-of-pocket costs are still many times higher than in publicly-funded primary care. Private health insurance, which is common especially in families with children, can be purchased to cover the costs. Wait times are short.

Publicly-funded primary care in Finland is organized by municipalities who form primary care areas (officially, health centers) on their own or in cooperation with other municipalities. Everyone has their designated health station, determined by the location of residence. In some areas, all health stations may be available on a visit-by-visit basis. Since 2014, all people have been able to choose their health station once a year, but these changes have been relatively rare. Municipal services are financed through transfers from the state, municipal taxation, copayments, and municipal bonds. The state guides municipal policies on copayments by setting which groups or services are exempted from copayments (in 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 copayment policies.

In July 2010, Finland adopted restricted prescription rights for nurses. Related to

that, the law on copayments was changed to allow primary care areas to charge a copayment for curative nurse visits. As a result, the law no longer specified the professions (e.g. doctors) whose appointments can be subjected to copayments. However, the decree still continued to explicitly mention only doctor appointments, potentially explaining why no areas instantly adopted the nurse visit copayment. Based on our data collection, the copayment was first introduced in January 2014. Many other areas also adopted it once they became aware of the possibility to collect more revenue: the staggered adoption is illustrated in terms of the number of municipalities and population in the top row in Figure 1. By the end of 2019, half the population lived in a municipality charging a copayment for curative nurse visits and a clear majority of the municipalities had adopted the copayment. Figure A1 shows the staggered adoption on maps, plotting policies by municipality at the end of a given year. Based on Figure A2, 80% of the municipalities with the copayment charged it only for the first three visits annually in Summer 2021, and by far the most common level for the per-visit copayment was approximately 11 euros, the population-weighted mean being 12 euros.

In July 2021, the government conducted a broader reform to the act to reduce barriers to access and health inequality. One of the changes was to exempt nurse visits from copayments. Consequently, more than 200 municipalities and almost three million people were affected (see the bottom row in Figure 1). Transfers to municipalities were increased to compensate for reduced copayment revenue. Figure A3 shows the pre-abolition policies by municipality on a map, also containing a bubble plot with bubble size proportional to the population 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. In analysis, we plan to rely on a parallel trends assumption (PTA) (either unconditional or conditioning on a linear pre-trend difference), but we note that the observed unbalance can make the unconditional PTA somewhat less plausible, especially in pandemic times which has affected urban and rural areas heterogeneously. We will later inspect the issue once we have access to all pre-abolition data.

The public primary care system is characterized by supply rigidity and an excess demand for labor. Cohorts in medical schools are fixed in size, and the public sector and the private sector compete for them. Primary care areas have challenges in hiring nurses at the current wage level determined by bargaining between the public sector employers and nurses' union. Resource decisions are done in a context where central and local governments have been running deficits since the financial crisis, which is expected to continue for years due to aging of the population. Consequently, our hypothesis is that the aggregate effects on service use are small but there may be heterogeneity along the income dimension.

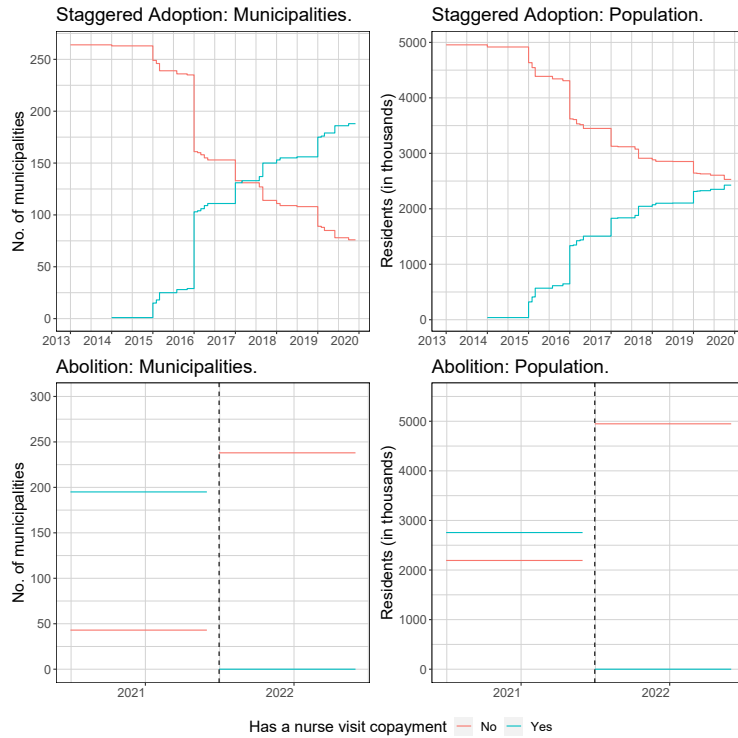


Figure 1: Staggered Adoption and Simultaneous Abolition of the Nurse Visit Copayment.

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 the staggered adoption contains those municipalities whose policies on copayments for curative nurse visits we observe in our data collection. Regarding the abolition, 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.

Several institutions protect the financially vulnerable from healthcare costs. There

are separate annual out-of-pocket caps for public healthcare services and prescription drugs of 692 euros and 592 euros in 2022. Patients themselves should monitor when the cap for healthcare services is reached. People with low enough income and little wealth can apply for a last-resort benefit called social assistance to get support for basic living costs such as out-of-pocket costs for public healthcare services and prescription drugs. The law on out-of-pocket costs in healthcare requires that, for some public services, patients can apply for an exemption of a lowered copayment if they are financially vulnerable. This right does not apply to nurse visit copayments, but some areas may still exempt individuals after application based on their criteria, e.g., the patient having received social assistance. A couple of primary care areas also provide more general exemptions to some low-income groups, such as people with the lowest pension or unemployment benefit.

3 Data

We link several Finnish national administrative registries by unique person IDs. Specifically, we use data on contacts in publicly funded primary care and in specialized healthcare, prescriptions, social assistance recipients, and socioeconomic characteristics of all individuals¹. Findata and Statistics Finland authorize data permissions for the microdata, and empirical work is conducted using Statistics Finland’s remote access system. Data on curative primary care nurse and GP visits will be available to us from 2011 to 2022, data on social assistance from 2012 to 2018, data on prescriptions from 2018 to 2022, and data on referrals to specialized healthcare from 2011 to 2022. However, we exclude years 2011-2012 from the primary care data and years 2011-2017 from the referral data due to

¹Primary care contacts and referrals to specialized healthcare are extracted from the Register of Primary Health Care Visits and the Care Register for Health Care, both administered by the Finnish Institute for Health and Welfare (THL). The data on social assistance recipients come from the Register of Social Assistance, administered by THL. Prescriptions are extracted from the Kanta Prescription Centre, administered by the Social Insurance Institution. The socioeconomic data comprise Statistics Finland’s FOLK modules “basic”, “family”, and “income”.

quality issues².

The socioeconomic data contain all individuals having a permanent residence in Finland at the end of a given year. With these socioeconomic data, we observe age and municipality of residence, which we use to link outcomes to copayment policies, and can construct a variable for equivalised family disposable income, which we use to extract individuals in the bottom 40% and in the top 40% of the income distribution. We estimate the results separately in these income groups throughout the study³. In the Online Appendix, we will also later show the main estimates on primary care outcomes by income quintile and decile to allow for a more flexible analysis on treatment effect heterogeneity by income (not yet done in this PAP). With the sociodemographic data, we also compute population sizes in subgroup-year cells. For years 2012 to 2021, we will use socioeconomic data from the end of the given year. For 2022, however, we will use the values from the end of 2021 due to a data release lag.

Additionally, we use publicly available data on each municipality’s primary care area in 2021⁴ and two THL’s publicly available social and healthcare organization registries (Topi and SOTE) that are linked to primary care contacts. We also create two tables linking areas and copayment policies. The first contains data at the municipality level on whether the municipality had adopted a copayment for curative nurse visits by the end of 2019 and on the possible adoption date. The information was collected by observing municipal documents, websites, and news in local media⁵. The other table, which is on pre-abolition

²Registering primary care contacts nationally started in 2011, and the coding rates in some variables were low at the beginning. Relevant for our study, a dummy distinguishing curative visits from preventive visits, which we need in data extraction, was missing from 7% of the contacts in 2012. Since 2013, that coding rate has been close to one. Regarding the data on referrals, the problem is that the share of contacts with the referral arrival date information included, which is crucial to us, has increased only gradually over time.

³For testing heterogeneity along income, we want two groups instead of more for simplicity and parsimony. Focusing on smaller groups than the bottom 40% and the top 40% has two disadvantages: larger variation in municipal outcomes due to smaller samples, and the fact that the share of social assistance recipients is larger in the bottom end of the income distribution, potentially attenuating estimates as social assistance can be applied to cover healthcare costs. Other researchers may choose another threshold than the 40% figure, but our pre-registered choice is to compare the bottom 40% to the top 40%.

⁴The data are collected by the Association of Finnish Municipalities.

⁵We thank the municipalities for sending us relevant documents in case we could not find the information

policies, was collected by observing the websites of primary care areas in Summer 2021 before the abolition⁶.

Our primary outcomes are the annualized number of curative nurse and GP visits per capita in publicly-funded primary care. The set of secondary outcomes depend on whether we analyze the staggered adoption or the later simultaneous abolition. When analyzing the adoption, we assess the effects on the share of individuals receiving social assistance and the annual sum of received basic social assistance⁷. Regarding the abolition, we examine the effects on the annualized number of prescriptions per capita written by public-sector organizations and on the annualized number of referrals per capita⁸. We refer to the above-mentioned outcomes when we use the following expressions: nurse and GP visits, share of social assistance recipients, sum of basic social assistance, prescriptions, and referrals.

We discuss in detail how we clean and construct our analysis data in Section A.1 in the Online Appendix, also motivating many of our choices for data construction. We only include those individual-by-year observations where the 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. Ultimately, we have an unbalanced panel at the municipality-by-time-period-by-income-decile-by-outcome level. Time period is month for all the other outcomes except for the sum of social assistance which is measured only annually. In analyzing the staggered adoption, we use primary care data from 1/2013 to 12/2019 and social assistance data from 1/2013 to 12/2018, restricting to pre-pandemic data⁹. Regarding the simultaneous abolition, we use data from a one-year bandwidth around the treatment

online.

⁶We thank Katja Ilmarinen who had done the same independently. We compared our table to Ilmarinen's and, thus, verified its contents.

⁷Social assistance is a means-tested last-resort benefit for households. Using data on family relations, we define that an individual received social assistance if the person lived in a family where someone is observed to receive social assistance. Similarly, we sum the amount of social assistance at the family level.

⁸Note, however, that we take only unique ID-date pairs of referrals, the date representing the date of arrival of the referral to specialized healthcare.

⁹The COVID-19 pandemic caused major supply and demand shocks in primary care in Finland, starting in March 2020.

date of July 1st, 2021¹⁰.

The panel is unbalanced for some outcomes because we have to exclude several observations due to quality issues in the primary care data and in the data on referrals. When the national primary care data collection started, not all areas were able to transfer data from their electronic health record systems (EHR) to the national registry. Later changes in the providers of EHR systems may also be visible in the data as a sudden but short drop to a near zero value in aggregate contacts. Regarding the data on referrals, the share of contacts containing information on the arrival date of the referral has improved over time, but is too low in the early 2010s. The details on how we identify those observations with data quality issues - these observations will consequently be excluded - are provided in the Online Appendix in Section A.1.

4 Staggered Adoption: Placebo Results

We start our analysis by focusing on the adoption of the copayment using a staggered difference-in-differences (DD) design with an irreversible treatment. As the baseline, we use a stacked event-by-event design (Gormley and Matsa, 2011; Cengiz et al., 2019) that is robust to biases in conventional two-way fixed effects (TWFE) regression models caused by the combination of staggered treatment timing and treatment effect heterogeneity. Theoretical results show that these biases exist with both static (Goodman-Bacon, 2021) and event study specifications (Sun and Abraham, 2021). Baker et al. (2022) find that they can really matter with real-world data and that stacking is a robust alternative in Monte Carlo simulations.

In stacking, 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 treatment window of 24 months before and 12 months after the copayment adoption¹¹. We only use data from

¹⁰When writing this PAP, we had data up to 12/2020. Therefore, we analyzed a placebo treatment that occurred earlier - in 2018 for primary care contacts and in 2019 for prescriptions and referrals.

¹¹Consequently, we avoid using earlier-treated units as a comparison for later treated units, which is the key problem in conventional static TWFE modeling (Goodman-Bacon, 2021).

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 primary care data quality issues (see Section A.1). In main analysis, we require a balanced panel in event (or relative) time. These event-specific datasets are then stacked (or pooled) together for estimation which we conduct using TWFE regression models but now with dataset-specific unit and time fixed effects. Due to large heterogeneity in municipality size, we weight by population in all regressions¹². Standard errors are clustered by municipality.

Using stacking as the baseline has several benefits. Like the conventional but potentially biased TWFE models, it is based on TWFE regression which can easily accommodate both static and dynamic specifications and triple differences models. The stacked regression estimator is efficient as it uses ordinary least squares to derive the weights on the clean 2x2 DD comparisons in cohort-specific datasets, trading off bias for efficiency, but the use of variance weighting may lead to estimates that are inconsistent for the sample-average ATT (Baker et al., 2022). We also provide extensive robustness checks by using the CS estimator by Callaway and Sant’Anna (2021)¹³.

Note that all of the following results are placebo results. We obtain them by using the actual adoption dates to construct policy paths, but randomize the sample municipalities to these paths. Consequently, treatment status (whether the area is ever treated or not) and the potential treatment date are, in expectation, unrelated to the outcome.

Pre-Trend Plots. Before presenting regression results, we first plot in Figure 2 the trends in curative nurse visits for the bottom 40% and the top 40% of the distribution of equivalised family disposable income in treatment and comparison municipalities using the stacked dataset. The same plots but for curative GP visits and social assistance outcomes are in figures A10 and A11, respectively. The (placebo) intervention does not appear to

¹²We discuss the choice of weighting in Section A.1. In robustness checks, we examine how sensitive the main results are to weighting municipalities uniformly.

¹³Conditional on observing diverging pre-trends in event study plots with the stacked data, we adjust the reporting of the results. In this case, the CS estimator, which has a less stringent parallel trends assumption than a stacked TWFE regression model assuming no pre-trend differences, will receive more weight in the main text. More on this is at the end of this section.

affect the outcome variables. Neither is there evidence of major pre-trend differences.

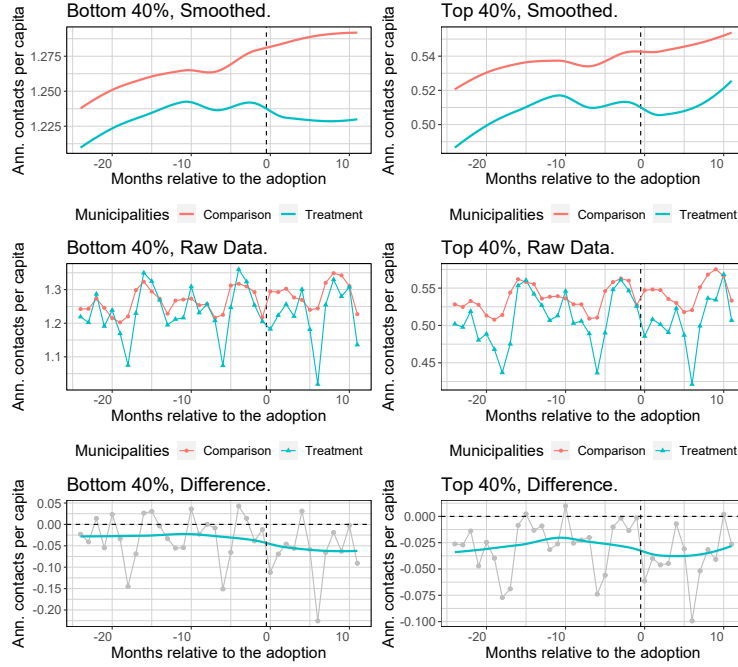


Figure 2: Adoption: Evolution in Nurse Visits.

Notes: These are placebo results - see Section 1.2. 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 top row contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle row, while the difference between treatment and comparison areas is depicted in the bottom row. Bottom 40% and top 40% refer to the distribution of equivalised family disposable income.

We continue assessing possible pre-trend differences by estimating a dynamic regression model on the stacked data, comparing the evolution of annualized contacts per capita between treated and unexposed municipalities. Our TWFE regression model includes a full set of treatment indicators for 24 and 12 months before and after the treatment¹⁴ and event-specific municipality and time fixed effects. Standard errors are clustered by municipality. The event study plots on nurse and GP visits are in Figure 3. For both outcomes and income groups, the leads appear to behave well by being close to zero. An exception is $t = -2$ for nurse visits in the bottom 40%, but that point estimate has wide

¹⁴The last pre-treatment period, namely $t = -1$, is omitted as a reference.

confidence intervals. Lags are somewhat farther from zero, but only two lags out of 48 in total are pointwise significant. The event study plots for our social assistance outcomes are in Figure A12 in the Online Appendix.

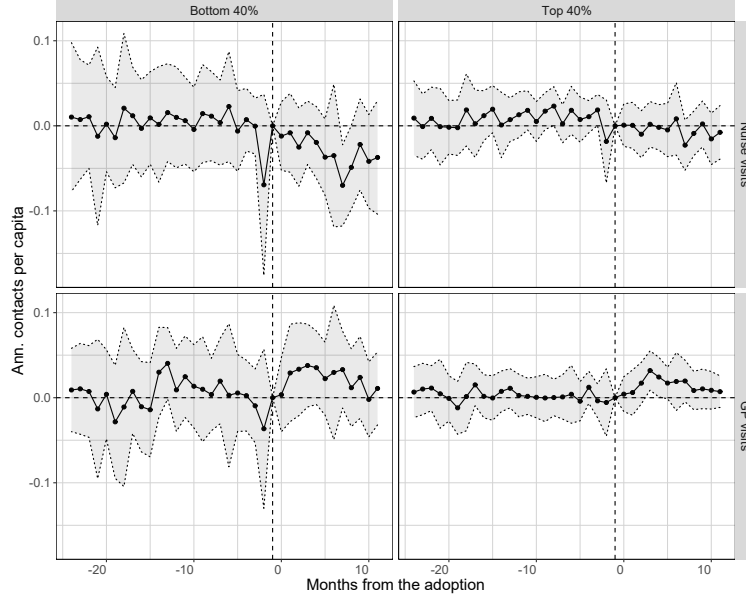


Figure 3: Adoption: Event-Study Plot on Primary Care Use with Stacked Data.

Notes: These are placebo results - see Section 1.2. 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. Our TWFE regression model includes a full set of treatment indicators for 24 and 12 months before and after the treatment and event-specific municipality and time fixed effects, comparing the evolution of annualized contacts per capita between treated and unexposed municipalities. The last pre-treatment period, namely $t = -1$, is omitted as a reference. Due to heterogeneity in municipality size, we weight by population. The standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalised family disposable income.

Main (Placebo) Results. We produce aggregated treatment effect estimates in the static DID framework on the stacked data by using a TWFE regression model that includes an indicator for post-treatment periods in treated municipalities and event-specific municipality and time fixed effects. Standard errors are clustered by municipality. We assume parallel trends. The results on annualized primary care contacts per capita are in Table 1. The point estimates are close to zero and insignificant. The results on social assistance outcomes are in Table A1 in the Online Appendix. They too are close to zero and

insignificant.

In both tables, we also provide the results from a modified specification allowing for pre-trend differences. That is, we replace the static treat x post indicator by lags of every post-treatment period for treated municipalities and by event-specific linear pre-trend differences in relative time between the treated and comparison municipalities. The average of the lags is reported in tables.

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

Metric	Nurse visits		GP visits	
	Bottom 40%	Top 40%	Bottom 40%	Top 40%
Level	1.235	0.509	1.361	0.614
Estimate (w/o trends)	−0.033	−0.013	0.019	0.012
Std. error	0.027	0.014	0.017	0.009
P-value	0.235	0.361	0.250	0.166
Change (%)	−2.644	−2.489	1.411	1.926
Estimate (with trends)	−0.021	−0.014	0.022	0.018
Change (%)	−1.692	−2.818	1.646	2.857
Events	19	19	19	19
Treated areas	150	150	150	150
All areas	246	246	246	246

Notes: These are placebo results - see Section 1.2. The dataset is stacked and balanced. 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 event-specific 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 event-specific linear pre-trend differences in relative time. The mean of the estimated dynamic effects is reported. Due to heterogeneity in municipality size, we weight by population. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equalised family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively.

Robustness Checks. We conduct a series of robustness checks, results being in the Online Appendix. In Panel A, Table A2 reports the stacked regression estimates on annualized primary care contacts per capita from using unbalanced event-specific datasets

that have more municipalities and municipality-month observations than the balanced datasets in main analysis. In Panel B, we present the estimates from using logarithm of annualized contacts per capita as the outcome. The conclusions remain unchanged.

As an alternative to stacking, we also use the CS estimator (Callaway and Sant’Anna, 2021). It too is robust in settings with staggered treatment timing and treatment effect heterogeneity. The key building block in Callaway and Sant’Anna (2021) is a group-time average treatment effect¹⁵, allowing for treatment effect heterogeneity over groups and time. The authors propose several two-step plug-in estimators for these group-time ATEs: first estimate nuisance functions and then plug the fitted values of the nuisance functions into the sample analogue of the group-time ATT. In this application, we use outcome regression, and weight by population and cluster standard errors by municipality. As with stacking, we include events with at least a 12-month follow-up. However, this time we use a balanced dataset in calendar time, not in relative time¹⁶. Using a balanced dataset in calendar time implies excluding those municipalities with issues in the primary care data quality. Our baseline is to exclude years 2013 and 2019 when analyzing primary care outcomes with the CS estimator to increase the number of sample municipalities given that we restrict to a balanced panel¹⁷. Regarding the social assistance outcomes, we use all data from 2012 to 2018. The data are aggregated to the municipality-by-time-period level for estimation.

The group-time ATEs can then be aggregated to construct measures of overall treatment effects such as event-study-type estimates or a single overall treatment effect estimate as in 2x2 DD designs. First, we provide event-study estimates on the primary care outcomes in Figure A13 and on the social assistance outcomes in Figure A14. Comparing

¹⁵Group is defined by the time period when units first receive the treatment.

¹⁶The CS algorithm, which we implemented with the R package *did*, is slow with unbalanced datasets.

¹⁷The exclusion of year 2013 means that we trade off one event and 12 months of data for a greater number of municipalities. In many cases, the primary care visit data quality issues reported in Section A.1 occurred early on in the panel. When 2013 is excluded, more municipalities have a balanced panel in the study period. Exclusion of 2019, on the other hand, lets us keep one large never-treated municipality that changed its electronic health records system in Spring 2019 and had consequently data quality issues.

these to respective figures 3 and A12 based on stacked regressions, we obtain similar results.

Next, we report a single estimate that is the weighted average of all group-time ATEs, weighted by group size¹⁸. Regarding the primary care outcomes, we estimate the effects on annualized visits per capita (in Figure A15) and on logarithm of annualized visits per capita (in Figure A16) in eight cases. We use either the never-treated or the not-yet-treated municipalities as the comparison group and exclude data from either 2013 or 2019 or both. To summarize, all the estimates are close to zero and insignificant. Results on the social assistance outcomes are in Figure A17: the estimates are close to zero and insignificant.

Heterogeneity by Income. Finally, we employ a triple differences (DDD) design with the stacked dataset to provide tests of treatment effect heterogeneity by income. Our hypothesis is that low-income individuals respond more to the copayment changes than high-income individuals. With this motivation, we compare the evolution of outcomes in the bottom 40% of the income distribution to that in the top 40% both in treatment and comparison areas. Parallel trends are assumed in ratios (Olden and Møen, 2022). If we use the (logarithm of the) number of annualized contacts as the outcome, the DDD estimates are unbiased if the pre-trend differences are similar in levels (in percentage terms) in the bottom 40% and in the top 40%. We estimate the effects using the following specification:

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

Here, α is an intercept, m , g , t , and e denote municipality, socioeconomic group, month, and event-specific dataset. $Treat$ is a dummy for treated municipalities, $Affected$ is a dummy

¹⁸Note that there are two differences to the estimand of the main analysis: the weighting procedure and the length of the follow-up. In the main analysis, the effects were estimated with a 12-month follow-up. Here, the follow-up varies by treatment group and is restricted only by the time each group participates in the treatment. This leads to putting more weight on the earlier-treated cohorts.

for the more affected group (the bottom 40%), $Post$ is a dummy for post-treatment periods, γ is the coefficient of interest, and ε is the error term. Note that the other coefficients except for the causal parameter of interest are event-specific. We weight by population and cluster standard errors by municipality.

The results on annualized contacts per capita and logarithmized annualized contacts per capita are in Table 2, using both balanced and unbalanced event-specific datasets. All the estimates are close to zero and insignificant.

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

Metric	Balanced datasets		Unbalanced datasets	
	Nurse Visits	GP Visits	Nurse Visits	GP Visits
A. Annualized contacts per capita				
Level	1.235	1.361	1.253	1.358
Estimate	-0.024	0.006	-0.014	0.012
Std. error	0.015	0.010	0.018	0.012
P-value	0.111	0.541	0.424	0.338
Change (%)	-1.959	0.453	-1.129	0.867
Events	19	19	19	19
Treated areas	150	150	174	174
All areas	246	246	263	263
B. Logarithmized annualized contacts per capita				
Estimate	-0.306	-0.688	-0.212	-0.182
Std. error	1.055	0.732	1.007	0.747
P-value	0.772	0.347	0.834	0.807
Events	18	18	19	19
Treated areas	116	129	174	174
All areas	206	224	263	263

Notes: These are placebo results - see Section 1.2. The dataset is stacked. We use Model 1. Estimates and standard errors are multiplied by 100 if the outcome is logarithm of annualized contacts per capita. Due to heterogeneity in municipality size, we weight by population. Standard errors are clustered by municipality.

Robustness to Weighting Municipalities Uniformly. We repeated the above analyses, but instead of population weighting, we weighted uniformly municipalities when using the CS estimator and municipality-by-income-decile observations when using the TWFE regression. The stacked TWFE regression (both DID and DDD) is robust to the weighting change, leading to estimates that are close to zero and insignificant. The CS estimates on nurse and GP visits are somewhat more sensitive to the weighting change, driving estimates farther from zero. Some of the estimates become significant. Figure A18 shows the event study plot with weighting municipalities uniformly in the CS estimator, while Figure A19 shows the respective aggregated estimates. The rest of the figures and tables are provided in the replication codes folder.

What If There Are Diverging Trends? The reporting above is planned assuming that pre-trend plots and event study plots do not hint of pre-trend differences, as we observed with the randomized placebo treatment assignment. If, however, our actual analysis data shows pre-trend differences, we will adjust our reporting in the following manner. First, we include in the event study plots a fitted linear pre-trend difference based on the lead estimates and extrapolated to the post-treatment periods, assuming that a linear fit well approximates the observed trend difference. These event study plots will then provide an important visual analysis of whether we observe any deviations of the outcome from an extrapolated linear pre-trend difference after the copayment adoption.

Second, we highlight results from the stacked TWFE DID model with a linear pre-trend difference and the CS estimator. We prefer the former method in a case where there is a clear pre-trend difference that is well approximated by a linear fit. The latter is preferable, if the pre-trends are not exactly flat nor can be well approximated with a linear curve. The difference between the parallel trends assumption these methods require is that the TWFE DID with a pre-trend difference assumes parallel trends in deviations of the outcome from a linear time trend while the CS estimator assumes parallel trends but only from the last pre-treatment period on. We then include the baseline CS results - highlighted

in Figure A15 - in main text in Table 1 like in Table 3 in Section 5. We will also highlight our preferred estimates based on pre-trend and event study plots in bold text in Table 1 as we already do in Table 3.

5 Simultaneous Abolition: Placebo Results

To analyze the effects of the simultaneous copayment abolition, we provide estimates using three methods: TWFE DID models with and without a linear pre-trend difference in the model and the CS estimator (Callaway and Sant’Anna, 2021). Each of them makes their own version of the parallel trends assumption in the absence of treatment. TWFE DID without a pre-trend difference assumes parallel trends throughout the whole study window. TWFE DID with a pre-trend difference allows for a linear trend difference, assuming parallel trends in deviations of the outcome from a linear time trend. The CS estimator assumes parallel trends but only from the last pre-treatment period on. Depending on what we observe in the pre-trend plots, we highlight to the reader our preferred estimator and estimate for each outcome. If the pre-trends are clearly flat, TWFE DID without a pre-trend difference is our preferred method. If, however, there is a clear linear pre-trend difference, we prefer TWFE DID with a pre-trend difference. If the pre-trends are not exactly flat nor can be well approximated with a linear curve, we highlight the estimates from the CS estimator. Lastly, our heterogeneity analyses are conducted using triple differences (DDD) models.

We use a one-year bandwidth around the treatment date and a balanced panel^{19,20}. The bandwidth is short, but we do not want to include months from the onset of the COVID-19 pandemic when the healthcare system faced major supply and demand shocks. Note that we assume that the effects of copayment adoption/abolition accumulate fully within a year. For this reason, we exclude those municipalities that adopted the copayment

¹⁹Nurses in six specialized healthcare districts out of 20 went on strike on April 1st, 2022. The strike may lead us to reduce our follow-up from 12 to 9 months. We discuss the strike in Section A.2.

²⁰An exception is the CS estimator which uses treatment group data only from the last pre-treatment period on.

less than 12 months before the start of the study window. In analyses, we weight by population size, and cluster standard errors by municipality.

Note that the following results are placebo results based on time placebos: we proceed as if the abolition had occurred on July 1st, 2018²¹. We use the same assignment to treatment and comparison as we will use in the actual analysis with the following exception: as in main analysis, we exclude municipalities that adopted the copayment less than 12 months before the start of the placebo study window, essentially excluding more municipalities the earlier the placebo intervention year is.

Pre-Trend Plots. Figure 4 plots the trends in curative nurse visits for the bottom 40% and the top 40% of the distribution of equivalised family disposable income in treatment and comparison municipalities. The same plots but for curative GP visits, prescriptions, and referrals to specialized healthcare are in figures A20, A21, and A22, respectively. The (placebo) intervention does not seem to affect the outcome variables. However, it seems like a strong assumption that the pre-trends were exactly parallel. Nurse visits clearly show a diverging trend prior to the placebo treatment, probably referrals and perhaps even prescription too. There is no such a clear pattern for GP visits, though. Consequently, we prefer TWFE DID with a linear pre-trend difference for nurse visits and TWFE DID without a pre-trend difference for GP visits. For prescriptions and referrals, we could consider both the CS estimator and TWFE DID with a linear pre-trend difference. The former, which is our preferred choice, allows us to easily test significance.

Main (Placebo) Results. The TWFE DID specification without a pre-trend difference contains a static indicator for post-treatment periods in treated municipalities, and municipality and time fixed effects. The TWFE DID 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

²¹We have prescription data only from 1/2018, and the quality of the referral data has improved over time. For these reasons, the placebo treatment starts on July 1st, 2019, for drug prescriptions and referrals to specialized healthcare.

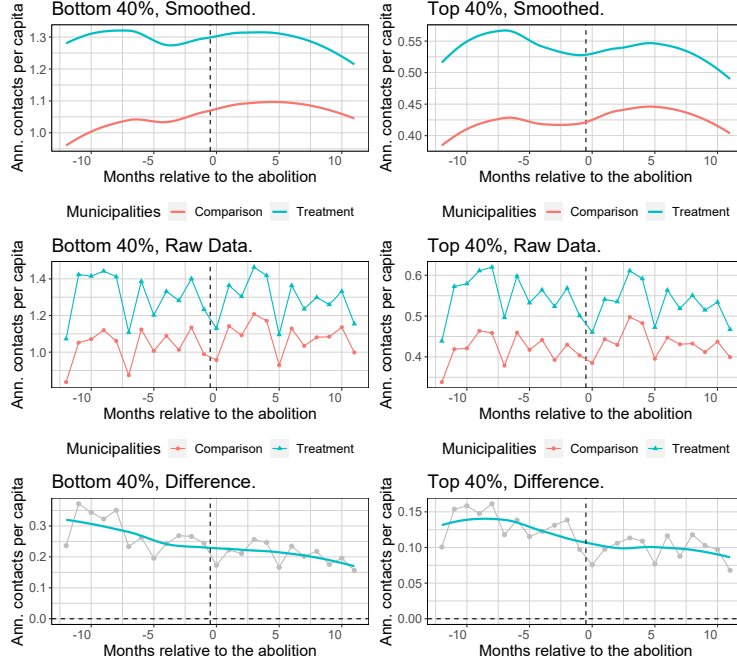


Figure 4: Abolition: Evolution in Nurse Visits.

Notes: These are placebo results: we proceed as if the placebo treatment started on July 1st, 2018. The outcome is the number of annualized curative nurse visits per capita. The top row contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle row, while the difference between treatment and comparison areas is depicted in the bottom row. Bottom 40% and top 40% refer to the distribution of equalised family disposable income.

treated and comparison areas. The average of the lags is reported. The CS estimator is applied similarly than in Section 4: we use outcome regression. With every method, we weight by population and cluster standard errors by municipality.

The results on annualized contacts per capita are in Table 3. Our preferred estimates for nurse visits are the ones from the TWFE DID specification with a linear pre-trend difference. The point estimates are close to zero, mapping to changes of less than 1.5% in absolute value. For GP visits, we prefer the estimates from the TWFE DID model without a pre-trend difference. These estimates are close to zero and insignificant, mapping to increases of less than 2.2%. For prescriptions and referrals, we prefer the CS estimator. The point estimates are close to zero and significant, the largest change being 1.0% in absolute value. The results on the logarithm of annualized contacts per capita are

in Table A3. Of the preferred estimates, only one is significant: the estimate on referrals for the top 40% of the income distribution.

Supplementary Analysis. We provide estimates of primary care outcomes from placebo experiments where we fix the treatment and comparison municipalities and the treatment date (July 1st) but proceed as if the treatment occurred in 2018 or 2019²². From both of these placebo experiments, we exclude municipalities that adopted the copayment in the study period or less than 12 months before the start of the study window²³. The implicit assumption is that the effects of copayment adoption have fully accumulated in a year, so that earlier adoptions do not confound our (placebo) analysis of the abolition. The estimates on primary care outcomes are in Figure A23. We use TWFE DID regression with and without a linear pre-trend difference. The point estimates from the specification without linear pre-trends regarding nurse visits are clearly negative and significant for the bottom 40% of the income distribution in 2018 and 2019, but the point estimates from the specification with linear pre-trends are close to zero. Based on the observed linear pre-trend in Figure 4, the specification with linear pre-trends is clearly more appropriate if one assumes that the pre-trend can be extrapolated to the post-treatment period in the absence of treatment. For the GP visits, however, the assumption of no pre-trends seems more plausible and, thus, we prefer the (placebo) estimates from the specification without pre-trends that are closer to zero than the estimates from the specification with pre-trends.

Heterogeneity by Income. In DDD estimation, we use a similar specification to Model 1, but now we only have a single event and, thus, only one event-specific dataset. The DDD model not only allows us to test heterogeneity by income, but it is also more robust to some violations of the parallel trends assumption. The reason is that the parallel trends assumption is in ratios (Olden and Møen, 2022). If we use the (logarithm of the) number

²²This adds little extra value in this PAP where the main (placebo) effects are estimated similarly, but this is the way we incorporate the time placebo analysis to the final report.

²³This restriction explains why we consider only 2018 and 2019 for placebo experiments. To get a sufficient amount of treatment municipalities, we need to include the cohort that adopted the copayment in 1/2016. The earliest placebo treatment year for this cohort is 2018.

Table 3: Abolition: Main Results.

Metric	Nurse Visits	GP Visits	Prescriptions	Referrals
A. Bottom 40%				
Level	1.309	1.377	5.919	0.686
Estimate (w/o trends)	-0.070	0.021	0.101	0.017
Std. error	0.032	0.020	0.059	0.009
Change (%)	-5.377	1.510	1.701	2.456
Municipalities	161	161	206	174
Estimate (with trends)	0.015	0.045	0.058	-0.007
Change (%)	1.115	3.297	0.983	-0.959
Estimate (CS)	-0.039	-0.050	-0.057	0.001
Std. error	0.034	0.026	0.073	0.009
Change (%)	-2.999	-3.610	-0.967	0.079
B. Top 40%				
Level	0.551	0.643	2.719	0.436
Estimate (w/o trends)	-0.034	0.014	0.063	0.009
Std. error	0.013	0.011	0.025	0.006
Change (%)	-6.171	2.124	2.299	1.962
Municipalities	161	161	206	174
Estimate (with trends)	-0.008	0.001	0.050	-0.014
Change (%)	-1.479	0.104	1.855	-3.246
Estimate (CS)	0.001	-0.014	0.006	-0.003
Std. error	0.014	0.013	0.031	0.005
Change (%)	0.111	-2.246	0.229	-0.650

Notes: These are placebo results: we proceed as if the placebo treatment started on July 1st, 2018 (nurse and GP visits) or 2019 (prescriptions and referrals). 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 and Sant’Anna, 2021). Due to heterogeneity in municipality size, we weight by population. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalised family disposable income. Outcomes are the annualized number contacts per capita. Our preferred estimate for each outcome is highlighted by bold text and is chosen based on pre-trend plots.

of annualized contacts as the outcome, the DDD estimates are unbiased if the pre-trend differences are similar in levels (in percentage terms) in the bottom 40% and in the top 40%. At least for the nurse visits (see Figure 4), the similar pre-trend differences in percentage terms is a more plausible assumption.

The results are in Table 4. For nurse visits, both estimates are insignificant, the one on the logarithm of annualized contacts being closer to zero. For GP visits, both estimates are insignificant, but the one on the annualized contacts is closer to zero. Regarding prescriptions and referrals, all of the point estimates are close to zero. Only the estimate on the logarithm of annualized referrals is significant. We also conduct the same time placebo exercise as with DD models and report the estimates on primary care outcomes in Figure A24.

Table 4: Abolition: DDD Comparisons.

Metric	Nurse Visits	GP Visits	Prescriptions	Referrals
A. Annualized contacts per capita				
Level	1.309	1.377	5.919	0.686
Estimate	-0.039	0.007	0.040	0.007
Std. error	0.021	0.012	0.051	0.006
P-value	0.070	0.551	0.427	0.217
Change (%)	-2.970	0.502	0.679	1.079
Municipalities	161	161	206	174
B. Logarithmized annualized contacts per capita				
Estimate	-0.807	-1.524	-0.902	2.083
Std. error	1.224	0.892	0.964	0.999
P-value	0.510	0.087	0.350	0.037
Municipalities	161	161	206	174

Notes: These are placebo results: we proceed as if the placebo treatment started on July 1st, 2018 (nurse and GP visits) or 2019 (prescriptions and referrals). We use Model 1 but without event-specific parameters as we analyze the simultaneous abolition. Estimates and standard errors are multiplied by 100 if the outcome is logarithm of annualized contacts per capita. Due to heterogeneity in municipality size, we weight by population. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalised family disposable income.

Robustness to Weighting Municipalities Uniformly. We repeated the above analyses, but instead of population weighting, we weighted uniformly municipalities when using the CS estimator and municipality-by-income-decile observations when using the TWFE regression. Having examined the pre-trend plots, the preferred estimation methods for a given outcome chosen in the main analysis based on the population-weighted data would no longer be the preferred methods with the data weighting municipalities uniformly. We would now prefer the TWFE regression without pre-trends for nurse and GP visits and referrals and the TWFE regression with a linear pre-trend for prescriptions. These DID estimates are in Table A4. Although the preferred estimates are insignificant and relatively close to zero (with the exception of referrals), they are now in absolute value farther from zero than the preferred estimates in the main analysis with population weights. The DDD estimates in Table A5 are relatively robust, the estimate on the logarithm of annualized referrals being the exception. The rest of the figures and tables are provided in the replication codes folder.

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

References

- Baker, A. C., Larcker, D. F., and Wang, C. C. Y. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395.
- Callaway, B. and Sant’Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.
- Card, D., Dobkin, C., and Maestas, N. (2008). The Impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare. *The American Economic Review*, 98(5):2242–2258.
- Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.
- Chandra, A., Gruber, J., and McKnight, R. (2010). Patient Cost-Sharing and Hospitalization Offsets in the Elderly. *American Economic Review*, 100(1):193–213.
- Chandra, A., Gruber, J., and McKnight, R. (2014). The impact of patient cost-sharing on low-income populations: Evidence from Massachusetts. *Journal of Health Economics*, 33:57–66.
- Fukushima, K., Mizuoka, S., Yamamoto, S., and Iizuka, T. (2016). Patient cost sharing and medical expenditures for the Elderly. *Journal of Health Economics*, 45:115–130.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*.
- Gormley, T. A. and Matsa, D. A. (2011). Growing out of trouble? Corporate responses to liability risk. *The Review of Financial Studies*, 24(8):2781–2821.
- Han, H.-W., Lien, H.-M., and 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(3):238–278.

- Iizuka, T. and Shigeoka, H. (2021). Is Zero a Special Price? Evidence from Child Healthcare. *American Economic Journal: Applied Economics* (forthcoming).
- Johansson, N., Jakobsson, N., and 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(8):1271–1280.
- Kondo, A. and Shigeoka, H. (2013). Effects of universal health insurance on health care utilization, and supply-side responses: Evidence from Japan. *Journal of Public Economics*, 99:1–23.
- Ma, Y. and Nolan, A. (2017). Public Healthcare Entitlements and Healthcare Utilisation among the Older Population in Ireland. *Health Economics*, 26(11):1412–1428.
- Nilsson, A. and Paul, A. (2018). Patient cost-sharing, socioeconomic status, and children’s health care utilization. *Journal of Health Economics*, 59:109–124.
- Olden, A. and Møen, J. (2022). The triple difference estimator. *The Econometrics Journal*.
- Olken, B. A. (2015). Promises and Perils of Pre-analysis Plans. *Journal of Economic Perspectives*, 29(3):61–80.
- Shigeoka, H. (2014). The Effect of Patient Cost Sharing on Utilization, Health, and Risk Protection. *American Economic Review*, 104(7):2152–2184.
- Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.
- WHO (2018). Building the economic case for primary health care: a scoping review. Technical report.

A Online Appendix

A.1 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 others 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.

We ultimately have 264 out of 293 municipalities in mainland Finland for a study on staggered adoption. Together, they had 5.0 million residents compared to the population of 5.5 million by the end of 2019. With respect to the simultaneous abolition, we have 249 municipalities with 5.0 million residents. Figure 1 illustrates the staggered adoption and the abolition in terms of treatment areas and treated population. Figures A1 (the staggered adoption) and A3 (the abolition) show the municipal policies on map. Figure A2 shows how the copayment level varied in Summer 2021.

Socioeconomic data: We exclude those ID-year observations where equivalised 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 equivalised family disposable income and sort individuals to 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 only. 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. Consequently, we start our analysis from 1/2013. The coding rate for profession containing both nurses and doctors has been steady at least since 2012, the other group containing other professions and missing values and constituting approximately 6% of the contacts.

We exclude weekend visits from analysis to reduce a potential bias stemming from changes in how urgent and emergency care visits are coded to the registers. During the study period, some primary care areas and hospitals have formed joint urgent care clinics, and these contacts may be coded either to the primary care registry or to the specialized healthcare registry. Duplicate contacts are dropped. That is, an individual cannot have more than one curative visit at 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 at private clinics which we want to exclude from the data. We do this by linking each visit in 2019-2022 to TOPI and SOTE organization registers that contain information on the provider of the visit. Both the organization registers are continuously updated. We have an annual cross-sectional dataset on TOPI and a cross-section on SOTE from early 2020. In 2019-2020, the linking of TOPI does not work for 4% of the rows while the same figure for SOTE is 1%. Then, we include

those visits whose provider a) has a TOPI service area code referring to health centers (120, 121, or 122) or b) is a public-sector organization in SOTE²⁴.

After having aggregated primary care contacts to municipality-month observations, we need to exclude several observations due to quality issues. Not all areas were able to transfer data from their electronic health record systems (EHR) to the national registry when the national primary care data collection started in 2011. Changes in the providers of EHR systems may also be visible in the data as a sudden drop to a near zero value in aggregate contacts.

To identify the suspiciously low values of primary care 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 of these means is our reference value with which we define an observation to be suspicious if its value is less than 40% of the largest mean. July, however, is not considered because many people - both professionals and patients - have holidays. We mark all municipality-year pairs to be excluded if they contain suspicious months. We conduct the algorithm separately before (from 1/2013 to 12/2019) and after (from 7/2020 to 6/2022) the onset of the COVID-19 pandemic.

Ultimately, 86 municipalities (out of 293) have suspiciously low values of primary care use in the pre-pandemic study period. The evolution of the sum of curative nurse and GP visits in these municipalities are illustrated in figures A4, A5, and A6. Pink segments highlight municipality-year pairs with suspiciously low values. Based on visual inspection, the algorithm appears to be good at detecting irregularities.

Regarding the analysis of the abolition, we do not have the actual data yet, but we illustrate how the algorithm works by focusing on a two-year window centered at July

²⁴There 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 at health centers should have a TOPI service area code referring to health centers.

1st, 2018, the date of our placebo intervention. In this case, only 13 municipalities have suspiciously low values of primary care use (Figure A7).

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 for 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. The abolition of the nurse visits copayment probably decreased both the number of social assistance recipients and the sum of 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 with primary care contacts. The algorithm comes up with 19 municipalities with weird 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 (excluding cancellations and edits of prescriptions), 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, but we have some doubts about the quality of the variable and decide 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 look for potentially missing values with the same algorithm as with primary care contacts, but the algorithm does not find any irregularities.

Referrals to specialized healthcare: We extract specialized healthcare contacts where the date of arrival of the referral is observed. When writing this PAP, the raw data appear to include only contacts that have led to actual events, such as outpatient visits,

hospitalizations, or surgeries. For the final report, we plan to update the data to include both contacts that led to actual events and contacts where the referral is processed and the patient is in the wait list. The referrals come from both public and private sector providers. We would like to restrict to referrals written by public sector providers, but coding rates in the relevant variable were too low pre-pandemic. With the updated data from 2020-2022, we will assess the quality again.

Of the referrals, we take unique ID-by-arrival-date observations before aggregating them to municipality-by-month level. That is, we do not count the aggregate number of referrals, but instead the number of distinct referral arrival dates. This is because the unique referral ID is rarely coded.

We observe major data quality issues with respect to the coding rate of referral arrival date in early 2010s. The number of extracted referral arrival dates almost doubled between 2012-2018 due to increasing coding rates. For this reason, we do not use the referral data to analyze the staggered copayment adoption in 2014-2019, but we plan to use the data to analyze the simultaneous abolition in 2021.

In this PAP, we analyze a placebo treatment in 2019. We look for potentially missing values with the same algorithm as with primary care contacts, but with a larger threshold that makes the algorithm more sensitive to irregularities. That is, we define municipality-month observations to be suspicious if its value is less than 60% of the largest mean (instead of 40%). In this case, 18 municipalities have suspiciously low values, illustrated in Figure A8.

The COVID-19 pandemic: At the onset of the pandemic, healthcare systems all around the world faced major supply and demand shocks. In Finland, a state of emergency was announced in March 2020. People decreased mobility and contacts due to containment policies and the fear of the virus, potentially even avoiding appropriate healthcare use. At the same time, elective surgeries and non-urgent appointments were being cancelled or postponed as the healthcare system allocated more labor to diagnose and treat patients with

COVID-19. As a result, Finland observed a sudden and drastic reduction in healthcare contacts, which is illustrated for curative primary care nurse visits in Figure A9. By the end of the summer, aggregate contacts had partially recovered. In the second half of 2020, the number of nurse visits was approximately -15% relative to the 2018-2019 baseline.

Regarding the copayment abolition, we exclude the period with aggregate primary care use much lower than normally as inclusion could create spurious pre-trend differences given that areas were differently impacted by the pandemic. Consequently, we choose to analyze the abolition using a two-year window centered on July 1st, 2021, the abolition date. Thus, the first pandemic wave between 3/2020 and 6/2020 is excluded.

Weighting by population. In the main analysis, we weight all regressions by population size in order to increase 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 more and focus on the bottom 40% or the top 40% of the income distribution. Therefore, it is obvious that the outcomes of small municipalities are 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 estimating the results with the TWFE regression (stacked or not). 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 in addition to the replication codes. 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.2 Potential Changes to the Pre-Analysis Plan

Methods. Our causal inference relies heavily on different versions of the parallel trends assumption (PTA), requiring that outcomes in treatment and comparison groups would have evolved similarly in the absence of treatment. Although the assumption is inherently untestable, event study plots are a conventional way to assess whether there is a clear pre-treatment difference in trends. Section 5 of this PAP illustrates that we may need to adjust our analysis and reporting depending on what we observe in pre-trend plots.

When analyzing both the adoption and the abolition, all result tables except for the DDD results contain estimates from a regression specification that includes a linear pre-trend difference. The point estimate is a mean of several lags, one for each post-treatment time period. Currently, we do not provide inference regarding the aggregated estimate, but we aim to add some bootstrap-based solution to estimate standard errors of these aggregated estimates in the final report. This would be a priority if the event-study plots lead us to prefer the specification that includes a linear pre-trend difference.

One open question is whether our algorithm to detect issues in the primary care data, that works well with pre-pandemic data, is satisfactory with data from the COVID-19 pandemic. Luckily, we expect that there were few of these issues in 2020-2022. Still, we may need to modify the algorithm to the pandemic times. We will make these potential changes before estimating the actual results.

Data and data quality. Three municipalities adopted a new EHR system on April 24th, 2021. Among them was Helsinki, the largest Finnish municipality, that provides many comparison individuals to the analysis of the copayment abolition. We need to verify

that the areas were able to continue transferring data to the national registry or otherwise exclude them from the analysis. Vantaa adopted the same EHR system in 2019 and could not transfer data right after the change.

The latest statistical year in our socioeconomic data is currently 2019. We expect that the socioeconomic data from the end of 2021 will be available by June 2023. In the final research report, we will use values from the end of 2021 for year 2022. Before we obtain the data from 2021, we use values from the end of 2020 for 2020-2022.

Follow-up length. On April 1st, 2022, approximately 25,000 nurses (the figure was released by their union) went on strike in six specialized healthcare districts out of 20 to demand higher wages. The strike was planned to extend to cover 13 specialized healthcare districts and 40,000 nurses by April 15th unless an agreement was reached earlier. Although the strike directly affects specialized healthcare and not public primary care, there are plausible indirect effects to the latter. First, the strike reduced the capacity of emergency departments at hospitals. Some patients visit these emergency departments even if they could be treated in public primary care. The strike plausibly increased demand for public primary care. Second, patients may find it harder to get a referral to specialized healthcare from public primary care, if the doctors take into account the situation in hospitals and their gatekeeping is stricter than in normal times. We do not know at the time of writing how the situation evolves. If the strike continues and extends, we may need to reduce our planned 12-month follow-up (July 2021 - June 2022) to a 9-month follow-up (July 2021 - March 2021), or use the shorter follow-up as a robustness check.

Additional figures and tables. We note that the government increased the maximum copayment for GP visits from 14.70 euros in 2014 to 16.10 euros in 2015 and to 20.90 euros in 2016. To our understanding, all municipalities except Helsinki, where no GP visit copayment is charged, made the first increase in 1/2015. Municipalities reacted differently to the latter increase: many made it instantly in 1/2016, some made it later, and a couple of areas have not made the increase by 2022. If the changes occur at the same time,

the GP visit copayment increases can increase the estimated behavioral effects of the nurse visit copayment adoption. At minimum, we will later provide a plot showing when the GP visit copayment increases occurred for each municipality relative to the nurse visit adoption. If considered important, we will conduct more supplementary analysis to address the topic.

In the main analysis and robustness checks, we estimate the results separately in the bottom 40% and the top 40% of the distribution of equivalised family disposable income. In the Online Appendix, we will also later show the main estimates on primary care outcomes by income quintile and decile to allow for a more flexible analysis on treatment effect heterogeneity by income (not yet done in this PAP).

In the Online Appendix of the final research report, we will list and report the changes in data construction and analysis that we take compared to this PAP. These changes may, e.g., be related to the above aspects, or they can be new supplementary analyses or robustness checks.

A.3 Additional Figures and Tables

Table A1: Adoption: Social Assistance Use.

Metric	Share receiving	Euros received
Level	3.416	187.087
Estimate (w/o trends)	0.005	-2.302
Std. error	0.053	3.118
P-value	0.925	0.461
Change (%)	0.146	-1.230
Estimate (with trends)	-0.005	-3.056
Change (%)	-0.137	-1.633
Events	16	5
Treated areas	152	112
All areas	264	264

Notes: These are placebo results - see Section 1.2. The dataset is stacked and balanced. 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 event-specific 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 event-specific linear pre-trend differences in relative time. The mean of the estimated dynamic effects is reported. Due to heterogeneity in municipality size, we weight by population. 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 received basic social assistance (in euros). With the latter outcome, we only include events that occurred on January 1st.

Table A2: 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	1.253	0.517	1.358	0.615
Estimate (w/o trends)	−0.024	−0.008	0.022	0.012
Std. error	0.026	0.013	0.018	0.009
P-value	0.356	0.528	0.211	0.154
Change (%)	−1.909	−1.615	1.638	2.021
Estimate (with trends)	−0.014	−0.013	0.031	0.021
Change (%)	−1.101	−2.474	2.263	3.372
Events	19	19	19	19
Treated areas	174	174	174	174
All areas	263	263	263	263
B. Balanced data, logarithmized outcome				
Estimate (w/o trends)	−1.546	−1.820	1.546	2.012
Std. error	2.068	2.497	1.289	1.499
P-value	0.456	0.467	0.232	0.181
Estimate (with trends)	−0.600	−3.069	2.809	3.905
Events	18	18	18	18
Treated areas	116	116	129	129
All areas	206	206	224	224

Notes: These are placebo results - see Section 1.2. The dataset is stacked. 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 event-specific 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 event-specific linear pre-trend differences in relative time. The mean of the estimated dynamic effects is reported. Due to heterogeneity in municipality size, we weight by population. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equalised family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively.

Table A3: Abolition: Logarithmized Outcome.

Metric	Nurse Visits	GP Visits	Prescriptions	Referrals
A. Bottom 40%				
Estimate (w/o trends)	-7.947	1.814	1.238	3.213
Std. error	2.970	1.709	0.979	1.583
P-value	0.008	0.290	0.208	0.044
Municipalities	161	161	206	174
Estimate (with trends)	-0.057	3.671	1.625	-0.399
Estimate (CS)	-5.686	-3.897	-1.780	-1.722
Std. error	2.981	2.182	1.244	1.290
B. Top 40%				
Estimate	-7.326	3.376	2.175	0.971
Std. error	2.901	2.084	1.034	1.594
P-value	0.013	0.107	0.037	0.543
Municipalities	161	161	206	174
Estimate (trends)	-3.571	1.733	4.158	-2.462
Estimate (CS)	-0.795	-2.568	-0.553	-3.059
Std. error	2.437	2.338	1.275	1.447

Notes: These are placebo results: we proceed as if the placebo treatment started on July 1st, 2018 (nurse and GP visits) or 2019 (prescriptions and referrals). 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 and Sant’Anna, 2021). Due to heterogeneity in municipality size, we weight by population. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalised family disposable income. Outcomes are the logarithm of the annualized number contacts per capita. Our preferred estimate for each outcome is highlighted by bold text and is chosen based on pre-trend plots.

Table A4: Abolition: Main Results, Uniform Weighting.

Metric	Nurse Visits	GP Visits	Prescriptions	Referrals
A. Bottom 40%				
Level	1.492	1.530	6.112	0.629
Estimate (w/o trends)	-0.040	-0.034	-0.115	-0.004
Std. error	0.066	0.041	0.067	0.008
Change (%)	-2.700	-2.234	-1.875	-0.626
Municipalities	161	161	206	174
Estimate (with trends)	-0.006	-0.042	-0.164	-0.004
Change (%)	-0.401	-2.751	-2.683	-0.632
Estimate (CS)	-0.099	-0.080	-0.196	-0.003
Std. error	0.064	0.045	0.079	0.014
Change (%)	-6.589	-5.253	-3.197	-0.491
B. Top 40%				
Level	0.726	0.839	3.166	0.430
Estimate (w/o trends)	-0.018	-0.033	-0.047	-0.022
Std. error	0.032	0.025	0.046	0.010
Change (%)	-2.437	-3.928	-1.484	-5.085
Municipalities	161	161	206	174
Estimate (with trends)	0.014	-0.009	0.009	-0.025
Change (%)	1.976	-1.026	0.284	-5.864
Estimate (CS)	-0.001	-0.030	-0.058	-0.027
Std. error	0.041	0.036	0.053	0.015
Change (%)	-0.133	-3.563	-1.825	-6.197

Notes: These are placebo results: we proceed as if the placebo treatment started on July 1st, 2018 (nurse and GP visits) or 2019 (prescriptions and referrals). 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 and Sant’Anna, 2021). The CS estimator weights municipalities uniformly; the TWFE regressions weights municipality-by-income-decile observations uniformly. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalised family disposable income. Outcomes are the annualized number contacts per capita. Our preferred estimate for each outcome is highlighted by bold text and is chosen based on pre-trend plots.

Table A5: Abolition: DDD Comparisons, Uniform Weighting.

Metric	Nurse Visits	GP Visits	Prescriptions	Referrals
A. Annualized contacts per capita				
Level	1.492	1.530	6.112	0.629
Estimate	−0.023	−0.001	−0.068	0.018
Std. error	0.041	0.026	0.056	0.011
P-value	0.584	0.962	0.226	0.106
Change (%)	−1.514	−0.081	−1.107	2.851
Municipalities	161	161	206	174
B. Logarithmized annualized contacts per capita				
Estimate	−2.837	−0.269	−0.596	7.881
Std. error	1.861	1.978	1.126	2.680
P-value	0.127	0.892	0.597	0.003
Municipalities	161	161	206	174

Notes: These are placebo results: we proceed as if the placebo treatment started on July 1st, 2018 (nurse and GP visits) or 2019 (prescriptions and referrals. We use Model 1 but without event-specific parameters as we analyze the simultaneous abolition. Estimates and standard errors are multiplied by 100 if the outcome is logarithm of annualized contacts per capita. Municipalities are weighted uniformly in regressions. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equalised family disposable income.

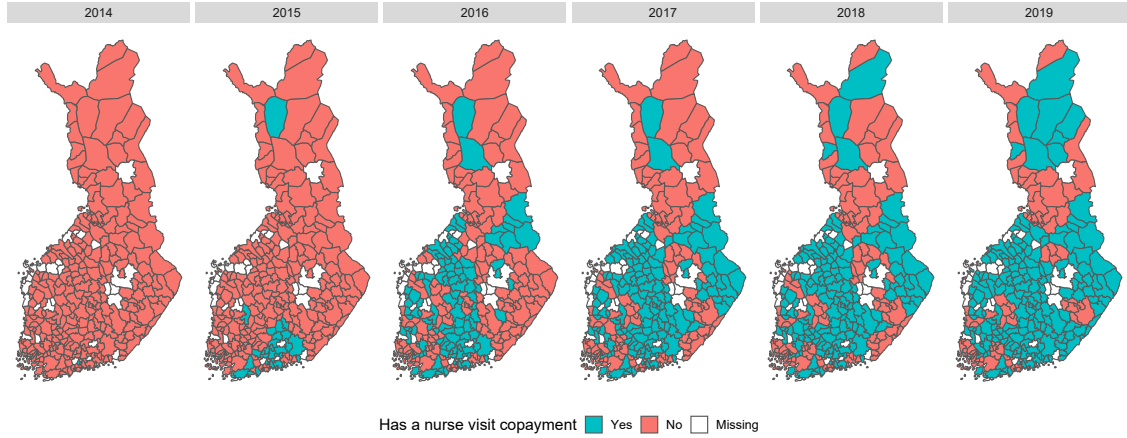


Figure A1: Staggered Adoption on Map.

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.

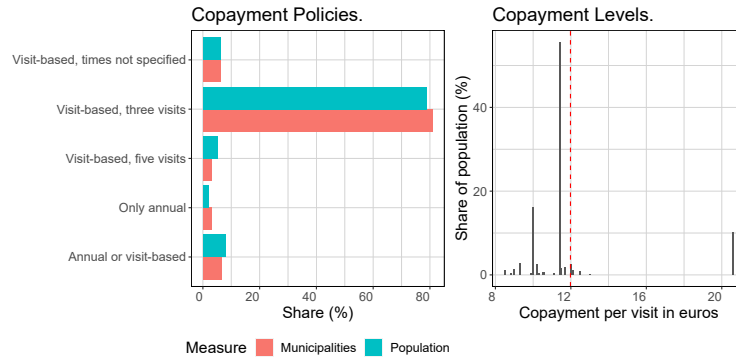


Figure A2: Copayment Levels and Policies in Summer 2021.

Notes: The plot shows how common different copayment options and levels were in Summer 2021 among the copayment municipalities of Section 5. 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.1 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.

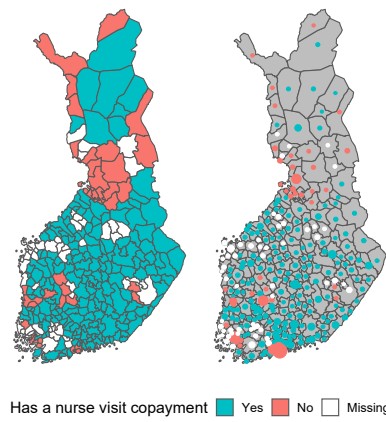


Figure A3: The Abolition on Map.

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.1 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.

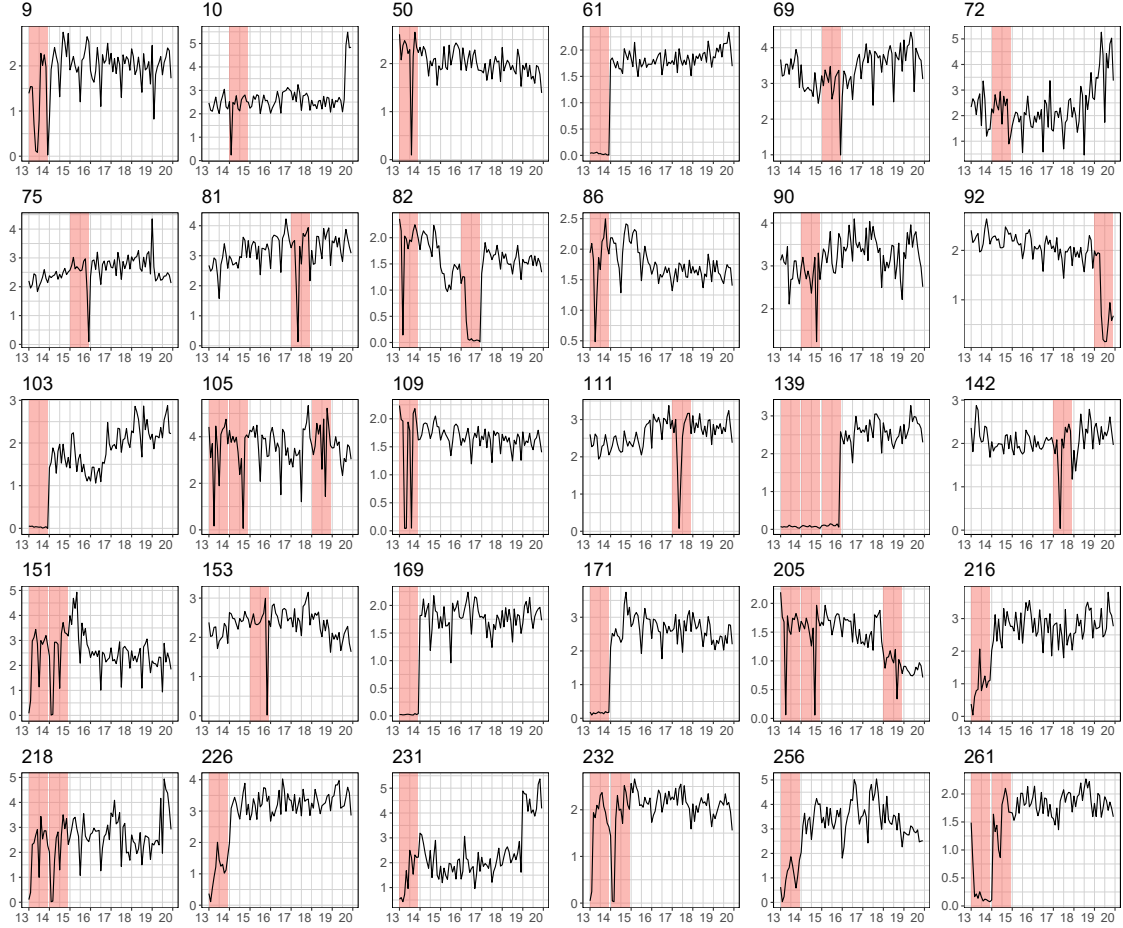


Figure A4: Adoption: Municipalities with Issues in the Primary Care Data, 1.

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. We first obtain a reference by deriving a distribution of mean contacts by permutationally excluding every combination of four consecutive months, then average, and ultimately choose the largest mean. We define a value suspiciously low if it is less than 40% of the largest mean. Municipality-year observations with suspicious values are highlighted by pink.



Figure A5: 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. We first obtain a reference by deriving a distribution of mean contacts by permutationally excluding every combination of four consecutive months, then average, and ultimately choose the largest mean. We define a value suspiciously low if it is less than 40% of the largest mean. Municipality-year observations with suspicious values are highlighted by pink.

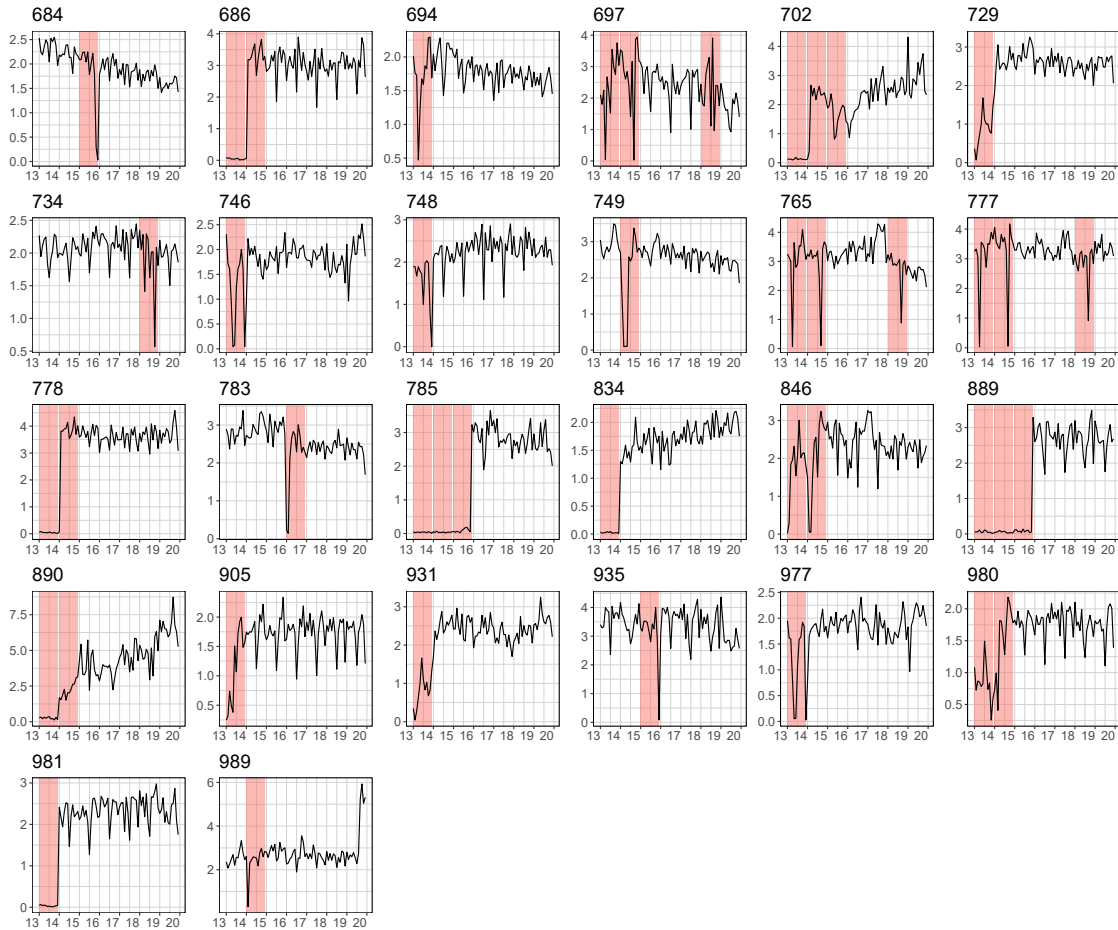


Figure A6: Adoption: Municipalities with Issues in the Primary Care Data, 3.

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. We first obtain a reference by deriving a distribution of mean contacts by permutationally excluding every combination of four consecutive months, then average, and ultimately choose the largest mean. We define a value suspiciously low if it is less than 40% of the largest mean. Municipality-year observations with suspicious values are highlighted by pink.

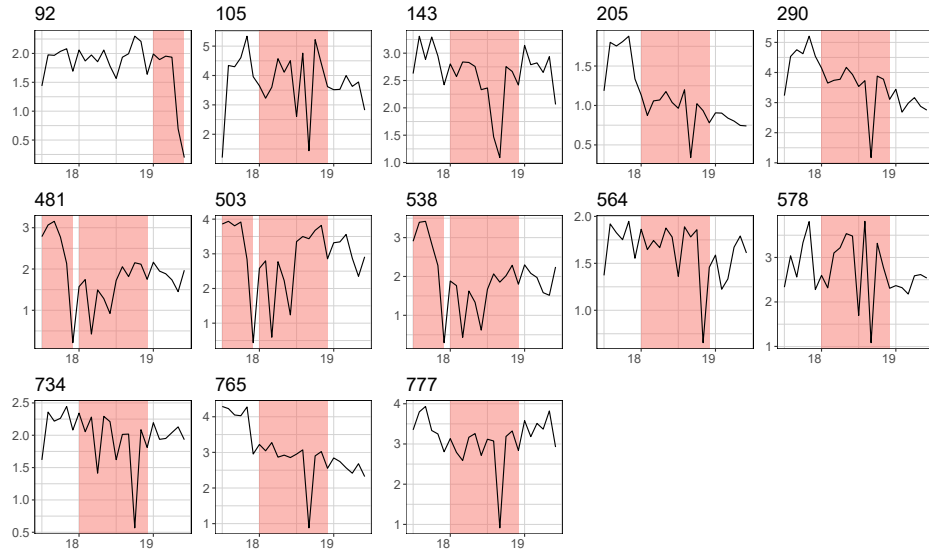


Figure A7: Abolition: Municipalities with Issues in the Primary Care Data.

Notes: Using data from a two-year window, centered on July 1st, 2018, we show the evolution in the annualized number of curative primary care visits (both nurse and GP visits; y axis) over time. We first obtain a reference by deriving a distribution of mean contacts by permutationally excluding every combination of four consecutive months, then average, and ultimately choose the largest mean. We define a value suspiciously low if it is less than 40% of the largest mean. Municipality-year observations with suspicious values are highlighted by pink.

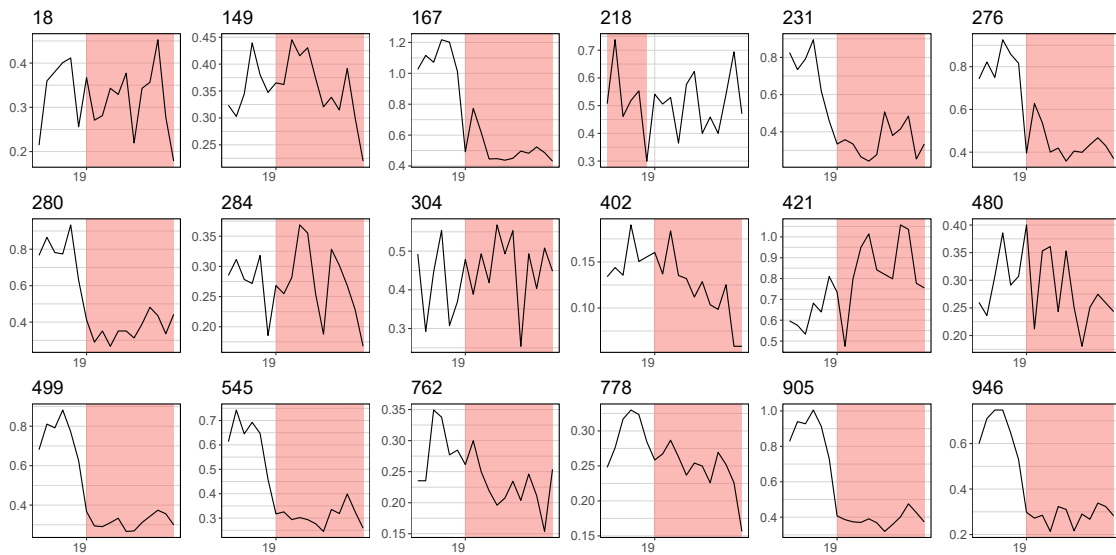


Figure A8: Abolition: Municipalities with Issues in the Referral Data.

Notes: Using data from a two-year window, centered on July 1st, 2019, we show the evolution in the annualized number of referrals to specialized healthcare (y axis) over time. Note however that we exclude 3/2020-6/2020 due to the COVID-19 pandemic. We first obtain a reference by deriving a distribution of mean contacts by permutationally excluding every combination of four consecutive months, then average, and ultimately choose the largest mean. We define a value suspiciously low if it is less than 60% of the largest mean. Municipality-year observations with suspicious values are highlighted by pink.

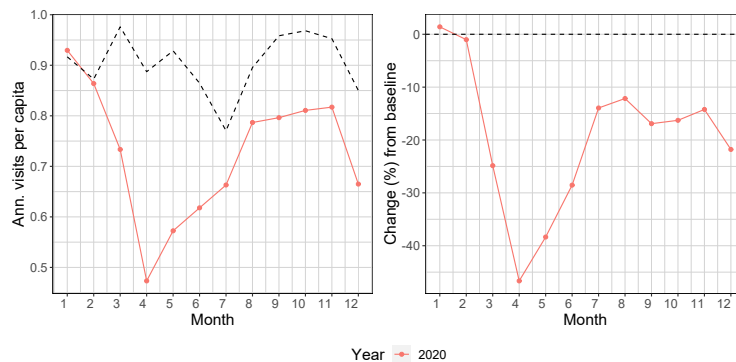


Figure A9: The Evolution of Nurse Visits during the COVID-19 Pandemic.

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. 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.

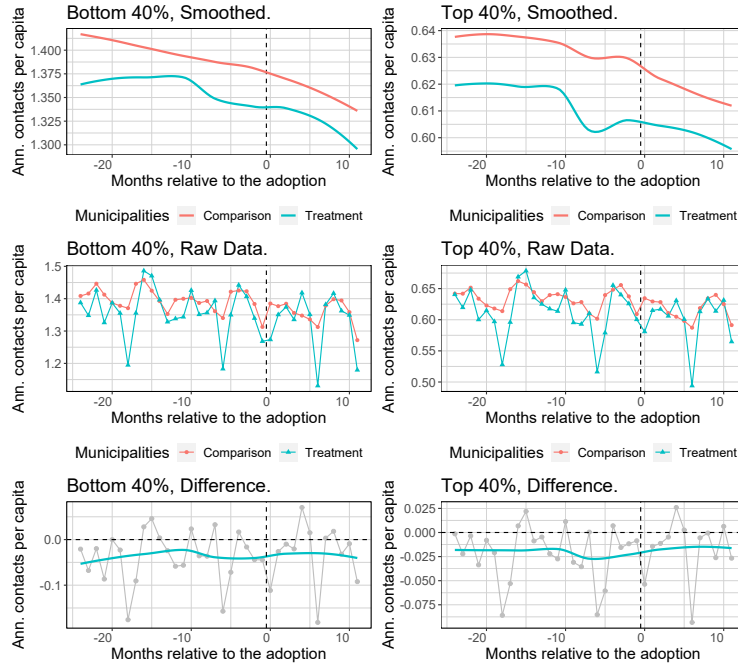


Figure A10: Adoption: Evolution in GP Visits.

Notes: These are placebo results - see Section 1.2. 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 top row contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle row, while the difference between treatment and comparison areas is depicted in the bottom row. Bottom 40% and top 40% refer to the distribution of equivalised family disposable income.

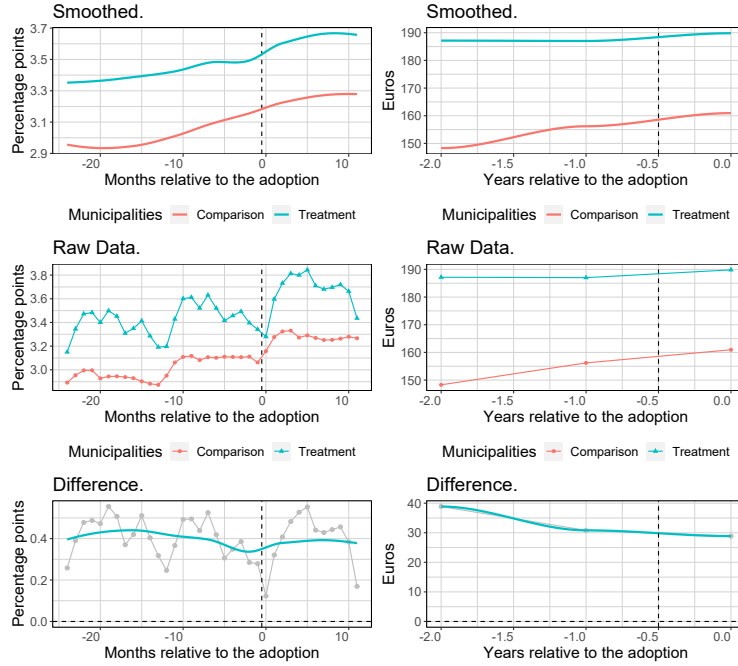


Figure A11: Adoption: Evolution in Social Assistance Recipients and Sums.

Notes: These are placebo results - see Section 1.2. The dataset is stacked and balanced. On the left, the outcome is the share of individuals in a family receiving social assistance (in percentages). On the right, 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 top row contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle row, while the difference between treatment and comparison areas is depicted in the bottom row.

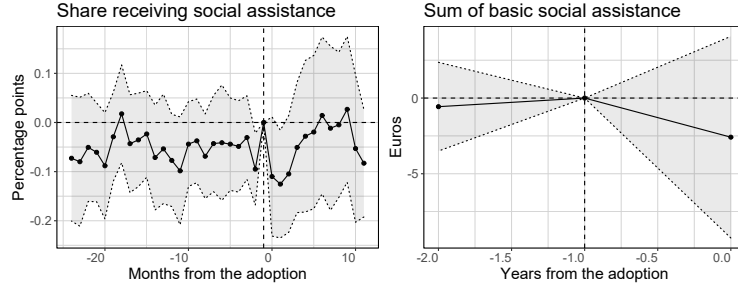


Figure A12: Adoption: Event-Study Plots on Social Assistance Use with Stacked Data.

Notes: These are placebo results - see Section 1.2. 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. Our TWFE regression model includes a full set of treatment indicators for 24 and 12 months before and after the treatment (or two and one years before and after the policy change for the sum of received social assistance) and event-specific municipality and time fixed effects, comparing the evolution of annualized contacts per capita between treated and unexposed municipalities. The last pre-treatment period, namely $t = -1$, is omitted as a reference. 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. The standard errors are clustered by municipality.

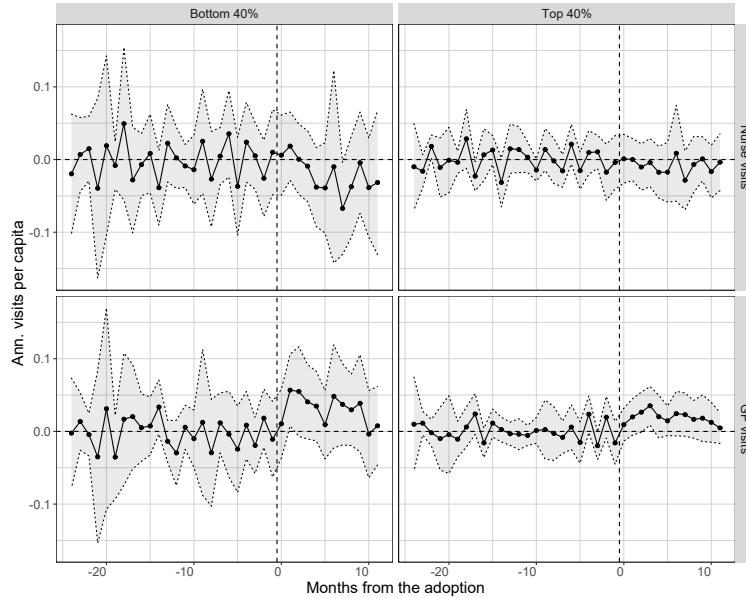


Figure A13: Adoption: Event-Study Plots Using the CS Estimator, Primary Care Use.

Notes: These are placebo results - see Section 1.2. The point estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. We use the CS estimator (Callaway and Sant'Anna, 2021) with outcome regression, weight by population, 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 equivalised family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively.

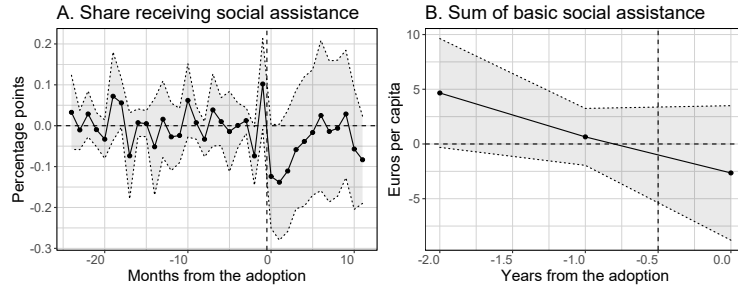


Figure A14: Adoption: Event-Study Plots Using the CS Estimator, Social Assistance Use.

Notes: These are placebo results - see Section 1.2. The point estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. We use the CS estimator (Callaway and Sant'Anna, 2021) with outcome regression, weight by population, 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.

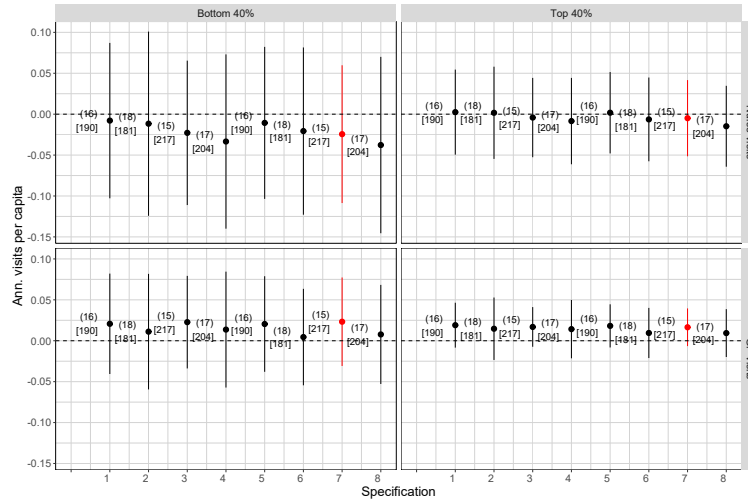


Figure A15: Adoption: the CS Estimator, Primary Care Use.

Notes: These are placebo results - see Section 1.2. The point estimates represent static effect estimates for the treatment group. We use the CS estimator (Callaway and Sant'Anna, 2021) with outcome regression, weight by population, and cluster standard errors by municipality. 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 equivalised 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 red.

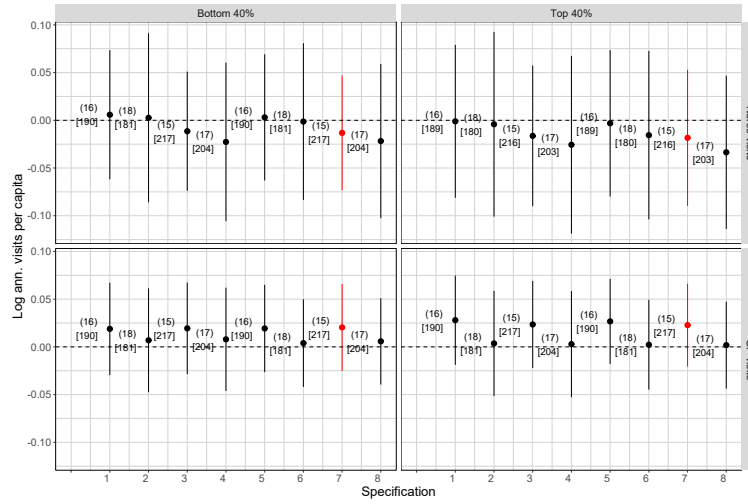


Figure A16: Adoption: the CS Estimator, Logarithmized Primary Care Use.

Notes: These are placebo results - see Section 1.2. The point estimates represent static effect estimates for the treatment group. We use the CS estimator (Callaway and Sant'Anna, 2021) with outcome regression, weight by population, and cluster standard errors by municipality. The estimates compare the evolution of logarithm of annualized contacts per capita between treated and unexposed municipalities. Bottom 40% and top 40% refer to the distribution of equivalised 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 red.

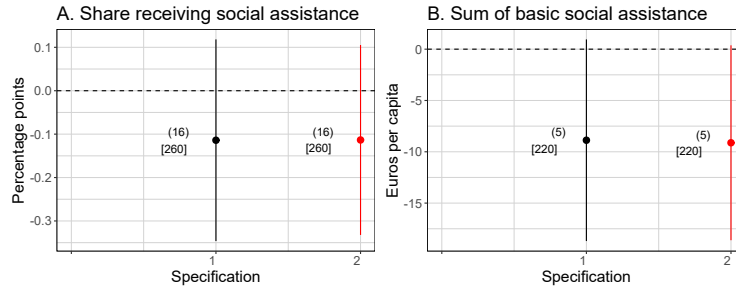


Figure A17: Adoption: the CS Estimator, Social Assistance Use.

Notes: These are placebo results - see Section 1.2. The point estimates represent static effect estimates for the treatment group. We use the CS estimator (Callaway and Sant'Anna, 2021) with outcome regression, weight by population, and cluster standard errors by municipality. The estimates compare the evolution of outcomes between treated and unexposed municipalities. Outcomes are the share of individuals in a family receiving social assistance (in percentages) and the annual some of received social assistance. Specifications (comparison units): 1) the never-treated, and 2) the not-yet-treated. The baseline is highlighted by red.

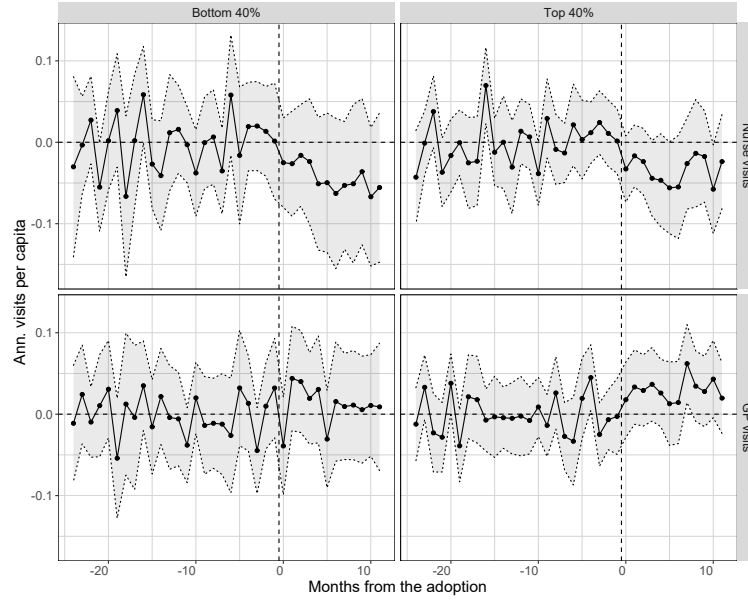


Figure A18: Adoption: Event-Study Plots Using the CS Estimator, Primary Care Use, Uniform Weighting.

Notes: These are placebo results - see Section 1.2. The point estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. We use the CS estimator (Callaway and Sant'Anna, 2021) with outcome regression, weight municipalities uniformly, 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 equivalised family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively.

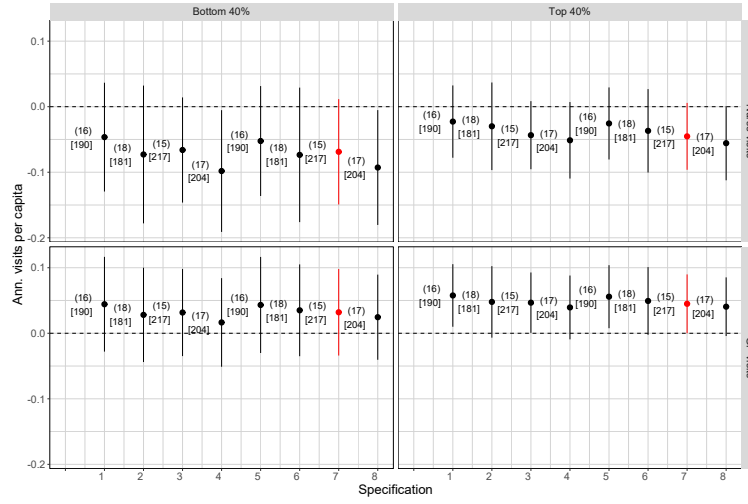


Figure A19: Adoption: the CS Estimator, Primary Care Use, Uniform Weighting.

Notes: These are placebo results - see Section 1.2. The point estimates represent static effect estimates for the treatment group. We use the CS estimator (Callaway and Sant'Anna, 2021) with outcome regression, weight municipalities uniformly, and cluster standard errors by municipality. 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 equivalised 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 red.

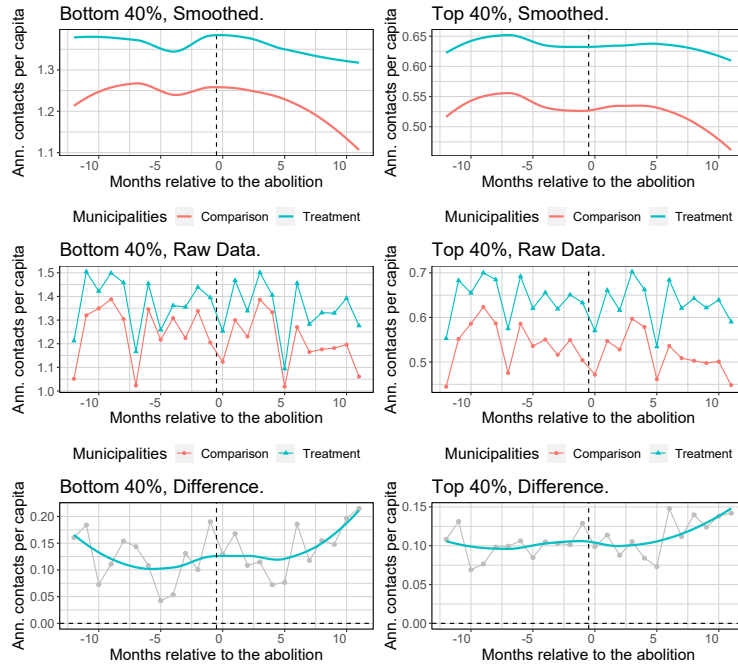


Figure A20: Abolition: Evolution in GP Visits.

Notes: These are placebo results: we proceed as if the placebo treatment started on July 1st, 2018. The outcome is the number of annualized curative GP visits per capita. The top row contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle row, while the difference between treatment and comparison areas is depicted in the bottom row. Bottom 40% and top 40% refer to the distribution of equivalised family disposable income.

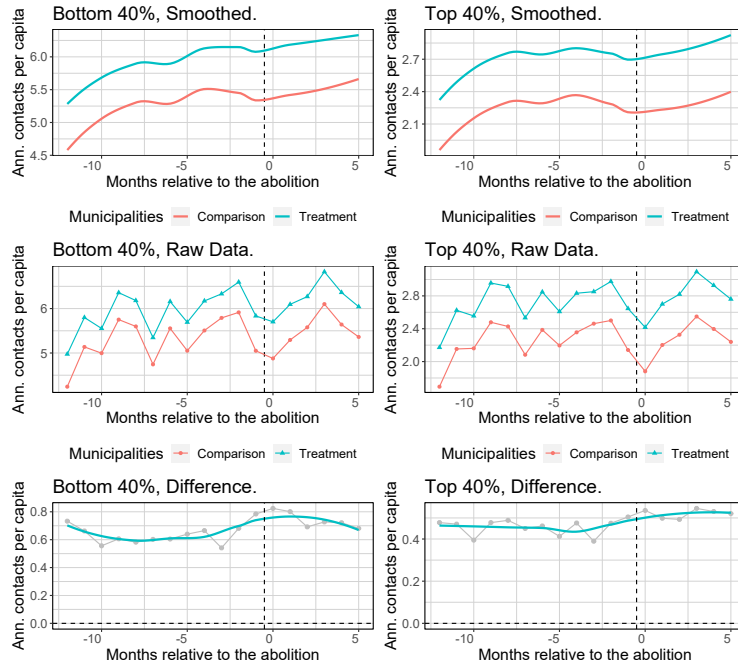


Figure A21: Abolition: Evolution Prescriptions.

Notes: These are placebo results: we proceed as if the placebo treatment started on July 1st, 2019. The outcome is the number of annualized prescriptions. The top row contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle row, while the difference between treatment and comparison areas is depicted in the bottom row. Bottom 40% and top 40% refer to the distribution of equivalised family disposable income.

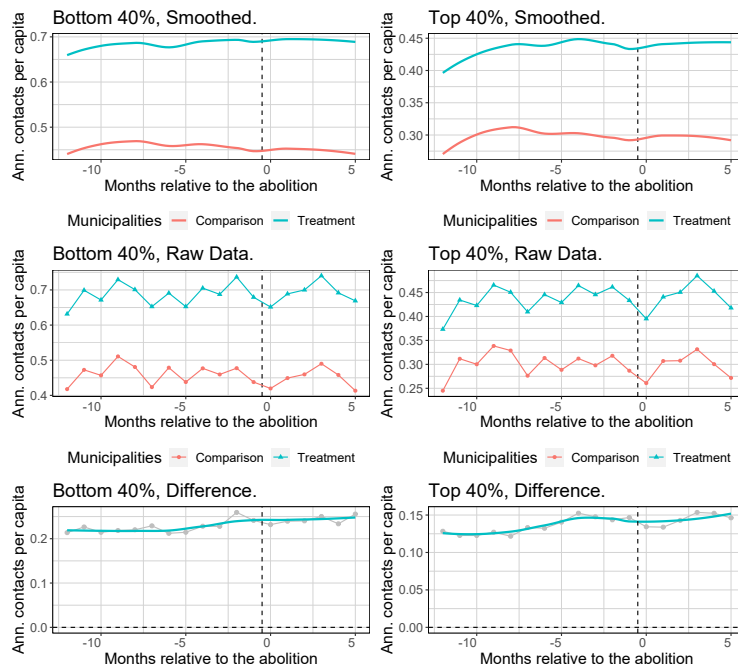


Figure A22: Abolition: Evolution in Referrals to Specialized Healthcare.

Notes: These are placebo results: we proceed as if the placebo treatment started on July 1st, 2019. The outcome is the number of annualized referrals (unique ID-by-referral-arrival-date observations) to specialized healthcare. The top row contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle row, while the difference between treatment and comparison areas is depicted in the bottom row. Bottom 40% and top 40% refer to the distribution of equivalised family disposable income.

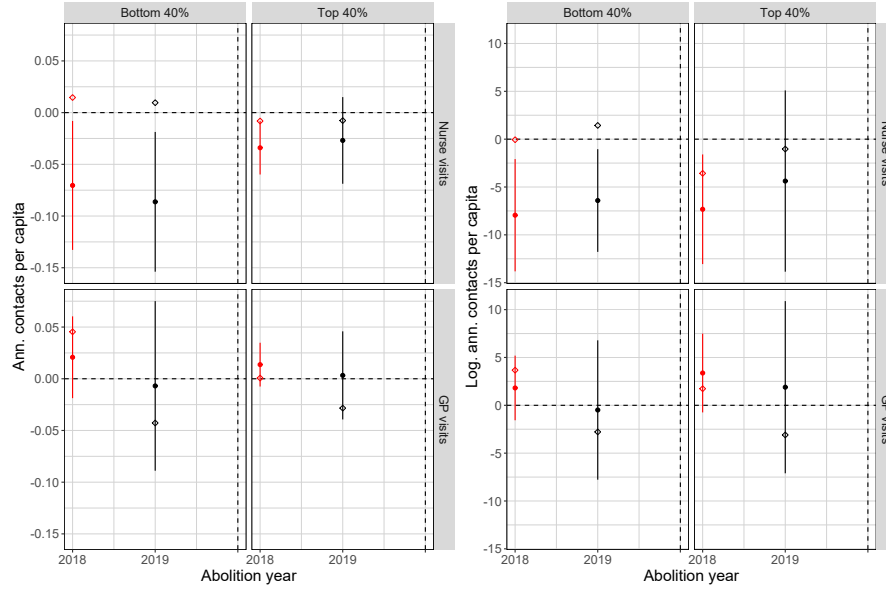


Figure A23: Abolition: DID Placebo Estimates, Primary Care Outcomes.

Notes: These are placebo results: we fix the treatment and comparison municipalities and the treatment date (July 1st) but proceed as if the treatment occurred in earlier years. The point estimates represent static effect estimates for the treatment group. First, we use a TWFE regression model that includes an indicator for post-treatment periods in treated municipalities, and municipality and time fixed effects. Point estimates are by circles with their confidence intervals. Then, we use another specification allowing for linear pre-trends by replacing 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 by squares. Due to heterogeneity in municipality size, we weight by population. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalised family disposable income. The main (placebo) estimate is highlighted by red. Estimates and standard errors are multiplied by 100 if the outcome is logarithm of annualized contacts per capita.

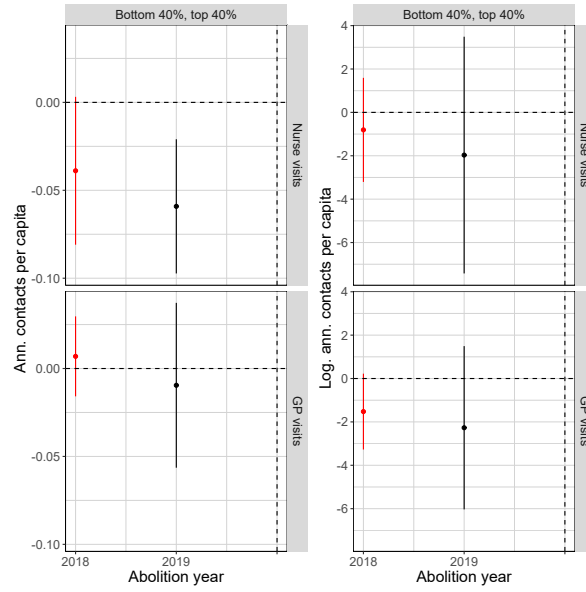


Figure A24: Abolition: DDD Placebo Estimates, Primary Care Outcomes.

Notes: These are placebo results: we fix the treatment and comparison municipalities and the treatment date (July 1st) but proceed as if the treatment occurred in earlier years. The point estimates represent static effect estimates for the treatment group. We use Model 1 but without event-specific parameters as we analyze a simultaneous abolition. Due to heterogeneity in municipality size, we weight by population. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalised family disposable income. The main (placebo) estimate is highlighted by red. Estimates and standard errors are multiplied by 100 if the outcome is logarithm of annualized contacts per capita.