The Online Appendix for the ”Effects of Nurse Visit Copayment on Primary Care Use: Do Low-Income Households Pay the Price?”

Tapio Haaga* † [a,b], Petri Böckerman[c,d], Mika Kortelainen[a,b], and Janne Tukiainen[a]

[a]Turku School of Economics, FI-20014 University of Turku
[b]Finnish Institute for Health and Welfare (THL), P.O. Box 30, FI-00271 Helsinki
[c]Jyväskylä University School of Business and Economics, P.O. Box 35, FI-40014 University of Jyväskylä
[d]Labour Institute for Economic Research LABORE, Arkadiankatu 7, FI-00100 Helsinki

February 2024

Abstract of the Main Paper

Nurses are increasingly providing primary care, yet the literature on cost-sharing has paid little attention to nurse visits. We employ a staggered difference-in-differences design to examine the effects of adopting a 10-euro copayment for nurse visits on the use of public primary care among Finnish adults. We find that the copayment reduced nurse visits by 9–10% during a one-year follow-up. There is heterogeneity by income in absolute terms, but not in relative terms. The spillover effects on general practitioner (GP) use are negative but small, with varying statistical significance. We also analyze the subsequent nationwide abolition of the copayment. However, we refrain from drawing causal conclusions from this due to the lack of credibility in the parallel trends assumption. Overall, our analysis suggests that moderate copayments can create a greater barrier to access for low-income individuals. We also provide an example of using a pre-analysis plan for retrospective observational data.

Keywords: Cost-sharing, copayment, primary care, difference-in-differences, pre-analysis plan, blind analysis
JEL codes: I18, I14, I13, H42, I11

*Corresponding author. E-mail address: tapio.haaga@utu.fi (T. Haaga)

†Pre-analysis plan and earlier working paper versions: https://osf.io/skuv9/. Replication codes: https://github.com/tapiohaa/ASMA3.

Contents

A.1 Staggered Adoption: Supplementary Results **A2**

A.2 Nationwide Copayment Abolition: Regression Results **A7**

A.3 Constructing the Analysis Data **A10**

A.4 Changes and Additions to the Pre-Analysis Plan **A17**

A.5 Parameter Values for Semi-Arc Elasticities **A25**

A.6 Additional Figures and Tables **A26**

A.1 Staggered Adoption: Supplementary Results

Estimates for all individuals (post-blind). We provide the results also for the whole sample population: pre-trend plots (Figure A12, post-blind), dynamic event-study plots using stacking (Figure A13, post-blind) and the CS estimator (Figure A14, post-blind), and aggregated static stacking (Table A3, post-blind) and CS (Figure A15, post-blind) estimates. We estimate that the nurse visit copayment reduced annualized nurse visits per capita by 0.09–0.10 visits, equivalent to a reduction of 9–10%, using stacking (one-year follow-up) and by 0.13–0.16 visits, equivalent to a reduction of 13–17%, using the CS estimator (a longer follow-up). Stacked effect estimates on GP visits imply a decrease of 3–5%, but the corresponding CS estimates are closer to zero and insignificant.

Estimates on social assistance use. We examine the impacts of the copayment adoption on 1) the share of individuals living in a family where someone received social assistance (extensive margin) and on 2) the annual sum of social assistance received by the family (intensive margin). We do not observe any clear effects in the trend plots (Figure A16) or the event-study plots (Figure A17). The stacked regression results are in Table A4, and the effect estimates on both outcomes are not statistically different from zero. The inclusion of a linear pre-trend difference in the model attenuates both estimates. Regarding robustness of the results, the CS event-study plots point to a decrease in the share of individuals receiving social assistance (Figure A18), but the corresponding aggregated estimates are not statistically different from zero (Figure A19). To summarize all the results, the effects on social assistance use are inconclusive: the point estimates are positive at the intensive margin and negative at the extensive margin but, in both cases, not statistically different from zero.

ED visits and unplanned hospitalizations for ACSCs. Our primary analyses focus on the effects of nurse visit copayments on the utilization of primary care and potential impact variations by income level. A logical follow-up question is, did the reduction in nurse visits result in undesirable offset effects on emergency department (ED) use or health? Here, we present results on two additional post-blind outcomes: ED visits and unplanned hospitalizations for ambulatory

care sensitive conditions (ACSC). Two main changes to the pre-specified main analyses are that we 1) estimate the impacts on all individuals to increase sample size and 2) use the CS estimator because it has a longer follow-up period than our baseline stacking analyses with a one-year follow-up. A longer follow-up period is reasonable because the adverse health impacts likely accumulate over time.

The CS event-study plots are in Figure A20 (post-blind). There appears to be no systematic pre-trend differences in either of the outcomes, but the estimates on ACSC hospitalizations are very noisy (these events are rare). In a one-year follow-up, the copayment adoption is associated with some statistically insignificant increases in ED visits and decreases in ACSC hospitalizations. The estimates are positive for both outcomes after two years, continuing to increase. However, we caution against strong interpretations far from the treatment, because the composition of areas changes and the effective sample size decreases the further we move from the adoption of the policy.

The static CS estimates are in Figure A21 (post-blind). For both outcomes, the effect estimates are positive but statistically insignificant, the point estimates suggesting a 2–6% increase in ED visits and a 4–10% increase in ACSC hospitalizations. Overall, given that the outcomes and chosen methods were determined post-blind, we view these statistically insignificant estimates as being more exploratory than confirmatory in nature.

Robustness of Main Difference-in-Differences Results. The stacking results on primary care use are robust to logarithmic outcomes and to unbalanced event-specific datasets that have more municipalities and observations than the balanced datasets in the main analysis (Table A5). Next, we report the results constructed using the CS estimator. Starting with event-study plots, the findings on nurse visits are similar using both stacking and the CS estimator. Regarding GP visits, however, the CS estimates show virtually no effect, while the stacked estimates suggest a small decrease (Figure A22 and Figure A23, post-blind).

Our static CS effect estimate averages over all group-time ATTs, weighted by group size. We estimate the effects on primary care use in eight cases, using either never-treated or

not-yet-treated municipalities as the comparison group and exclude data from either 2013 or 2019 or both. The CS estimates differ from the stacked estimates in two ways (Figure A24 and Figure A25). First, the estimated reductions in nurse visits are larger. We estimate that the nurse visit copayment reduced annualized nurse visits per capita by 0.18–0.23 visits, or by 13–17% when using logarithmized outcomes, for individuals at the bottom 40% and by 0.08–0.10 visits, or by 13–16% when using logarithmized outcomes, at the top 40%. Different estimands and accumulating effects plausibly explain the difference. The stacked results are based on a one-year follow-up, while the CS estimand averages over all group-time ATEs, having for most units a longer follow-up. Second, the estimated effects on GP visits are all negative but close to zero and insignificant.

Means and standard deviations of the outcomes. Table A6 (post-blind) complements Table 1 of the main paper by including population-weighted means and standard deviations of the outcome variables based on the stacked dataset, as well as standardized effect sizes constructed based on the estimated standard deviations.

Follow-up length, and cohort-specific effects. To examine how the estimates depend on the follow-up length, Figure A26 (post-blind) shows the CS event-study plots on the number of annualized primary care visits per capita for all observable durations. The effect size for nurse visits appears to increase steadily for two years before attenuating.[1] Motivated by this finding, Table A7 (post-blind) shows the stacked results using a 24-month follow-up. Compared to the main results (Table 1), the effect sizes increase slightly for nurse visits and decrease for GP visits, and more than that with the specification allowing for a linear pre-trend difference. Finally, Figure A27 (post-blind) shows CS estimates by treatment cohort. There is variation in the magnitude of the point estimates and in the width of the confidence intervals. It is challenging to determine whether this variation stems from noise or actual heterogeneity in the treatment effect. For nurse visits, a clear majority of the point estimates are negative: e.g., 12 are negative (9 significant) and 2 positive for the bottom 40% (0 significant).

[1] Note that the composition of areas changes in relative time as the sample size decreases and uncertainty increases the further we are from the event.

Weighting municipalities uniformly. We repeated the adoption analyses, but instead of population weighting, we uniformly weighted municipalities when using the CS estimator and municipality-by-income-decile observations when using stacked TWFE regressions. The pre-treatment healthcare use is now higher as small municipalities have a larger weight. Qualitatively, the main findings are robust to the weighting scheme: the copayment adoption is associated with a reduction in nurse use in the treated areas. The estimates for GP visits are negative but small, and the estimates on social assistance outcomes are inconclusive. The uniformly-weighted stacking results on annualized primary care contacts per capita are reported in Table A8 (post-blind). We estimate that the nurse visit copayment reduced nurse visits by 0.17 visits (11.2%) in the bottom 40% and by 0.07 visits (9.4%) in the top 40%. In contrast, the population-weighted estimates in Table 1 show smaller decreases: 0.13 visits (9.3%) in the bottom 40% and 0.05 visits (8.0%) in the top 40%. The rest of the figures and tables are provided in the replication codes folder.

The Roth and Sant'Anna (2023) estimator (RS), and the later-treated as comparisons. We use not-yet-treated municipalities as comparisons in the main stacked analysis, and causal inference relies on the parallel trends assumption. However, if only later-treated municipalities are the comparisons, we consider as plausible a stronger assumption of quasi-random treatment timing. This assumption would allow us to use the estimator proposed by Roth and Sant'Anna (2023). The key benefit of the RS estimator is that it can produce considerably more precise estimates in a staggered setting than the earlier methods (e.g., the CS DD estimator).

As a post-blind complementary analysis, we use later-treated municipalities as comparisons and estimate the effects using the RS estimator with the *staggered* R package. We continue to include events that have at least 12 follow-up months. We use a balanced panel in calendar time and exclude the treatment cohorts that have only a single treated municipality, as required by the software. Following the assumption of quasi-random treatment timing, municipalities are weighted equally and not by population. We use the refined variance estimator

proposed by Roth and Sant'Anna (2023) to calculate standard errors.[2] As with the CS estimator, we use 2013 or 2014 as the panel start year and 2018 or 2019 as the end year. Given that we restrict the analysis to balanced panels (in calendar time), shorter observation windows tend to have more treatment cohorts to study but less monthly observations for a given included municipality.

Three estimands proposed by Roth and Sant'Anna (2023) are considered. The key building block is $ATE(t,g)$, defined as the average treatment effect on the outcome in period t of being first-treated in period g relative to not being treated at all. Our first estimand, "simple", is a simple average of the $ATE(t,g)$ weighted by cohort size. Our second estimate, "cohort", first averages the $ATE(t,g)$ for each cohort g before averaging these cohort effects weighting by cohort size. Finally, "calendar" first averages the $ATE(t,g)$ for each time period t weighting by cohort size before taking a mean of the calendar effects.

Table A9 (post-blind) contains the RS results on annualized primary care visits per capita. Our main finding of the nurse visit copayment reducing nurse use is qualitatively robust, but the pattern of heterogeneity by income level differs from the pre-specified analyses.

We estimate that the nurse visit copayment reduced annualized nurse visits per capita by 0.12–0.17 visits, equivalent to a reduction of 8–11%, for the bottom 40% of the income distribution and by 0.09–0.14 visits, equivalent to a reduction of 12–18%, for the top 40%. There is no longer clear heterogeneity in absolute terms (levels) by income level. Consequently, the effects in relative terms appear even larger for the higher end of the income distribution than for low-income individuals. With respect to GP visits, the effect estimates vary around zero and are mostly insignificant. The effects show changes of $+0.07$ to -0.06 annualized GP visits ($+5\%$ to -4%) for the bottom 40% of the income distribution and of $+0.02$ to -0.03 visits ($+2\%$ to -4%) for the top 40%. Table A10 (post-blind) reports the same results but using a logarithmic outcome.

[2] In 12 cases out of 96, the software code produces a warning that a more conservative Neyman-style variance estimate is less than an estimated adjustment factor. In these cases, we consequently use the unrefined Neyman-style variance estimator discussed in Roth and Sant'Anna (2023).

A.2 Nationwide Copayment Abolition: Regression Results

The analyses presented in Section 6 of the main paper reveal that the parallel trends assumption is not credible in the context of the copayment abolition. Consequently, reporting the pre-registered regression analyses in the main paper does not inform much on the utilization effects resulting from the abolition of copayments in this case. The results of the regressions planned in the PAP are instead reported here in this appendix for the sake of transparency. Uncertainty about the identification assumption is considerable. For this reason, we argue that uncertainty about the point estimates conditional on the identification assumption being valid seriously underestimates the overall uncertainty about the causal relationship of interest. We therefore emphasize caution in interpreting the regression estimates: they provide evidence neither for the existence nor the absence of utilization effects.

Trend plots. Trend plots on the number of drug prescriptions, a secondary outcome, are in Figure A28.

Main regression analyses. We use the following three methods that differ by the parallel trends assumption required, weighted by population size. We view none of the three assumptions superior to the rest based on observed pre-trend patterns in Section 6. First, the TWFE DD specification without a pre-trend difference contains a static indicator for the post-treatment periods of the treated municipalities (D_{mt}), and municipality (m) and time (t) fixed effects:

$$y_{mt} = \alpha_m + \gamma_t + \delta^{DD} D_{mt} + \varepsilon_{mt}. \tag{A1}$$

Model A1 assumes parallel trends throughout the observation window. Second, we modify the TWFE DD specification to allow for linear pre-trend differences:

$$y_{mt} = \alpha_m + \gamma_t + \theta d_m t + \sum_{l=0}^{10} \mu_l D^l_{mt} + \varepsilon_{mt}. \tag{A2}$$

D^l_{mt} is a dummy for the treated areas for observations l months from the copayment abolition.

Regarding the linear pre-trends, d_m is a dummy for the municipality being treated, and t denotes time relative to the abolition. We report the mean of μ_l over l as our point estimate. Model A2 assumes parallel trends in deviations of the outcome from a linear time trend.

Third, the CS estimator assumes parallel trends but only from the last pre-treatment period onward.

The results for annualized contacts per capita are in Table A11. As expected, the point estimates vary noticeably depending on the method due to different parallel trends assumptions. For nurse visits, the estimates of the impacts of the nurse visit copayment abolition vary between -0.01 and $+0.09$ annualized visits (from -1.0% to $+7.9\%$) for the bottom 40% and between $+0.04$ and $+0.06$ annualized visits (from $+7.8\%$ to $+12.2\%$) for the top 40%. The estimates on GP visits and, especially, on drug prescriptions are close to zero. For nurse visits, there is no heterogeneity by income in absolute terms, but in relative terms the estimates appear to be larger for the top 40%. Table A12 shows the corresponding estimates but from using the logarithm of annualized contacts per capita as the outcome. The estimates for nurse and GP visits vary depending on the method, while the estimates for nurse visits appear largest – at least for the top 40% of the income distribution.

Time-placebo analyses. Our pre-analysis plan (Haaga et al., 2022) reports placebo estimates where we fix the treatment and comparison municipalities and the treatment date (July 1st) but proceed as if the treatment occurred in 2018 or 2019 – this was the method with which we wrote the statistical plan. In this sense, the PAP contains (unblinded) time placebo estimates. These placebo results (e.g., PAP Table 3) show that the estimates can be sensitive to the specific version of the parallel trends assumption even in normal times, but these issues were much more prominent when analysing the data from the actual observation window in 2020–2022 for this paper.

Heterogeneity by income. The aim here was to formally test whether there is treatment effect heterogeneity between the income groups. We use a similar DDD specification to Model 2, but now we only have a single event and, thus, only one event-specific dataset and much fewer

parameters to be estimated:

$$
\begin{aligned}
y_{mgt} = \alpha &+ \beta_1 Treat_m + \beta_2 Bottom40_g + \beta_3 Post_t + \beta_4 Treat_m \times Bottom40_g \\
&+ \beta_5 Treat_m \times Post_t + \beta_6 Bottom40_g \times Post_t \\
&+ \gamma Treat_m \times Bottom40_g \times Post_t + \varepsilon_{mgt}.
\end{aligned} \tag{A3}
$$

Here, subscripts m, g, and t denote municipality, income group, and time (month). $Treat_m$ is an indicator for the treated municipalities, and $Post_t$ indicates the post-abolition periods. $Bottom40_g$ is a dummy for the bottom 40% of the income distribution (0 for those in the top 40%), and γ is the coefficient of interest.

The DDD results are in Table A13. At face value, these results suggest that nurse use *decreased* by 5–7% at the bottom 40% of the income distribution from their baseline mean compared to the top 40% after the copayment abolition, but, as discussed above, we emphasize considerable caution in interpreting these estimates.

Weighting municipalities uniformly. The remaining two abolition regression tables registered in the PAP repeat the above analyses, but instead of using population weighting, we weighted uniformly municipality-by-income-decile observations. The DD estimates on annualized contacts per capita are in Table A14 and the corresponding DDD estimates are in Table A15. Also these results highlight that the DD estimates for nurse visits are highly sensitive to the method (and the parallel trends assumption).

A.3 Constructing the Analysis Data

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

After the above restrictions, we have 264 out of 293 municipalities in mainland Finland for the study of the staggered adoption. In total, they had 5.0 million residents compared to the Finnish population of 5.5 million. Regarding the simultaneous abolition, we have 249 municipalities with 5.0 million residents. Figure 1 illustrates the staggered adoption in terms of treatment areas and treated population, and Figure A4 does the same for the abolition. Figure A2 (the staggered adoption) and Figure A5 (the abolition) show the municipal policies graphically. Figure A3 shows the variation of the copayment level in Summer 2021.

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

Socioeconomic data: From the full population residing in Finland at year's end, we exclude those ID-year observations where equivalized family disposable income is exactly zero (less than 1% of the rows) and only include those observations where an individual is aged 25 years or more in order to exclude minors, who are exempted from the copayment, and students, who

have access to student healthcare. Without any other restrictions on the data, this leaves us with approximately four million individuals out of the population of 5.5 million. With the population remaining after the above two restrictions, we compute the distribution of the equivalized family disposable income and sort individuals into income deciles.

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

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

Since 2019, the register also contains outpatient contacts in private clinics that we aim to exclude. We do this by linking each visit in 2019–2022 to TOPI and SOTE organization registers

that contain information on the visit provider. Both registers are continuously updated. We have an annual cross-sectional dataset on TOPI and two cross-sections on SOTE from early 2020 and early 2022. We use SOTE from early 2020 for years 2019 and 2020 and SOTE from early 2022 for years 2021 and 2022. In 2019–2022, the linking of TOPI does not work for 4% of the extracted visits while the same figure for SOTE is 1%. Then, we include those visits whose provider a) has a TOPI service area code that refers to health centers (120, 121, or 122) or b) is a public sector organization in SOTE.[3]

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

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

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

90 municipalities (out of 293) have suspiciously low or high values of primary care use in the pre-pandemic study period using a threshold of $X = 0.4$. The evolution of the sum of curative nurse and GP visits in these municipalities is illustrated in Figure A29, Figure A30, and Figure A31. Gray segments highlight municipality-year pairs with suspiciously low values. Based on visual inspection, the algorithm appears to be good at detecting irregularities. Regarding the abolition analyses, we set a higher threshold of 0.55 to make the algorithm more sensitive. 8 municipalities have suspicious values (Figure A32). In the replication codes folder, we illustrate primary care use in the remaining municipalities for which we find no suspicious observations.

Figure A33 (post-blind) examines the relationship between the yearly indicator for the quality issues (as defined above) and the copayment adoption in time relative to the policy change, between the treated and comparison areas. In essence, Figure A33 (post-blind) is similar to Figure 2, but it uses the annual indicator for quality issues as the outcome instead of excluding the municipalities with quality issues. The figure shows that the population-weighted share of municipalities with quality issues is clearly higher in the treatment group *for months* -13 to -24 *before the treatment*, but there are no major differences in a *one-year* window around the treatment. When the data from 2013–2014 are excluded, no relevant differences remain. Thus, the quality issues that disproportionately affect policy groups originate from the years 2013–2014.

Thus, the quality issues were more prominent in the treated municipalities, which are smaller by their population size and more rural than the never-treated municipalities, in 2013–2014, during the early years of national data collection. This observed selection and our code to account for quality issues should not bias the point estimates, only change the composition of municipalities used in event-specific datasets. Consistent with this hypothesis, not accounting for quality issues in analysis does not appear to change the results: Figure A34 (post-blind) shows the trend plots for nurse visits and all individuals based on stacked data where no municipalities are excluded due to quality issues. Its counterpart is Figure A12 (post-blind), which is otherwise similar but in which the sample is smaller (and arguably of higher quality) because we exclude those municipality-year observations with quality issues and simultaneously require balanced event-specific datasets.

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

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

Prescriptions: We extract prescriptions written at public-sector units, containing both initializations of drug therapies and renewed prescriptions. Ideally, we would have wanted to consider only drug therapy initializations as patients can apply for renewals online or at pharmacies without a GP visit. However, we had some doubts about the quality of the variable and decided to include all prescriptions. The exclusion of private sector prescriptions might be a concern in a small subset of municipalities, primarily smaller ones, that have outsourced their primary care services to private providers. We search for potentially missing values with the same algorithm as with primary care contacts. The algorithm does not find any irregularities even if we use a higher threshold of 0.6, making it more sensitive to outliers.

ED visits and unplanned hospitalizations for ACSCs. These data for 2013–2019 originate from the Care Register for Health Care. In 2019, there was a change in variables

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

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

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

The plausibly increased precision due to population weighting comes with a trade-off. First, the ATT estimates may not generalize to the whole country if the variation essentially stems from the largest municipalities. The ATTs may not need to be homogeneous by municipality

size. Second, an institutional change (for example, a change in the EHR system) in one large municipality can more easily bias the estimates than in a case where the variation arises from a large number of municipalities, with each municipality receiving the same weight.

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

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

The purpose of this section is to make it easier for the reader to get a sense of what has changed *ex post* and why, by detailing each relevant change one by one. It would be cumbersome for the reader to compare without any guidance both the PAP and the main paper and also their corresponding *R* codes. *Changes* can mean two things: either we changed our *codes* or kept the codes intact but changed the presentation in the *paper*. The paper distinguishes tables and figures that were not pre-registered by labeling them as "post-blind" in the text as well as in the table and figure notes.

There is some gray area regarding 1) what counts as a change that needs to be reported and 2) does any deviation from the PAP make a specific figure or table post-blind. First, suppose that we wish to add a new figure and mark it as post-blind. We subsequently rewrite a function that extracts and aggregates data because the function needs more functionality for us to be able to produce the desired figure. In this case, this appendix would discuss the addition of the figure but not the change to the underlying function. In general, this appendix does not discuss changes that we believe are very minor in a sense that they should not affect the analysis, such as tidying parts of the code. As a result, this appendix provides the information that we, as readers, would want to see ourselves. In the end, what information is enough is subjective. Large language models (LLM) can make it easier to compare all changes in the replication codes of the PAP and this version.

Second, consider Figure A29 that illustrates how our algorithm performs to detect quality issues in the primary care data. A similar figure was in the PAP. The algorithm has changed a bit, which is discussed below, but the underlying idea of the figure remains the same. In this case, we discuss the change in the algorithm in this section, but do not view the figure as post-blind.

Based on the experiences of writing this paper and using a PAP-based workflow for the first time, we recommend that researchers 1) write narrower PAPs than we did, limiting to only primary outcomes and the most important confirmatory statistical tests, 2) make a populated PAP ("results of the PAP") available, a document that does not deviate from the PAP, before further changes, and 3) write a paper and adjust the analysis (e.g., by adding exploratory analysis or robustness checks) taking into account what was learned from the pre-registered confirmatory

analyses. We now believe that this kind of workflow would make the reporting of the results easier, the distinction between the PAP and ex-post analyses clearer, and leave more room for exploratory analyses which complement the confirmatory findings.

Data: the role of nurse visits (addition). We added Figure A1 (post-blind) to show a scatter plot and histogram of curative nurse and GP visits by municipality.

Data: detecting quality issues in the data (change and addition). In the PAP, the algorithm which we use to detect quality issues in the primary care data was designed to find periods with abnormally low health care use. Here, we modify it so that it can detect periods with abnormally high values as well, which we argue is a reasonable change. The implications can be seen for municipality no. 10 in Figure A29: now the spike in late 2019 is detected as an abnormally high value. We also use a higher threshold of 0.55 for the abolition analyses, making the algorithm more sensitive to outliers. Using a higher threshold of 0.6 for prescriptions does not change the number of excluded observations.

We also added Figure A33 (post-blind) to examine the relationship between the yearly indicator for quality issues (as defined in Section A.3) and the copayment adoption in time relative to the policy change, between the treated and comparison areas. Moreover, we added Figure A34 (post-blind) to evaluate whether our methods to account for quality issues affect the effect estimates for nurse visits and for all individuals.

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

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

Data: does data quality improve when the copayment is adopted (addition). Once

the copayment is adopted, an area needs to distinguish between curative and preventive nurse visits and be able to count the number of curative nurse visits. This may impact the quality of the data, potentially leading to the changes in the number of recorded curative nurse visits even if the underlying utilization remains unchanged. We checked in Figure A7 (post-blind) that the number of preventive nurse visits does not increase in the treated municipalities after the copayment adoption, even if the number of curative nurse visits decreases considerably.

Reporting the results (change). We have made several changes to the presentation of results to make the article more organized and easier to follow following the feedback we received in the peer review process. The main paper now i) better summarizes the contents of key appendix material and ii) has less direct references to appendix tables so that a reader would not feel a need to jump back and forth to appendix tables all the time. Some of the most important changes are listed and discussed in more detail below in this section. These changes make the reading experience of the main paper much better and the paper more independent from the appendix, but they come at a cost: it is harder to read the PAP and the main paper side by side to compare their similarities and differences, because the structure of the main paper has somewhat changed.

Results: estimates by income decile (addition). We committed in the PAP to show the main estimates for the primary care outcomes by income quintile and decile to allow for a more flexible analysis of treatment effect heterogeneity by income. Specifically, we now estimate and show in Figure 4 (post-blind) the adoption results on the number of nurse visits by income decile using stacking with balanced event-specific datasets as in the main analysis. However, we do not estimate the abolition results by income deciles, as the abolition results are highly sensitive to a specific version of the parallel trends assumption. We do not present (nor have estimated) the effects by income quintile (a less flexible measure than deciles) to keep the length of the paper reasonable.

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

Adoption: estimates on social assistance use (change). In the PAP, we registered two

social assistance outcomes as secondary outcomes, the corresponding figures and tables being in the PAP appendix but explicitly referred to in the main paper. We changed the way social assistance results are presented in the main paper by 1) summarizing these results in one paragraph in Section 5 and 2) not providing explicit references to the corresponding figures and tables in the main paper anymore. These changes aim to make the results section more organized and easier to follow, without a reader having to jump so much back and forth to appendix tables.

Adoption: weighting municipalities uniformly (addition). We re-estimated the results weighting our municipality-by-income-decile observations uniformly instead of population-weights as a robustness check. In the PAP, the idea was generally speaking to report the population-weighted estimates in the report and its appendixes and provide the uniformly-weighted results in the replication folder with the exception of presenting some uniformly-weighted results, which we view important to show to the reader, in the online appendix. Regarding the adoption analyses, we added Table A8 (post-blind; uniformly-weighted stacked baseline results) to the appendix. Two figures from the PAP (uniformly-weighted CS event-study plots and static CS estimates) are no longer in the appendix, but they are provided in the replication folder as are the rest of the uniformly-weighted results. In the PAP, the uniformly-weighted results were discussed in the main text, but we moved this discussion to the Online Appendix to shorten the main body of the text.

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

using all sample individuals (Table A3, Figure A12, Figure A13, Figure A14, and Figure A15, all post-blind). For the abolition analyses, the estimates for all individuals are in the replication files.

Adoption: changes to GP visit copayments (addition). In the PAP, we committed to add a plot showing when the GP visit copayment increases were made relative to the adoption of the nurse visit copayment. Details were fixed after the PAP registration, and this plot is in Figure A6 (post-blind).

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

Adoption: logarithmized results (addition). In the PAP, we showed the key results (static stacked baseline results, static CS results, and the DDD results) both in levels and logs. The logarithmized results were (mostly) in the appendix. After registering the PAP, we also added event-study plots (stacked in Figure A9 and CS in Figure A23, both post-blind) using logarithmized outcomes. The purpose was to visually examine whether the pre-treatment trends in the outcomes were parallel in both levels and logs.

Adoption: the follow-up length and cohort-specific effects (addition). The observed differences in estimates from our pre-registered stacking and CS specifications raised the possibility that the effects are increasing by the length of the follow-up period. To examine this, we added a CS event-study plot on the annualized number of primary care visits in Figure A26 (post-blind) and static stacked results using a 24-month follow-up (Table A7, post-blind). Additionally, Figure A27 (post-blind) shows the CS results for each treatment cohort.

Adoption: are social assistance recipients mainly in the first income decile (addition). We added a figure (post-blind) to the replication folder to answer the question. The

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

Adoption: the Roth and Sant'Anna (2023) estimator and later-treated comparisons (addition). We use not-yet-treated municipalities as comparisons in the main stacked analyses. As a robustness check, we exclude the never-treated from analyses and thus use the later-treated for comparison. After this change, it is a plausible assumption that treatment timing is quasi-random. Making this assumption, we estimate the effects on primary care use with the estimator proposed by Roth and Sant'Anna (2023) that can be much more precise than the earlier DD based alternatives (e.g., the CS estimator). The results are in Table A9 (post-blind) in levels and in Table A10 (post-blind) in logs.

Adoption: ED visits and unplanned hospitalizations for ACSCs (addition). We also estimated the effects on emergency department (ED) visits and unplanned hospitalizations for ambulatory care sensitive conditions (ACSC), both post-blind outcomes. We estimate the impacts on all individuals to increase sample size and use the CS estimator because it has a longer follow-up than our baseline stacking analyses with a one-year follow-up. The longer follow-up is reasonable because the negative health impacts likely accumulate over time. The CS event-study plots are in Figure A20 (post-blind) and the static estimates in Figure A21 (post-blind).

Adoption: effects for diabetes and hypertension patients (addition). We also examine treatment effect heterogeneity with respect to having received a drug prescription in 2018–2019 with an ATC code referring to diabetes or hypertension (A10, C02–C03, and C07–C09), proxying a diagnosis of these conditions. The stacking results are in Table A2 (post-blind).

Adoption: means and standard deviations of the outcomes (addition). Based on the feedback we received in the peer review process, we added Table A6 (post-blind) which includes means and standard deviations of the outcome variables, complementing Table 1 of the main paper.

The role of adoption vs. abolition analyses (change). In the PAP, the plan was to

equally examine and report both the staggered adoption and the later abolition of the copayment. The final research paper focuses much more on the staggered adoption because its design is much more credible, and we moved parts of the abolition analyses (specifically, regression results) to the Online Appendix. The adoption analyses are based on several events occurring at different times, thus exploiting variation both in time and area. This reduces the probability that external factors systematically bias the results. In contrast, the abolition analyses rely only on one policy change that occurred during the COVID-19 pandemic when the contemporary changes in primary care use were large, caused by both demand-side and supply-side factors. The PAP should have more explicitly argued that the adoption represents a more promising design *ex ante*. *Ex post*, we have serious doubts about the validity of the parallel trends assumption for the abolition setting based on pre-treatment patterns of outcomes. This lead us to focus more on the staggered adoption setting which we argue is a credible and good design.

Abolition: follow-up length and socioeconomic data (change). In contrast to what was planned, the reported abolition analyses use slightly different/inferior data. Regarding socioeconomic data, we use data from 2020 for years 2021–2022. We could have added the year 2021 (expected to be available by June 2023) as planned, but did not. Regarding outcomes, the pre-registered follow-up for the abolition analyses was 12 months. In this version, we use an 11-month follow-up since we currently have data only until the end of 5/2022. We have balanced benefits and costs when deciding not to get the most recent data for the analysis. The benefits would be small given that the pre-treatment evolution in outcomes led us to view the parallel trends assumption as not credible. Getting the most recent data would not change this. On the other hand, there are costs in terms of time and money, and also waiting times, for requesting the most recent batch of outcome data. We have for now prioritized other aspects when refining the paper and its analyses.

Abolition: bounding violations of parallel trends (addition). We used the methods proposed by Rambachan and Roth (2023) to assess how sensitive the abolition results are to small post-treatment violations of parallel trends. Figure 6 (post-blind) shows that they indeed

are sensitive. There is major uncertainty about the validity of the identification assumption. The implication for us is that little can be learned about the effects of the abolition.

Abolition: regression analysis to the appendix (change). This is the most visible change from the PAP. The trend plots (Figure 5) and Figure 6 (post-blind) led us to seriously question the validity of the parallel trends assumption for the abolition analyses. The planned regression tables reported in the main text in the PAP (Table A11 and Table A13 of this paper) are rather irrelevant if we cannot credibly make the parallel trends assumption. For this reason, we moved the abolition regression tables to the Online Appendix where readers can see them. We emphasize great caution in interpreting the point estimates in any way as uncertainty about the identification assumption is so large.

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

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

A.5 Parameter Values for Semi-Arc Elasticities

To obtain the price of care after the copayment adoption, we multiply a per-visit copayment of 10 euros by 0.8, which is chosen to present the probability that the patient pays the copayment. The copayment is not always charged. As Figure A3 shows, by far the most common policy is to charge it for the first three visits annually. Finally, we divide the multiplication by the total cost of the visit. For our baseline, we use the estimated average cost of 35 euros in 2017 in public healthcare (Mäklin & Kokko, 2020). To obtain the upper bound, we use a total cost of 61 euros. This is the total cost of a 20-minute nurse appointment in a major private provider in July 2022 (66 euros), adjusted to the 2017 prices.

Regarding the quantities, we use Table A3 (post-blind) as the baseline. These estimates present the stacked results for all individuals in a one-year follow-up. Specifically, we average over the two specifications in Panel A that assume parallel trends and use contacts per capita as the outcome: the pre-treatment mean is thus 1.000 annualized nurse visits, and the effect estimate is -0.089. To obtain the upper bound, we fix the pre-treatment mean of 1.000 annualized visits, but use a -15% reduction as the effect estimate, which is based on Figure A15 (post-blind) showing the CS results for all individuals. In this case, the follow-up time is on average longer and depends on the municipality.

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

A.6 Additional Figures and Tables

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

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

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

Table A2: Adoption: DD Comparisons, Primary Care Use, Diabetes or Hypertension.

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

Notes: This table was not pre-registered and is post-blind. The dataset is stacked. Our baseline model is Model 1, but we also use its extension (Model 1.2) allowing for differential linear pre-trend ("trends"). Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. The table examines treatment effect heterogeneity with respect to having received a drug prescription ("has prescription") in 2018–2019 with an ATC code referring to diabetes or hypertension (A10, C02–C03, and C07–C09), proxying a diagnosis of these conditions. Outcomes are the annualized number of curative nurse and GP visits, respectively. Estimates for logarithmized outcomes are multiplied by 100.

Table A3: Adoption: DD Comparisons, Primary Care Use, All Individuals.

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

Notes: This table was not pre-registered and is post-blind. The dataset is stacked. Our baseline model is Model 1, but we also use its extension (Model 1.2) allowing for differential linear pre-trend ("trends"). Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Outcomes are the annualized number of curative nurse and GP visits (or their logarithm). Depending on the column, event-specific datasets are either balanced or unbalanced. Estimates for logarithmized outcomes are multiplied by 100.

Table A4: Adoption: Social Assistance Use.

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

Notes: The dataset is stacked and balanced. Our baseline model is Model 1, but we also use its extension (Model 1.2) allowing for differential linear pre-trend ("trends"). Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Outcomes are the share of individuals in a family receiving social assistance (in percentages) and the annual sum of of the family's received basic social assistance (in euros) per capita. With the latter outcome, we only include events that occurred on January 1st.

Table A5: Adoption: DD Comparisons, Primary Care Use, Robustness Checks.

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

Notes: The dataset is stacked. Our baseline model is Model 1, but we also use its extension (Model 1.2) allowing for differential linear pre-trend ("trends"). Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively. Estimates for logarithmized outcomes are multiplied by 100.

Table A6: Adoption: Means, Standard Deviations, and Difference-in-Differences Estimates of Effects of Nurse Visit Copayment on Primary Care Visits.

	Bottom 40%		Top 40%	
Metric	Before	After	Before	After
A. Nurse visits				
Treated mean	1.373	1.298	0.603	0.574
Treated SD	0.555	0.536	0.298	0.278
Control mean	1.182	1.235	0.487	0.505
Control SD	0.518	0.537	0.262	0.272
Effect		-0.127		-0.048
Relative effect (%)		-9.252		-8.020
Std. effect size		-0.229		-0.162
B. GP visits				
Treated mean	1.450	1.362	0.705	0.660
Treated SD	0.435	0.455	0.288	0.289
Control mean	1.345	1.320	0.593	0.582
Control SD	0.395	0.390	0.251	0.247
Effect		-0.056		-0.033
Relative effect (%)		-3.879		-4.669
Std. effect size		-0.129		-0.114

Notes: This table was not pre-registered and is post-blind. The table contains population-weighted means and standard deviations of the annualized number of public primary care visits per capita, estimated separately for the bottom 40% and the top 40% of the distribution of equivalized family disposable income, before and after the introduction of the nurse visit copayment. Outcomes are the annualized number of curative nurse and GP visits per capita. The table also contains our baseline stacked difference-in-differences effect estimates from Model 1 for the impact of introducing the nurse visit copayment – the same estimates are in Table 1 of the main paper. Relative effects are calculated by dividing effect estimates by baseline means among the treated and multiplying by 100. Standardized effects sizes are constructed by dividing effect estimates by baseline standard deviations among the treated. The dataset is stacked, balanced, and contains a one-year follow-up. Due to heterogeneity in municipality size, we weight by population size.

Table A7: Adoption: DD Comparisons, Primary Care Use, 24-Month Follow-Up.

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

Notes: This table was not pre-registered and is post-blind. The dataset is stacked and balanced. Our baseline model is Model 1, but we also use its extension (Model 1.2) allowing for differential linear pre-trend ("trends"). In contrast to the main analysis, we use a 24-month follow-up. Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively.

Table A8: Adoption: DD Comparisons, Primary Care Use, Uniform Weighting.

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

Notes: This table was not pre-registered and is post-blind. The dataset is stacked and balanced. Our baseline model is Model 1, but we also use its extension (Model 1.2) allowing for differential linear pre-trend ("trends"). We weight our municipality-by-income-decile observations uniformly. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number of curative nurse and GP visits, respectively.

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

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

Notes: This table was not pre-registered and is post-blind. We use the Roth and Sant'Anna (2023) estimator (RS) and estimate the effects on annualized number of curative nurse and GP visits per capita. We report point estimates, standard errors (in parenthesis), and the effect as a percentage change (in square brackets). Pre-treatment mean for the treated is averaged over 12 pre-treatment months. Later-treated municipalities are used as comparisons. Events that have at least 12 follow-up months are included. We use a balanced panel in calendar time and drop treatment cohorts that have only a single treated municipality. Municipalities are weighted equally. We use the refined variance estimator proposed by Roth and Sant'Anna (2023) for standard errors unless the software produces a warning (12 cases out of 96) that a more conservative Neyman-style variance estimate is less than an estimated adjustment factor. In these cases, we consequently use the unrefined Neyman-style variance estimator discussed in Roth and Sant'Anna (2023). We show the results for four study windows. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Three aggregated estimands proposed by Roth and Sant'Anna (2023) are reported. "Simple" is a simple average of the $ATE(t,g)$ weighted by cohort size. "Cohort" first averages the $ATE(t,g)$ for each cohort g before averaging these cohort effects weighting by cohort size. "Calendar" first averages the $ATE(t,g)$ for each time period t weighting by cohort size before taking a mean of the calendar effects.

Table A10: Adoption: the RS estimator, Logarithmized Primary Care Use.

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

Notes: This table was not pre-registered and is post-blind. We use the Roth and Sant'Anna (2023) estimator (RS) and estimate the effects on logarithmized annualized number of curative nurse and GP visits per capita. We report point estimates and standard errors (in parenthesis). Later-treated municipalities are used as comparisons. Events that have at least 12 follow-up months are included. We use a balanced panel in calendar time and drop treatment cohorts that have only a single treated municipality. Municipalities are weighted equally. We use the refined variance estimator proposed by Roth and Sant'Anna (2023) for standard errors unless the software produces a warning (12 cases out of 96) that a more conservative Neyman-style variance estimate is less than an estimated adjustment factor. In these cases, we consequently use the unrefined Neyman-style variance estimator discussed in Roth and Sant'Anna (2023). We show the results for four study windows. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Three aggregated estimands proposed by Roth and Sant'Anna (2023) are reported. "Simple" is a simple average of the $ATE(t,g)$ weighted by cohort size. "Cohort" first averages the $ATE(t,g)$ for each cohort g before averaging these cohort effects weighting by cohort size. "Calendar" first averages the $ATE(t,g)$ for each time period t weighting by cohort size before taking a mean of the calendar effects.

Table A11: Abolition: Main DD Estimates.

Metric	Nurse Visits	GP Visits	Prescriptions
A. Bottom 40%			
Level	1.075	1.057	5.929
Municipalities	241	241	249
Estimate (w/o trends, Model A1)	−0.011	−0.011	−0.029
Std. error	0.033	0.023	0.029
Change (%)	−0.978	−1.067	−0.490
Estimate (with trends, Model A2)	0.085	0.053	−0.090
Change (%)	7.862	5.033	−1.511
Estimate (CS)	0.035	−0.013	−0.081
Std. error	0.027	0.027	0.042
Change (%)	3.272	−1.218	−1.371
B. Top 40%			
Level	0.526	0.513	2.718
Municipalities	241	241	249
Estimate (w/o trends, Model A1)	0.041	0.005	0.018
Std. error	0.023	0.009	0.017
Change (%)	7.788	0.943	0.666
Estimate (with trends, Model A2)	0.064	0.026	−0.044
Change (%)	12.180	5.045	−1.624
Estimate (CS)	0.041	−0.002	0.001
Std. error	0.021	0.013	0.024
Change (%)	7.809	−0.367	0.019

Notes: We suggest caution in interpreting the point estimates as we do not view the parallel trends assumption as credible. The following methods are used: 1) a standard TWFE DD model (Model A1), 2) its modification allowing for linear pre-trend differences (Model A2), and 3) the CS estimator with outcome regression (Callaway & Sant'Anna, 2021) described in Section 4. We do not view any of the methods (read: any of the parallel trends assumptions) as superior to the rest. Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the annualized number contacts per capita.

Table A12: Abolition: DD Estimates, Logarithmized Outcome.

Metric	Nurse Visits	GP Visits	Prescriptions
A. Bottom 40%			
Municipalities	240	241	249
Estimate (w/o trends, Model A1)	0.233	0.143	−0.008
Std. error	2.718	2.727	0.574
P-value	0.932	0.958	0.989
Estimate (with trends, Model A2)	6.104	6.862	0.362
Estimate (CS)	3.838	−1.573	−0.938
Std. error	2.323	2.637	0.751
B. Top 40%			
Municipalities	240	241	249
Estimate (w/o trends, Model A1)	7.226	1.748	0.341
Std. error	4.268	2.508	0.709
P-value	0.092	0.487	0.632
Estimate (with trends, Model A2)	7.682	9.261	2.043
Estimate (CS)	7.070	−1.052	0.150
Std. error	4.564	2.890	1.022

Notes: We suggest caution in interpreting the point estimates as we do not view the parallel trends assumption as credible. The following methods are used: 1) a standard TWFE DD model (Model A1), 2) its modification allowing for linear pre-trend differences (Model A2), and 3) the CS estimator with outcome regression (Callaway & Sant'Anna, 2021) described in Section 4. We do not view any of the methods (read: any of the parallel trends assumptions) as superior to the rest. Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Outcomes are the logarithm of annualized number contacts per capita, and the estimates are multiplied by 100.

Table A13: Abolition: DDD Estimates.

Metric	Nurse Visits	GP Visits	Prescriptions
A. Annualized contacts per capita			
Level	1.075	1.057	5.929
Estimate	−0.051	−0.016	−0.047
Std. error	0.020	0.017	0.024
P-value	0.012	0.330	0.053
Change (%)	−4.790	−1.524	−0.795
Municipalities	241	241	249
B. Logarithmized annualized contacts per capita			
Estimate	−7.038	−1.556	−0.347
Std. error	2.602	1.487	0.588
P-value	0.007	0.295	0.555
Municipalities	240	241	249

Notes: We suggest caution in interpreting the point estimates as we do not view the parallel trends assumption as credible. Model A3 is used. Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Estimates for logarithmized outcomes are multiplied by 100.

Table A14: Abolition: Main DD Estimates, Uniform Weighting.

Metric	Nurse Visits	GP Visits	Prescriptions
A. Bottom 40%			
Level	1.253	1.217	6.185
Municipalities	241	241	249
Estimate (w/o trends, Model A1)	−0.010	−0.042	−0.098
Std. error	0.037	0.032	0.068
Change (%)	−0.768	−3.432	−1.585
Estimate (with trends, Model A2)	0.197	0.051	−0.060
Change (%)	15.692	4.218	−0.966
Estimate (CS)	0.101	0.061	−0.039
Std. error	0.048	0.040	0.078
Change (%)	8.048	5.008	−0.630
B. Top 40%			
Level	0.704	0.688	3.159
Municipalities	241	241	249
Estimate (w/o trends, Model A1)	0.016	−0.038	−0.078
Std. error	0.033	0.017	0.034
Change (%)	2.237	−5.569	−2.481
Estimate (with trends, Model A2)	0.209	−0.004	−0.091
Change (%)	29.666	−0.592	−2.868
Estimate (CS)	0.090	0.035	0.003
Std. error	0.042	0.024	0.058
Change (%)	12.606	4.992	0.087

Notes: We suggest caution in interpreting the point estimates as we do not view the parallel trends assumption as credible. The following methods are used: 1) a standard TWFE DD model (Model A1), 2) its modification allowing for linear pre-trend differences (Model A2), and 3) the CS estimator with outcome regression (Callaway & Sant'Anna, 2021) described in Section 4. We do not view any of the methods (read: any of the parallel trends assumptions) as superior to the rest. In contrast to the main analysis, the CS estimator weights municipalities uniformly, and 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 equivalized family disposable income. Outcomes are the annualized number contacts per capita.

Table A15: Abolition: DDD Estimates, Uniform Weighting.

Metric	Nurse Visits	GP Visits	Prescriptions
A. Annualized contacts per capita			
Level	1.253	1.217	6.185
Estimate	−0.025	−0.003	−0.020
Std. error	0.026	0.023	0.055
P-value	0.323	0.879	0.721
Change (%)	−2.025	−0.283	−0.317
Municipalities	241	241	249
B. Logarithmized annualized contacts per capita			
Estimate	−3.717	1.606	1.284
Std. error	2.331	1.483	0.982
P-value	0.111	0.279	0.191
Municipalities	240	241	249

Notes: We suggest caution in interpreting the point estimates as we do not view the parallel trends assumption as credible. Model A3 is used. In contrast to the main analysis, we weight municipality-by-income-decile observations uniformly. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income. Estimates for logarithmized outcomes are multiplied by 100.

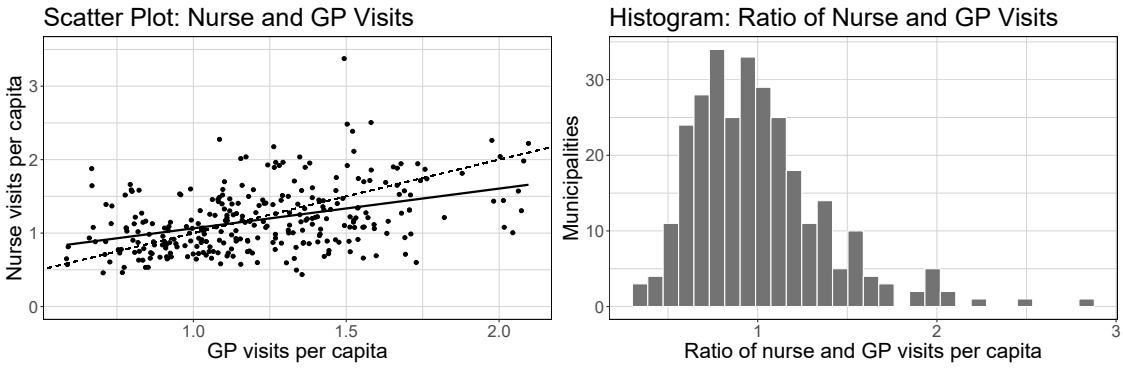

Figure A1: Curative Nurse and GP Visits by Municipality.

Notes: This figure was not pre-registered and is post-blind. On the left, the figure shows the scatter plot of curative GP and nurse utilization by municipality using data from 2013–2019. The black line represents the fitted OLS curve and the dashed line a 1-to-1 correspondence. On the right, the figure shows the histogram of the ratio of curative nurse visits to curative GP visits, using municipal-level data from 2013–2019.

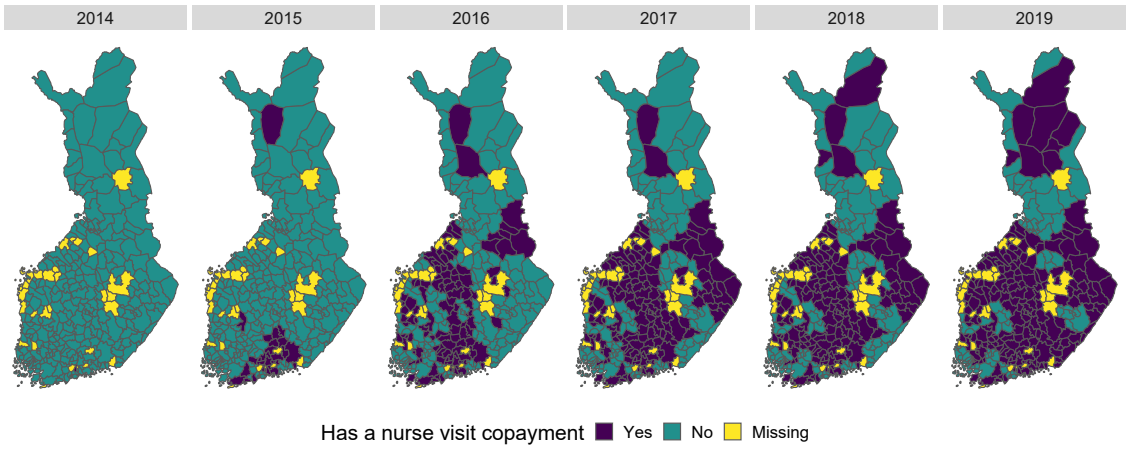

Figure A2: Staggered Adoption on Map.

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.

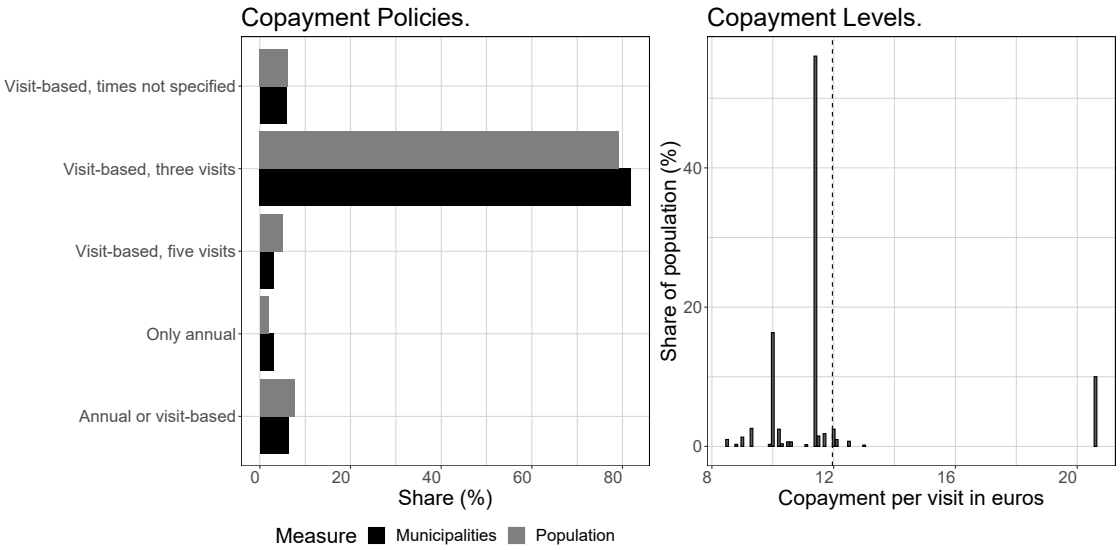

Figure A3: Copayment Levels and Policies in the Summer of 2021.

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

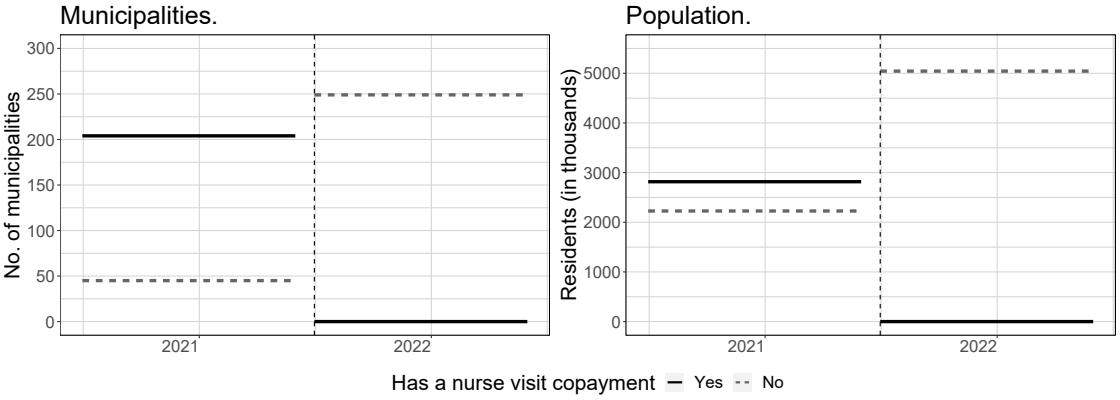

Figure A4: Simultaneous Abolition of the Nurse Visit Copayment.

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

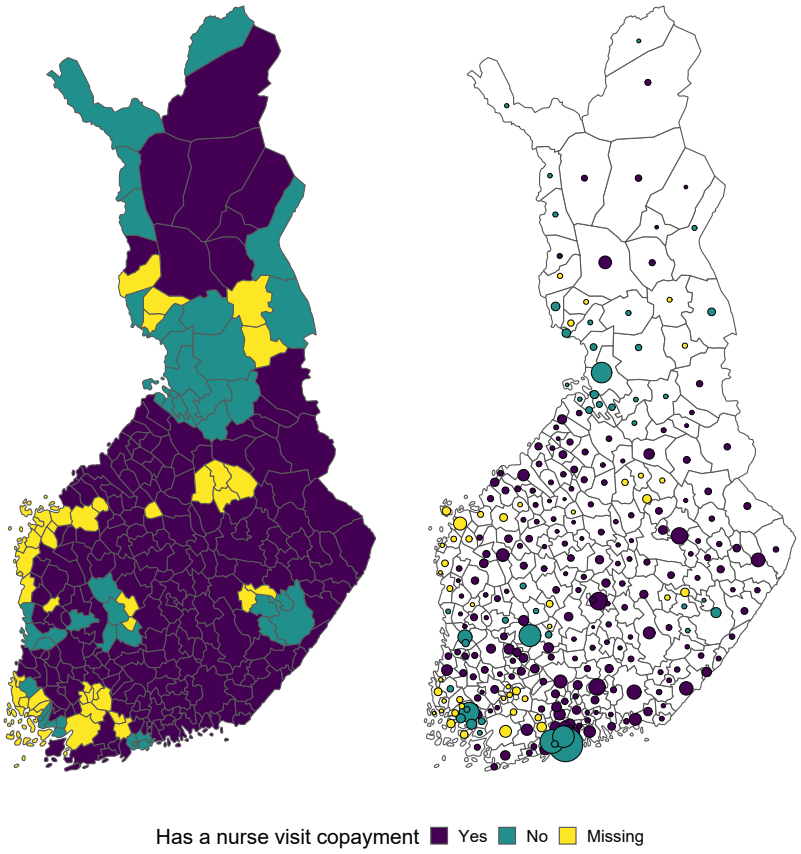

Figure A5: The Abolition on Map.

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

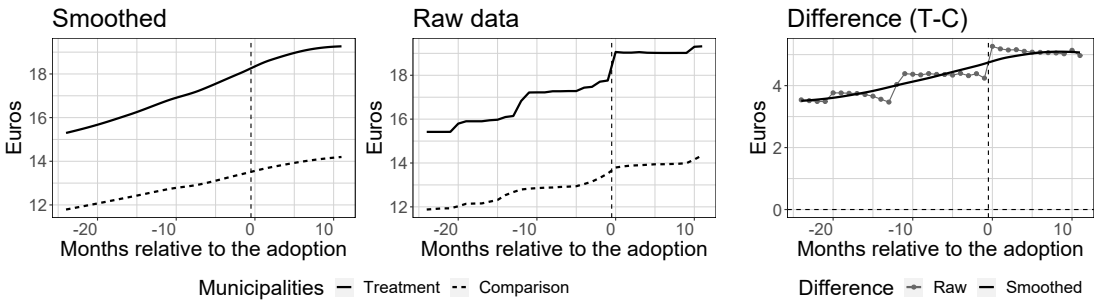

Figure A6: Adoption: Evolution in GP Visit Copayments.

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

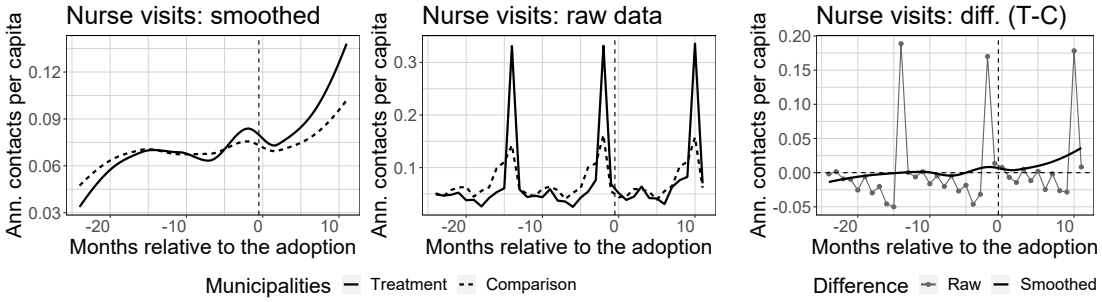

Figure A7: Adoption: Evolution in Preventive Nurse Visits, All Individuals.

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

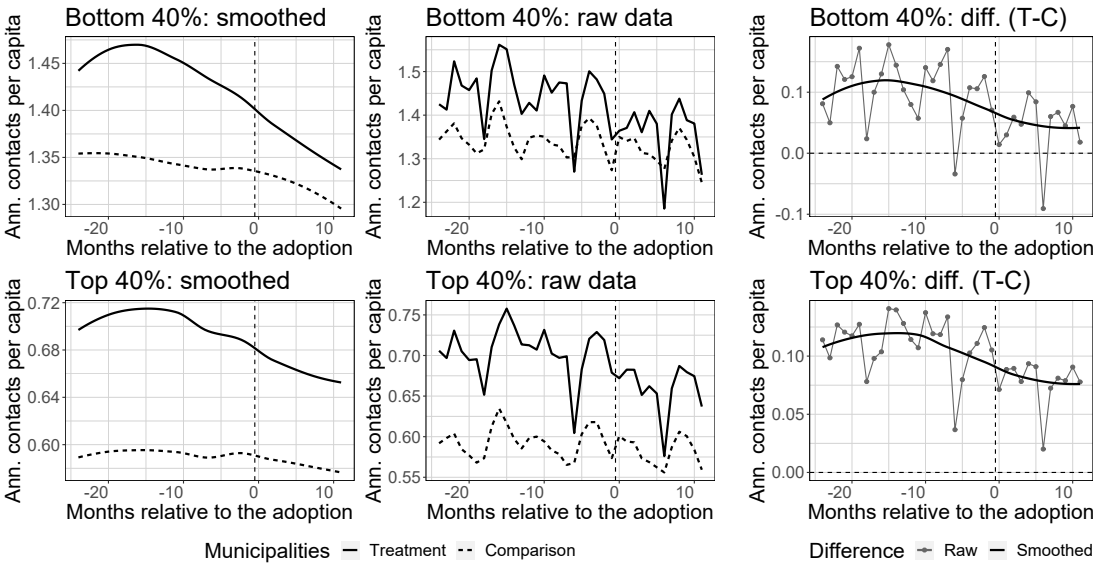

Figure A8: Adoption: Evolution in GP Visits.

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

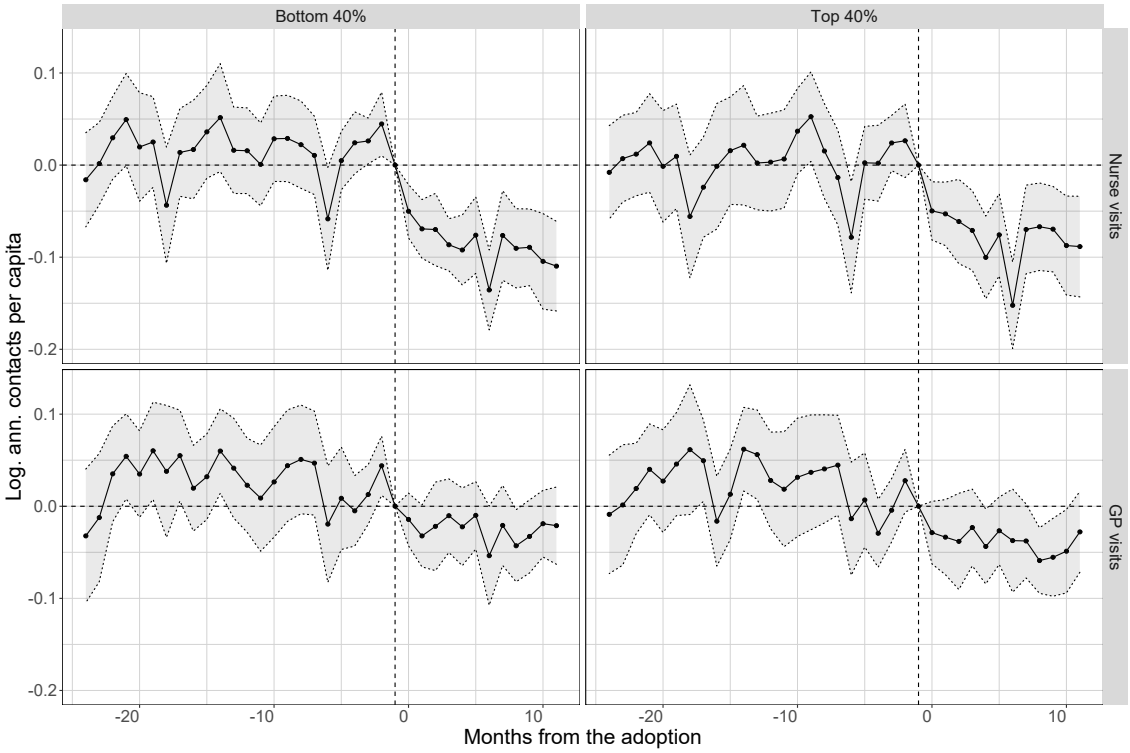

Figure A9: Adoption: Event-Study Plot on Logarithmized Primary Care Use with Stacked Data.

Notes: This figure was not pre-registered and is post-blind. We use Model 1.1. The point estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The dataset is stacked, outcome logarithmized, and event-specific datasets balanced. We compare the evolution of annualized contacts per capita between treated and unexposed municipalities. Due to heterogeneity in municipality size, we weight by population size. Standard errors are clustered by municipality. Bottom 40% and top 40% refer to the distribution of equivalized family disposable income.

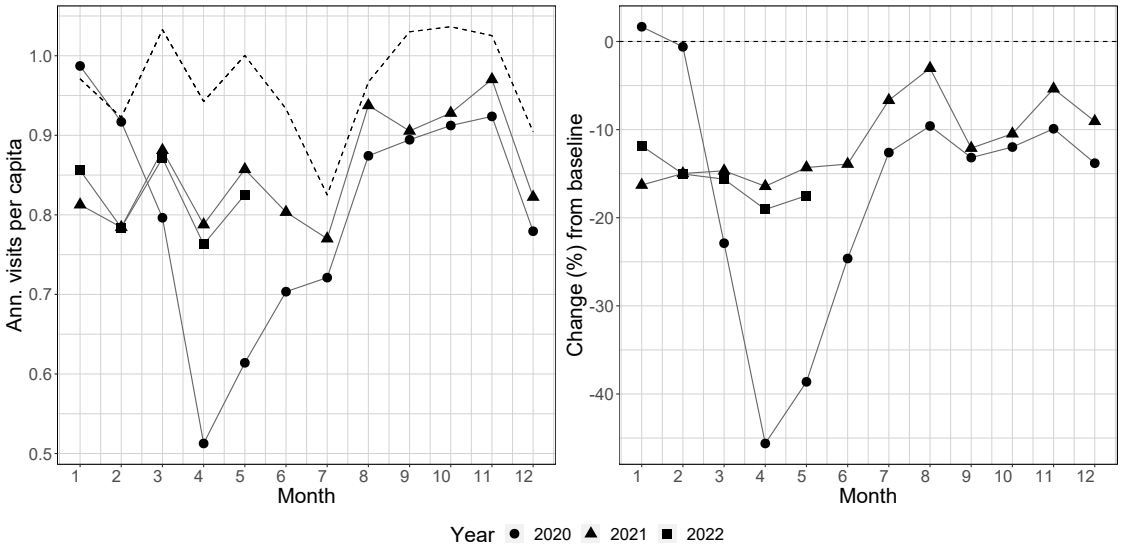

Figure A10: The Evolution of Nurse Visits during the COVID-19 Pandemic.

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

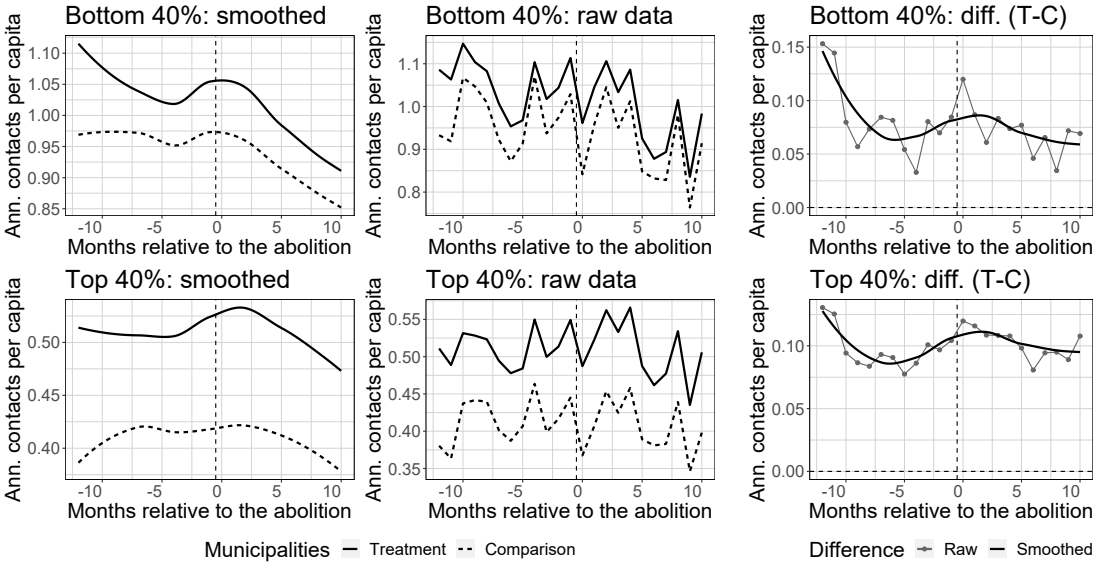

Figure A11: Abolition: Evolution in GP Visits.

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

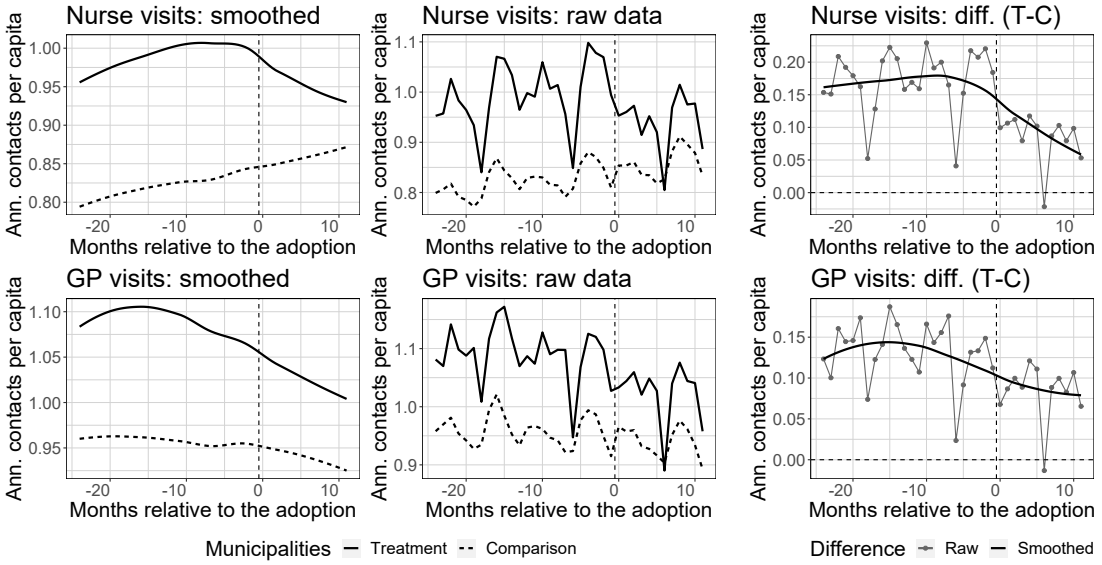

Figure A12: Adoption: Evolution in Outcomes, All Individuals.

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

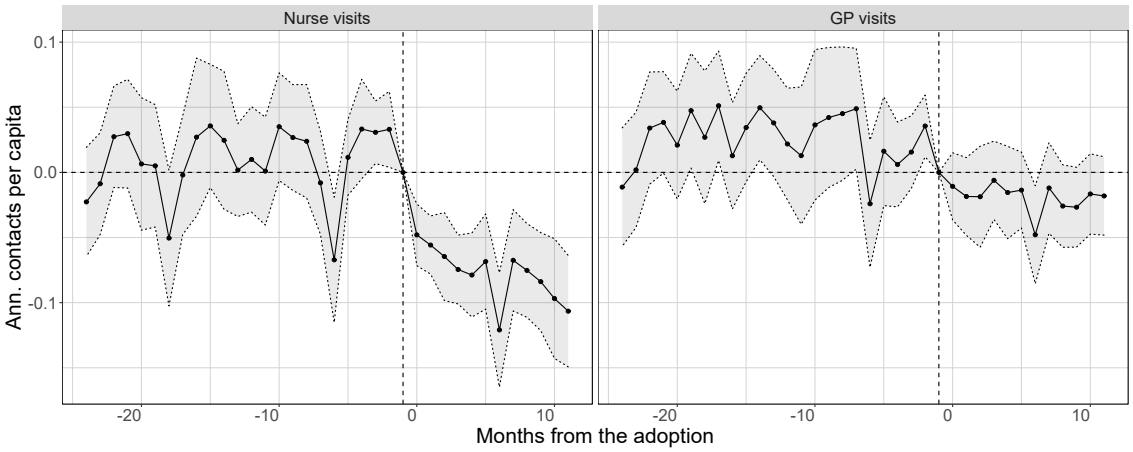

Figure A13: Adoption: Stacked Event-Study Plot on Primary Care Use, All Individuals.

Notes: This figure was not pre-registered and is post-blind. We use Model 1.1. The point estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The dataset is stacked, and event-specific datasets balanced. We compare the evolution of annualized contacts per capita between treated and unexposed municipalities. Due to heterogeneity in municipality size, we weight by population size. The standard errors are clustered by municipality.

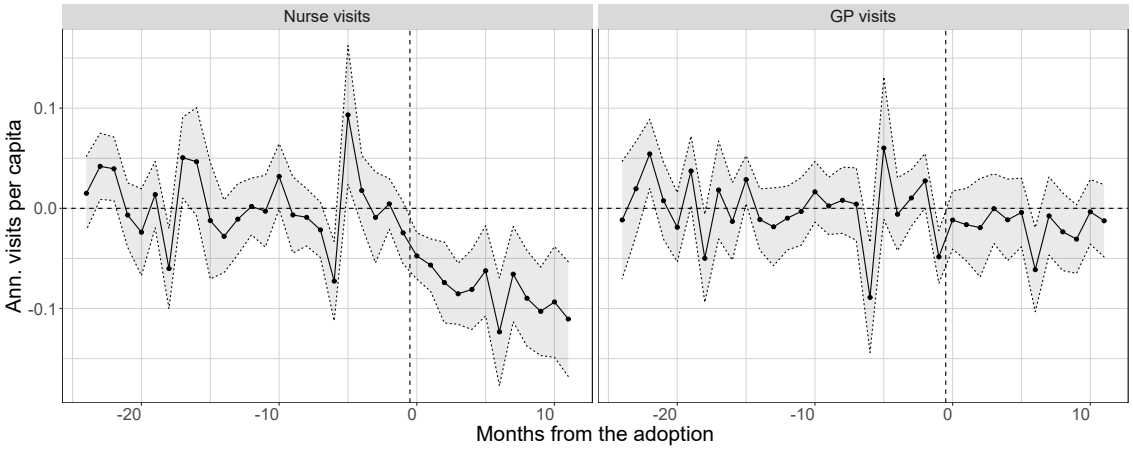

Figure A14: Adoption: the CS Event-Study Plots, Primary Care Use, All Individuals.

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

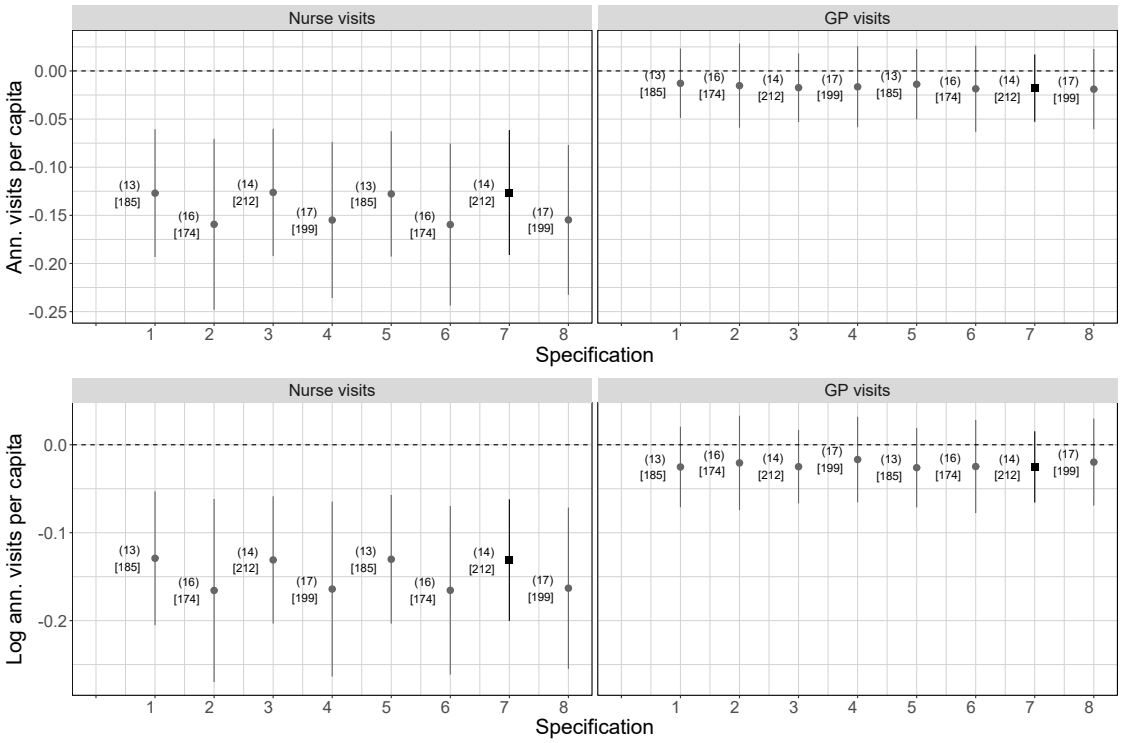

Figure A15: Adoption: the CS Estimator, Primary Care Use, All Individuals.

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

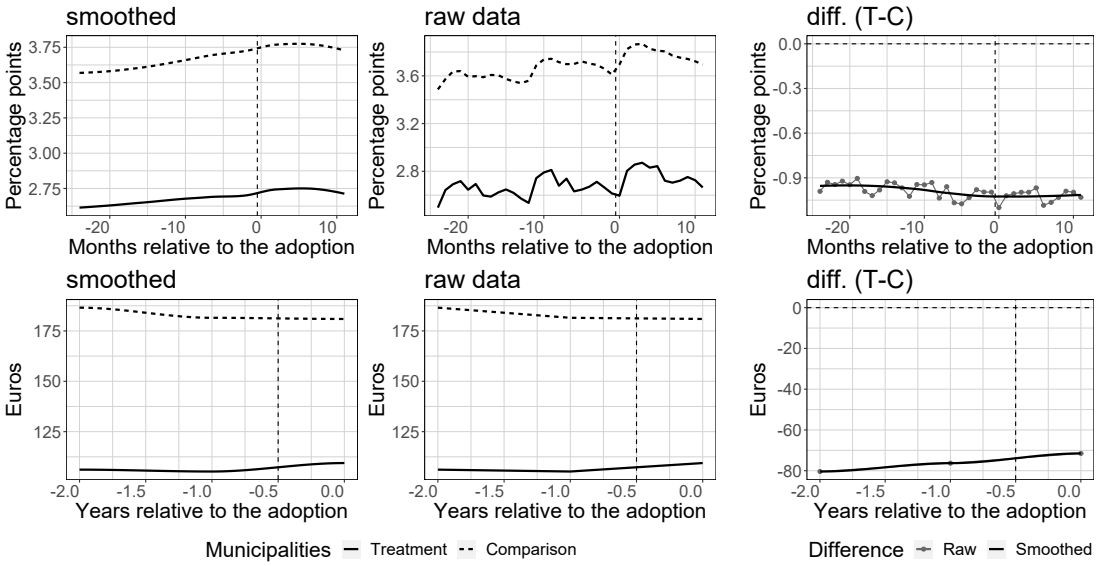

Figure A16: Adoption: Evolution in Social Assistance Recipients and Sums.

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

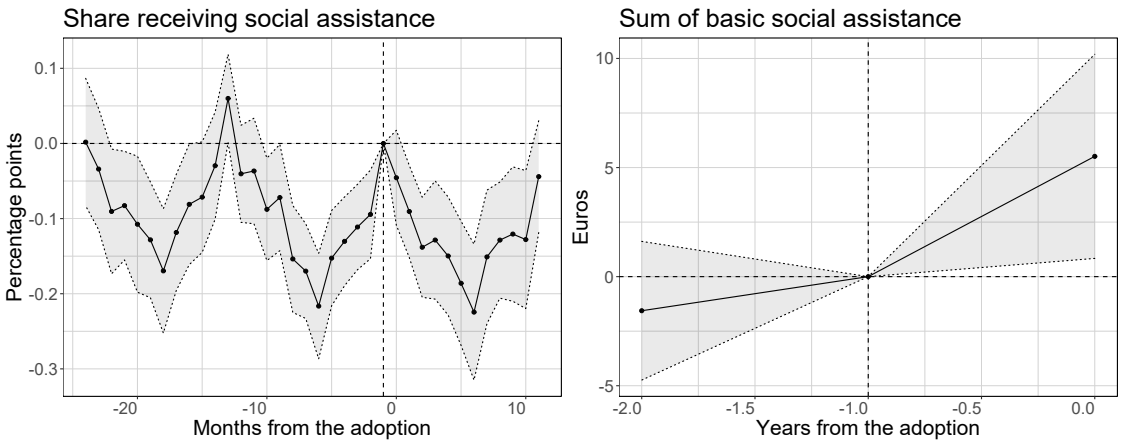

Figure A17: Adoption: Event-Study Plots on Social Assistance Use with Stacked Data.

Notes: We use Model 1.1. The point estimates represent effect estimates for the treatment group as a function of time relative to the copayment adoption. The dataset is stacked, and event-specific datasets balanced. We compare the evolution of outcomes between treated and unexposed municipalities. With respect to the annual data on the sum of received social assistance, we include only events that occurred on January 1st. Due to heterogeneity in municipality size, we weight by population size. The standard errors are clustered by municipality.

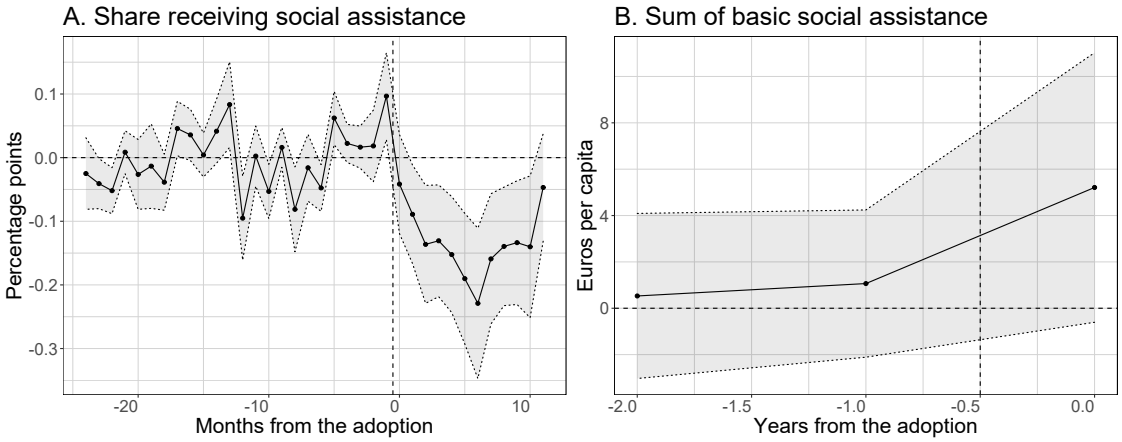

Figure A18: Adoption: Event-Study Plots Using the CS Estimator, Social Assistance Use.

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

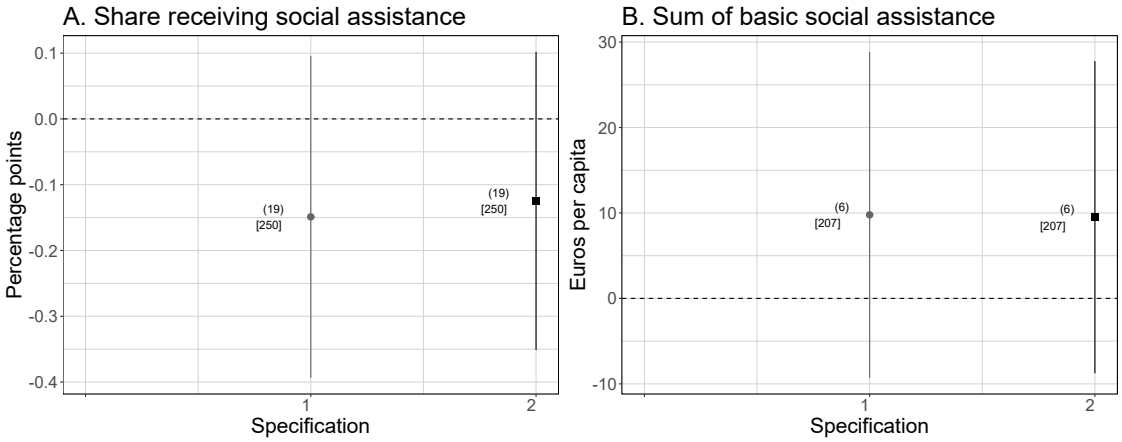

Figure A19: Adoption: the CS Estimator, Social Assistance Use.

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

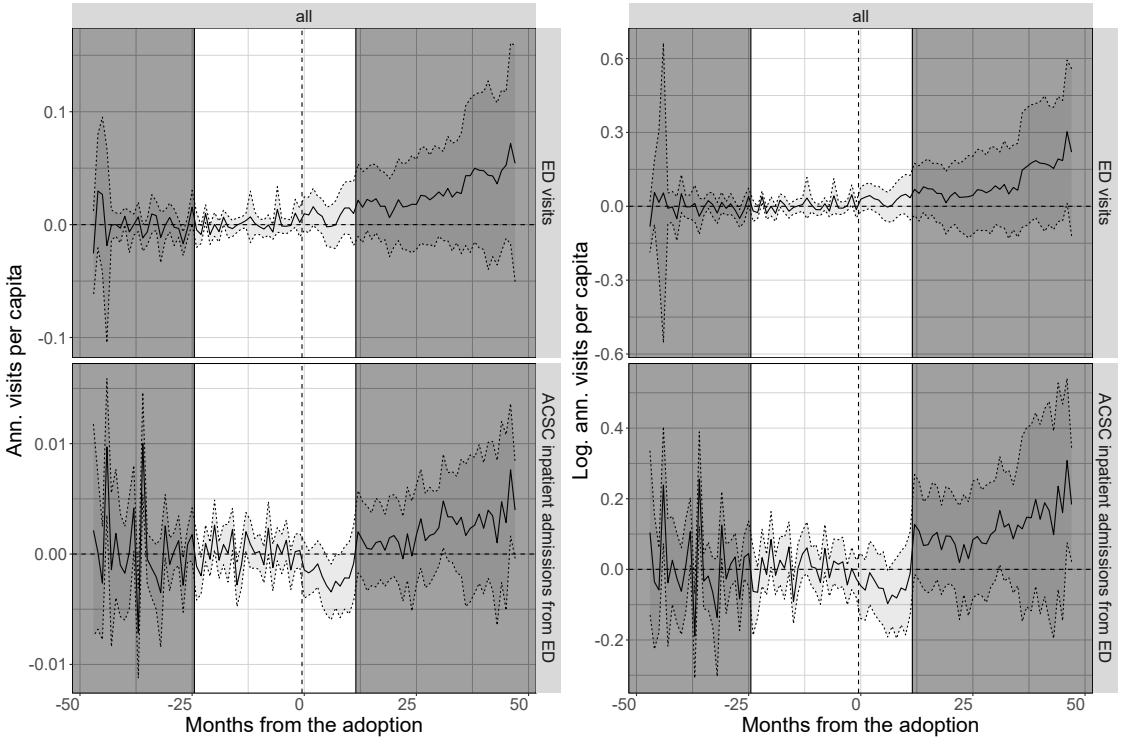

Figure A20: Adoption: Event-Study Plots Using the CS Estimator, ED Visits and Unplanned Hospitalizations for ACSCs.

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

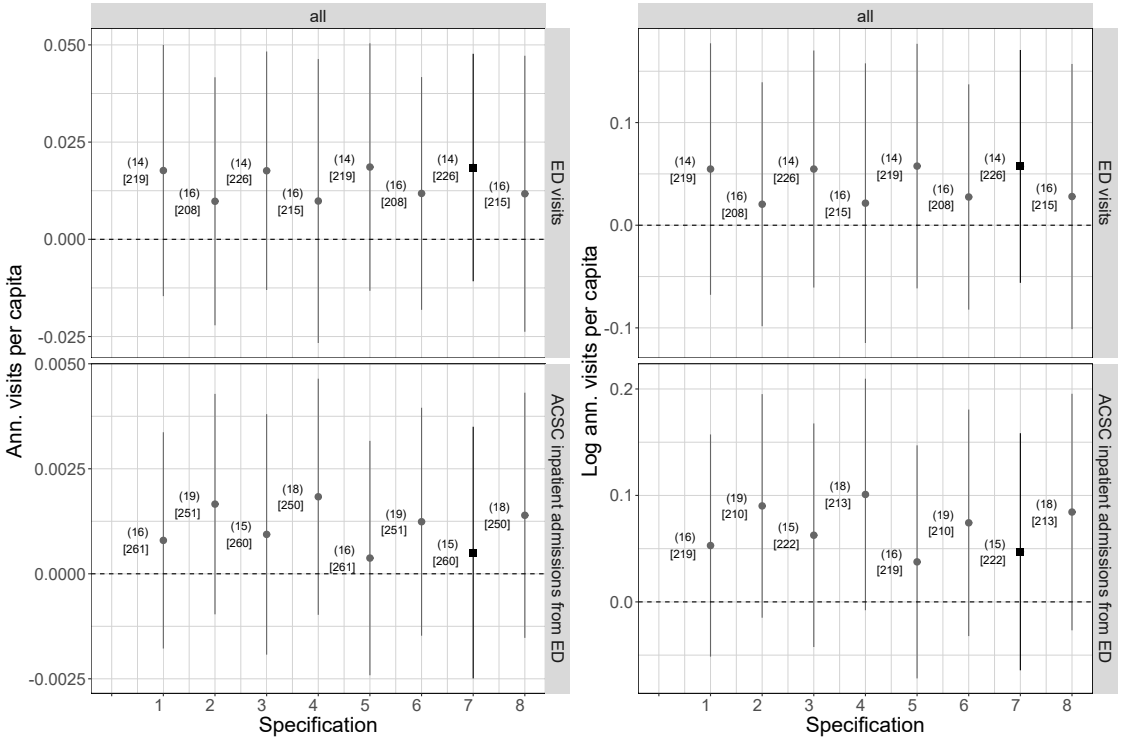

Figure A21: Adoption: the CS Estimator, ED Visits and Unplanned Hospitalizations for ACSCs.

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

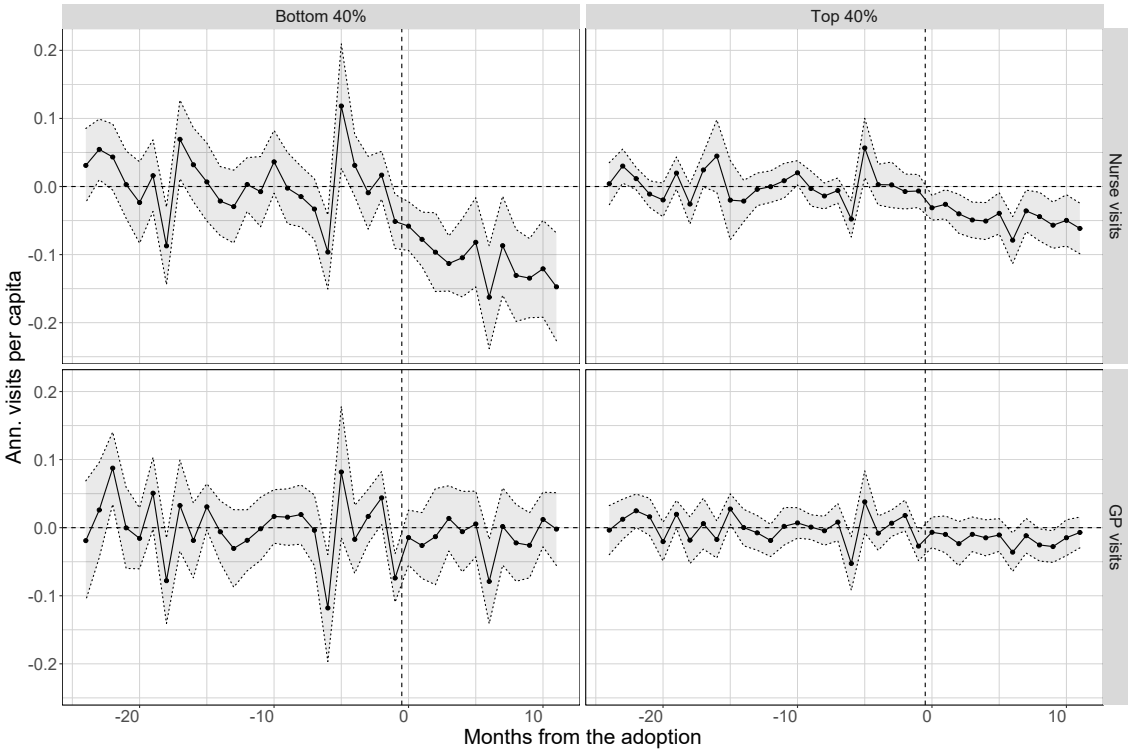

Figure A22: Adoption: Event-Study Plots Using the CS Estimator, Primary Care Use.

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

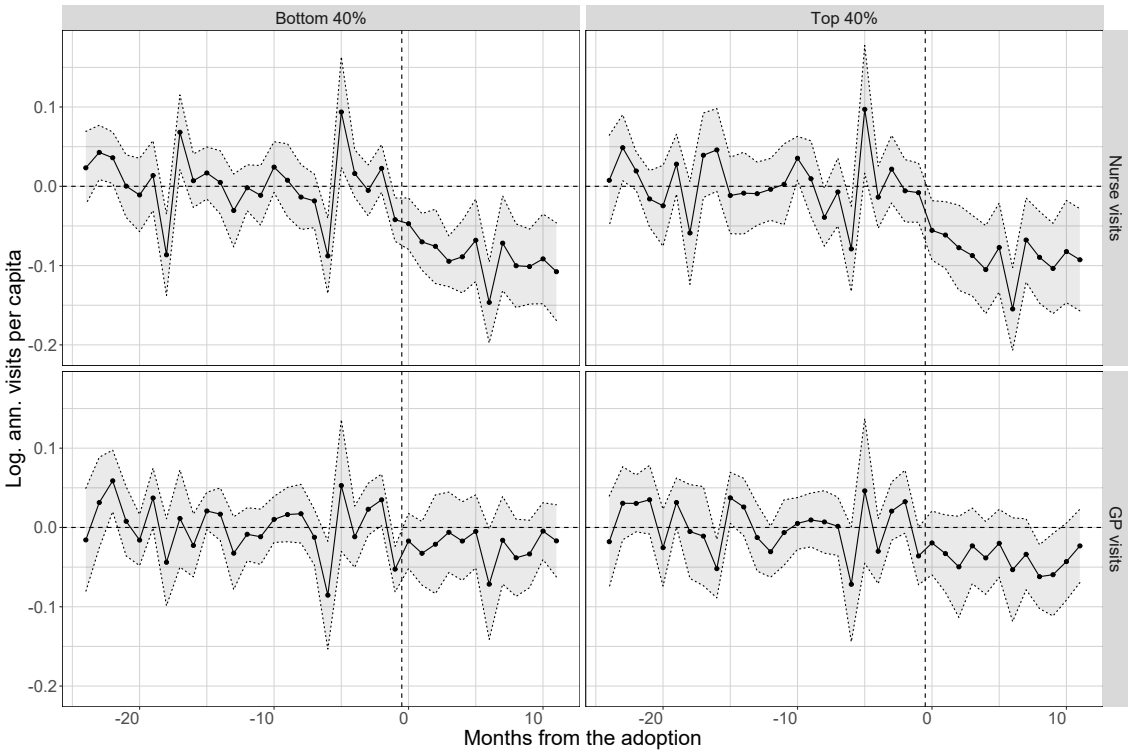

Figure A23: Adoption: Event-Study Plots Using the CS Estimator, Logarithmized Primary Care Use.

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

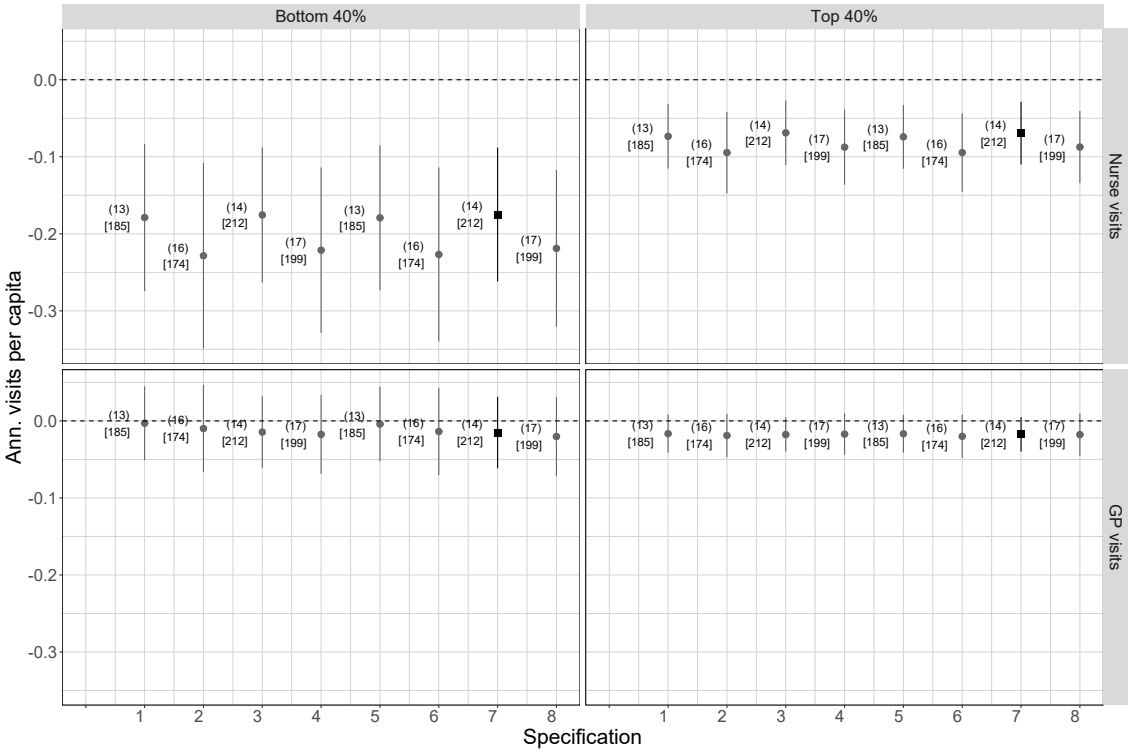

Figure A24: Adoption: the CS Estimator, Primary Care Use.

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

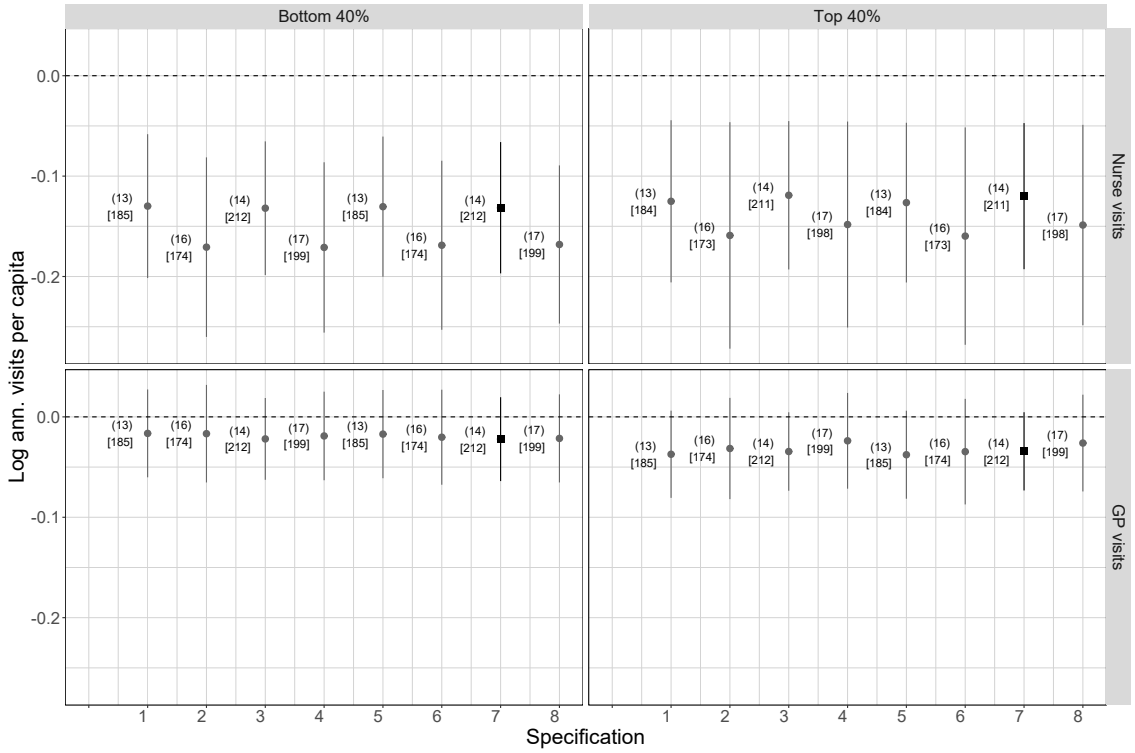

Figure A25: Adoption: the CS Estimator, Logarithmized Primary Care Use.

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

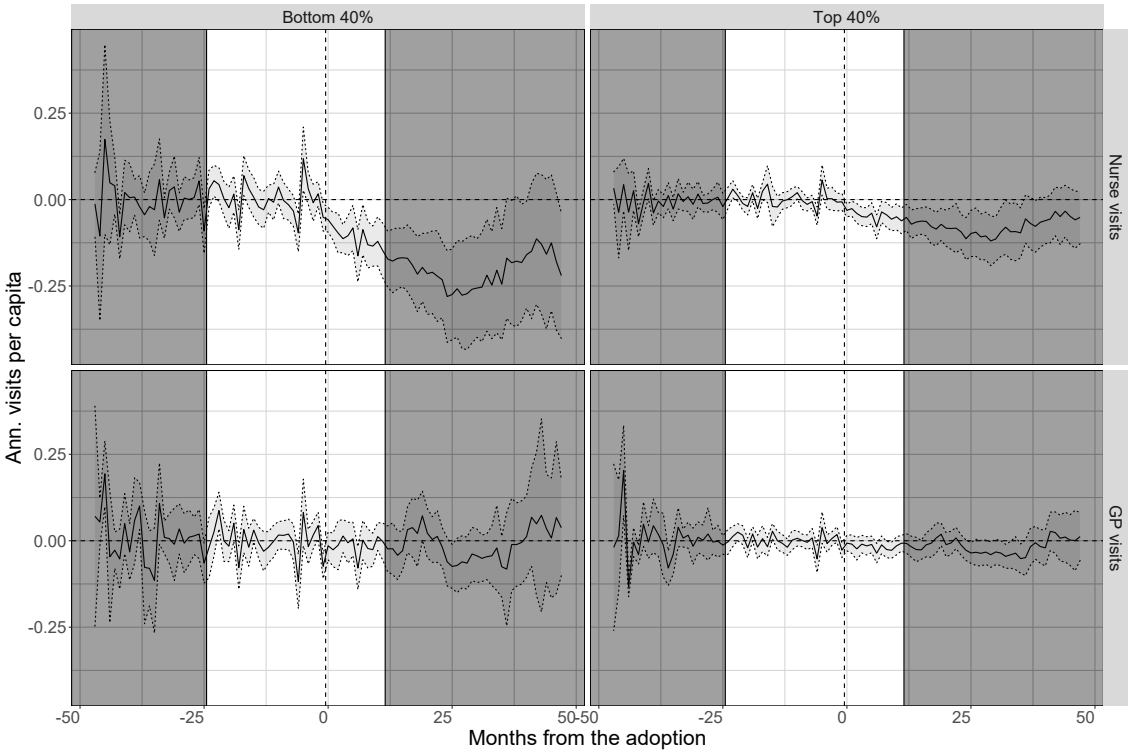

Figure A26: Adoption: Event-Study Plots Using the CS Estimator, Primary Care Use, All Leads and Lags.

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

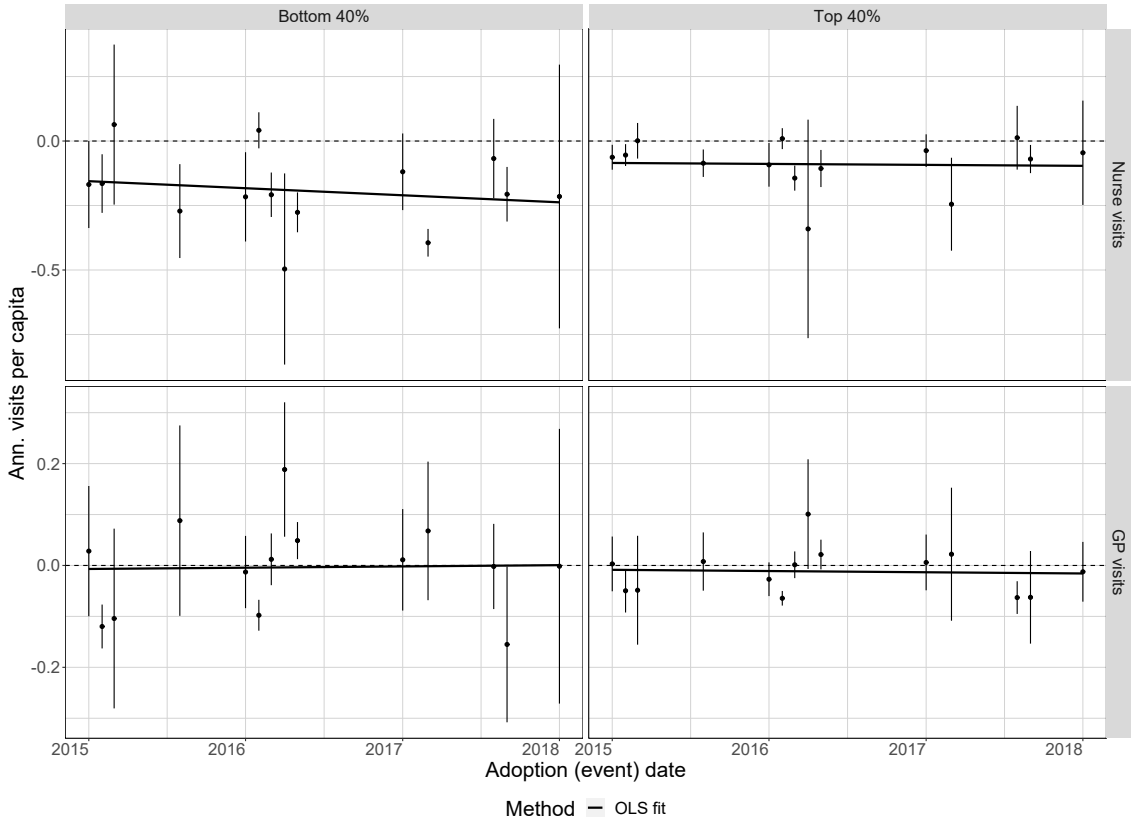

Figure A27: Adoption: Cohort-Specific CS Results, Primary Care Use.

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

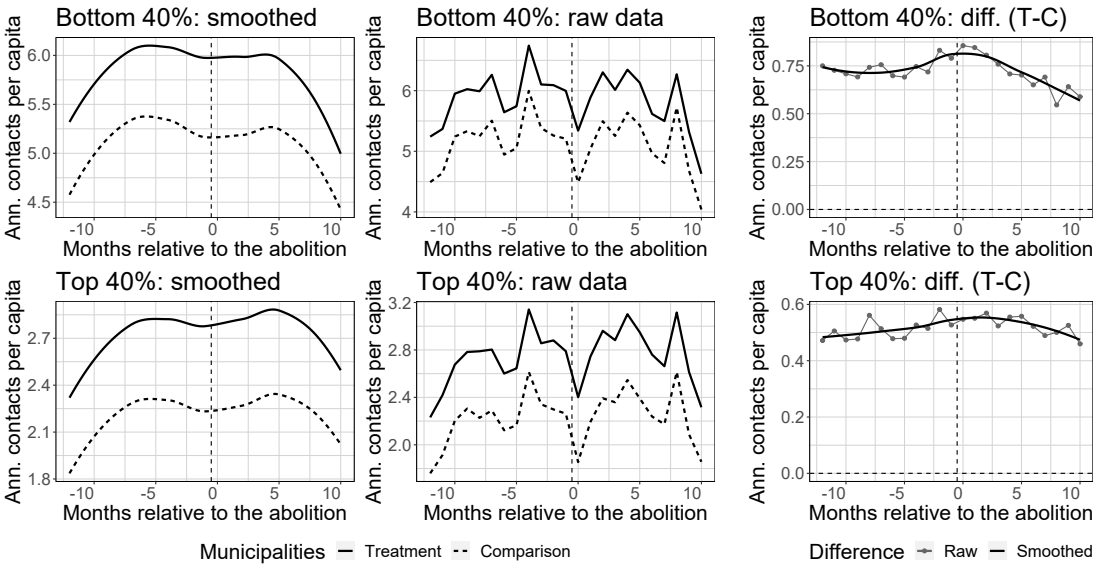

Figure A28: Abolition: Evolution Prescriptions.

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

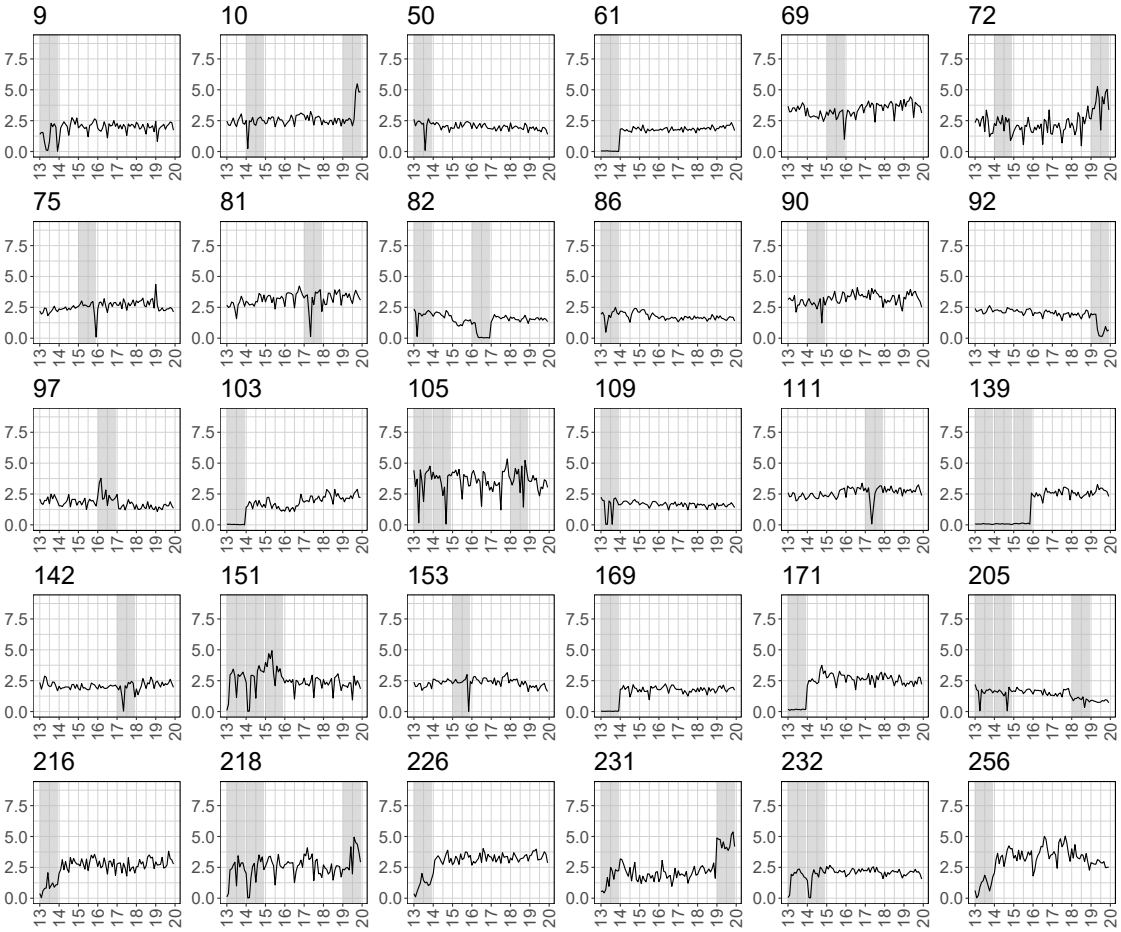

Figure A29: Adoption: Municipalities with Issues in the Primary Care Data, 1.

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

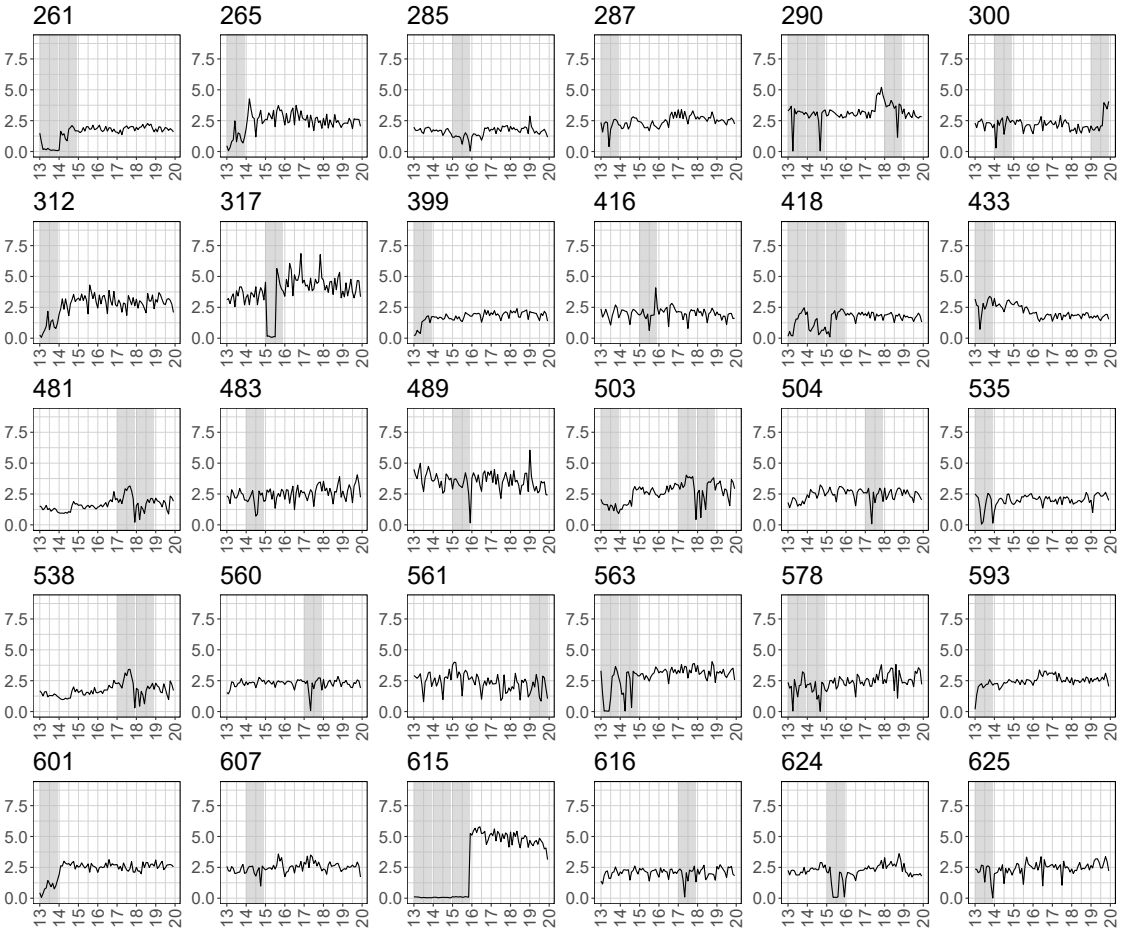

Figure A30: Adoption: Municipalities with Issues in the Primary Care Data, 2.

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

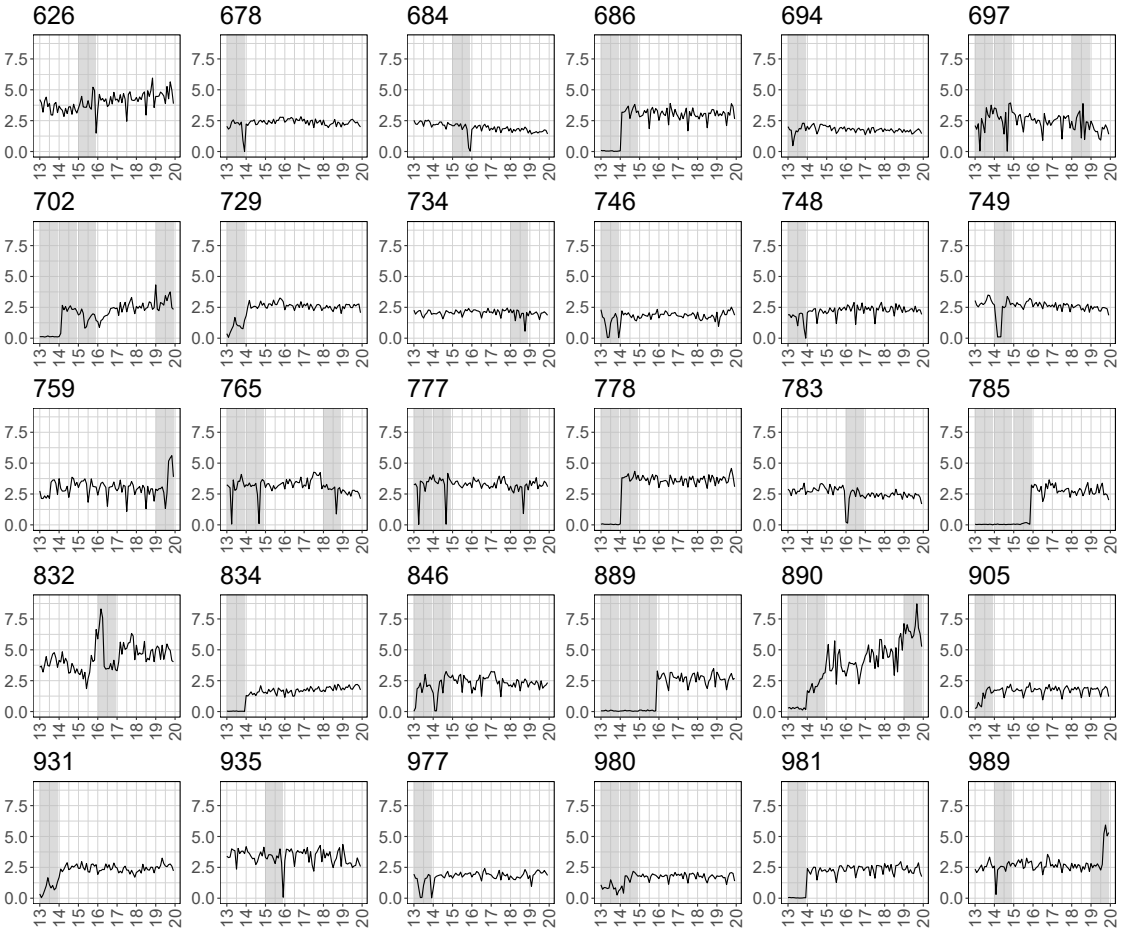

Figure A31: Adoption: Municipalities with Issues in the Primary Care Data, 3.

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

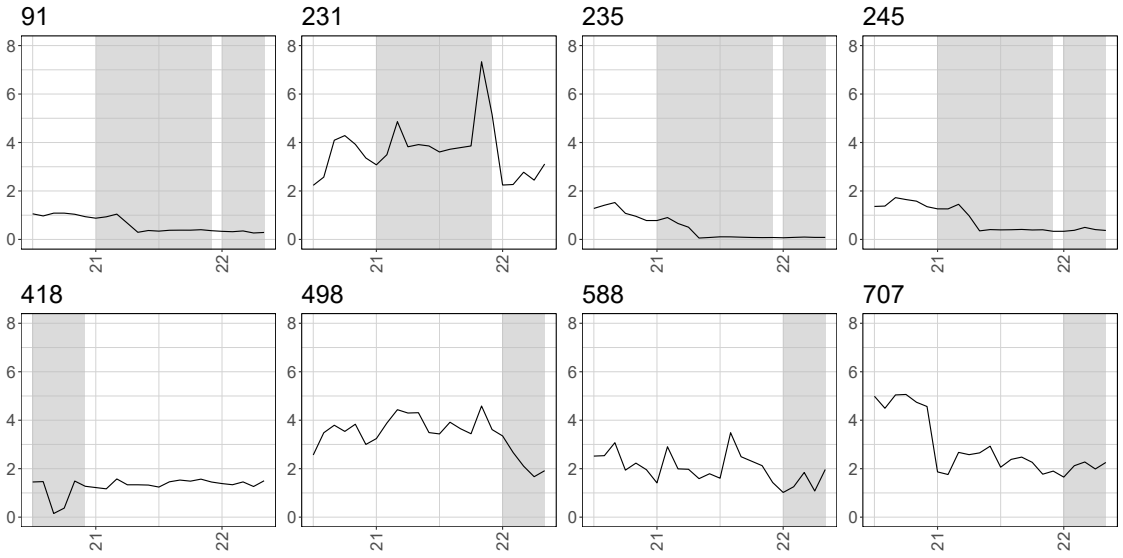

Figure A32: Abolition: Municipalities with Issues in the Primary Care Data.

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

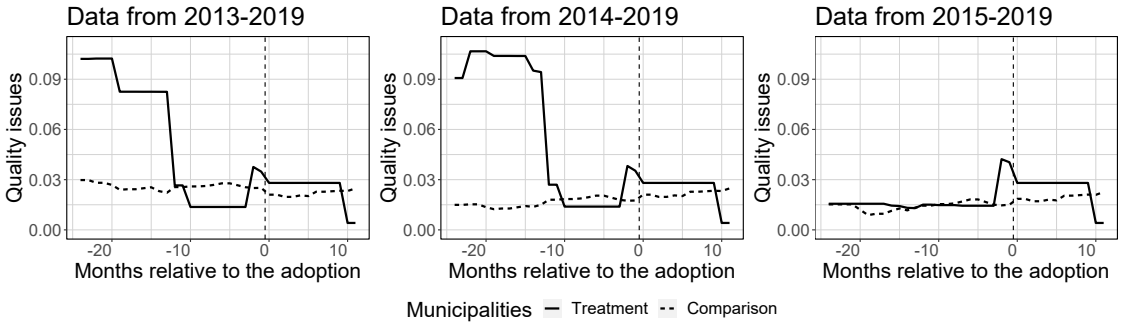

Figure A33: Relationship between Copayment Adoption and Quality Issues.

Notes: This figure was not pre-registered and is post-blind. The dataset is stacked, and no observations are excluded for quality issues. The outcome is the yearly indicator for having quality issues, defined as described in Section A.3. We weight by population due to heterogeneity in municipality size. Treatment municipalities adopted the nurse visits copayment at time 0 in relative time.

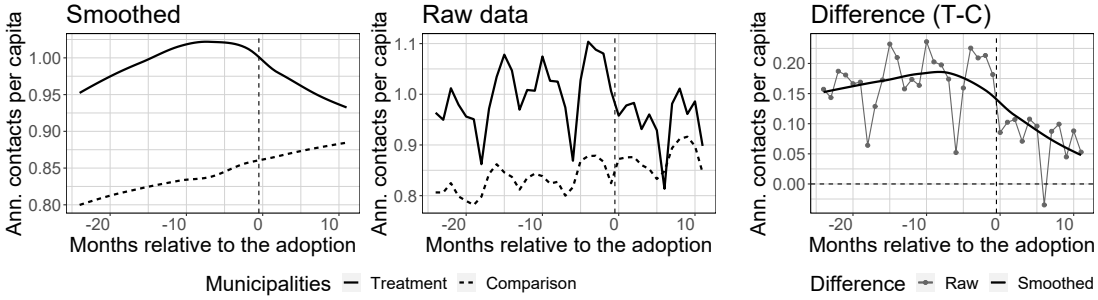

Figure A34: Adoption: Evolution in Nurse Visits, All Individuals, Not Accounting for Quality Issues.

Notes: This figure was not pre-registered and is post-blind. The dataset is stacked, and no observations are excluded for quality issues. The outcome is the number of annualized curative nurse visits per capita. We weight by population due to heterogeneity in municipality size. Treatment municipalities adopted the nurse visits copayment at time 0 in relative time. The left column contains smoothed conditional means, fitted with local linear regression. The raw data is illustrated in the middle column, while the difference between treatment and comparison areas is depicted in the right column.

References

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*, 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Haaga, T., Böckerman, P., Kortelainen, M., & Tukiainen, J. (2022). Cost sharing and primary care use: Evidence from staggered copayment adoption and later abolition. A pre-analysis plan. https://doi.org/10.17605/OSF.IO/FV2GA

Mäklin, S., & Kokko, P. (2020). Terveyden- ja sosiaalihuollon yksikkökustannukset Suomessa vuonna 2017. https://urn.fi/URN:ISBN:978-952-343-493-6

Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *The Review of Economic Studies*. https://doi.org/10.1093/restud/rdad018

Roth, J., & Sant'Anna, P. H. C. (2023). Efficient estimation for staggered rollout designs. *Journal of Political Economy Microeconomics*. https://doi.org/10.1086/726581