Face Masks Considerably Reduce Covid-19 Cases in Germany

—

A synthetic control method approach

Timo Mitze[(a)], Reinhold Kosfeld[(b)], Johannes Rode[(c)] and Klaus Wälde[(d),1]

(a) University of Southern Denmark, RWI and RCEA, (b) University of Kassel, (c) TU Darmstadt
(d) Johannes Gutenberg University Mainz, CESifo and Visiting Research Fellow IZA

July 28, 2020

Classification: Physical Sciences (statistics) and Biological Sciences (medical sciences)

Keywords: Covid-19, public health measures, face masks, synthetic control method

Autor Contributions: All authors contributed equally to the paper.

Abstract: We use the synthetic control method to analyze the effect of face masks on the spread of Covid-19 in Germany. Our identification approach exploits regional variation in the point in time when wearing of face masks became mandatory in public transport and sales shops. Depending on the region we consider, we find that face masks reduced the number of newly registered SARS-CoV-2 infections between 15% and 75% over a period of 20 days after their mandatory introduction. Assessing the credibility of the various estimates, we conclude that face masks reduce the daily growth rate of reported infections by around 47%.

Significance Statement: Mitigating the spread of Covid-19 is the objective of most governments of this world. It is of utmost importance to understand how effective various public health measures are. We study the effectiveness of face masks. We employ public regional data about reported SARS-CoV-2 infections for Germany. As face masks became mandatory at different points in time across German regions, we can employ statistical methods to compare the rise in infections in regions with masks and regions without masks. Weighing various estimates, we conclude that 20 days after becoming mandatory, face masks have reduced the number of new infections by around 45%. As economic costs are close to zero compared to other public health measures, masks seem to be a cost-effective means to combat Covid-19.

[1] Timo Mitze (ORCID 0000-0003-3799-5200), Department of Business and Economics, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark, tmitze@sam.sdu.dk. Reinhold Kosfeld (ORCID 0000-0002-8815-6039), Universität Kassel, Institute of Economics, Nora-Platiel-Str. 4, 34127 Kassel, Germany, rkosfeld@uni-kassel.de. Johannes Rode (ORCID 0000-0002-3134-0759), Department of Law and Economics, Technische Universität Darmstadt, Hochschulstraße 1, 64289 Darmstadt, Germany, rode@vwl.tu-darmstadt.de. Klaus Wälde (corresponding author, ORCID 0001-5717-0383), Gutenberg School of Management and Economics, Johannes-Gutenberg-Universität Mainz, Jakob-Welder-Weg 4, D-55131 Mainz, Germany, waelde@uni-mainz.de. This is a revised version of (1) strengthening our earlier findings.

1 Introduction

Many countries have experimented with several public health measures to mitigate the spread of Covid-19. One particular measure that has been introduced are face masks. It is of obvious interest to understand the contribution made by such a measure in reducing infections.

The effect of face masks on the spread of infections has been studied for a long time. The usefulness in the clinical context is beyond dispute. There is also evidence that they helped in mitigating the spread of earlier epidemics such as SARS 2003 or influenza. The impact of face masks worn in public on the spread of Covid-19 has to be systematically analyzed yet. This is the objective of this paper.

There is a general perception in Germany that the mandatory use of face masks in public reduces Covid-19 incidences considerably. This perception comes mainly from the city of Jena. After face masks became mandatory between April 1st and April 10, 2020, the number of new infections fell almost to zero. Jena is not the only region in Germany, however, that introduced face masks. Six further regions made masks compulsory before the introduction at the federal state level. Eventually, face masks became mandatory in all federal states between April 20 and 29, 2020 (see appendix A for background).

We quantify the effectiveness of face masks by employing the synthetic control method (SCM, 2–5). Our identification approach exploits this regional variation in the point in time when face masks became mandatory. We use data for 401 regions in Germany (municipal districts) to estimate the effect of this particular policy intervention on the development of registered infections with Covid-19. We consider the timing of mandating face covering as an exogenous event to the local population: Masks were imposed by local authorities and were not the outcome of some process in which the population was involved. We compare the Covid-19 development in various regions to their synthetic counterparts. The latter are constructed as weighted average of control regions that are structurally similar to treated regions. Structural dimensions considered include prior Covid-19 cases, the demographic composition and the local health care system.

A detailed analysis of the timing of all public health measures in the regions we study guarantees that we correctly attribute our findings to face masks (and not erroneously to other public health measures). We also employ a standard SIR model and undertake a novel analysis of the distribution of the lag between infection and reporting date. This allows us to provide a precise interpretation of our empirical effectiveness measure and to pin down the point in time when the effects of face masks should be visible in the data.

We find statistically significant and sizeable support for the general perception that the public wearing of face masks in Jena strongly reduced the number of incidences. We obtain a synthetic control group that closely follows the Covid-19 trend before the introduction of mandatory masks in Jena. The difference between Jena and this group becomes significant thereafter. Our findings indicate that the early introduction of face masks in Jena has resulted in a drop in newly registered Covid-19-cases of around 75% after 20 days. Put simply: If the control region observes 100 new infections over a period of 20 days, the mask-region only observes 25 cases. This drop is greatest, by more than 90%, for the age group 60 years and above. Our results are robust to different sensitivity checks, among which placebo-in-space and placebo-in-time analyses.

As a means to verify the generalizability of our findings for Jena, we move from a single to a multiple treatment approach and estimate average treatment effects of introducing face masks on the spread of Covid-19 for all regions that have introduced masks by April 22 (approximately 8% of all German regions). Although the estimated average treatment effect is smaller compared to the one found for Jena, it still significant and sufficiently large to support our point that wearing face masks is an effective and cost-efficient measure for fighting Covid-19. When we summarize all of our findings in one single measure (see appendix B.5), we conclude that the daily growth rate of Covid-19 cases in the treatment group falls by around 47% due to mandatory mask-wearing relative to the synthetic control group.[2]

Our findings can be aligned with earlier evidence on face masks, public health measures and the epidemic spread of Covid-19 although consolidated scientific knowledge is limited (see appendix D). While there is a growing consensus from clinical studies that face masks significantly reduce the transmission risk of SARS-CoV-2 and Covid-19 (7, 8), non-clinical evidence on the effectiveness of face masks is still largely missing. (9) survey evidence on the population impacts of a widespread community mask use and stress that "*no randomized control trials on the use of masks <…> has been published*". The study which is "*the most relevant paper*" for (9) is one that analyzed *"exhaled breath and coughs of children and adults with acute respiratory illness*" (10, p. 676), i.e. used a clinical setting. Concerning the effect of masks on community transmissions, the survey needs to rely on pre-Covid-19 studies.

(11) are among the first to estimate the population impact of face masks on SARS-CoV-2 infections.[3] The authors track the development of Covid-19 in three pandemic epicenters Wuhan, Italy, and New York City between January 23 and May 9, 2020 and find sizable mitigating effects of face masks on epidemic spread. While their study offers important novel insights of the population effects of face masks, a methodical limitation is that estimates are only carried out in a "before-after" manner with no use of a strict control group approach. This may limit the causal interpretation of results. We therefore follow the spirit of (5) and provide the first causal evidence identifying the population impact of mandatory face masks on the spread of Covid-19.

2 Results: The effects of face masks on the spread of Covid-19

2.1 Results for Jena

Face masks became mandatory in Jena in three steps between April 1st and 10. The most important measure (in the sense of having the largest impact measured in terms of social contacts) requires face masks in public transports and shops and entered into force on April 6 (again, see appendix A for detailed information). We therefore center our discussion on this date.

Panel A in Figure 1 shows the SCM results for the introduction of face masks in Jena on April 6. The visual inspection of the development of cumulative Covid-19 cases shows that the trend

[2] The main channel through which masks reduce transmission of SARS-CoV-2 is the limiting effect for the spread of exhaled air, as argued by (6). (6) and (7) argue that aerosols (as opposed to larger droplets) are filtered only by high-quality masks. Droplets are also filtered by home-made masks.

[3] (8) conduct a systematic review and meta-analysis. They do not report a study (see their table 1) that analyzes the entire population of a country.

development of the synthetic control group is very similar to Jena before the treatment indicating a good fit.[4] The difference in the cumulated registered Covid-19 cases between Jena and its corresponding synthetic control group after the start of the treatment on April 6 can be interpreted as the treatment effect on the treated (see appendix B.4 for (post-)estimation details).

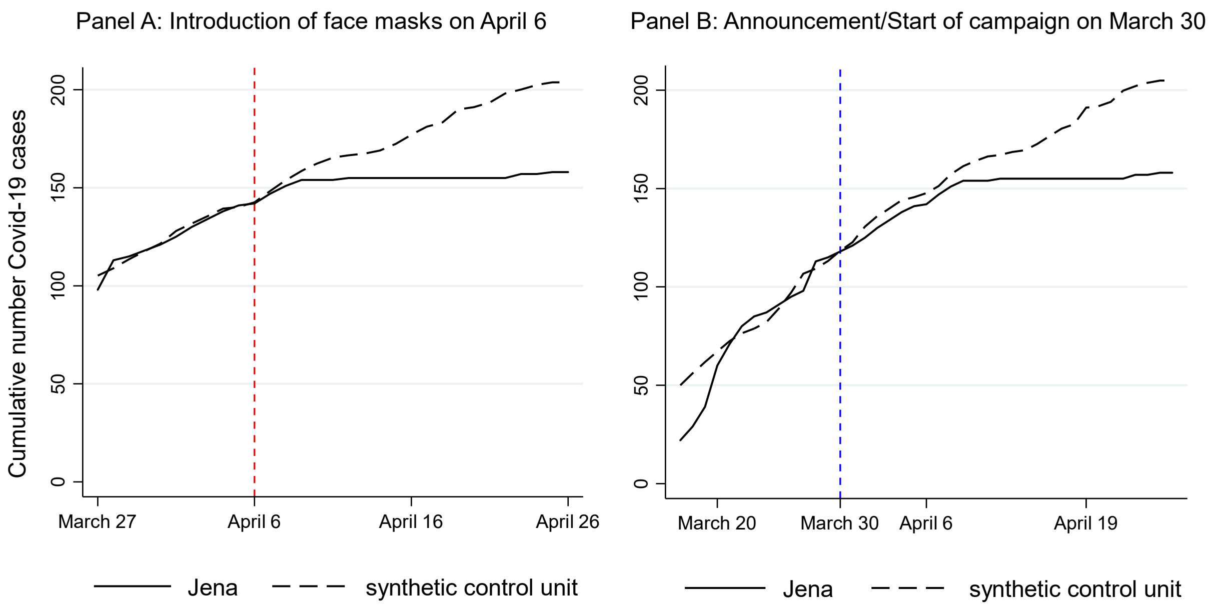

Figure 1: Treatment effects of mandatory face masks in Jena on April 6 and its announcement on March 30

Panel A in the figure clearly shows a gradually widening gap in the cumulative number of Covid-19 cases between Jena and its synthetic control group. The size of the effect 20 days after the start of the treatment (April 6) amounts to a decrease in the number of cumulative Covid-19 cases of 23%, which corresponds to a drop in newly registered cases of roughly 75%. Expressed differently, the daily growth rate of the number of infections decreases by 1.28 percentage points per day (see appendix B.5 for computational details and an overview of all measures). If we look at the estimated differences by age groups, Table 5 in the appendix indicates that the largest effects are due to the age group of persons aged 60 years and above. Here the reduction in cumulative cases even exceed 50%, which corresponds to a drop in newly registered SARS-CoV-2 infections by more than 90%. The significant drop can be explained with strict introduction of face masks in elderly and nursing homes, which already started on April 2. For the other two age groups the decrease in the number of cumulative cases lies between 10 and 20%.

If we consider a median time lag of 10.5 days from infection until registration (see appendix A.3), the occurrence of a gradually widening gap between Jena and its synthetic control in the first week after the introduction of mandatory face masks seems fast. One might conjecture that an announcement effect has played a role. As shown in appendix B.8.1, online searches for (purchasing) face masks peaked on April 22nd, when it was announced that face masks

[4] As a measure for the quality of the fit between the treated region and its synthetic control group, the pre-treatment root mean square prediction error (RMSPE) can be calculated and compared to a reference case. For Jena the pre-treatment RMSPE is 3.145, which is considerably lower than a RMSPE of 6.669, which has been calculated as the average RMSPE for all other 400 regions and their synthetic controls in the pre-treatment period until April 6. This points to the relatively good fit of the synthetic control group for Jena in this period.

would become compulsory in all German federal states.[5] Another peak of online searches, almost as large (70% of the peak of April 22nd), appeared on March 31st. This marks the date of the regulation making masks compulsory between April 1st and April 10 in Jena. The regulation was accompanied by a campaign "Jena zeigt Maske" communicating the necessity to wear face masks in public that started on March 30. [6]

Panel B in Figure 1 plots the estimated effect size when we set the start of the treatment period to the start of the campaign on March 30. The visual inspection of the difference between Jena and its synthetic control group points to the presence of a small anticipation effect. Yet, the gap to the synthetic control significantly widens only approximately 10-12 days after the announcement and then strongly grows over time. As this temporal transmission channel appears plausible given a median time lag between infection and registration of almost equal length, we take this as first evidence for a face mask-effect in the reduction of SARS-CoV-2 infections. Appendix B.6 reports SCM results by age groups.

2.2 Robustness checks

Obviously, the estimated differences in the development of Jena vis-à-vis the synthetic Jena is only consistently estimated if our SCM approach delivers robust results. Accordingly, we have carried out several tests to check for the sensitivity of our findings.

Cross-validation tests. One important factor is that our results are not sensitive to changes in the choice of predictor variables. We therefore perform cross-validation checks by modifying the length of the training and validation period before the start of the treatment. Importantly, we do not find a systematic downward bias of our baseline specification compared to alternative specifications with longer lag structures and accordingly shorter trainings periods (see appendix B.10). Given that regional Covid-19 cases developed very dynamically and non-linearly in this period, this is an important finding in terms of the robustness of our results.

Changing the donor pool. This may be equally important as our baseline specification includes the region of Heinsberg in the donor pool used to construct the synthetic Jena (with a weight of 4.6%; compare Table 11). As Heinsberg is one of the German regions that was significantly affected by the Covid-19 pandemic during the Carnival season, one may expect that this leads to an overestimation of the effects of face masks. Accordingly, appendix B.9 presents estimates for alternative donor pools. Again, we do not find evidence for a significant bias in our baseline specification. By tendency, the treatment effect becomes larger, particularly if we compare Jena only to other regions in Thuringia (to rule out macro-regional trends) and to a subsample of larger cities (*kreisfreie Städte*), which reduces the degree of latent regional heterogeneity, for instance, with regard to social interactions. Both subsamples exclude Heinsberg. We also run SCM for subsamples excluding Thuringia (to rule out spatial spillover effects) and for East and West German regions only (again to test for specific macro regional trends). Generally, these sensitivity tests underline the robustness of the estimated treatment effect for Jena.

Placebo-in-space tests. These tests check whether other cities that did not introduce face masks on April 6 have nonetheless experienced a similar decline in the number of registered Covid-19

---

[5] For a German-wide news report see, e.g., https://www.tagesschau.de/inland/corona-maskenpflicht-103.html. Last accessed 14 July 2020.

[6] See local newspaper reports, for instance, at: https://www.jenaer-nachrichten.de/stadtleben/13069-jena-zeigt-maske-kampagne-f%C3%BCr-mundschutz-startet. Last accessed 18 July 2020.

cases. If this had been the case, the treatment effect may be driven by other latent factors rather than face masks. Such latent factors may, for instance, be related to the macro-regional dynamics of Covid-19 in Germany. Therefore, appendix B.11 reports pseudo-treatment effects for similarly sized cities in the federal state of Thuringia assuming that they have introduced face masks on April 6 although –in fact– they did not. As the figure shows, these cities show either a significantly higher or similar development of registered Covid-19 compared to their synthetic controls. This result provides further empirical support for a relevant effect in the case of Jena.

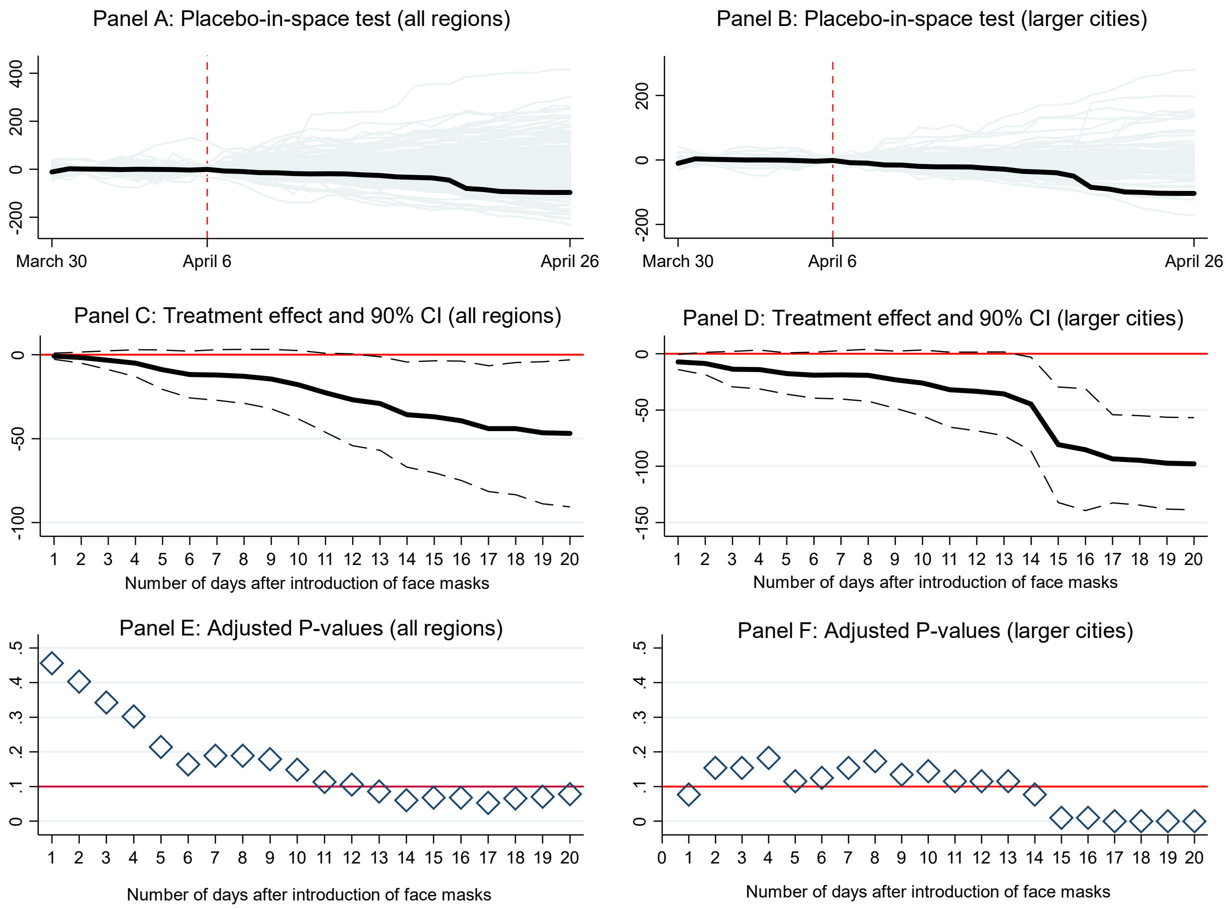

Figure 2: Comprehensive placebo-in-space tests for the effect of face masks on Covid-19 cases

Notes: Treatment effects in Panel A to Panel D are measured in terms of registered Covid-19 cases; subgraphs in Panel A and B exclude the following regions with a large absolute number of registered Covid-19 cases: Hamburg (2000), Berlin (11000), Munich (9162), Cologne (5315) and Heinsberg (5370). In line with (3, we only include placebo effects in the pool for inference if the match quality (pre-treatment RMSPE) of the specific control regions is smaller than 20 times the match quality of the treated unit; *p*-values reported in Panel E and F are accordingly adjusted for the pre-treatment match quality (see 12).

As a more comprehensive test, we run placebo-in-space tests for all other regions that did not introduce face masks on April 6 or closely afterwards. Again, we estimate the same model on each untreated region, assuming it was treated at the same time as Jena. The empirical results in Figure 2 indicate that the reduction in the reported number of Covid-19 cases in Jena clearly exceeds the trend in most other regions – both for the overall sample in Panel A and the sub-sample of large cities (*kreisfreie Städte*) in Panel B.

As outlined above, one advantage of these tests is that they allow us to conduct inference on the significance of the estimated treatment effects for Jena. Accordingly, Panel C and Panel D visualize the estimated treatment effects together with 90% confidence intervals, which have been calculated on the basis of (one-sided) pre-treatment match quality adjusted *p*-values as reported in Panel E and Panel F of Figure 2.[7] The latter indicate the probability that the reduction in the number of Covid-19 cases was observed by chance given the distribution of pseudo-treatment effects in the other German regions (see 12). For both sample distributions, the reported confidence intervals and underlying *p*-values indicate that the reduction in the number of Covid-19 cases was not a random event in Jena can be attributed to the introduction of face masks two weeks after the start of the treatment. Again, this timing is in line with our above argument that a sufficiently long incubation time and testing lags need to be considered in the evaluation of treatment effects.[8]

Placebo-in-time tests. As for the case of placebo-in-space tests it is important for the validity of results that we do not observe significant treatments effect for Jena prior to the introduction of face masks on April 6 or its announcement on March 30. To rule out such anticipation effects, we have systematically reviewed all general decrees published by the local administration in Jena. Of particular interest are those decrees that significantly differ with respect to their timing from those at the federal state level in Thuringia.

Looking at the figure in appendix A.2, Jena and Thuringia passed at least 40 public health measures before end of April 2020. Jena implemented 27 of those 40 either earlier than Thuringia or on its own. Examples of earlier implementation include the closing of bars, cafés and restaurants or quarantine rules for travelers returning back home. Relevant regions included foreign countries but also other German federal states among which Bavaria, Baden-Wurttemberg and North-Rhine Westphalia. Measures imposed by Jena only include the complete closing of hotels (in contrast to closing of hotels for tourism only in Thuringia) and a curfew (which lasted for only two weeks, though).

As these major health decrees were accompanied by smaller ones on an almost daily basis until March 20, we run a series of SCM estimations using each day between March 14 and 20 as (pseudo-)treatment period.[9] The results for the full donor pool including all other German regions and the subsample of larger cities are shown in Panel A and Panel B of Figure 3. Results are reported until March 30 when the mandatory introduction of face masks was announced. The visual inspection of the relative development of Covid-19 cases in Jena vis-à-vis its synthetic

[7] We follow the method proposed in (13) to calculate confidence intervals from *p*-values. As pointed out in (14), the interpretation of confidence intervals and *p*-values is restricted to the question of whether or not the estimated effect of the actual treatment is large relative to the distribution of placebo effects.

[8] We analyze a measure that is introduced for the first time in this region. One might conjecture that our estimation measures both the true effect of a face mask but also any other change in behavior (washing hands, limiting interactions, staying more at home etc.) that was triggered by this policy. This change in behavior is known as the Hawthorn effect. Individuals in this pioneer region might take the crisis more seriously than in the other areas. Although German health authorities had been strongly recommending such behavioral changes in daily life since mid-March, we cannot fully rule out this mixing of effects. Mobility data for federal states in appendix B.8.2 shows that federal states moved in a relatively coordinated way in this respect. Unfortunately, mobility data for Jena is not easily available.

[9] Alternatively, we have also tested for pseudo-treatment effects in Jena over a period of 20 days before the introduction of face masks. This period is equally split into a pre- and pseudo post-treatment period. As Panel B in Figure 17 shows, there is no strong deviation from the path of the synthetic control group.

Jena does not indicate a clear treatment effect in terms of reducing Covid-19 cases prior to April 1st. The results are particularly clear-cut for the sample of larger cities in Panel B indicating that earlier public health measures alone have not significantly suppressed the number of Covid-19 cases in Jena in the first two weeks after their introduction.

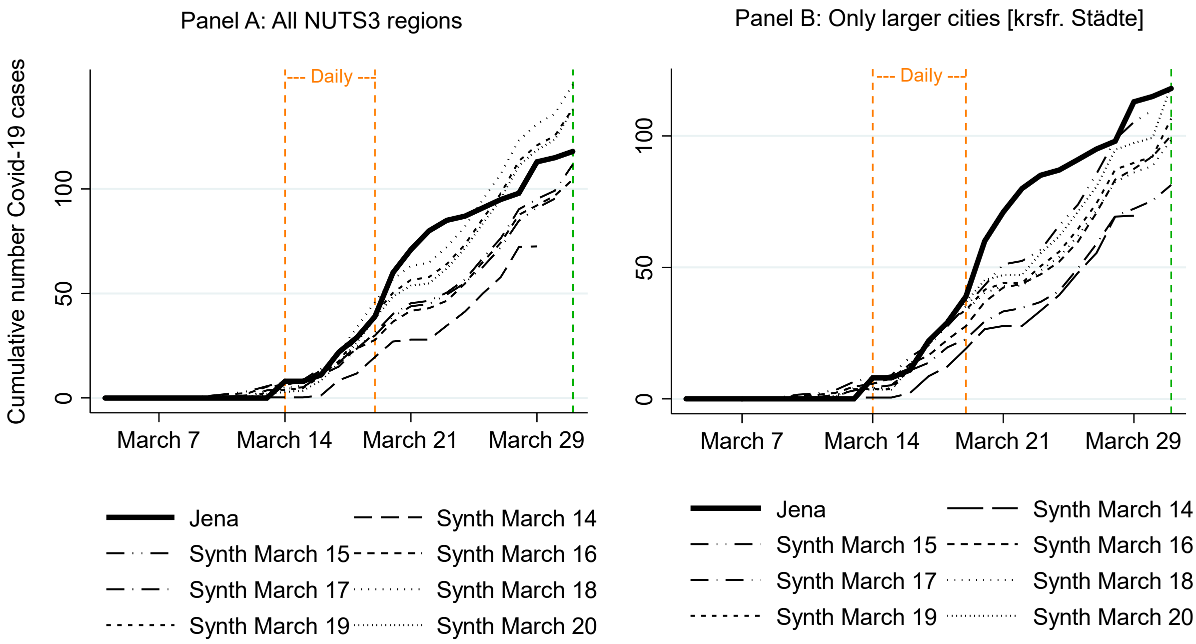

Figure 3: Placebo-in-time tests for period March 14 to March 20

Notes: All SCM estimates are based on the same predictor settings as for the baseline specification.

The reported trajectories of synthetic Jena in Panel A of Figure 3 leave us with some degree of ambiguity, though. To explicitly test for a potential trend reversal in the development of Covid-19 cases prior to the introduction of face masks, we further run an alternative robustness test on the basis of incremental difference-in-difference (DiD) estimation. The DiD estimator is particularly well-suited to estimate dynamic treatment effects in the context of limited information about the exact length of transmission lags before individual measures show measurable effects (a detail description of is given in appendix E). As the results clearly show, treatment effects from public health measures in Jena in terms of a reduction in Covid-19 cases only become statistically significant roughly two weeks after the introduction of face masks on April 6. If we resort to the estimated incubation and reporting lag as shown in appendix A.3, this result supports our main SCM findings that the relative reduction in the cumulative number of Covid-19 cases is mainly attributable to the timing of introducing face masks. The incremental DiD results also support our main SCM findings in terms of the magnitude of the treatment effect.

2.3 Results for other regions

Jena may be a unique case. We therefore also study treatment effects for other regions that have antedated the general introduction of face masks in Germany. Further single unit treatment analyses are shown in appendix C. SCM estimation for multiple treated units is studied in two ways. The first sample covers the full set of municipal districts and accordingly includes a total of 32 treated units. The second focusses on the subsample of larger cities (*kreisfreie Städte*) of which 8 are treated units. Treated regions introduced face masks by April 22 at the latest. The donor pool of control regions is specified in such a way that the minimum time lag in the introduction of face masks between treated and control regions ranges between 5 and 13 days.

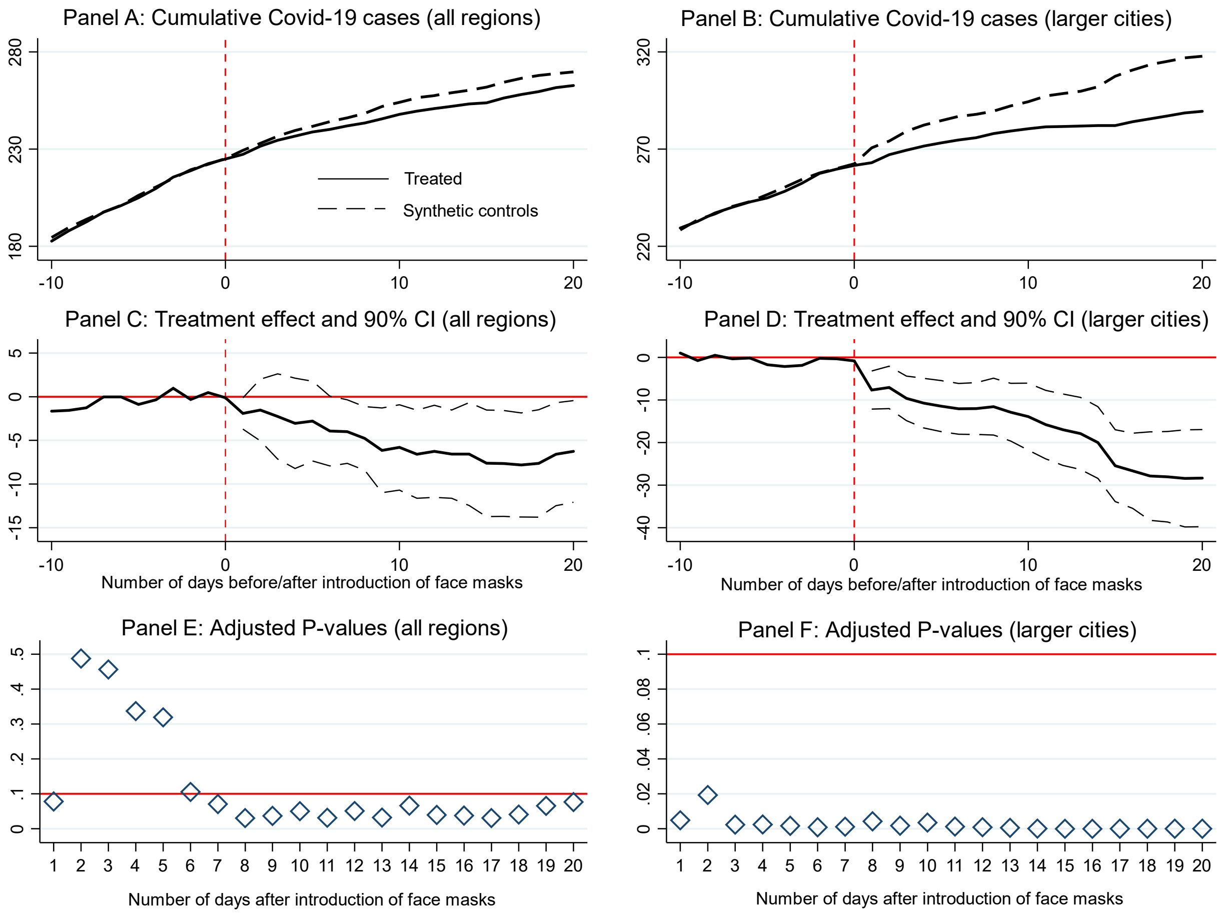

Figure 4: Average treatment effects for introduction of face masks (multiple treated units)

Notes: Panel A and Panel B measure the average number of cumulative Covid-19 cases in the respective groups; Panel C and Panel D visualize treatment effects in terms of the estimated reduction in the cumulative number of Covid-19 cases together with 90% confidence intervals. The reported *p*-values in Panel E and Panel F are adjusted for the pre-treatment match quality (see Figure 2 for details); inference has been conducted on the basis of a random sample of 1,000,000 placebo averages.

The results, visible in Figure 4, point to a significant face mask-effect in the reduction of SARS-CoV-2 infections over a period of 20 days after the introduction. The temporal evolution of the average number of cumulative Covid-19 cases for treated regions and their corresponding synthetic control groups are shown in Panel A and Panel B of Figure 4, respectively. The reported 90% confidence intervals in Panel C and Panel D calculated on the basis of adjusted *p*-values shown in Panel E and Panel F indicate that the estimated treatment effects are not random for both samples. While treatment effects of face masks turn significant after roughly one week for the overall sample, the emergence of a reduction in the subsample of larger cities is fast and points to early anticipation effects of face masks in urban areas, particularly during the period when local economies were gradually reopened after April 20.

Importantly, however, the trend development for larger cities as shown in Figure 4 not only indicates a drop in the number of newly registered Covid-19 cases around the immediate timing of the introduction of face masks, but also points to the presence of dynamic treatment effects as the average gap between treated regions and their synthetic control groups widens over time. This hints at the role played by mandatory face masks in avoiding a new wave of new infections once the economy and labor market is re-opened. As Panel B in Figure 4 highlights, such an avoidance effect may be particularly important in larger cities with higher population density and accordingly higher intensity of social interaction.[10]

Taken together, over a period of 20 days, we observe an average reduction of 28.4 cases between treated and control regions in the context of urban areas. Relative to the average number of cumulative Covid-19 cases on May 11 in control regions (317.9), this amounts to a reduction of 8.9% in the cumulative number of Covid-19 cases and a reduction of 51.2% in newly registered case. The difference in the daily growth rate of the number of infections correspondingly amounts to 0.46 percentage points. For the full sample, this difference is estimated to be 0.13 percentage points (see again appendix B.5 for an overview of all measures and section 4 for theoretical background). This smaller magnitude in the latter sample including all municipal districts has to be evaluated against the background of a considerable degree of structural heterogeneity, for instance, related to the composition of the local population but also the local Covid-19 spread. We argue that the latter should thus be interpreted as a lower bound for the true treatment effects.

3 Discussion

We set out by analyzing the effect of face masks on the spread of Covid-19 for a comparative case study of the city of Jena. Our quasi-experimental control group approach using SCM shows that the introduction of face masks on April 6 reduced the number of newly registered Covid-19 cases over the next 20 days by 75% relative to the synthetic control group. Comparing the daily growth rate in the synthetic control group with the observed daily growth rate in Jena, the latter shrinks by around 70% due to the introduction of face masks. This is a sizeable effect. The introduction of mandatory face masks and the associated signal to the local population to take the risk of person-to-person transmissions seriously apparently helped considerably in reducing the spread of Covid-19. Looking at average treatment effects for all other regions puts this result in some perspective. The reduction in the daily growth rate of infections amounts to 14% only. By contrast, when we focus on larger cities, we find a reduction in the daily growth rate of infections by roughly 47%.

What would we reply if we were asked what the effect of introducing face masks would have been if they had been made mandatory all over Germany? The answer depends, first, on which of the percentage measures we found above is the most convincing and, second, on the point in time when face masks are made compulsory. The second aspect is definitely not only of academic interest but would play a major role in the case of a second wave.

We believe that the reduction in the daily growth rates of infections between 47% and 70% is our best estimate of the effects of face masks. Arguments in favor of the high 70% stress that Jena introduced face masks before any other region did so. It announced face masks as the first region in Germany while in our post-treatment period hardly any other public health measures were introduced or eased. Hence, it provides the most clear-cut quasi-experimental setting for studying its effects. Second, as described below in section 4, Jena is a fairly representative region of Germany in terms of Covid-19 cases. Third, the smaller treatment effects observed in

[10] This is perfectly in line with (7) given the reduction in aerosols and droplets via using masks.

the multiple treatment analysis may also result from the fact that –by the time that other regions followed the example of Jena– behavioral adjustments in Germany's population had already taken place. Wearing face masks gradually became more common and more and more people started to adopt their usage even when it was not yet required. The results for the subsample of larger cities are, however, quantitatively similar to Jena.

Arguments for the lower 47% state that the stronger impact of face masks on the infectious in Jena may thereby partly be driven by a Hawthorn effect. The population in this pioneer region might have reacted very strongly to the mandatory introduction of face masks by taking the other imposed public health measures and hygiene rules (washing hands, limiting interactions, staying more at home etc.) more seriously.

Concerning the point in time (or better, the point of the epidemic cycle) when face masks become mandatory, all of our estimates might actually be modest. The daily growth rates in the number of infections when face masks were introduced in Jena was around 2-3%. These are low growth rates compared to the early days of the epidemic in Germany, where daily growth rates lay above 50% (15). One might therefore conjecture that the effects might have been even greater if masks had been introduced earlier.

This timing effect might also explain the difference between Jena estimates and lower estimates for other regions. By the time Jena introduced face masks on April 6, the general trend development of Covid-19 cases was still relatively dynamic across German regions. In mid-April, when other regions followed the example of Jena and introduced face masks before the general introduction at the federal state level, overall daily growth rates were already lower.

We simultaneously stress the need for further complementary analyses. First, Germany is only one specific country. Different regulations, norms or climatic conditions might change the empirical picture for other countries. Second, we ignored the impact of the number of tests on reported infections. While we do not believe that this matters for Germany as rules for testing are homogenous across regions, this might play a bigger role for international comparisons. Third, we have ignored spatial dependencies in the epidemic diffusion of Covid-19. This might also matter. Finally, there are various types of face masks. We cannot identify differential effects since mask regulations in German regions do not require a certain type. In any case, given the low economic costs of face masks compared to other public health measures, a cost-benefit view clearly speaks in favor of face masks.

4 Method and data

Method. Six regions in Germany (municipal districts, equivalent to the EU nomenclature of territorial units for statistics, NUTS, level 3 categorization) made face masks mandatory before their respective federal states. They are displayed in Figure 5. The figure also shows differences across federal states in the timing of introducing mandatory face masks.

To identify treatment effects from introducing face masks, we apply the synthetic control method (SCM) for single and multiple treated units (see appendix B.1 for more background).

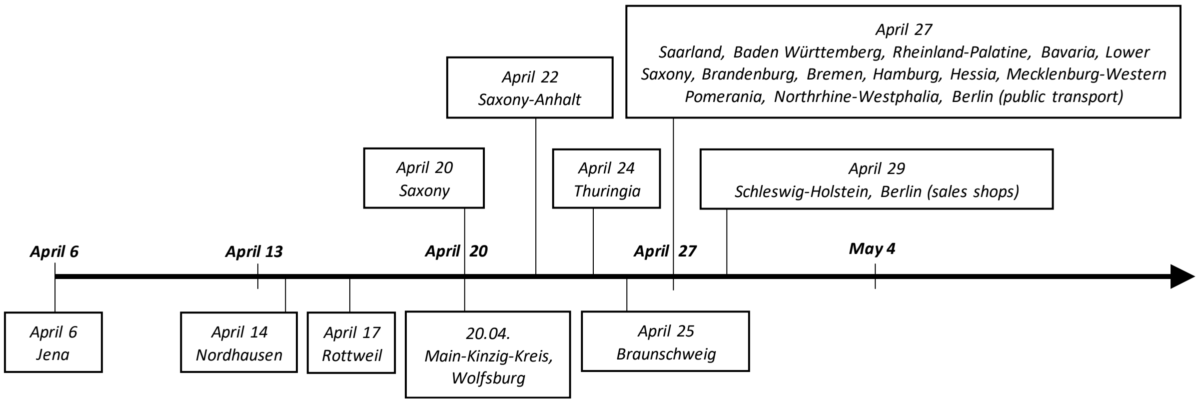

Figure 5: The timing of mandatory mask wearing in federal states (top) and individual regions (below)

Data. We use the official German statistics on reported Covid-19 cases from the Robert Koch Institute (16). We build a balanced panel for 401 NUTS level 3 regions and 105 days spanning the period from January 28 to May 11, 2020 (42,105 observations). We use the cumulative number of registered Covid-19 cases in each district and the number of cumulative Covid-19 cases per 100,000 inhabitants as main outcome variable. We estimate overall effects for this variable together with disaggregated effects by age groups (persons aged 15-34 years, 35-59 years and 60+ years). We also employ regional data to inter alia identify control regions. The table in appendix B.2 shows summary statistics.

Face masks are clearly not the only public health measures to mitigate the spread of Covid 19. Identification of the face mask effect therefore need to take the timing of other public health measures into account. To this end, we built a database for all public health measures in Jena and Thuringia and for face masks in all other federal states. See appendix A.1 and A.2 for details.

Conceptional background. To facilitate the interpretation of our findings, we employ a standard SIR model with three states: Susceptible, infectious and removed. (See appendix A.4 for more details.) Imagine we study a region where face masks are not mandatory. The time path $I(t)$ of infections individuals in this (synthetic) control group is displayed in Figure 6 below as $I_{control}(t)$. The time path for I^{ever}(t) in the control group is denoted by $I_{control}^{ever}$(t). Now consider the introduction of mandatory face masks at T (set to 29.5 in our figure below). [11] Mandatory masks reduce the infection rate (via a parameter r in the SIR model). Given a (median) delay of D^m between infection and reporting to authorities (estimated at 10.5 days in appendix A.3), we model this delay by effectively reducing r at $T+D^m$. Hence, as of $T+D^m$, the number of infectious individuals falls faster, see "face masks $I_{mask}(t)$", and the number of individuals ever infected rises less quickly, as visible when looking at I_{mask}^{ever}(t). Note the qualitative similarity between the yellow and pink curve here and the corresponding curves in panel A of Figure 1 and panels A and B of Figure 4 in the main text.

[11] We chose T=29.5 as this yields a date when masks show an effect in the data on $T+D^m$=40 where the epidemic is already beyond its peak in our simple model. This is consistent with Jena where the incidence has already been declining when face masks became mandatory.

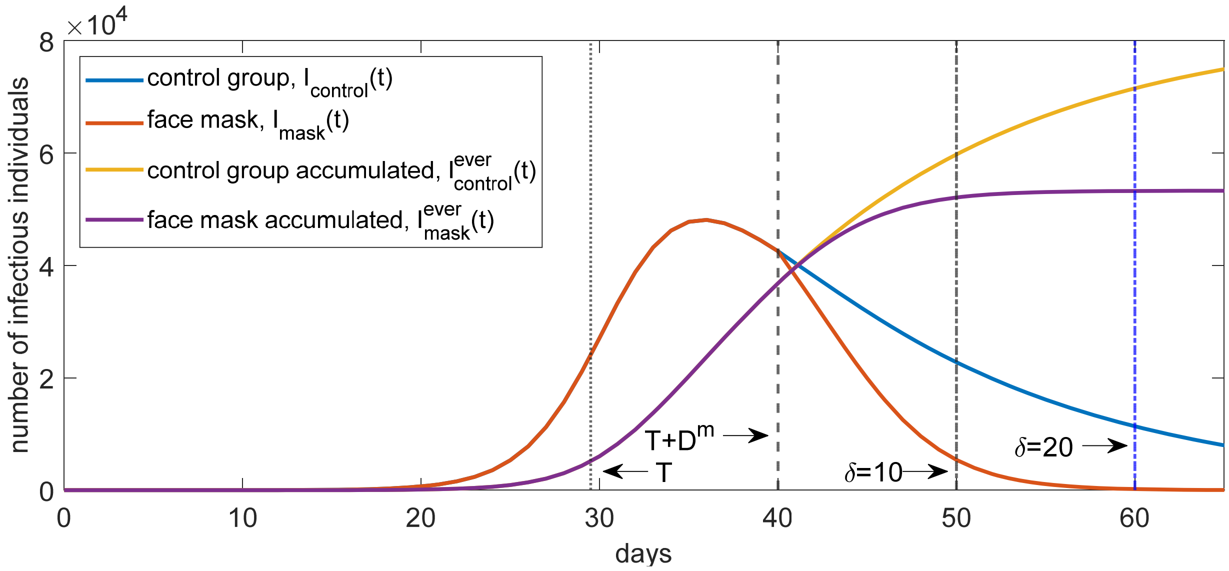

Figure 6: Theoretical effects of face masks on the number of infectious individuals *I(t)* and on the accumulated number of infectious individuals $I^{ever}(t)$

Now imagine we want to quantify the effect of face masks. The model suggests that the effect of face masks can be described by the reduction in the total number of individuals ever infected. As an example, consider time $T+D^m+\delta$, i.e. δ days after face masks became effective. The difference between the control region and the face-mask region is given by $I^{ever}_{control}(T+D^m+\delta)$-$I^{ever}_{mask}(T+D^m+\delta)$. Hence, the introduction of face-masks reduced the number of Covid-19 cases by

$$reduction\ over\ \delta\ days\ = \frac{I^{ever}_{control}(T+D^m+\delta)- I^{ever}_{mask}(T+D^m+\delta)}{I^{ever}_{control}(T+D^m+\delta)- I^{ever}_{control}(T+D^m)} * 100\%. \qquad (1)$$

This equation produces the numbers we report to quantify the effects of face masks. Appendix B.5 describes our measures based on daily growth rates.

5 Acknowledgements

The authors are grateful for an almost uncountable number of worldwide comments on the earlier version of this paper (1), both from colleagues from many disciplines, from public administration and from the general public. They considerably helped in improving this analysis. We would especially like to thank Enikö Bán, Soeren Enkelmann, Jan Franke, Manfred Hempfling, Christof Kuhbandner, Falk Laser and Philip Savage. Carolin Kleyer and David Osten provided excellent research assistance.

References

1. T. Mitze, R. Kosfeld, J. Rode, K. Wälde, Face masks considerably reduce Covid-19 cases in Germany - A synthetic control method approach. *IZA Discussion Paper No. 13319* (2020).
2. A. Abadie, J. Gardeazabal, The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review* **93**, 113–132 (2003).
3. A. Abadie, A. Diamond, J. Hainmueller, Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association* **105**, 493–505 (2010).

4. A. Abadie, Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *[Article prepared for the] Journal of Economic Literature (*2019).

5. C. A. W. Bruhn, S. Hetterich, C. Schuck-Paim, E. Kürüm, R. J. Taylor, R. Lustig, E. D. Shapiro, J. L. Warren, L. Simonsen, D. M. Weinberger, Estimating the population-level impact of vaccines using synthetic controls. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 1524–1529 (2017).

6. C. J. Kähler, R. Hain, Fundamental protective mechanisms of face masks against droplet infections. *Journal of Aerosol Science* **148**, 105617 (2020).

7. K. A. Prather, C. C. Wang, R. T. Schooley, Reducing transmission of SARS-CoV-2. *Science*. 10.1126/science.abc6197 (2020).

8. D. K. Chu, E. A. Akl, S. Duda, K. Solo, S. Yaacoub, H. J. Schünemann, A. El-harakeh, A. Bognanni, T. Lotfi, M. Loeb, A. Hajizadeh, A. Bak, A. Izcovich, C. A. Cuello-Garcia, C. Chen, D. J. Harris, E. Borowiack, F. Chamseddine, F. Schünemann, G. P. Morgano, G. E. U. Muti Schünemann, G. Chen, H. Zhao, I. Neumann, J. Chan, J. Khabsa, L. Hneiny, L. Harrison, M. Smith, N. Rizk, P. Giorgi Rossi, P. AbiHanna, R. El-khoury, R. Stalteri, T. Baldeh, T. Piggott, Y. Zhang, Z. Saad, A. Khamis, M. Reinap, Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet* **395**, 1973–1987 (2020).

9. J. Howard, A. Huang, Z. Li, Z. Tufekci, V. Zdimal, H.-M. van der Westhuizen, A. von Delft, A. Price, L. Fridman, L.-H. Tang, V. Tang, G. L. Watson, C. E. Bax, R. Shaikh, F. Questier, D. Hernandez, L. F. Chu, C. M. Ramirez, A. W. Rimoin, Face Masks Against COVID-19: An Evidence Review. 10.20944/preprints202004.0203.v1 (2020).

10. N. H. L. Leung, D. K. W. Chu, E. Y. C. Shiu, K.-H. Chan, J. J. McDevitt, B. J. P. Hau, H.-L. Yen, Y. Li, D. K. M. Ip, J. S. M. Peiris, W.-H. Seto, G. M. Leung, D. K. Milton, B. J. Cowling, Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nature medicine* **26**, 676–680 (2020).

11. R. Zhang, Y. Li, A. L. Zhang, Y. Wang, M. J. Molina, Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 14857–14863 (2020).

12. S. Galiani, B. Quistorff, The Synth_Runner Package: Utilities to Automate Synthetic Control Estimation Using Synth. *The Stata Journal* **17**, 834–849 (2017).

13. D. G. Altman, J. M. Bland, How to obtain the confidence interval from a P value. *BMJ (Clinical research ed.)* **343**, d2090 (2011).

14. A. Abadie, A. Diamond, J. Hainmueller, Comparative Politics and the Synthetic Control Method. *American Journal of Political Science* **59**, 495–510 (2015).

15. K. Wälde, Corona-Blog, Einschätzung vom Freitag, 20. März 2020: https://www.macro.economics.uni-mainz.de/corona-blog/.

16. Robert Koch Institute, Covid-19 Infektionen, General Website (NPGEO Corona Hub): https://npgeo-corona-npgeo-de.hub.arcgis.com/.

Online-Appendix

for

Face Masks Considerably Reduce Covid-19 Cases in Germany

—

A synthetic control method approach

Timo Mitze[(a)], Reinhold Kosfeld[(b)], Johannes Rode[(c)] and Klaus Wälde[(d)]

(a) University of Southern Denmark, RWI and RCEA, (b) University of Kassel, (c) TU Darmstadt
(d) Johannes Gutenberg University Mainz, CESifo and Visiting Research Fellow IZA

A Timing of public health measures and visibility in data

A.1 Timing of the introduction of mandatory face masks

Face mask were introduced in two ways in federal states. One measure relates to public transports and shops, the other to services for which a distance of 1.5 meters cannot be guaranteed. The points in time differ, however. An overview is in the next figure.

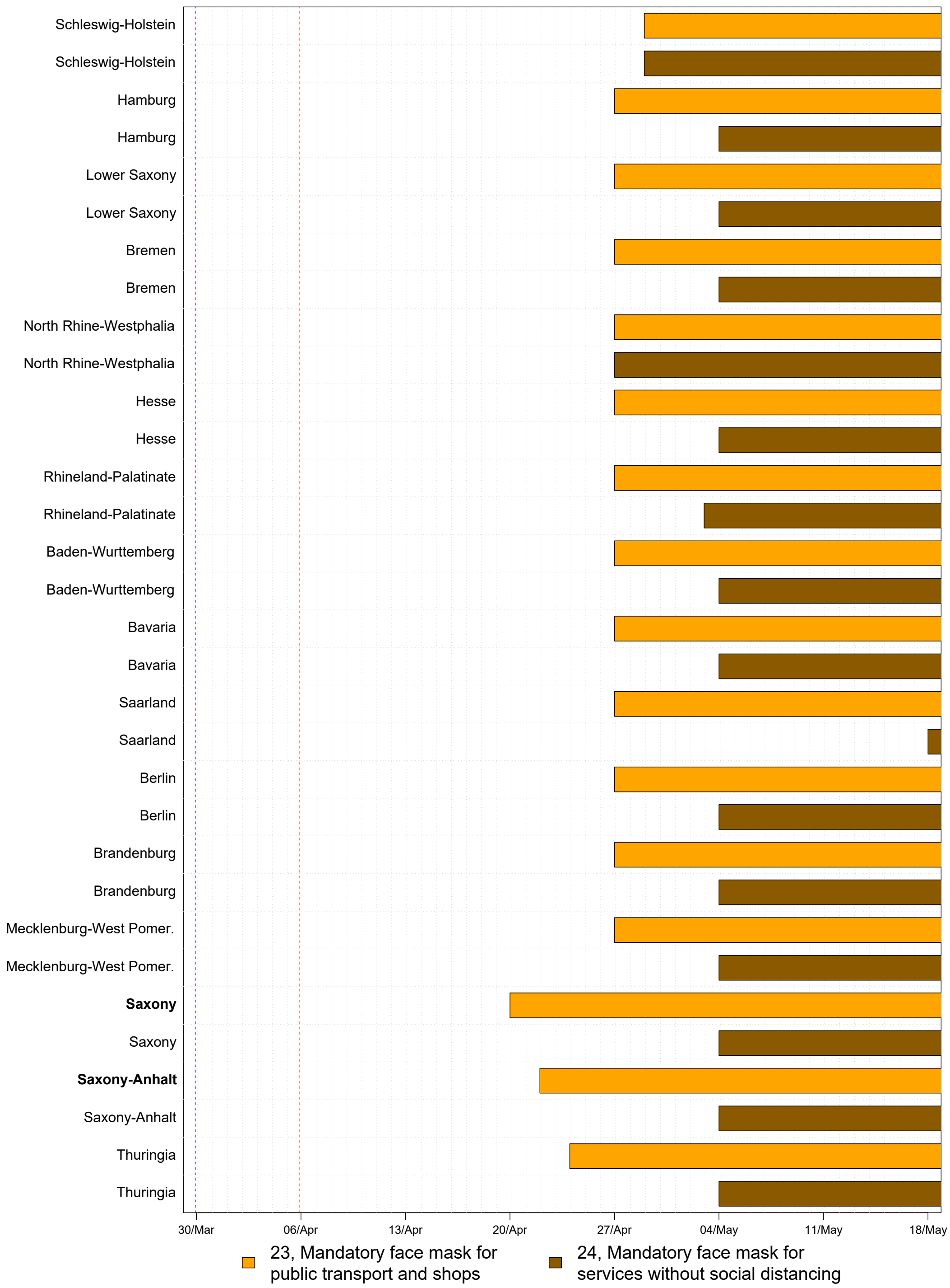

Figure 7: Time line of making face masks mandatory across federal states

We found two exceptions to this general principle of two measures. Thuringia only introduced face masks for public transports and shops. Bavaria introduced face masks for public transports and shops first as a recommendation ("should be worn") on April 20. This was corrected by making face masks mandatory ("have to be worn") on April 27. We display the latter in the figure. We do not believe that adding Bavaria to the treatment group (by assuming that "should be" was already understood by the public as "have to be") would considerably change our findings.

For clarity, we present the dates for federal states in the following table. This table also displays regions such as Jena that introduced face masks earlier than the federal state to which they belong.

Table 1: When face masks became compulsory in federal states and municipal districts

federal state	**public transport**	**services w/o distancing**	**individual NUTS3 region**	**mandatory face masks**	**difference in days to fed. state**
Baden-Wurttemberg	27.04.2020	04.05.2020	Landkreis Rottweil	17.04.2020	10
Bavaria	27.04.2020	04.05.2020			
Berlin	27.04.2020	04.05.2020			
Brandenburg	27.04.2020	04.05.2020			
Bremen	27.04.2020	04.05.2020			
Hamburg	27.04.2020	04.05.2020			
Hesse	27.04.2020	04.05.2020	Main-Kinzig-Kreis	20.04.2020	7
Mecklenburg-West Pomer.	27.04.2020	04.05.2020			
Lower Saxony	27.04.2020	04.05.2020	Wolfsburg	20.04.2020	7
			Braunschweig	25.04.2020	2
North Rhine-Westphalia	27.04.2020	27.04.2020			
Rhineland-Palatinate	27.04.2020	03.05.2020			
Saarland	27.04.2020	18.05.2020			
Saxony	20.04.2020	04.05.2020			
Saxony-Anhalt	22.04.2020	04.05.2020			
Schleswig-Holstein	29.04.2020	29.04.2020			
Thuringia	24.04.2020	-	Jena	06.04.2020	18
			Nordhausen	14.04.2020	10

A.2 The timing of other public health measures

As it is not enough to take only dates into account when face masks became mandatory, we provide an overview of the timing of other public health measures. This will show that our results capture the effects of face masks and not of other public health measures. Figure 8 shows the points in time when measures entered into force in Jena. All measures for Thuringia are also binding for Jena.[12] As Jena introduced three regulations concerning face masks, they became mandatory in three steps. April 1st saw the introduction of face masks for services where

[12] We are grateful to Jan Franke for many explanations related to public health measures in Jena and Thuringia.

a distance of 1.5 meters cannot be kept. On April 6, masks became mandatory for public transports, shops, food deliveries stores and offices of craftsmen and service providers. As of April 10, masks also became mandatory at work and in public buildings, as long distance of 1.5 m cannot be kept. (See also the box on the next page.) Measures of April 1st and 6 are measures also employed by federal states subsequently (see section A.1). The measure of 10 April was employed only by Jena (at least in this wording).

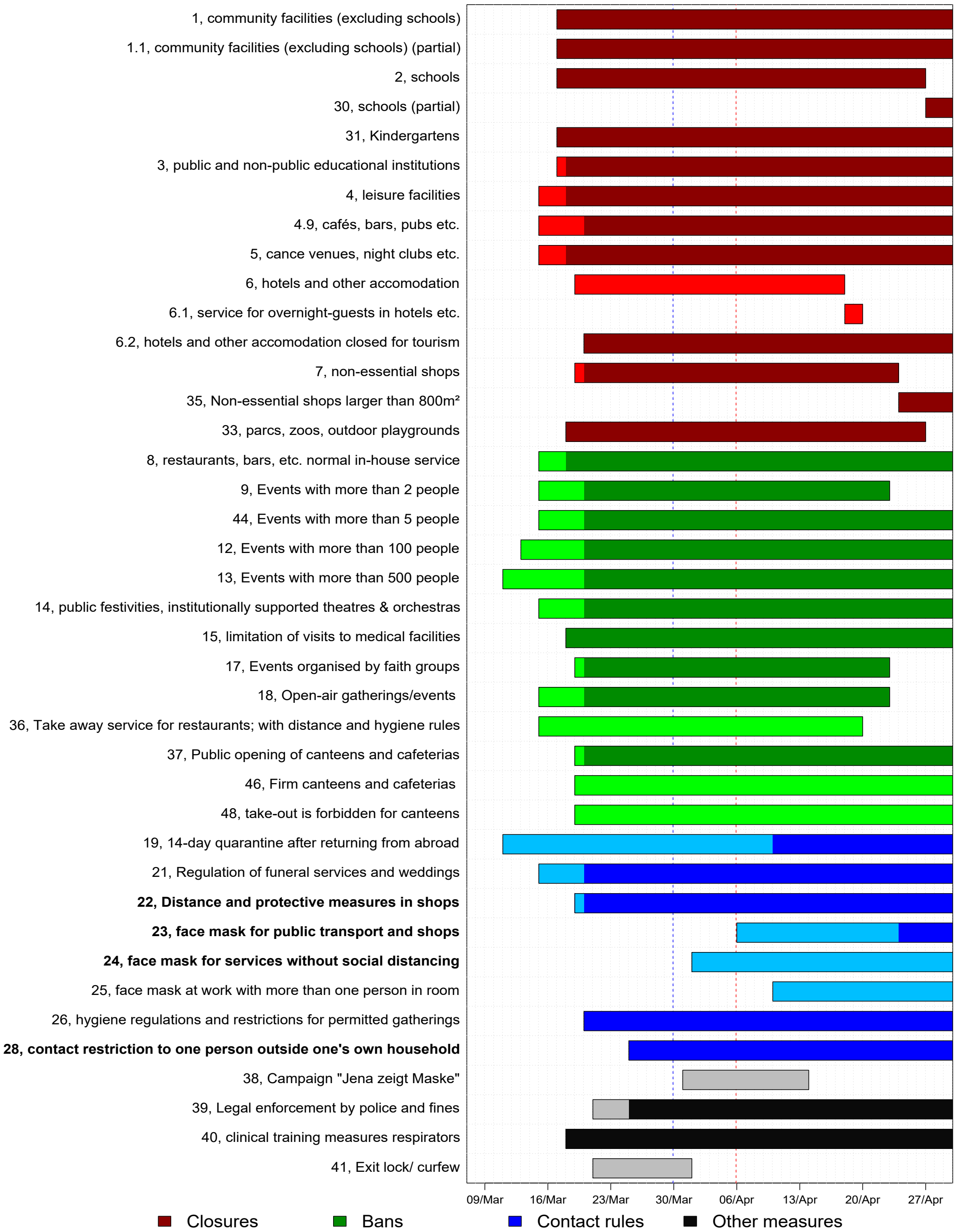

Figure 8: Time line of public health measures in Jena. Light bars indicated measures in force only in Jena, dark bars indicate measures in force in Thuringia (and thereby also in Jena)

Most importantly for our strategy to quantify the effect of face masks, we note that the regulation closest in time, apart from the campaign "Jena zeigt Maske", entered into force on March 25 (number 28, contact restriction). After face masks became mandatory, only exit strategies were implemented. Measure 6.1 that restricts service for over-night guests in hotels is part of an exit strategies that allows hotels to reopen (measure 6) provided hotels do not provide service to over-night guests.[13]

This picture proves that there are no measures relevant for public health implemented in Jena that could affect the spread of Covid-19 around the time when face masks where introduced. We therefore conclude that it is face masks itself whose effect we measure in the main text.

Due to the enormous interest in our study, both within Germany and worldwide, we reproduce here the regulation that makes face masks mandatory in Jena. The regulation is dated March 31st, 2020 and enters in force on April 1st, 2020.

Box 1: The regulation concerning face masks in Jena (source: Öffentliche Bekanntmachung der Stadt Jena, 31. 03. 2020, Vollzug des Gesetzes zur Verhütung und Bekämpfung von Infektionskrankheiten beim Menschen)

13. Jedermann hat bei Vorliegen der nachfolgend genannten Voraussetzungen einen Mund-Nasen-Schutz zu tragen. Anerkannt ist jeder Schutz, der aufgrund seiner Beschaffenheit geeignet ist, eine Ausbreitung von übertragungsfähigen Tröpfchenpartikeln durch Husten, Niesen, Aussprache zu verringern, unabhängig von einer Kennzeichnung oder zertifizierten Schutzkategorie (ausreichend sind daher auch aus Baumwolle selbstgeschneiderte Masken, Schals, Tücher, Buffs etc.)

a) Diese Verpflichtung gilt ab sofort für folgende Bereiche:
   - Die Inanspruchnahme und Erbringung von Dienstleistungen, bei denen sich der Mindestabstand von 1,5 m nicht durchgängig einhalten lässt.

b) Weiterhin gilt diese Verpflichtung ab dem 06.04.2020 für folgende Bereiche:
   - die Nutzung des öffentlichen Personennahverkehrs im Stadtgebiet Jenas,
   - das Betreten von geöffneten Verkaufsstellen,
   - das Betreten von Orten zur Abgabe von Speisen und Getränken zum Mitnehmen bzw. Ausliefern,
   - das Betreten der Diensträume von Handwerkern und Dienstleistern.

c) Schließlich gilt diese Verpflichtung ab dem 10.04.2020 für folgende Bereiche:
   - der Aufenthalt in geschlossenen Räumen mit mindestens einer anderen Person (insbesondere auch die Arbeitsstätte), ausgenommen hiervon ist der private Wohnbereich oder wenn im Raum pro Person mindestens 20 qm zur Verfügung stehen und der Mindestabstand von 1,5 m sichergestellt ist,
   - generell im öffentlichen Raum, wo eine Unterschreitung des Mindestabstands von 1,5 m nicht dauerhaft sichergestellt ist (dies gilt nicht bei Bewegung unter freiem Himmel, insbesondere Spazierengehen und Sport).

A.3 When are effects of public health measures visible in the data?

Imagine a public health measure is implemented on a certain day and that it is effective. When should we see the effects in the data? This delay between measure and statistical visibility depends on the usual incubation period and on the reporting delay. The incubation period is well-studied and has a median of 5.2 days and 95% of all delays lie in the range of around 2 to 12 days. They seem to be approximately log-normally distributed (1, 2). The reporting delay is not as well-studied. It consists of a delay due to diagnosis, testing and reporting of the test: A person with symptoms needs to decide to go to a GP in order to obtain a diagnosis. With typical symptoms, a test is undertaken, and the result needs to be reported to authorities. Formally, let D_I denote a random variable that describes the incubation period. Let D_R denote a second random variable that describes the delay between perceptible symptoms and reporting to authorities of a positive SARS-CoV2 test. We are interested in the distributional properties of the

[13] Note that measures 6 and 6.1 were implemented in Jena only (hence the light color). The corresponding measure 6.2 in Thuringia (dark red) closed hotels for tourism only.

overall delay defined as D = D_I+D_R. We will take the median of D as our measure for how long it takes before effects of public health measures are visible in the data.

Luckily, 3 provides information on the date of reporting and on the day of first symptoms (for around 80% of all reported Covid-19 cases). The difference between these two dates gives a vector of realizations of the random variable D_R. In total, we have 103,171 observations.

Findings for incubation. 2 and 1 describe the delay between infection and symptoms, i.e. the incubation period, by a lognormal distribution. To be precise about parameters in what follows, a lognormal distribution of a random variable X has the density $f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)}\ for\ x > 0$, where σ is the dispersion parameter and μ the scale parameter. The mean, median and variance are given by

$$E\ X = e^{\mu+\frac{\sigma^2}{2}}, m = e^{\mu},\ Var\ X\ =\ \left[e^{\sigma^2} - 1\right]e^{2\mu+\sigma^2}.$$

(2) report m=5.1 and that 95% of all cases lie between 2.2 and 11.5 days. The latter reads, more formally $\int_{2.2}^{11.5} f(x)dx = .95$. We numerically compute the parameters σ from this equation and obtain σ=0.4149. The scale parameter is given by μ=ln 5.1= 1.63.

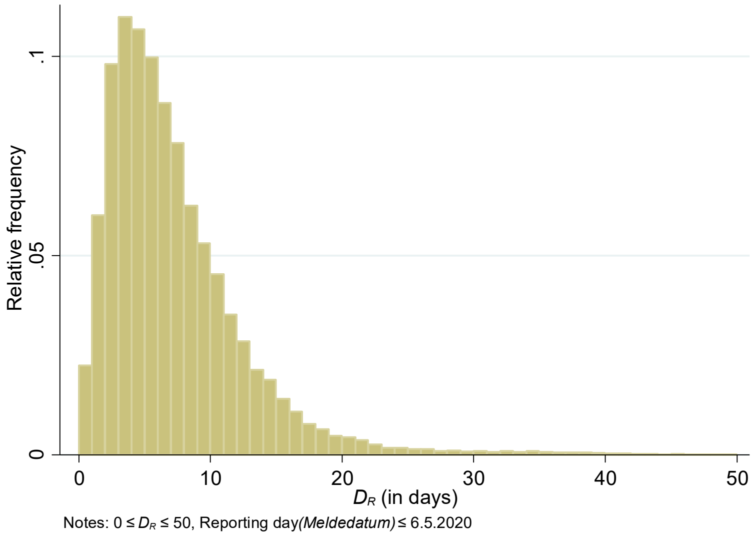

Figure 9: Histogram of delay between first symptoms and reporting

Findings for reporting. For illustration purposes, we plot a histogram of realizations of D_R in Figure 9. The mean, median (50% percentile), variance and standard deviation of D_R are reported in the next table.

Table 2: Descriptive statistics for the reporting delay D_R

Mean	Median	Variance	Standard deviation
6.80	6	30.92	5.56

Note: In the RKI data set, there are 119,917 observations with information on day of infection (until reporting day May 6, 2020). We focus on 118,618 with D_R≥0.

Merging the two. We consider the duration between infection and reporting as one random variable. We call it total delay *D* and it consists of the sum of incubation and reporting delay, $D = D_I + D_R$. Obviously, the mean is $ED = ED_I + ED_R$ and the variance reads $\mathrm{Var}D = \mathrm{Var}D_I + \mathrm{Var}D_R$ if we are willing to assume independence between the two random variables. As we do not believe that diagnosis or reporting lags are influenced by the length of the incubation period, we believe that this is a weak assumption.

As we need more information than the first two moments for our analysis, we now derive the distribution of *D*, i.e. the distribution of a sum of two random variables. We denote it by $F_D(\delta)$, i.e. $F_D(\delta)=Prob(D\leq\delta)$. We ask what the probability is that $D<\delta$ where δ is some constant. We continue to assume that D_I and D_R are independent random variables. The corresponding densities are $f(\delta_I)$ and $g(\delta_R)$, respectively. This probability is given by

$$Prob(D_I + D_R \leq \delta) = \int_0^\delta \left[\int_0^{\delta-\delta_I} f(\delta_I)g(\delta_R)d\delta_R\right] d\delta_I,$$

having the usual interpretation: when we are interested in values below or equal to δ, we let δ_I run from 0 to δ and δ_R from 0 to δ- δ_I such that the sum of the two is always smaller or equal to δ. Integrating over the joint density (which is a product given independence) gives the desired probability. This integral gives us the distribution $F_D(\delta)$ we were looking for.

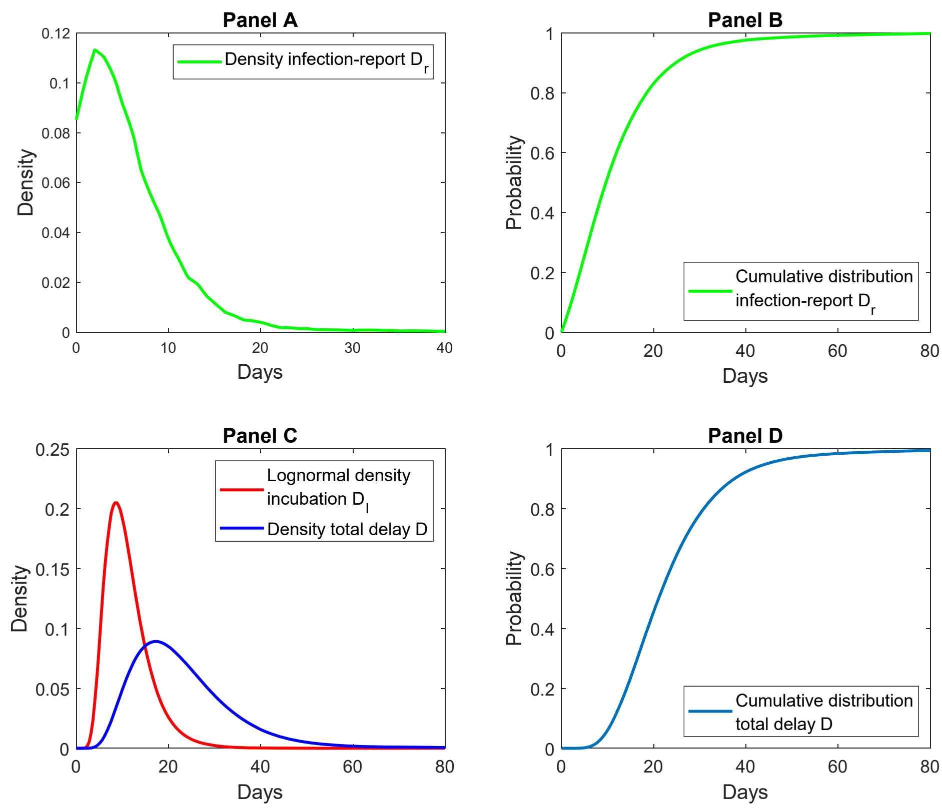

Figure 10: Density of the total delay D

If we needed a density $f_D(\delta)$, we could compute the derivative of this expression with respect to δ. This would give the usual convolution expression,

$$f_D(\delta) = \frac{dF_D(\delta)}{d\delta} = \frac{d}{d\delta}\int_0^{\delta}\left[\int_0^{\delta-\delta_I} f(\delta_I)g(\delta_R)d\delta_R\right]d\delta_I = \int_0^{\delta-\delta} f(\delta)g(\delta_R)d\delta_R + \int_0^{\delta}\frac{d}{d\delta}\left[\int_0^{\delta-\delta_I} f(\delta_I)g(\delta_R)d\delta_R\right]d\delta_I = \int_0^{\delta} f(\delta_I)g(\delta-\delta_I)d\delta_I.$$

Keeping in mind that we work with the assumption that $f(\delta_I)$ is the density of the exponential distribution and that $g(\delta_R)$ is the density corresponding to the histogram in Figure 9 above, we can easily compute the density numerically. Figure 10 provides a visual impression.
Our data imply a mean of 11.7 days and a median of 10.5 days. This provides a basis for studies (e.g. 4) that need to assume a certain delay between infection and visibility in the data. [14] Our findings show that a delay of two to three weeks is too large. The percentiles of the total delay are in the following table.

Table 3: Percentiles of total delay *D*

1	**2.5**	**5**	**10**	**25**	**50**	**75**	**80**	**90**	**95**	**97,5**	**99**
3.42	4.09	4.78	5.70	7.65	10.52	14.30	15.41	18.74	22.22	26.29	34.23

A.4 Visibility in data II – Conceptual background

Conceptional background. We now present a standard SIR model. Let the (expected) number of individuals in the state of being susceptible at a point in time *t* be denoted by *S(t)*, the number of infectious individuals is *I(t)* and the number of removed is R(t).[15] The number of susceptible falls according to $\dot{S}(t) = -rI(t)S(t)$, where *r* is a constant and $rI(t)$ can be called the individual infection rate. Denoting the sum of individual recovery and death rate by a constant *a*, the number of infectious individuals changes according to $\dot{I}(t) = rI(t)S(t) - aI(t)$. Finally, the number of removed (recovered or death) individuals rises over time according to $\dot{R}(t) = aI(t)$. The number of individuals that have ever been infectious between the beginning of the epidemic in 0 and some point in time *t* amounts to $I^{ever}(\mathrm{t}) = \int_0^t rI(x)S(x)dx$. This number is the theoretical counterpart to the number of Covid-19 cases reported by health authorities worldwide. This model is used for our conceptional discussion in the main part of the paper.

We could also wish to inquire into the long-run effects of face masks. In this case, we would have to solve the underlying SIR model for the long-run, i.e. for when the epidemic is over. There are two issues. First, the future course of the epidemic is unknown given uncertainty about the availability of pharmaceutical solutions. Second, the long-run number of susceptible individuals depends on model parameters and can be larger than zero (8, 9). The SIR model therefore does not automatically end with herd immunity.[16] If the outflow from I(t) is larger than the inflow, the epidemic ends. To judge these long-run effects of face masks one would have to ignore potential pharmaceutical solutions and structurally estimate parameters of a much more elaborated SIR model. We therefore present the effects of masks by the measure proposed above in equation (1).

---

[14] We are grateful to Christof Kuhbandner for discussions of this point.
[15] More elaborate models designed for Covid-19 exist (e.g. 5–7). The simple model employed here is, however, sufficient for our interpretation purposes.
[16] In this case, any public health intervention would only delay the epidemic but not reduce the long-run total number of infections. See e.g. (5, 6) for a discussion.

B Background and additional estimates for SCM application to Jena

B.1 The synthetic control method

General background. Our methodical choice is motivated as follows: First, the design of SCM to "*estimate the effects of <…> interventions that are implemented at an aggregate level affecting a small number of large units (such as cities, regions, or countries)*" (10, p. 3) clearly matches with our empirical setup. Compared to standard regression analyses, SCM is particularly well suited for comparative case studies with only one treated unit or a very small number thereof (11, 12). Second, the method is flexible, transparent and has become a widely utilized tool in the policy evaluation literature (13) and for causal analyses in related disciplines (see, e.g., 14, for an overview of SCM in health economics, 15, for a biomedical application).[17]

SCM identifies synthetic control groups for the treated unit(s) to build a counterfactual. In our case, we need to find a group of structurally similar regions that has followed the same Covid-19 trend as treated unit(s) before mandatory masks have been introduced in the latter. This control group would then most likely have had the same behavior as treated unit(s) in the absence of the mask obligation. We can then use this group to 'synthesize' the treated unit and conduct causal inference.

The synthetic control group is thereby constructed as an estimated weighted average of all regions in which masks did not become compulsory earlier on. Historical realizations of the outcome variable and several other predictor variables that are relevant in determining outcome levels allow us to generate the associated weights, which result from minimizing a pre-treatment prediction error function (see 10, 11, 17 for methodical details).

Implementation. The implementation of the SCM is organized as follows. As baseline analysis, we focus on the single treatment case for the city of Jena for three reasons. First, as shown in Figure 1, Jena was the pioneer region for mandatory face masks in public transport and sales shops on April 6. This results in a lead time of 18 days relative to mandatory face masks in the surrounding federal state Thuringia on April 24. By April 29, all German regions had introduced face masks. A sufficiently long lag between mandatory face masks in the treated unit vis-à-vis control regions is important for effect identification.

Second, the timing of the introduction of face masks in Jena is not affected by other overlapping public health measures related to the Covid-19 spread. To support this claim, we looked at all regulations (totaling almost 50) that were implemented in Jena between the beginning of March 2020 and end of April.[18] We also looked at all regulations in Thuringia as they are binding for Jena. A graphical illustration of the timing of the various measures and related discussion is in appendix A.1. As all other measures are more than 10 days away from masks becoming mandatory, we can be certain that we measure the effects of face masks.

[17] 16 employ the SCM to estimate the effect of the shelter-in-place order for California in the development of Covid-19. The authors find *inter alia* that around 1600 deaths from Covid-19 were avoided by this measure during the first four weeks.

[18] The first public health measure in Germany to mitigate the spread of Covid-19 dates from March 10 in North-Rhine Westphalia and prohibited meetings with more than 1000 participants. This measure was also implemented by many other federal states, including Thuringia one day after. See 18 for more background.

Third, Jena is in various ways a representative German city suitable for studying the Covid-19 development: On April 5, which is one day before face masks became compulsory in Jena, the cumulative number of registered Covid-19 cases in Jena was 144. This is very close to the median of 155 registered cases per region in Germany. Similarly, the cumulative number of Covid-19 incidences per 100,000 inhabitants was 126.9 in Jena compared to a mean value of 119.3 in Germany (compare Figure 11 in appendix B.3).

In our baseline configuration of the SCM, we construct the synthetic Jena by including the number of cumulative Covid-19 cases (measured one and seven days before the start of the treatment) and the number of newly registered Covid-19 cases (in the last seven days prior to the start of the treatment) as autoregressive predictor variables. The chosen lag structure shall ensure that the highly dynamic Covid-19 development is properly captured. We use cross-validation tests to check the sensitivity of the SCM results when we impose a longer lag structure. The autoregressive predictors are complemented by cross-sectional data on the region's demographic and basic health care structure to control for confounding factors at the regional level.

Although the case study of Jena can be framed in a clear identification strategy, the Covid-19 spread in a single municipality may still be driven by certain particularities and random events that may prevent a generalization of estimated effects. We therefore also test for treatment effect in regions that introduced face masks after Jena but still before they became compulsory all across Germany. Importantly, here we extend the single treatment approach to the analysis of multiple treated units by considering all regions in the treated group that introduced face masks by April 22. This results in a total of 32 regions out of which 8 are larger cities (*kreisfreie Städte*).

All other regions apart those located in Thuringia (April 24) and Schleswig-Holstein (April 29) introduced face masks on April 27. We employ this staggered introduction to study the effects of mandatory masks up to May 11, which gives us a time window of 20 days to measure treatment effects. We end our analysis on May 11 to avoid a potential underestimation of treatment effects since by that day all control regions had face masks in use for 14 days. This cut-off date is important as we expect that differences in the epidemic spread between treated and control regions would disappear afterwards if we assume a median incubation period of 5.2 days (see 1, 2) and a similar reporting lag. This overall time lag between the infection and registration in the data is also crucial for the interpretation of our results and we discuss it in detail in appendix A.3.

Although SCM appears to be a natural choice for our empirical identification strategy, we are aware that its validity depends on important practical requirements including the availability of a proper comparison group, the absence of early anticipation effects or interference from other events (10, 19, 20). In the implementation of the single and multiple treatment SCM we check for these pitfalls through different sensitivity checks and placebo tests. In our baseline case study for Jena (and similarly of the multiple treatment approach), we deal with these issues in as follows:

1. We make sure that regions used to create the synthetic control, i.e. the donor pool, are not affected by the treatment. We eliminate the immediate geographical neighbors of Jena from the donor pool to rule out spatial spillover effects. We also exclude those regions for

which anticipation effects may have taken place because face masks became compulsory in quick succession to Jena.

2. We account for early anticipation effects in Jena. Specifically, we take the announcement that face masks will become compulsory one week before their *de facto* introduction as an alternative start of the treatment period.[19]
3. We have screened the introduction and easing of other public health measures to account for all potential interferences taking place during our period of study. Significant public health measures that have been introduced in Jena but not (or only temporally delayed) in the federal state of Thuringia or other German regions will be tested for their intervening effect on the introduction of face masks by means of placebo-in-time test and auxiliary difference-in-difference regressions.
4. We apply cross-validation tests to check for sensitivities related to changes in historical values in the outcome variables used as predictors. We also test for the sensitivity of the results when changing the composition of regions in the donor pool for computing the synthetic control group.
5. As a mode of inference in the SCM framework, we run comprehensive placebo-in-space tests, which are based on the estimation of placebo treatment effects for each control region in which masks did not become compulsory early on.

Hence, our mode of inference relies on permutation tests and follows the procedures suggested by (19) and applied, for example, by (21 or (22. Estimated placebo treatments for control regions should be small, relative to the treated region(s). We calculate significance levels for the test of the hypothesis that mandatory face masks did not significantly reduce the number of reported Covid-19 cases. This provides us with a set of (one-sided) *p*–values for each day, which capture the estimated treatment effect on reported Covid-19 cases from placebo regions. The *p*-values are derived from a ranking of the actual treatment effect within the distribution of placebo treatment effects for each day after the start of the treatment (i.e., introduction of face masks). We follow the suggestion in (23 and compute adjusted *p*-values taking the pre-treatment match quality of the placebo treatments into account.[20] We also use the set of *p*-values to compute confidence intervals for treatment effects to visualize the significance and precision of the estimated effects (25).

B.2 Summary statistics

The RKI collects data from local health authorities and provides updates on a daily basis (available via API). We use the cumulative number of registered Covid-19 cases in each district as main outcome variable.[21] As an alternative outcome variable, we also use the cumulative incidence rate, i.e. the number of cumulative Covid-19 cases per 100,000 inhabitants.

[19] We use March 30 as the day of the announcement when several local media reports covered the introduction of face masks on April 6. The general decree of the local administration in Jena has been published on March 31.

[20] We conduct all estimations in STATA using "Synth" and "Synth Runner" packages (23, 24). Data and estimation files can be obtained from the authors upon request.

[21] We are aware of the existence of hidden infections. As it appears plausible to assume that they are proportional to observed infections across regions, we do not believe that they affect our results. We chose the date of reporting (as opposed to date of infections) because not all reported infections include information about the date of infection.

Table 4: Summary Statistics of Covid-19 indicators (outcome variables) and predictors characterizing the regional demographic structure and basic health care system

	Mean	**S.D.**	**Min.**	**Max.**
PANEL A: Data on registered Covid-19 cases				
[1] Newly registered cases per day	3.91	10.24	0	310
[2] Cumulative number of cases	147.70	327.12	0	6066
[3] Cum. cases [2] per 100,000 inhabitants	73.50	120.38	0	1542.69
PANEL B: Regional demographic structure and local health care system				
Population density (inhabitants/km^2)	534.79	702.40	36.13	4,686.17
Population share of highly educated* individuals (in %)	13.07	6.20	5.59	42.93
Share of females in population (in %)	50.59	0.64	48.39	52.74
Average age of females in population (in years)	45.86	2.11	40.70	52.12
Average age of males in population (in years)	43.17	1.83	38.80	48.20
Old-age dependency ratio (persons aged 65 years and above per 100 of population age 15-64)	34.34	5.46	22.40	53.98
Young-age dependency ratio (persons aged 14 years and below per 100 of population age 15-64)	20.54	1.44	15.08	24.68
Physicians per 10,000 of population	14.58	4.41	7.33	30.48
Pharmacies per 100,000 of population	27.01	4.90	18.15	51.68
Settlement type (categorial variable$^{\$}$)	2.59	1.04	1	4

Notes: * = International Standard Classification of Education (ISCED) Level 6 and higher; $ = included categories are 1) larger cities *(kreisfreie Großstädte)*, 2) urban districts *(städtische Kreise)*, 3) rural districts *(ländliche Kreise mit Verdichtungsansätzen)*, 4) sparsely populated rural districts *(dünn besiedelte ländliche Kreise)*.

Table 1 also presents summary statistics of our other predictor variables. We focus on factors that are likely to describe the regional number and dynamics of reported Covid-19 cases. Obviously, past values of (newly) registered Covid-19 cases are important to predict regional trajectories of Covid-19 cases over time in an autoregressive manner. In addition, we argue that a region's demographic structure, such as the overall population density and age structure, and its basic health care system, such as the regional endowment with physicians and pharmacies per population, are important factors for characterizing the local context of Covid-19. Predictor variables are obtained from the INKAR online database of the Federal Institute for Research on Building, Urban Affairs and Spatial Development. We use the latest year available in the database (2017). We consider it likely that regional demographic structures only gradually vary over time such that they can be used to measure regional differences during the spread of Covid-19 in early 2020.

B.3 Covid-19 cases and cumulative incidence rate in Jena and Germany on April 5

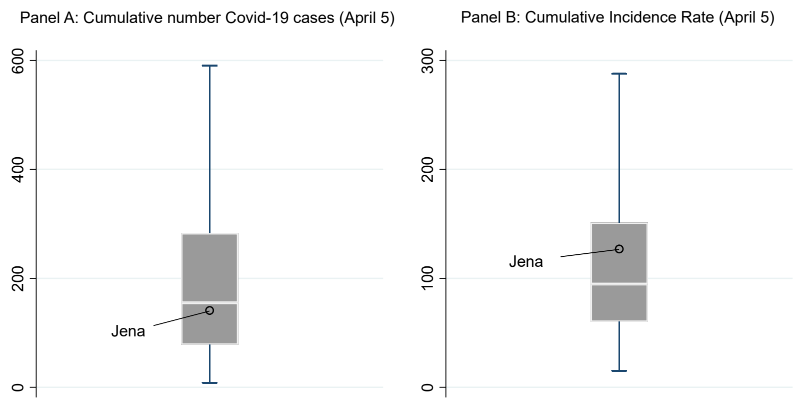

Figure 11: Box plots for distribution of cumulative Covid-19 cases for all 401 German regions on April 5

B.4 Prediction error (RMSPE) and control regions

This appendix shows the balancing properties of the SCM approach together with the root mean square percentage error (RMSPE) as a measure for the quality of the pre-treatment prediction. Weights of control regions are in the subsequent table.

Table 5: Pre-treatment predictor balance and RMSPE for SCM in Figure 1

Treatment:		**Introduction of face masks**		**Announcement/ start of campaign**
	Jena	Synthetic control group	Jena	Synthetic control group
Cumulative number of registered Covid-19 cases (one and seven days before start of treatment, average)	129.5	129.2	93	92.7
Number of newly registered Covid-19 cases (last seven days before the start of the treatment, average)	3.7	3.8	5	5.2
Population density (Population/km^2)	38.4	22.8	38.4	26.3
Share of highly educated population (in %)	968.1	1074.3	968.1	947.9
Share of female in population (in %)	50.1	50.1	50.1	50.1
Average age of female population (in years)	43.5	43.7	43.5	43.9
Average age of male population (in years)	40.5	40.6	40.5	40.8
Old-age dependency ratio (in %)	32.1	29.3	32.1	29.8
Young-age dependency ratio (in %)	20.3	19.6	20.3	19.5
Physicians per 10,000 of population	20.5	19.8	20.5	20.8
Pharmacies per 100,000 of population	28.8	28.7	28.8	28.6
Settlement type (categorial variable)	1	1.3	1	1.9
RMSPE (pre-treatment)	**3.145**		**4.796**	

Donor pool includes all other German NUTS3 regions except the two immediate neighboring regions of Jena (Weimarer Land, Saale-Holzland-Kreis) as well as the regions Nordhausen and

Rottweil which introduced face masks in rapid succession to Jena on April 14 and April 17, respectively.

Table 6: Sample weights for regions in synthetic control group

Introduction of face masks (Panel A in Figure 2)		
ID	**NUTS 3 region**	**Weight**
13003	Rostock	0.326
6411	Darmstadt	0.311
3453	Cloppenburg	0.118
7211	Trier	0.117
6611	Kassel	0.082
5370	Heinsberg	0.046

Note: Donor pools corresponds to SCM estimation in Panel A of Figure 2. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

B.5 Growth rates

Jena has 142 registered cases on April 6 compared to an estimated number of 143 cases in the synthetic control group. On April 26 Jena counts 158 cases and the synthetic control group reports 205 (again estimated) cases. The daily growth rate in Jena is denoted by Δx_{Jena} and can be computed from 142 $[1+\Delta x_{Jena}]^{20} = 158$. The daily growth rate in the control group is denoted by $x_{control}$ and can be computed from 143 $[1+\Delta x_{control}]^{20} = 205$. Hence, the introduction of the face mask is associated with a decrease in the number of infections of ($\Delta x_{control} - \Delta x_{Jena}$) percentage points per day. Analogously, we also calculate differences in the daily growth rates for our SCM analysis including multiple treated units. The results are summarized in the following table.

Table 7: Summary of treatment effects of face mask introduction in Germany

Difference between treated region(s) and synthetic control group(s)	**Single Treatment (Jena)**	**Multiple treatments (all districts)**	**Multiple treatments (larger cities)**
Absolute change in cumulative number of Covid-19 cases over 20 days	-46.9	-7.0	-28.4
Percentage change in cumulative number of Covid-19 cases over 20 days	-22.9%	-2.6%	-8.9%
Percentage change in newly registered Covid-19 cases over 20 days	-75.6%	-15.7%	-51.2%
Difference in daily growth rates of Covid-19 cases in percentage points	-1.28%	-0.13%	-0.46%
Reduction in daily growth rates of Covid-19 cases (in percent)	70.6%	14.0%	47.3%

All indicators in this table are compiled in an Excel-file available online. See e.g. https://www.macro.economics.uni-mainz.de/klaus-waelde/ongoing-work-and-publications/

B.6 SCM results by age groups

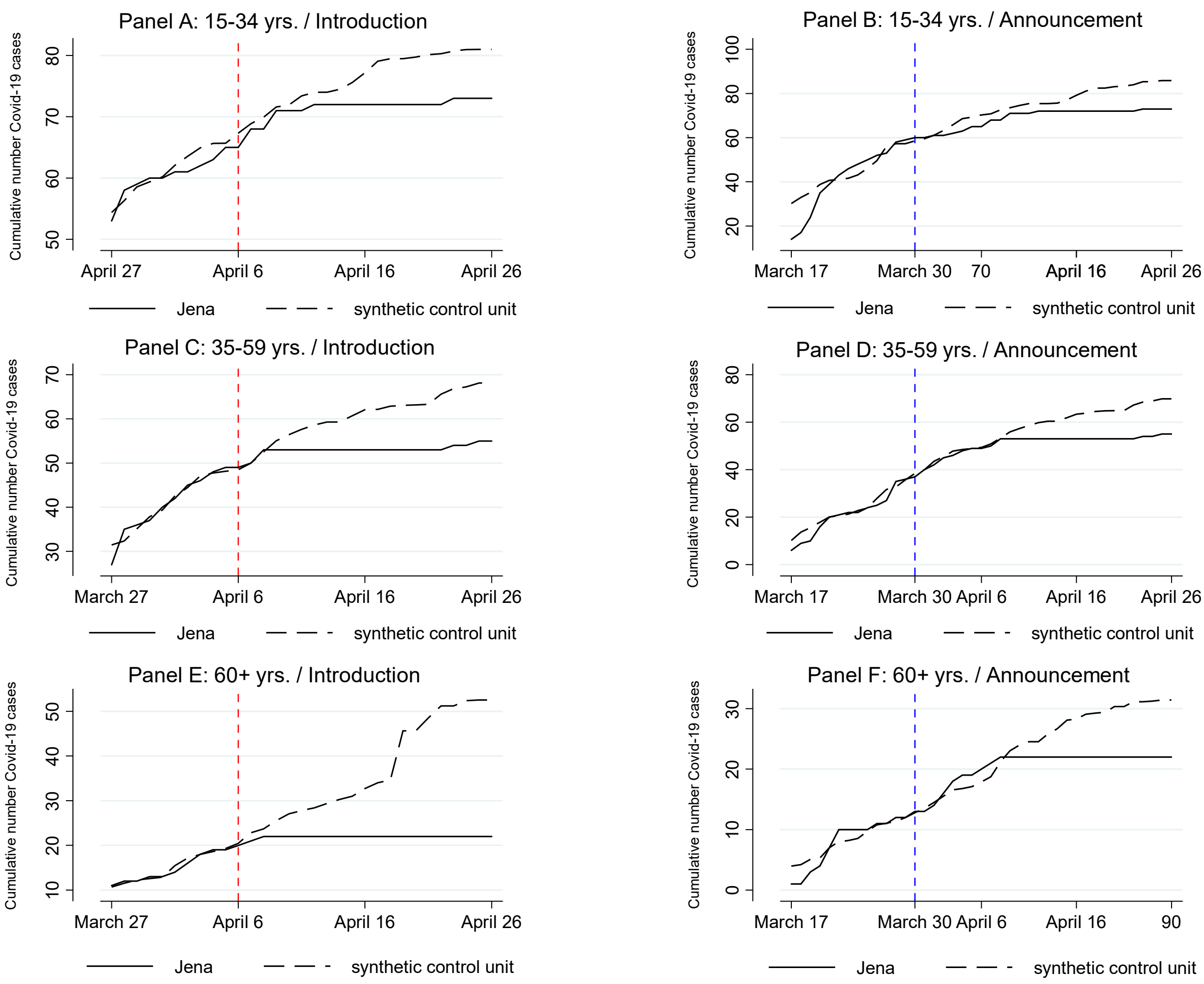

Figure 12: Treatment effects for introduction and announcement of face masks in Jena

Notes: Predictor variables are chosen as for the baseline specification shown in Figure 2; see main text.

Table 8: Sample weights in donor pool for synthetic Jena (cumulative Covid-19 cases; by age groups)

Age Group 15-34 years			**Age Group 35-59 years**			**Age Group 60 years and above**		
ID	**NUTS 3 region**	**Weight**	**ID**	**NUTS 3 region**	**Weight**	**ID**	**NUTS 3 region**	**Weight**
1001	Flensburg	0.323	6411	Darmstadt	0.528	6411	Darmstadt	0.522
7211	Trier	0.207	16055	Weimar	0.16	16055	Weimar	0.244
13003	Rostock	0.184	14511	Chemnitz	0.15	7316	Neustadt a.d. Weinstraße	0.068
5370	Heinsberg	0.142	8221	Baden-Baden	0.07	9562	Erlangen	0.06
3453	Cloppenburg	0.107	6434	Hochtaunus-kreis	0.062	3356	Osterholz	0.056
6413	Offenbach am Main	0.038	8435	Bodenseekreis	0.029	5515	Münster	0.027
			5370	Heinsberg	0.001	9188	Starnberg	0.022

Note: Donor pools corresponds to SCM estimations in Figure 12. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

B.7 Effects on cumulative number of infections per 100,000 inhabitants

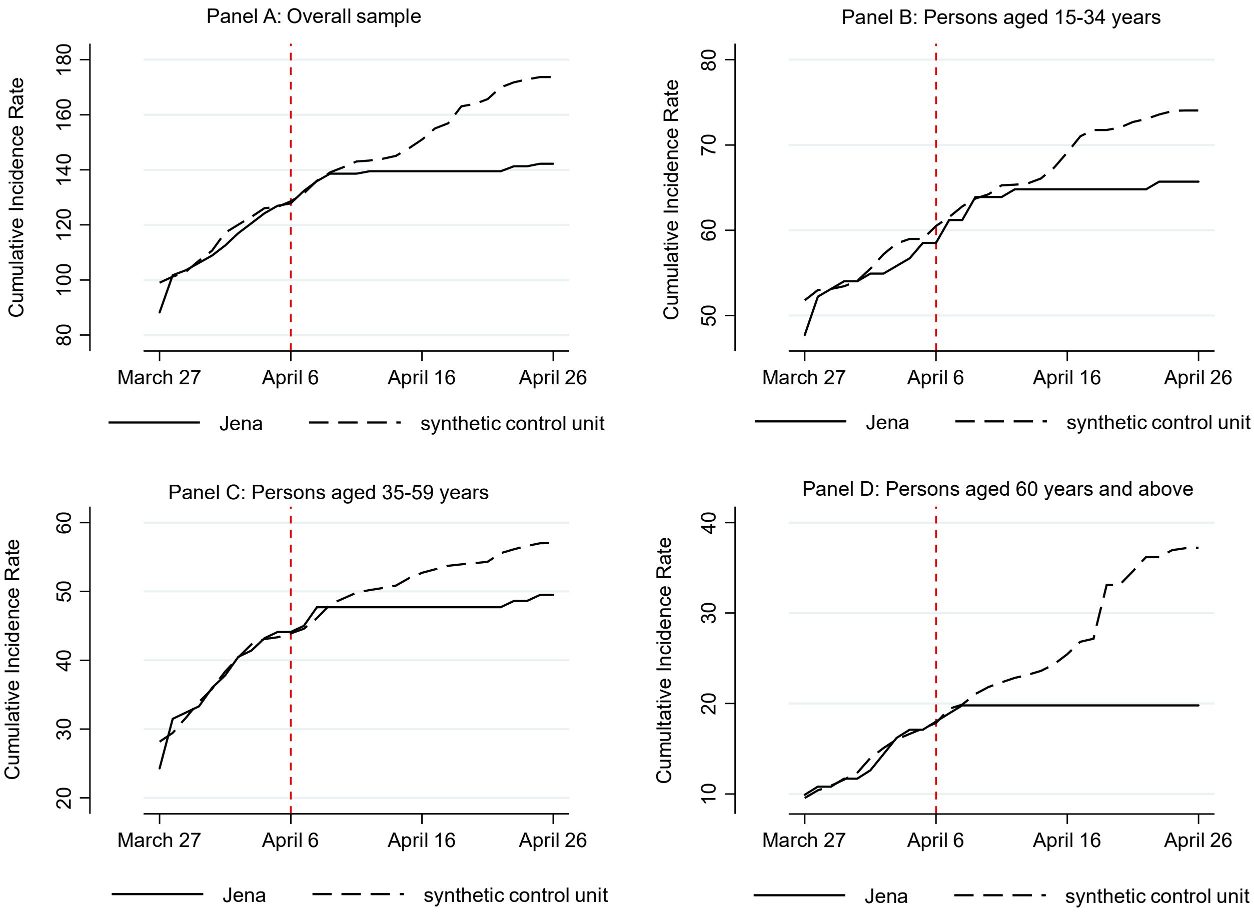

Figure 13: Treatment effects for introduction of face masks on cumulative incidence rate

Notes: See Table 1 for a definition of the incidence rate. Predictor variables are chosen as for baseline specification shown in Figure 2; see main text.

Table 9: Sample weights in donor pool for synthetic Jena (cumulative incidence rate)

ID	**NUTS 3 region**	**Weight**
6411	Darmstadt	0.46
15003	Magdeburg	0.171
5370	Heinsberg	0.133
13003	Rostock	0.093
5515	Münster	0.066
11000	Berlin	0.035
12052	Cottbus	0.032
6611	Kassel	0.011

Note: Donor pools corresponds to SCM estimation in Figure 13. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

Table 10: Sample weights in donor pool for synthetic Jena (cumulative incidence rate; by age groups)

Age Group 15-34 years			Age Group 35-59 years			Age Group 60 years and above		
ID	NUTS 3 region	Weight	ID	NUTS 3 region	Weight	ID	NUTS 3 region	Weight
5370	Heinsberg	0.377	6411	Darmstadt	0.419	6411	Darmstadt	0.448
13003	Rostock	0.288	14511	Chemnitz	0.184	14612	Dresden	0.313
1001	Flensburg	0.14	14612	Dresden	0.154	9188	Starnberg	0.071
6611	Kassel	0.138	8221	Heidelberg	0.138	16054	Suhl	0.069
11000	Berlin	0.058	9188	Starnberg	0.088	5515	Münster	0.06
			5370	Heinsberg	0.016	8221	Heidelberg	0.039

Note: Donor pools corresponds to SCM estimations in Figure 13. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

B.8 Announcement and mobility

B.8.1 Google trends and announcement effects

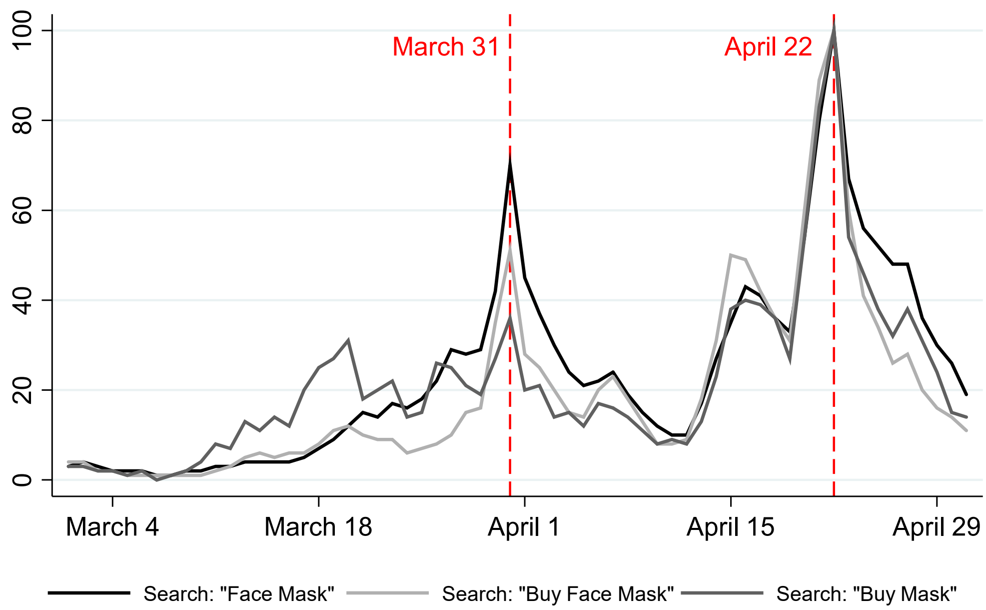

Figure 14: Online search for face masks and purchase options according to Google Trends

Note: Online search for keywords (in German) as shown in the legend as Face Mask ("Mund.-Nasen-Schutz"), Buy Face Mask ("Mundschutz kaufen") and Buy mask ("Maske kaufen"); alternative keywords show similar peaks but with a lower number of hits; based on data from Google Trends (2020).

B.8.2 Mobility trends across German federal states

Figure 15 shows the trend development in overall mobility patterns across German federal states between Feb 17 and May 18, 2020. Data source are the "Covid-19 Community Mobility reports" published by Google LLC (2020). The data track the frequency of visits to different places covered in Google maps on a daily basis compared to a baseline. The latter is set as median value for the corresponding weekday during Jan 3 and Feb 6, 2020. To arrive at a compact measure of regional mobility, we have aggregated data over the different place categories: retail and recreation, groceries and pharmacies, parks, transit stations and workplaces. Given the high volatility of daily data, Figure 13 displays weekly averages. The mobility trends show a clear common pattern: With public health measures taken across all federal states to restrict professional and social contacts (RSC), mobility sharply declined in mid-March. It stayed low for most of the following weeks and only gradually increased from mid-April onwards when first actions to lift RSC and to re-open the economy have been taken (see 18 for details). Importantly, during the timing of the mandatory introduction of face masks in Jena on April 6, no significant change in mobility patterns across federal states can be observed, which potentially confounds our empirical estimates. Although mobility data are increasingly used to study the effects of public health measures, the inspection of the Google data urges to use such data only very carefully in comparative studies at the countries/regional level given a the generally high volatility and significant outliers. This is also recognized by Google LLC (2020).[22]

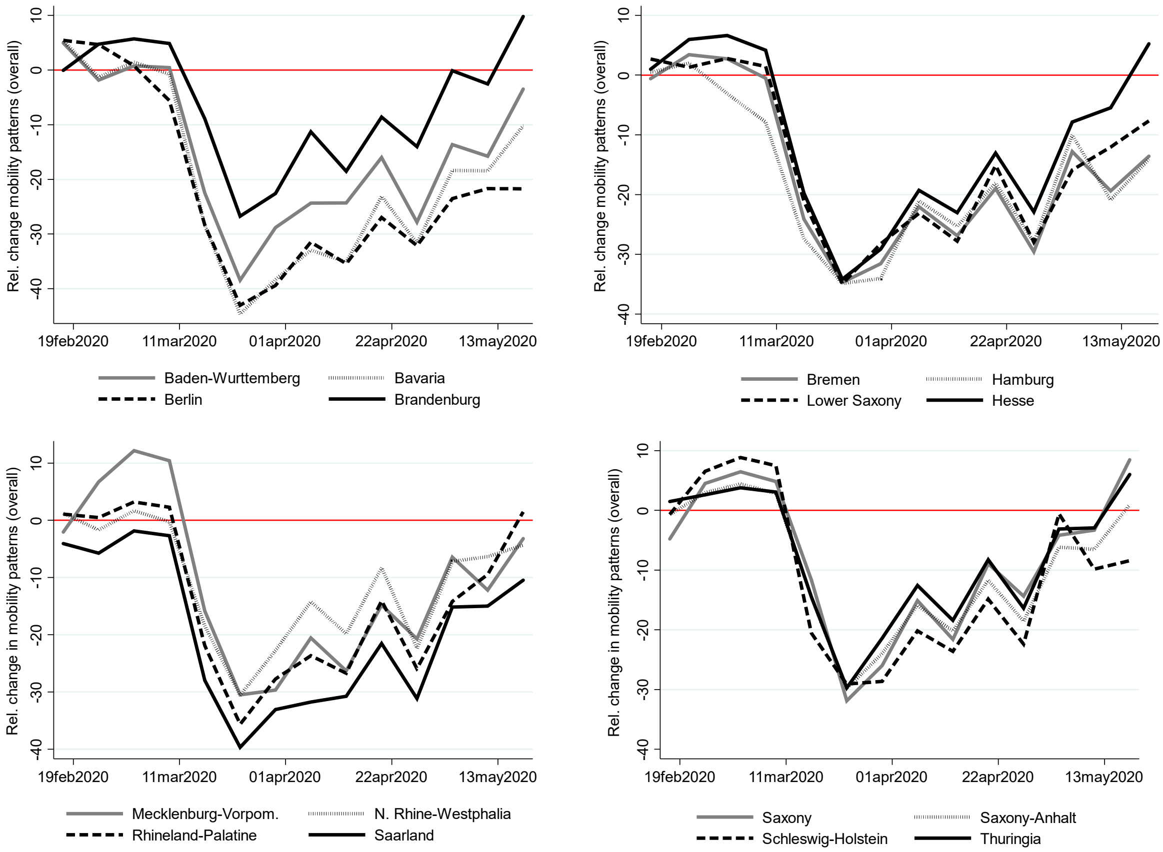

Figure 15: Overall trend in mobility patterns across German federal states (Feb 17 to May 18, 2020)

Source: Google LLC (2020). https://www.google.com/covid19/mobility/ Accessed: 04.06.2020

[22] For details see: https://www.google.com/covid19/mobility/data_documentation.html?hl=en.

B.9 Changes in donor pool for synthetic Jena

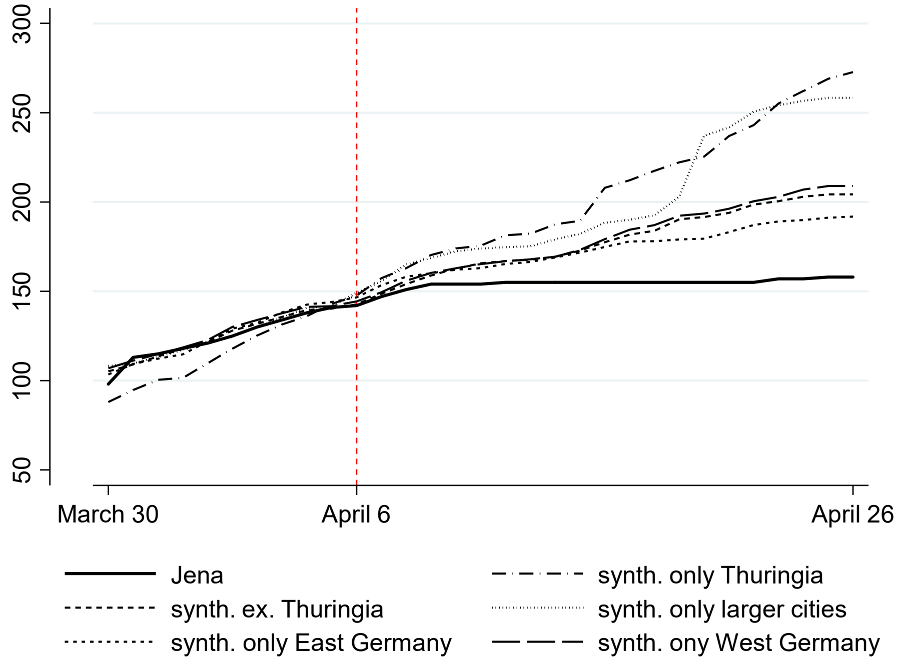

Figure 16: Treatment effects for changes in donor pool used to construct synthetic Jena

Notes: See main text for a detailed definition of the respective donor pools. Predictor variables are chosen as for baseline specification shown in Figure 2; see main text.

Table 11: Sample weights for alternative donor pools used to construct synthetic Jena

	Only Thuringia			**Excluding Thuringia**			**Only larger cities**	
ID	**NUTS 3 region**	**Weight**	**ID**	**NUTS 3 region**	**Weight**	**ID**	**NUTS 3 region**	**Weight**
16076	Greiz	0.533	13003	Rostock	0.318	6411	Darmstadt	0.504
16051	Erfurt	0.467	6411	Darmstadt	0.302	13003	Rostock	0.304
			7211	Trier	0.129	5113	Essen	0.192
			3453	Cloppenburg	0.122			
			6611	Kassel	0.083			
			5370	Heinsberg	0.046			
	Only East Germany			**Only West Germany**				
ID	**NUTS 3 region**	**Weight**	**ID**	**NUTS 3 region**	**Weight**			
16051	Erfurt	0.865	6411	Darmstadt	0.242			
14612	Dresden	0.124	3402	Emden	0.198			
11000	Berlin	0.011	6611	Kassel	0.169			
			7211	Trier	0.168			
			4012	Bremerhaven	0.167			
			5370	Heinsberg	0.057			

Note: Donor pools corresponds to SCM estimations in Figure 16. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

B.10 Cross validation and additional placebo-in-time test

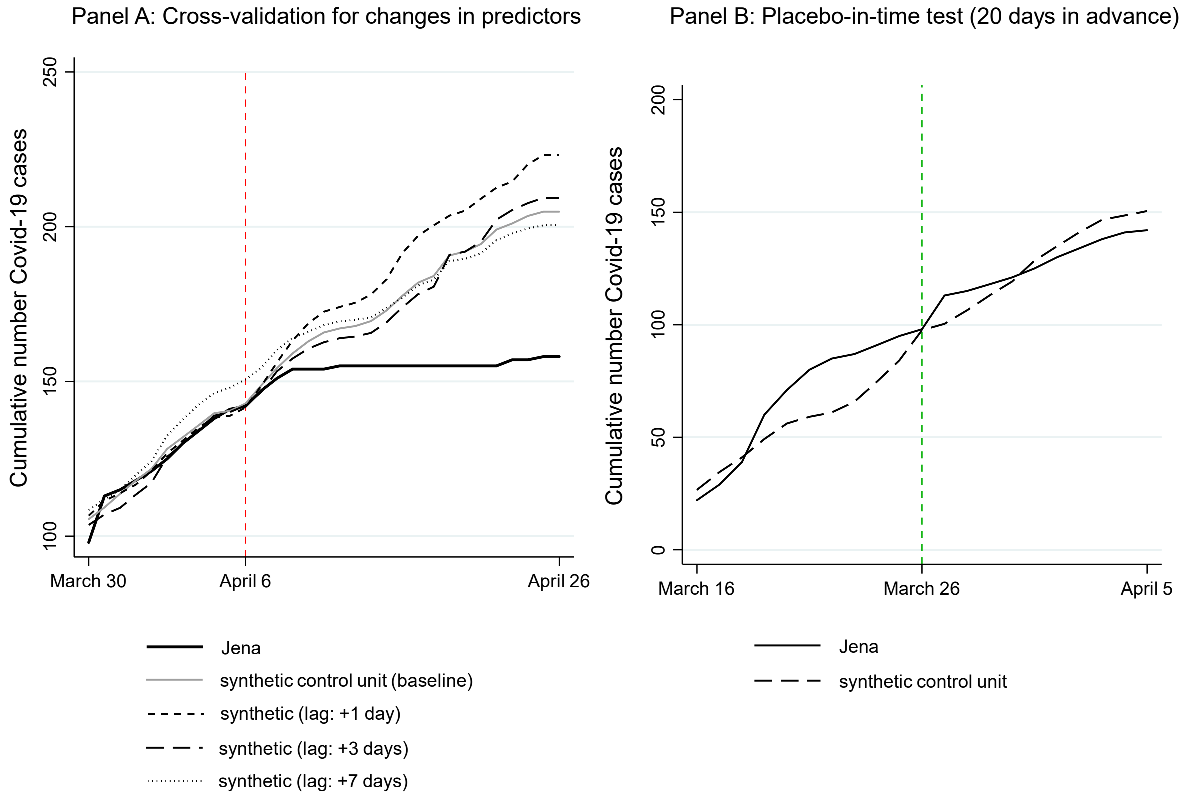

Figure 17: Cross-validation for changes in predictor variables and placebo-in-time test

Notes: In Panel A the baseline specification for the synthetic control group uses historical values of the outcome variable in the following way: i) number of cumulative Covid-19 cases (measured one and seven days before the start of the treatment), ii) the number of newly registered Covid-19 cases (in the last seven days prior to the start of the treatment); the alternative specifications lag these values by 1, 3 and 7 days. In Panel B pseudo-treatment effects for Jena are calculated over a period of 20 days before the introduction of face masks. This period is equally split into a pre- and pseudo post-treatment period.

B.11 Place-in-space tests for other major cities in Thuringia

For the placebo tests in the other cities in Thuringia the same set of predictors as for Jena (Figure 2) has been applied. The reported regions cover all *kreisfreie Städte* plus Gotha (*Landkreis*). The cities Weimar, Suhl and Eisenach have been aggregated since the number of reported Covid-19 is low in these cities.

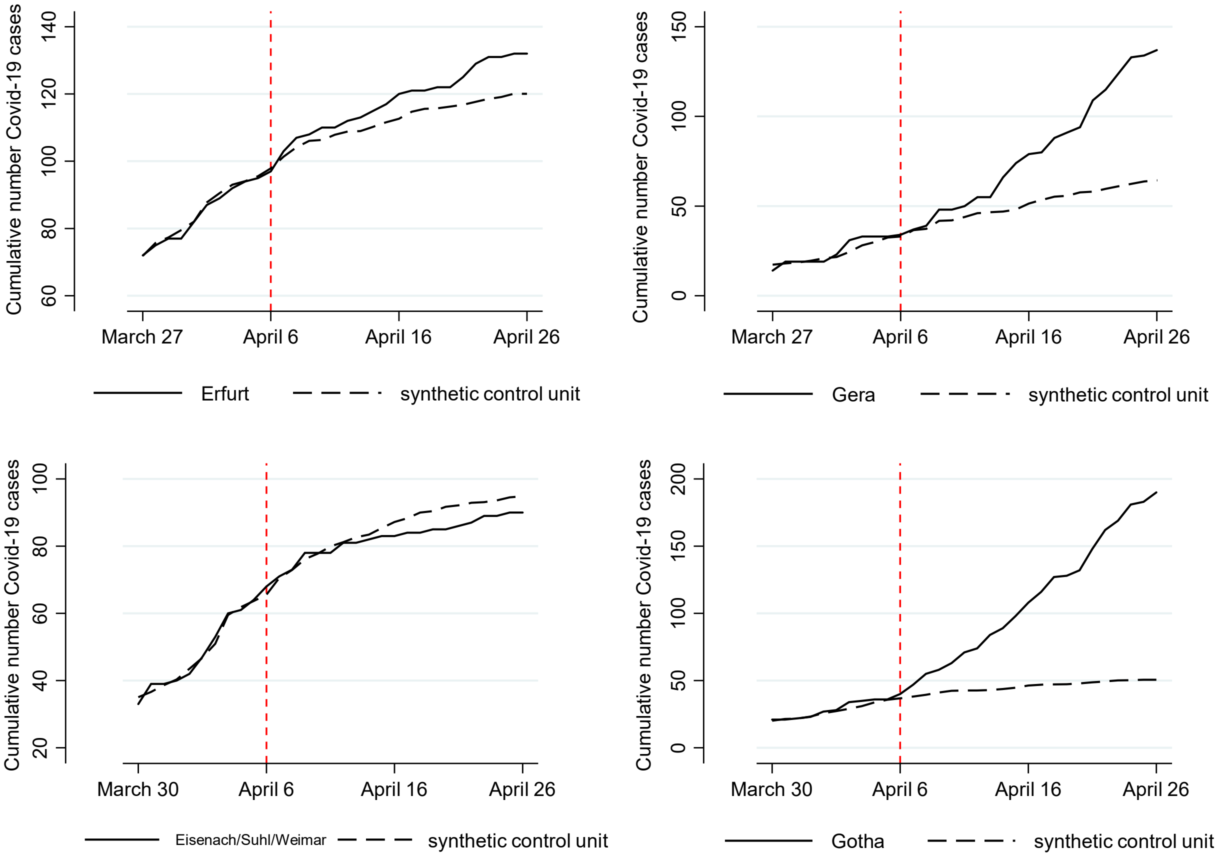

Figure 18: Placebo tests for the effect of face masks in other cities in Thuringia on April 6.

Table 12: Sample weights in donor pool for synthetic control groups (other cities in Thuringia)

	Erfurt			**Gera**	
ID	**NUTS 3 region**	**Weight**	**ID**	**NUTS 3 region**	**Weight**
13003	Rostock	0.28	15001	Dessau-Roßlau	0.501
16055	Weimar	0.244	16054	Suhl	0.222
3356	Osterholz	0.212	7318	Speyer	0.162
7313	Landau in der Pfalz	0.154	8231	Pforzheim	0.061
6413	Offenbach am Main	0.078	7311	Frankenthal (Pfalz)	0.046
5370	Heinsberg	0.029	8211	Baden-Baden	0.005
5515	Münster	0.004	9662	Schweinfurt	0.003
			14521	Erzgebirgskreis	0.001

Note: Donor pools corresponds to SCM estimations in Figure 18. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

Table 13 (cont'd): Sample weights in donor pool for synthetic control groups (other cities in Thuringia)

Weimar/Suhl/Eisenach			Gotha		
ID	**NUTS 3 region**	**Weight**	**ID**	**NUTS 3 region**	**Weight**
15001	Dessau-Roßlau	0.263	15081	Altmarkkreis	0.23
12052	Cottbus	0.236	16077	Altenburger Land	0.164
13004	Schwerin	0.202	15086	Jerichower	0.161
9361	Amberg	0.177	3402	Emden	0.111
14626	Görlitz	0.069	16071	Weimarer Land	0.108
9363	Weiden i.d. Opf.	0.036	16074	Saale-Holzland-Kreis	0.063
14521	Erzgebirgskreis	0.008	16061	Eichsfeld	0.058
9184	München	0.005	16070	Ilm-Kreis	0.055
6411	Darmstadt	0.005	3453	Cloppenburg	0.027
			15003	Magdeburg	0.017
			4012	Bremerhaven	0.007

Note: Donor pools corresponds to SCM estimations in Figure 18. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

C Single treatment analysis in other German cities and regions

In addition to Jena, we estimated treatment effects in Nordhausen (Thuringia, April 14), Rottweil (Baden Württemberg, April 17), Main-Kinzig-Kreis (Hessia, April 20), and Wolfsburg (Lower Saxony, April 20).

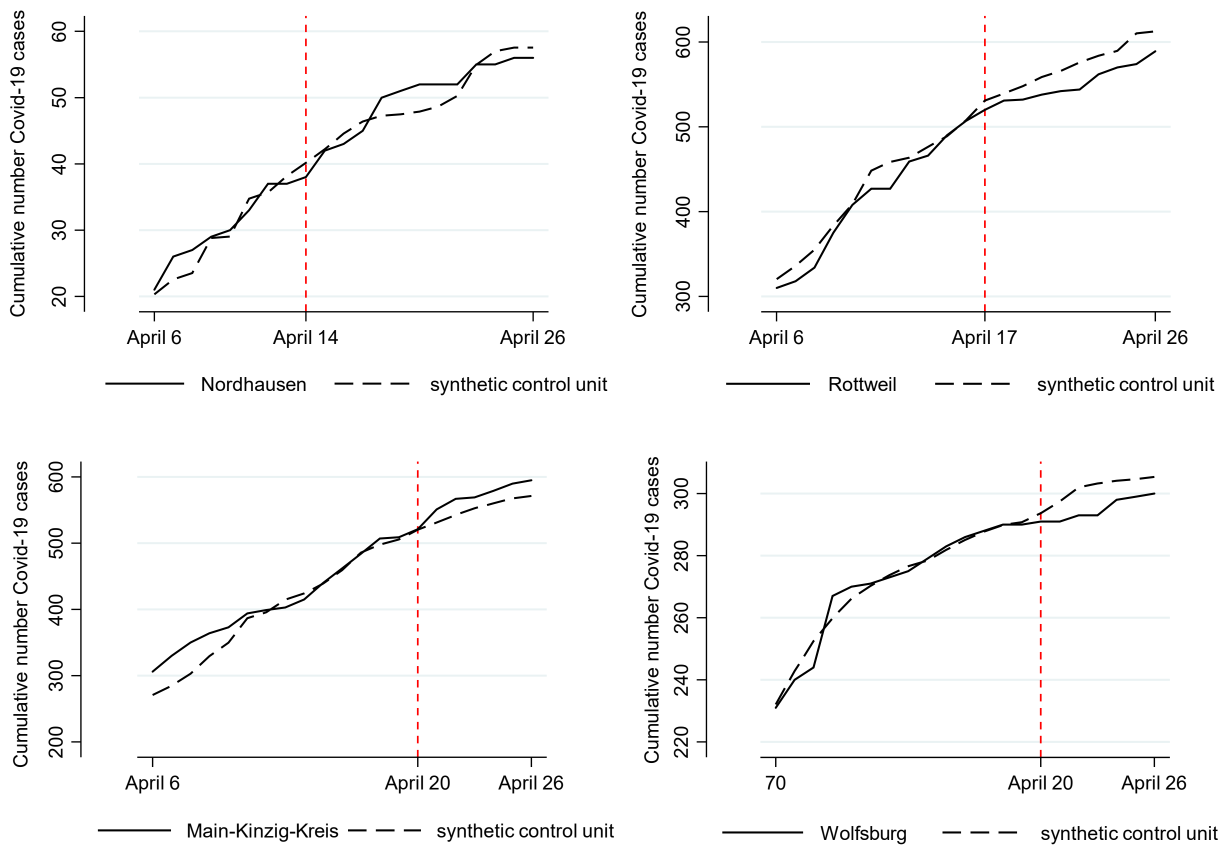

Figure 19: Treatment effects for introduction of face masks in other cities

We ignore Braunschweig here as the introduction of face masks became effective only two days in advance of its federal state. Predictor variables are chosen as for overall specification shown in Figure 1. As the figure shows, the result is 2:1:1. Rottweil and Wolfsburg display a positive effect of mandatory mask wearing, just as Jena. The results in Nordhausen are very small or unclear. In the region of Main-Kinzig, it even seems to be the case that masks increased the number of cases relative to the synthetic control group. As all of these regions introduced masks after Jena, the time period available to identify effects is smaller than for Jena. The effects of mandatory face masks could also be underestimated as announcement effects and learning from Jena might have induced individuals to wear masks already before they became mandatory. Finally, the average pre-treatment RMSPE for these four regions (7.150) is larger than for the case of Jena (3.145). For instance, in the case of the region of Main-Kinzig it is more than three times as high (9.719), which indicates a lower pre-treatment fit. The obtained treatment effects should then be interpreted with some care as the pre-treatment estimation error could also translate into the treatment period. In order to minimize the influence of a poor pre-treatment fit for some individual regions, in the main text, we therefore compare the results for Jena with SCM estimates for multiple treated units.

Table 14: Sample weights in donor pool for synthetic controls (other treated NUTS3 regions)

	Nordhausen			**Rottweil**	
ID	**NUTS 3 region**	**Weight**	**ID**	**NUTS 3 region**	**Weight**
16069	Hildburghausen	0.228	8327	Tuttlingen	0.324
6636	Werra-Meißner-Kreis	0.209	5966	Olpe	0.216
16064	Unstrut-Hainich-Kreis	0.168	8136	Ostalbkreis	0.2
16054	Suhl	0.109	16071	Weimarer Land	0.063
3402	Emden	0.093	14521	Erzgebirgskreis	0.06
12073	Uckermark	0.071	3102	Salzgitter	0.043
12053	Frankfurt (Oder)	0.07	16061	Eichsfeld	0.035
3354	Lüchow-Dannenberg	0.051	9187	Rosenheim	0.031
			9279	Dingolfing-Landau	0.025
			3455	Friesland	0.003
	Main-Kinzig-Kreis			**Wolfsburg**	
ID	**NUTS 3 region**	**Weight**	**ID**	**NUTS 3 region**	**Weight**
8136	Ostalbkreis	0.193	8212	Karlsruhe	0.357
1062	Stormarn	0.168	8221	Heidelberg	0.189
5966	Olpe	0.113	8211	Baden-Baden	0.158
6433	Groß-Gerau	0.105	10046	St. Wendel	0.128
9473	Coburg	0.092	14511	Chemnitz	0.071
5562	Recklinghausen	0.063	5117	Mülheim an der Ruhr	0.059
7313	Landau in der Pfalz	0.059	5315	Köln	0.028
9171	Altrötting	0.056	15003	Magdeburg	0.007
7338	Rhein-Pfalz-Kreis	0.047	9663	Würzburg	0.004
6437	Odenwaldkreis	0.041			
8236	Enzkreis	0.041			
3159	Göttingen	0.023			

Note: Donor pools corresponds to SCM estimations in Figure 19. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

D A brief survey of research on public health measures against Covid-19

D.1 General overview

Consolidated scientific knowledge on Covid-19 and public health measures taken to fight its epidemic spread, though rapidly evolving, is still limited. Our approach goes in line with various studies that have already tried to better understand the effect of public health measures on the spread of Covid-19 (5, 6, 26–32). However, these earlier studies all take an aggregate approach in the sense that they look at implementation dates for a certain measure and search for subsequent changes in the national incidence. There are some prior analyses that take a regional focus (7) but no attention is paid to the effect of policy measures.[23]

There are also many cross-country analyses, both in a structural SIR (susceptible, infectious and removed) sense (34) and with an econometric focus on forecasting the future development of the Covid-19 pandemic (35). Others draw parallels between earlier pandemics and Covid-19 (36). These studies do not explicitly take public health measures into account. Some studies discuss potential effects of public health measures and survey general findings (37–39) but do not provide direct statistical evidence on specific measures.

The synthetic control method (SCM) has been applied by (16) to estimate the effect of the shelter-in-place order for California, USA, in the development of Covid-19. The authors find *inter alia* that around 1600 deaths from Covid-19 have been avoided by this measure during the first four weeks. (40) use SCM to study the case of Sweden as one of the few countries without a lock down. The results indicate that the infection dynamics in the synthetic control group (constructed from a donor pool of other European countries) does not systematically differ from the actual dynamics in Sweden. Bases on Google mobility data, the authors further find that Swedes adjusted their activities in similar ways as in the synthetic control group even without a mandated lock down.

D.2 Evidence for face masks

At present, more and more clinical evidence is presented indicating that face masks catch infectious particles that occur when speaking, coughing, or sneezing. This reduces the risk of infecting another person (41, 42). The effects of face masks have been systematically surveyed by (43) and (44). (44) mainly present evidence on the effect of face masks during non-Covid epidemics (influenza and SARS). (45) reports that they "*did not find any studies that investigated the effectiveness of face mask use in limiting the spread of COVID-19 among those who are not medically diagnosed with COVID-19 to support current public health recommendations*".

In addition to medical aspects (like transmission characteristics of Covid-19 and filtering capabilities of masks), (43 survey evidence on mask efficiency and on the effect of a population. They first stress that "*no randomized control trials on the use of masks <…> has been published*". The study which is "*the most relevant paper*" for (43 is one that analyzed *"exhaled breath and*

[23] In a short note, (33) apply panel methods based on time dummies to understand the relative importance of various public health measures. They employ data at the federal state level and not at the regional level. As a detailed model description is not available, an appreciation of results is difficult at this point.

coughs of children and adults with acute respiratory illness" (46, p. 676), i.e. used a clinical setting. Concerning the effect of masks on community transmissions, the survey needs to rely on pre-Covid-19 studies.

Only very recently, first non-clinical observational studies on the effectiveness of face masks have been published. The work that is most closely related to our approach is (47), who estimate the effects of public health measures on the spread of Covid-19 in the three pandemic epicenters Wuhan, Italy, and New York City over the period January 23 to May 9, 2020. The authors find sizable effects for the introduction of face masks indicating that this public health measure alone reduced the number of infections by over 78,000 in Italy from April 6 to May 9 and by over 66,000 in New York City from April 17 to May 9.

The authors adopt an empirical identification strategy that utilizes the successive implementation of individual public health measures and estimate linear time trends for the period before the introduction of face masks in Italy and New York City. The difference between these trends and actual Covid-19 cases is interpreted as the mitigating effect of mandated face covering. Although the authors argue that their trend projections are reasonable considering the excellent linear correlation for the data prior to the onset of mandated face covering, a limitation is that their study does not employ a strict control group approach and conducts inference on in a "before-after" comparison, which may not suffice to rule out all confounding factors.[24]

(48) use household data for 335 families in Beijing with at least one confirmed Covid-19 case to study factors that influence disease transmission within families. The authors track the rate of secondary transmissions over the two weeks of follow-up from onset of the primary case within the family. Findings suggest that transmission was significantly reduced by frequent use of chlorine or ethanol-based disinfectant in households and family members (including the primary case) wearing a face mask at home before the primary case developed the illness. The authors motivate their findings for wearing face masks early one by the fact that the viral load is highest in the 2 days before symptom onset and on the first day of symptoms, and up to 44% of transmission is during the pre-symptomatic period.

Finally, (49) use a simulation study to assess the role of face masks on the epidemic spread with or without other public health measures being simultaneously in place. Their findings indicate that that face masks can effectively mitigate the epidemic spread if they are used by the public all the time (not just from when symptoms first appear). The simulated effects are the greatest when the adoption rate of wearing face masks in the public is 100 percent and when it is combined with an early lock-down situation. When interpreting their simulation results, the authors stress that accurate experimental evidence for potential control interventions would be needed to fully evaluate the effect of face masks.

[24] Although the authors compare their findings for Italy and New York City with global Covid-19 trends in the world and in the United States, the lack of a suitable comparison groups cannot rule out that some unobserved factors in Italy and New York City other than the introduction of face masks have driven the estimated trend reversal.

E Incremental difference-in-difference estimates for the timing of treatment effects

One difficulty in the empirical identification of treatment effects of face masks relates to the fact that Jena has introduced several public health measures to fight the local spread of Covid-19 in rapid procession over time. An overview is in Figure 8 above. We emphasized that some of these measures in Jena (light colors) deviate from their general introduction at the federal state level (dark colors). These anticipated measures may be taken as a signal for the severity of the pandemic and may, accordingly, have induced behavioral changes of the local population even before face masks became compulsory. To test for the strength of such dynamic treatment effects over time, we complement our SCM approach by conducting incremental difference-in-difference (IDiD) estimation (50; see 51, for a general discussion of the use of difference-in-difference estimation to identify causal effects of Covid-19 policies).

We discuss significant additional measures taken by the local health authorities in Jena to suppress the spread of Covid-19 in the main text. These additional measures started on March 11 with the leading closure of restaurants, sport and fitness centers etc. We use this date to define a baseline treatment dummy, which takes a value of one for Jena from March 14 onwards and is zero before that day. We include this treatment dummy in a fixed effect (FE) regression model, which uses the (log-transformed) cumulative number of Covid-19 cases as outcome variable. Starting from this baseline treatment specification, we run a series of regressions, which add a second treatment dummy to the model. The latter takes a value of one for Jena from day *m* onwards and is zero before that day. We allow *m* to vary between March 15 and April 25. The overall sample length is set to May 6.

The main idea of the proposed IDiD approach is to see whether we observe a general treatment effect with the start of public health measures on March 14. On top, we see since when we potentially observe an additional effect, which relates to specific public measures introduced during the time interval. Again, as outlined in appendix A.3, we need to account for the time lag resulting from an incubation period and a reporting lag to health authorities. Formally the *m*-th equation for the set of *m*=(1,…,*M*) regressions thereby takes the following form

$$covid_{i,t} = \beta \times \Delta covid_{i,t-1} + \gamma \times base_{i,t} + \delta_m \times add_{i,t}^m + D_{Weekday} + \mu_i + \Psi_{k(t)} + e_{i,t}$$

where $covid_{i,t}$ denotes the (log-transformed) cumulative number of registered Covid-19 cases in municipal district *i* at day *t* with $i = 1, \dots, N$ and $t = 1, \dots, T$. $\Delta covid_{i,t}$ is the number of newly registered Covid-19 cases at day *t-1*. $base_{i,t}$ refers to the baseline treatment dummy and $add_{i,t}^m$ is the additional treatment dummy from day *m* onwards. Further, μ_i are municipal district-fixed effects, $D_{Weekday}$ is a set of binary dummies for the different days of the week and $\Psi_{k(t)}$ are time-fixed effects for each with *k=1,…,K* calendar week in the sample period. $e_{i,t}$ denotes the model's *i.i.d.* error term. We are mostly interested in estimating γ and δ_m, which sum up to the overall treatment effect of public health measures in Jena taken from March 14 onwards.

We estimated the FE-based IDiD model by means of weighted least square (WLS), where weights are generated from a first step Probit regression with $base_{i,t}$ as the outcome variable. We estimate the Probit model as a cross-sectional specification for March 14 and includes values of newly registered Covid-19 cases before March 14 as well as the set of structural regional characteristics as shown in Table 1 in appendix B.2. Hence, in analogy to our SCM approach, the main idea for our two-step approach is to give those control regions a larger sample weight, which have similar characteristics as Jena before the baseline treatment starts (51). This may overcome the problem of heteroscedasticity associated with difference-in-difference estimation if there are very few (or even only one) treatment group (see 52 for a general discussion of inference in DiD models with few treated groups and heteroscedasticity). The resulting two-step estimator is also known as conditional difference-in-difference estimator (53).

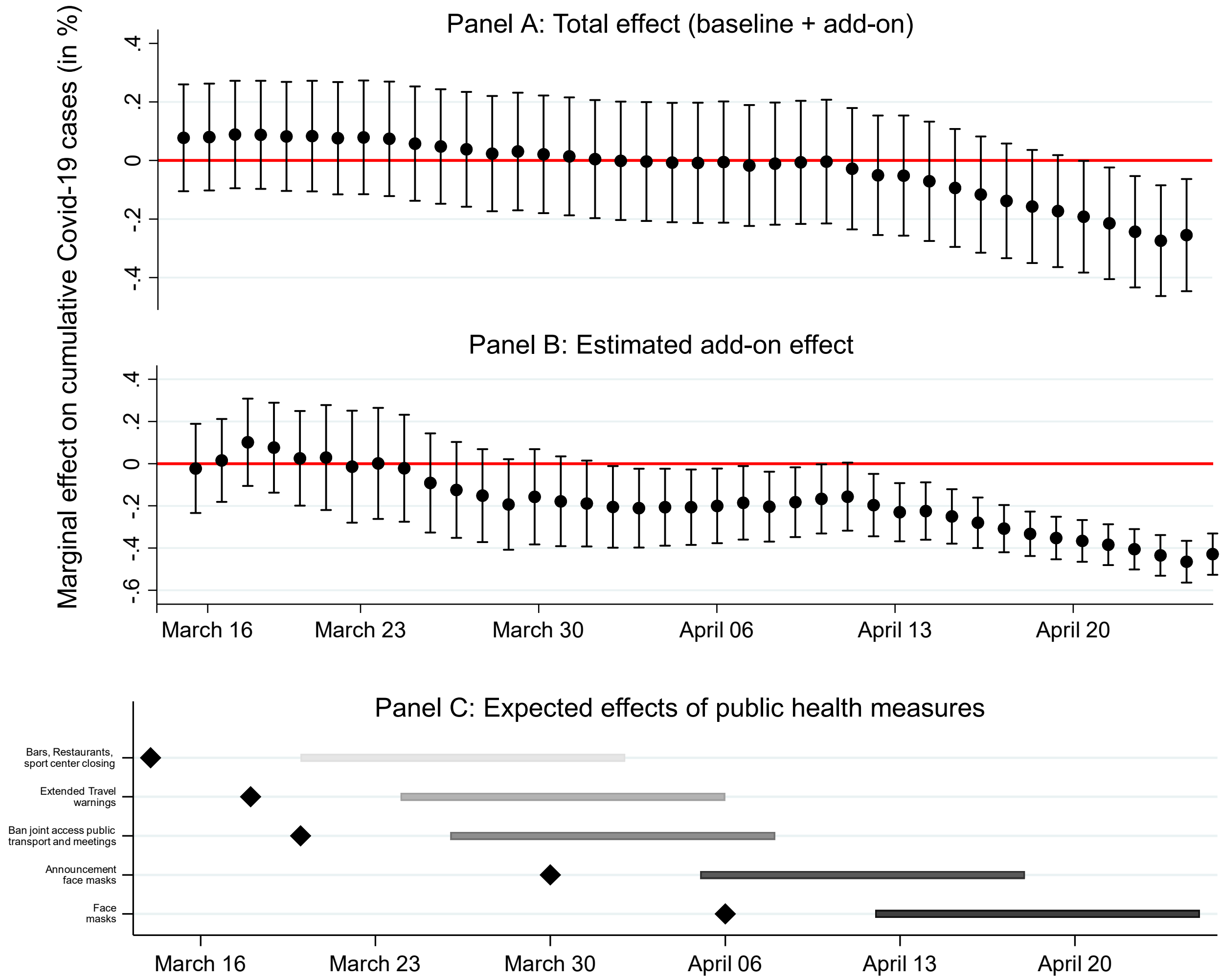

Figure 20: Estimated effects from incremental difference-in-difference (IDiD) model

Notes: We calculate point estimates and standard errors for the total treatment effect $(\gamma + \delta_m)$ on the basis of the Delta method. In Panel A and Panel B solid lined indicate 95% confidence intervals for reported point estimates. Standard errors in the FE-model are clustered at the municipal district level. In Panel C markers indicate the start of a specific public health measures; bars indicate the range of expected effects taking an incubation period and reporting delay into account.

Figure 20 shows the second-step IDiD regression results for the total treatment effect $(\gamma + \delta_m)$ in Panel A and the add-on treatment effect (δ_m) in Panel B. Panel C shows the expected timing of effects for different public health measures if we consider a total delay *D* of 19 days for the incubation period and an associated reporting lag. Estimations are based on a sample of 20 regions (19 controls with positive sample weights plus Jena) during the sample period January 28 until May 6 (with a total number of 1,980 observations). We find that the total treatment effect for public health measures in Jena relative to the control group only becomes significant roughly two weeks after the introduction of face masks on April 6. This strongly overlaps with expected effects stemming from the announcement and introduction of compulsory face masks in Jena (as shown in Panel C). In terms of the magnitude of the effect, we find a reduction in the cumulative number of Covid-19 cases by roughly 20%. Both findings are in line with our baseline SCM approach.

While Panel B of Figure 20 shows that we find marginally significant add-on effects from early April on, their magnitude is not sufficient to translate into a significant reduction in the number of Covid-19 cases vis-à-vis the set of control regions. Only from April 13 onwards, thus roughly one week after the introduction of face masks, the add-on treatment effect becomes gradually stronger in magnitude and statistically significant. If we resort to the total delay *D* as estimated in appendix A.3, this result further supports our SCM findings that the relative reduction in the cumulative Covid-19 cases is mainly attributable to the announcement/introducing face masks.

Table 15: Control regions included in the IDiD estimation

ID	**NUTS3 region**
2000	Hamburg
3101	Braunschweig
3102	Salzgitter
3103	Wolfsburg
5315	Köln
5515	Münster
6411	Darmstadt
6412	Frankfurt am Main
7315	Mainz
8111	Stuttgart
8212	Karlsruhe
8221	Heidelberg
8222	Mannheim
9161	Ingolstadt
9562	Erlangen
14511	Chemnitz
14612	Dresden
14713	Leipzig
16051	Erfurt

Notes: Selection of regions is based on Probit regression with the baseline treatment dummy in Jena on March 14 as outcome variable (see text in this appendix for details). In the FE-specification reported in Figure 20, we have set sample weights for selected control regions equal to one; alternative specifications with changing weights deliver very similar results and are not explicitly reported here (regression outputs can be obtained from the authors).

References for the appendix

1. N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-M. Jung, B. Yuan, R. Kinoshita, H. Nishiura, Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *Journal of clinical medicine* **9** **(**2020).
2. S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, J. Lessler, The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of internal medicine* **172**, 577–582 (2020).
3. Robert Koch Institute, Covid-19 Infektionen, General Website (NPGEO Corona Hub): https://npgeo-corona-npgeo-de.hub.arcgis.com/.
4. U. Buchholz, S. Buda, K. Prahm, Abrupter Rückgang der Raten an Atemwegserkrankungen in der deutschen Bevölkerung. *Epidemiologisches Bulletin* **2020**, 7–9 (2020).
5. Donsimoni, J. R., R. Glawion, B. Plachter & K. Wälde, Projecting the Spread of COVID-19 for Germany. *German Economic Review* **21**, 181–216 (2020).
6. Donsimoni, J. R., R. Glawion, B. Plachter, K. Wälde & C. Weiser, Should Contact Bans Be Lifted in Germany? A Quantitative Prediction of Its Effects. *CESifo Economic Studies (*2020).
7. S. Khailaie, T. Mitra, A. Bandyopadhyay, M. Schips, P. Mascheroni, P. Vanella, B. Lange, S. Binder, M. Meyer-Hermann, *Estimate of the development of the epidemic reproduction number Rt from Coronavirus SARS-CoV-2 case data and implications for political measures based on prognostics* (2020).
8. T. Harko, F. S.N. Lobo, M. K. Mak, Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Applied Mathematics and Computation* **236**, 184–194 (2014).
9. A. A. Toda, Susceptible-infected-recovered (SIR) dynamics of Covid-19 and economic impact. *Covid Economics*, 43–63 (2020).
10. A. Abadie, Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *[Article prepared for the] Journal of Economic Literature (*2019).
11. A. Abadie, J. Gardeazabal, The Economic Costs of Conflict: A Case Study of the Basque Country. *American Economic Review* **93**, 113–132 (2003).
12. S. O. Becker, S. Heblich, D. Sturm, The impact of public employment: Evidence from Bonn.
13. S. Athey, G. W. Imbens, The State of Applied Econometrics: Causality and Policy Evaluation. *Journal of Economic Perspectives* **31**, 3–32 (2017).
14. N. Kreif, R. Grieve, D. Hangartner, A. J. Turner, S. Nikolova, M. Sutton, Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units. *Health economics* **25**, 1514–1528 (2016).
15. H. Pieters, D. Curzi, A. Olper, J. Swinnen, Effect of democratic reforms on child mortality: a synthetic control analysis. *The Lancet Global Health* **4**, e627-e632 (2016).

16. A. Friedson, D. McNichols, J.J. Sabia, D. Dave, Did California's Shelter-In-Place Order Work? Early Coronavirus-Related Public Health Effects. *IZA Discussion Paper No. 13160 (*2020).

17. A. Abadie, A. Diamond, J. Hainmueller, Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association* **105**, 493–505 (2010).

18. K. Wälde, Corona-Blog, Einschätzung vom Freitag, 20. März 2020: https://www.macro.economics.uni-mainz.de/corona-blog/.

19. E. Cavallo, S. Galiani, I. Noy, J. Pantano, Catastrophic Natural Disasters and Economic Growth. *Review of Economics and Statistics* **95**, 1549–1561 (2013).

20. N. F. Campos, F. Coricelli, L. Moretti, Institutional integration and economic growth in Europe. *Journal of Monetary Economics* **103**, 88–104 (2019).

21. P. Eliason, B. Lutz, Can fiscal rules constrain the size of government? An analysis of the "crown jewel" of tax and expenditure limitations. *Journal of public economics* **166**, 115–144 (2018).

22. L. Hu, R. Kaestner, B. Mazumder, S. Miller, A. Wong, The Effect of the Affordable Care Act Medicaid Expansions on Financial Wellbeing. *Journal of public economics* **163**, 99–112 (2018).

23. S. Galiani, B. Quistorff, The Synth_Runner Package: Utilities to Automate Synthetic Control Estimation Using Synth. *The Stata Journal* **17**, 834–849 (2017).

24. Alberto Abadie & Alexis Diamond & Jens Hainmueller, SYNTH: Stata module to implement Synthetic Control Methods for Comparative Case Studies: Boston College Department of Economics, revised 09 May 2020. *Statistical Software Components (*2011).

25. S. Firpo, V. Possebom, Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets. *Journal of Causal Inference* **6 (**2018).

26. W. Bock, B. Adamik, M. Bawiec, V. Bezborodov, M. Bodych, J. P. Burgard, T. Goetz, T. Krueger, A. Migalska, B. Pabjan, T. Ozanski, E. Rafajlowicz, W. Rafajlowicz, E. Skubalska-Rafajlowicz, S. Ryfczynska, E. Szczurek, P. Szymanski, *Mitigation and herd immunity strategy for COVID-19 is likely to fail* (2020).

27. M. V. Barbarossa, J. Fuhrmann, J. H. Meinke, S. Krieg, H. V. Varma, N. Castelletti, T. Lippert, *The impact of current and future control measures on the spread of COVID-19 in Germany* (2020).

28. G. Bonaccorsi, F. Pierri, M. Cinelli, A. Flori, A. Galeazzi, F. Porcelli, A. L. Schmidt, C. M. Valensise, A. Scala, W. Quattrociocchi, F. Pammolli, Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 15530–15535 (2020).

29. N. G. Davies, A. J. Kucharski, R. M. Eggo, A. Gimma, W. J. Edmunds, T. Jombart, K. O'Reilly, A. Endo, J. Hellewell, E. S. Nightingale, B. J. Quilty, C. I. Jarvis, T. W. Russell, P. Klepac, N. I. Bosse, S. Funk, S. Abbott, G. F. Medley, H. Gibbs, C. A. B. Pearson, S. Flasche, M. Jit, S. Clifford, K. Prem, C. Diamond, J. Emery, A. K. Deol, S. R. Procter, K. van Zandvoort, Y. F. Sun, J. D. Munday, A. Rosello, M. Auzenbergs, G. Knight, R. M. G. J. Houben, Y. Liu, Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for

hospital services in the UK: a modelling study. *The Lancet Public Health* **5**, e375-e385 (2020).

30. J. Dehning, J. Zierenberg, F. P. Spitzner, M. Wibral, J. P. Neto, M. Wilczek, V. Priesemann, Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science (New York, N.Y.)*. 10.1126/science.abb9789 (2020).

31. C. Gros, R. Valenti, L. Schneider, K. Valenti, D. Gros, *Containment efficiency and control strategies for the Corona pandemic costs* (2020).

32. T. Hartl, K. Wälde, E. Weber, Measuring the impact of the German public shutdown on the spread of Covid-19. *Covid Economics* **2020**, 25–32.

33. T. Hartl, E. Weber, Welche Maßnahmen brachten Corona unter Kontrolle?

34. Chen, X. & Z. Qiu, Scenario analysis of non-pharmaceutical interventions on global Covid-19 transmissions. *Covid Economics*, 46–67 (2020).

35. A. Ritschl, Visualising and forecasting Covid-19. *Covid Economics* **2020**, 79–105.

36. Barro, Robert J. and Ursua, José F and Weng, Joanna, The Coronavirus and the Great Influenza Epidemic - Lessons from the 'Spanish Flu' for the Coronavirus's Potential Effects on Mortality and Economic Activity.

37. A. Wilder-Smith, C. J. Chiew, V. J. Lee, Can we contain the COVID-19 outbreak with the same measures as for SARS? *The Lancet. Infectious Diseases*. 10.1016/S1473-3099(20)30129-8 (2020).

38. R. M. Anderson, H. Heesterbeek, D. Klinkenberg, T. D. Hollingsworth, How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet* **395**, 931–934 (2020).

39. N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. Okell, S. van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. Walker, P. Winskill, C. Whittaker, C. Donnelly, S. Riley, A. Ghani, *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand* (Imperial College London, 2020).

40. Born, B., A. Dietrich & G. Müller, *Do lockdowns work? A counterfactual for Sweden.: CEPR Discussion Paper No. 14744, Centre for Economic Policy Research* (2020).

41. K. A. Prather, C. C. Wang, R. T. Schooley, Reducing transmission of SARS-CoV-2. *Science*. 10.1126/science.abc6197 (2020).

42. D. K. Chu, E. A. Akl, S. Duda, K. Solo, S. Yaacoub, H. J. Schünemann, A. El-harakeh, A. Bognanni, T. Lotfi, M. Loeb, A. Hajizadeh, A. Bak, A. Izcovich, C. A. Cuello-Garcia, C. Chen, D. J. Harris, E. Borowiack, F. Chamseddine, F. Schünemann, G. P. Morgano, G. E. U. Muti Schünemann, G. Chen, H. Zhao, I. Neumann, J. Chan, J. Khabsa, L. Hneiny, L. Harrison, M. Smith, N. Rizk, P. Giorgi Rossi, P. AbiHanna, R. El-khoury, R. Stalteri, T. Baldeh, T. Piggott, Y. Zhang, Z. Saad, A. Khamis, M. Reinap, Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet* **395**, 1973–1987 (2020).

43. J. Howard, A. Huang, Z. Li, Z. Tufekci, V. Zdimal, H.-M. van der Westhuizen, A. von Delft, A. Price, L. Fridman, L.-H. Tang, V. Tang, G. L. Watson, C. E. Bax, R. Shaikh, F. Questier, D. Hernandez, L. F. Chu, C. M. Ramirez, A. W. Rimoin, Face Masks Against COVID-19: An Evidence Review. 10.20944/preprints202004.0203.v1 (2020).

44. T. Greenhalgh, M. B. Schmid, T. Czypionka, D. Bassler, L. Gruer, Face masks for the public during the covid-19 crisis. *BMJ (Clinical research ed.)* **369**, m1435 (2020).

45. K. M. Marasinghe, *Concerns around public health recommendations on face mask use among individuals who are not medically diagnosed with COVID-19 supported by a systematic review search for evidence* (2020).

46. N. H. L. Leung, D. K. W. Chu, E. Y. C. Shiu, K.-H. Chan, J. J. McDevitt, B. J. P. Hau, H.-L. Yen, Y. Li, D. K. M. Ip, J. S. M. Peiris, W.-H. Seto, G. M. Leung, D. K. Milton, B. J. Cowling, Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nature medicine* **26**, 676–680 (2020).

47. R. Zhang, Y. Li, A. L. Zhang, Y. Wang, M. J. Molina, Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 14857–14863 (2020).

48. Y. Wang, H. Tian, L. Zhang, M. Zhang, D. Guo, W. Wu, X. Zhang, G. L. Kan, L. Jia, D. Huo, B. Liu, X. Wang, Y. Sun, Q. Wang, P. Yang, C. R. MacIntyre, Reduction of secondary transmission of SARS-CoV-2 in households by face mask use, disinfection and social distancing: a cohort study in Beijing, China. *BMJ global health* **5 (**2020).

49. R. O. J. H. Stutt, R. Retkute, M. Bradley, C. A. Gilligan, J. Colvin, A modelling framework to assess the likely effectiveness of facemasks in combination with 'lock-down' in managing the COVID-19 pandemic. *Proc. R. Soc. A.* **476**, 20200376 (2020).

50. P. Dolton, C. R. Bondibene, J. Wadsworth, The UK National Minimum Wage in Retrospect*. *Fiscal Studies* **31**, 509–534 (2010).

51. A. Goodman-Bacon, J. Marcus, Using Difference-in-Differences to Identify Causal Effects of COVID-19 Policies. 10.18148/srm/2020.v14i2.7723 (2020).

52. B. Ferman, C. Pinto, Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity. *Review of Economics and Statistics* **101**, 452–467 (2019).

53. J. Heckman, H. Ichimura, J. Smith, P. Todd, Characterizing Selection Bias Using Experimental Data. *Econometrica* **66**, 1017 (1998).

Content