



Face masks considerably reduce COVID-19 cases in Germany

Timo Mitze^{a,1}, Reinhold Kosfeld^{b,1}, Johannes Rode^{c,1}, and Klaus Wälde^{d,e,1,2}

^aDepartment of Business and Economics, University of Southern Denmark, 5230 Odense, Denmark; ^bInstitute of Economics, Universität Kassel, 34127 Kassel, Germany; ^cDepartment of Law and Economics, Technische Universität Darmstadt, 64289 Darmstadt, Germany; ^dGutenberg School of Management and Economics, Johannes-Gutenberg-Universität Mainz, D-55131 Mainz, Germany; and ^eIZA Institute of Labor Economics, D-53113 Bonn, Germany

Edited by Andrea Rinaldo, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, and approved November 10, 2020 (received for review July 28, 2020)

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 shops. Depending on the region we consider, we find that face masks reduced the number of newly registered severe acute respiratory syndrome coronavirus 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%.

COVID-19 | public health measures | face masks | synthetic control method

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 of facial protection for clinical personnel is beyond dispute, even though there are many questions left open (1). There is also evidence that face masks helped in mitigating the spread of earlier epidemics such as SARS 2003 (severe acute respiratory syndrome 2003) or influenza (see *SI Appendix, section E* for a brief literature survey). The impact of face masks worn in public on the spread of COVID-19 has yet to be systematically analyzed. 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 stems mainly from the city of Jena. After face masks became mandatory between 1 April and 10 April 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 20 April and 29 April 2020 (see *SI Appendix, section A* for background).

We quantify the effectiveness of face masks by employing the synthetic control method (SCM; refs. 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 mandatory 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 averages 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 (susceptible–infected–removed) model and undertake an 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 d. Put simply, if the control region observes 100 new infections over a period of 20 d, the mask region observes only 25 cases. This drop is greatest, by more than 90%, for the age group 60 y and above. Our results are robust to different sensitivity checks, among which are 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

Significance

Mitigating the spread of COVID-19 is the objective of most governments. 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 severe acute respiratory syndrome coronavirus 2 infections for Germany. As face masks became mandatory at different points in time across German regions, we can compare the rise in infections in regions with masks and regions without masks. Weighing various estimates, we conclude that 20 d 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.

Author contributions: T.M., R.K., J.R., and K.W. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

¹T.M., R.K., J.R., and K.W. contributed equally to this work.

²To whom correspondence may be addressed. Email: waelde@uni-mainz.de.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2015954117/-DCSupplemental>.

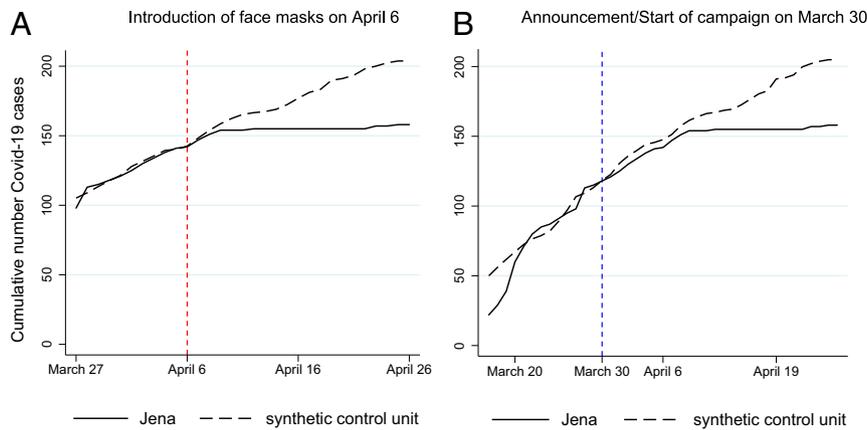


Fig. 1. Treatment effects of the mandatory introduction of face masks (6 April) and of its announcement (30 March) in Jena. The figure shows the development of the cumulative number of COVID-19 cases in Jena (treated region, black solid line) compared to a synthetic control group (gray dashed line) over time. (Details on the construction of the synthetic control group and SCM estimation are given in *Method and Data*.) Both panels distinguish between a pretreatment and treatment period. In *A*, the treatment period starts on 6 April when face masks became mandatory in public transport and shops. The start of the treatment period is indicated by the dashed vertical line (red). In *B*, the treatment period is set to begin on 30 March as starting date of the local campaign in Jena to wear face masks in public. The start of the treatment period is indicated by the dashed vertical line (blue). The panels show that the conclusion is independent of the starting dates: Face masks strongly reduced the number of COVID-19 cases in Jena.

on the spread of COVID-19 for all regions that introduced masks by 22 April (~8% of all German regions). Although the estimated average treatment effect is smaller compared to the one found for Jena, it is still statistically 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 (*SI Appendix, section D.2*), 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.*

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 (*SI Appendix, section E*). While there is a growing consensus from clinical studies that face masks significantly reduce the transmission risk of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and COVID-19 (7, 10), nonclinical evidence on the effectiveness of face masks is still largely missing. Ref. 11 surveys evidence on the population impacts of a widespread community mask use and stresses that “no randomized control trials on the use of masks . . . has been published.” The study which is “the most relevant paper” for ref. 11 is one that analyzed “exhaled breath and coughs of children and adults with acute respiratory illness” (ref. 12, p. 676), that is, used a clinical setting. Concerning the effect of masks on community transmissions, the survey needs to rely on pre-COVID-19 studies.

Ref. 13 is among the first to estimate the population impact of face masks on SARS-CoV-2 infections.† The authors track the development of COVID-19 in three pandemic epicenters, Wuhan, Italy, and New York City, between 23 January and 9 May 2020 and find sizable mitigating effects of face masks on epidemic spread. While their study offers important insights into

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 ref. 4 and provide causal evidence identifying the population impact of mandatory face masks on the spread of COVID-19.

Results: The Effects of Face Masks on the Spread of COVID-19

All results are obtained by applying the synthetic control method. It is described briefly in *Method and Data* and in more detail in *SI Appendix, section B*.

Results for Jena. Face masks became mandatory in Jena in three steps between 1 and 10 April. 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 6 April (see *SI Appendix, section A* for detailed information). We therefore center our discussion on this date.

Fig. 1*A* shows the SCM results for the introduction of face masks in Jena on 6 April. The visual inspection of the development of cumulative COVID-19 cases shows that the trend development of the synthetic control group is very similar to Jena before the treatment, indicating a good fit.‡ The difference in the cumulated registered COVID-19 cases between Jena and its corresponding synthetic control group after the start of the treatment on 6 April can be interpreted as the treatment effect on the treated [see *SI Appendix, section C.3* for (post)estimation details].

Fig. 1*A* 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 d after the start of the treatment (6 April) 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 *SI Appendix, section D.2* for computational details and an overview of all

*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 ref. 6. Refs. 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. Earlier work includes ref. 8 that was recently extended by ref. 9. Ref. 9 finds that all face covers (without an outlet valve) “reduce the front flow through jet by more than 90%.” As surgical and hand-made masks generally do not tightly fit, they generate backward and downward jets.

†Ref. 10 conducts a systematic review and meta-analysis. They do not report a study (see their table 1) that analyzes the entire population of a country.

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

measures). If we look at the estimated differences by age groups, *SI Appendix, Fig. S7* indicates that the largest effects occur for individuals above 60 y of age. Here the reduction in cumulative cases even exceeds 50%, which corresponds to a drop in newly registered SARS-CoV-2 infections by more than 90%. The significant drop can be explained by the introduction of face masks in elderly and nursing homes, which had already started on 1 April. 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 d from infection until registration (see *SI Appendix, section 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 played a role. As shown in *SI Appendix, section C.6.1*, online searches for (purchasing) face masks peaked on 22 April, when it was announced that face masks would become compulsory in all German federal states.[§] Another peak in online searches, almost as large (70% of the peak of 22 April), appeared on 31 March. This marks the date of the regulation making masks compulsory between 1 and 10 April in Jena. The regulation was accompanied by a campaign “*Jena zeigt Maske*,” communicating the necessity to wear face masks in public, that started on 30 March.[#]

Fig. 1*B* plots the estimated effect size when we define the start of the treatment period by the start of the campaign on 30 March. 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 ~10 to 12 d after the announcement and then grows considerably 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.

Robustness Checks. Obviously, the estimated difference in COVID-19 development in Jena vis-à-vis the synthetic Jena is only convincing if 1) the requirements of the SCM are fulfilled and 2) potentially concurrent policies other than masks can be ruled out. The role of the first set of robustness checks consists of understanding the sensitivity to the length of the preintervention period (*Cross-Validation Tests*) and to the composition of the control pool (*Changing the Donor Pool*).^{||} In a second step, we rule out unobserved macro effects shared by many regions (*Placebo-in-Space Tests*) and test for anticipation effects potentially caused by other public health measures (*Placebo-in-Time Tests*). The second step also comprises difference-in-difference estimation as a further test for latent concurrent policies that go beyond observable policies (displayed in *SI Appendix, Fig. S2*).

Cross-validation tests. We study the sensitivity of our estimates with respect to the length of the training and validation period before the start of the treatment. We accordingly alter the imposed lag structure for predictors that have a time dimension, that is, the number of cumulative and newly registered COVID-19 cases. The set of time-constant predictors is kept unchanged by this test. As shown in *SI Appendix, section C.7*, we do not find a systematic estimation bias of our baseline SCM specification compared to alternative ones with longer lag structures (up to 7 d) and accordingly shorter training periods. Given that regional COVID-19 cases developed very dynamically and nonlinearly in the pretreatment period, this is an important finding in terms of the robustness of our results.

We also test for the sensitivity of the estimated treatment effects to changes in the set of time-constant predictors. We do so by sequentially excluding individual variables from the set of predictors. As shown in *SI Appendix, section C.7*, the estimated trajectories for the respective synthetic control groups follow a very similar trend. All of them identify a reduction in the number of cumulative COVID-19 cases in Jena vis-à-vis the synthetic Jena that widens over time. We conclude from these results that using the full set of predictors is the most reasonable approach. **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 *SI Appendix, Table S5*). 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, *SI Appendix, section C.8* 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 macroregional trends) and to a subsample of larger cities (*kreisfreie Städte*). The latter comparison 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 6 April have nonetheless experienced a similar decline in the number of registered COVID-19 cases. If this had been the case, the treatment effect might have been driven by other latent factors rather than by face masks. Such latent factors may, for instance, be related to the macroregional dynamics of COVID-19 in Germany. Therefore, *SI Appendix, section C.9* reports pseudo-treatment effects for similarly sized cities in the federal state of Thuringia assuming that they had introduced face masks on 6 April—although, in fact, they did not. As the figure illustrates, these cities show either a significantly higher or a similar development of registered COVID-19 cases compared to their synthetic controls. This result provides further empirical support for a relevant effect in the case of Jena.

As a more comprehensive test, we run placebo-in-space tests for all other regions that did not introduce face masks on 6 April or closely afterward. Again, we estimate the same model on each untreated region, assuming it was treated at the same time as Jena. The empirical results in Fig. 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 Fig. 2*A* and the subsample of large cities (*kreisfreie Städte*) in Fig. 2*B*.

One advantage of these tests is that they allow us to conduct inference on the significance of the estimated treatment effects for Jena. Accordingly, Fig. 2*C* and *D* visualize the estimated treatment effects together with 90% confidence intervals. Intervals have been calculated on the basis of (one-sided) *P* values (pretreatment match quality adjusted) reported in Fig. 2*E* and *F*.^{**} 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 ref. 18). In both panels, the reported confidence intervals and underlying *P* values indicate that the reduction in

[§]For a German-wide news report see, for example, ref. 14.

[#]See local newspaper reports, for instance ref. 15.

^{||}Further requirements, that are less central to our application, are listed in *Method and Data* and are discussed in *SI Appendix, section B*.

^{**}We follow the method proposed in ref. 16 to calculate confidence intervals from *P* values. As pointed out in ref. 17, 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.

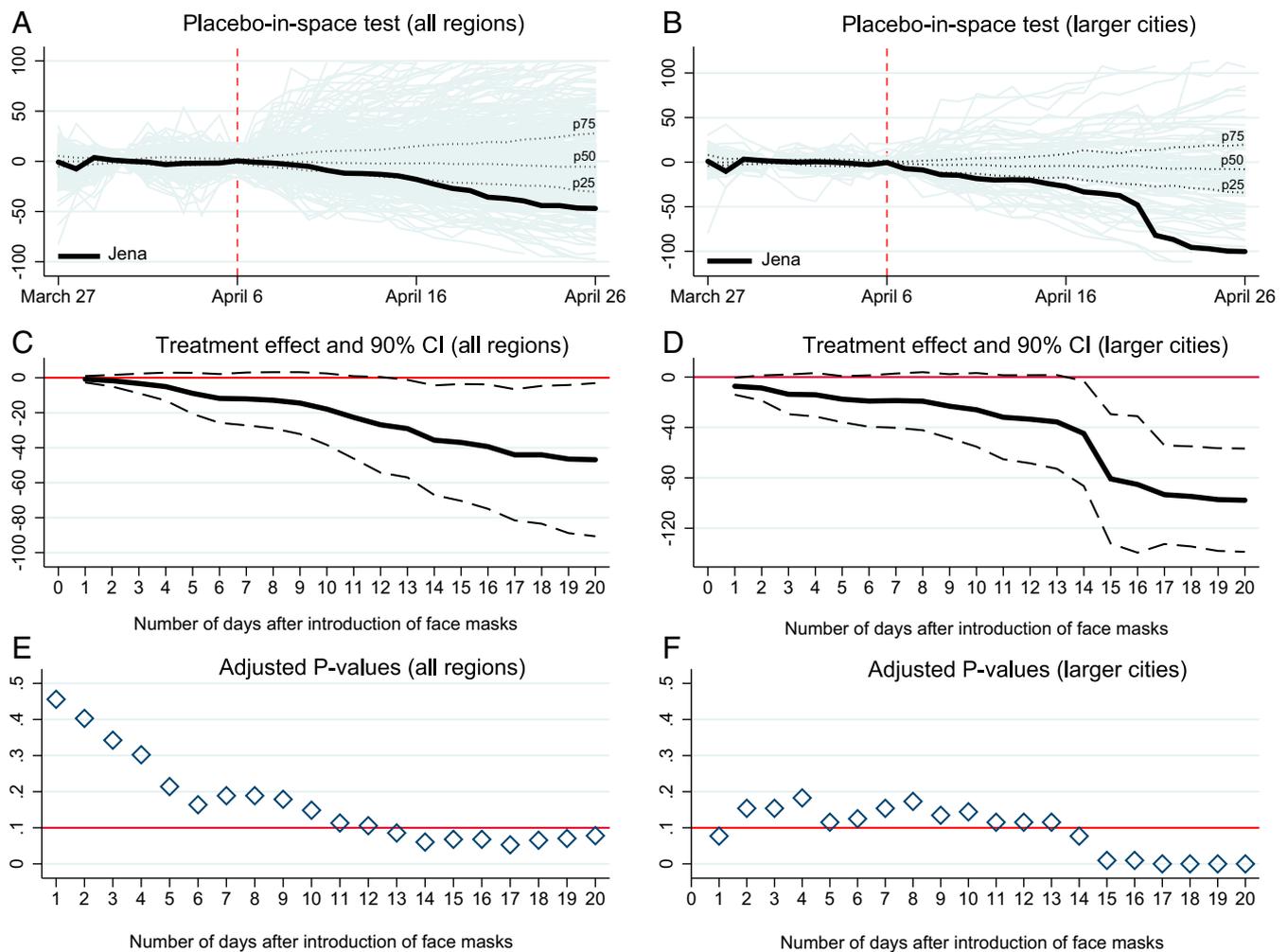


Fig. 2. Comprehensive placebo-in-space tests for the effect of face masks on COVID-19 cases. The figure compares the difference in the cumulative number of COVID-19 cases between Jena and its synthetic control group with the differences between all other regions in the donor pool and their respective controls over time. Differences have been calculated for the treatment period after 6 April when Jena de facto introduced mandatory face masks. For all other regions, the reported differences in the treatment period constitute pseudo-treatment effects since these regions had not introduced face masks back then. In *A*, the black solid line measures the treatment effect for Jena. The solid gray lines measure the pseudo-treatment effects for all other German regions in the donor pool (see *Method and Data* for details on the specification of the donor pool). Plots are shown for donor regions with an acceptable pretreatment root-mean-square prediction error (RMSPE), that is, less than 10 times the RMSPE of Jena; dotted lines indicate the median (p50), 25th percentile (p25), and 75th percentile (p75) of pseudo-treatment effects. In *B*, the donor pool is reduced to comprise only larger cities (*kreisfreie Städte*). *C* and *D* also plot the treatment effect for Jena and 90% confidence intervals (gray dashed lines) for the full sample of regions and the subsample of larger cities, respectively. Confidence intervals are constructed on the basis of pseudo *P* values as shown in *E* and *F* for the first 20 d after the introduction of face masks in Jena. These *P* values are adjusted for the pretreatment match quality (see *Method and Data* for details). The red horizontal line in *E* and *F* indicates a threshold *P* value of 0.1.

the number of COVID-19 cases was not a random event. In larger cities (Fig. 2*D*), the reduction due to the introduction of face masks is clearly visible 2 wk 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.^{††}

^{††}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 at home more, 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 *SI Appendix, section C.6.2* show that federal states moved in a relatively coordinated way in this respect. Unfortunately, mobility data for Jena are not easily available.

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 treatment effects for Jena prior to the introduction of face masks on 6 April or its announcement on 30 March. 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 *SI Appendix, Fig. S2*, Jena and Thuringia passed at least 40 public health measures before the end of April 2020. Jena implemented 27 of these 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 home. Relevant regions included foreign countries but also other German federal states, among which were Bavaria, Baden-Wurttemberg, and North-Rhine Westphalia. Measures

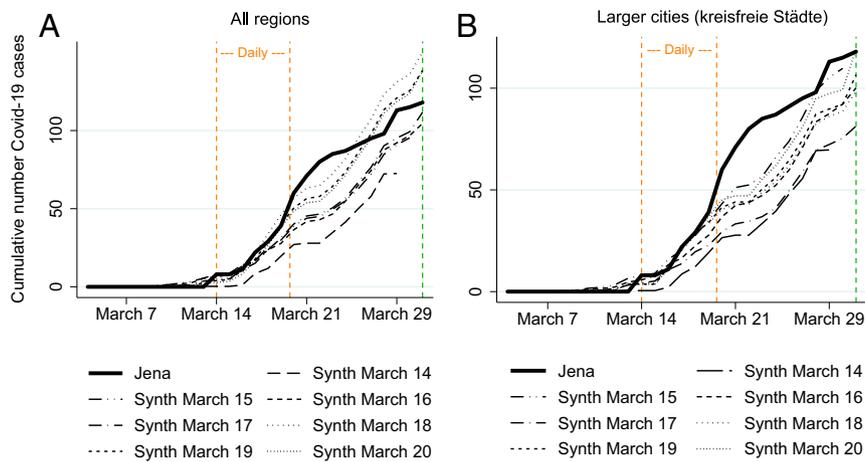


Fig. 3. Placebo-in-time tests for (pseudo) treatment effects in the period 14 to 20 March. The figure shows the empirical development of the cumulative number of COVID-19 cases in Jena (treated region, black solid line) and the estimated development in different synthetic control groups over time. The key difference between these synthetic control groups lies in the starting point of the (pseudo) treatment period. The starting point varies on a daily basis between 14 and 20 March. In *A*, the full sample of German regions is used as the donor pool for the construction of synthetic control groups (see *Method and Data* for details on the specification of the donor pool). Vertical dashed lines (orange) indicate the time corridor in which the respective (pseudo) treatment periods start. In *B*, the donor pool is reduced to comprise only larger cities (*kreisfreie Städte*). Again, vertical dashed lines (orange) indicate the time corridor in which the respective (pseudo) treatment periods start.

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 2 wk, though).

As these major health decrees were accompanied by smaller ones on an almost daily basis until 20 March, we run a series of SCM estimations using each day between 14 and 20 March as a (pseudo) treatment period.^{‡‡} The results for the full donor pool including all other German regions and the subsample of larger cities are shown in Fig. 3 *A* and *B*. Results are reported until 30 March when the mandatory introduction of face masks was announced.

To take an example, consider the graph with the long dashes in Fig. 3*A* denoted “Synth March 14.” It shows the cumulative number of COVID-19 cases for a synthetic control group for Jena extracted from the full donor pool under the assumption that some treatment had started on 14 March. Similarly, “Synth March 15” (and so on) show the evolution of COVID-19 cases for a starting date of 15 March. Hence, each SCM estimation starts at a different point in time.

The visual inspection of the relative development of COVID-19 cases in Jena vis-à-vis its synthetic Jena does not indicate a clear treatment effect in terms of reducing COVID-19 cases prior to 1 April. The results are particularly clear-cut for the sample of larger cities in Fig. 3*B*, indicating that earlier public health measures alone did not significantly suppress the number of COVID-19 cases in Jena in the first 2 wk after their introduction.

Difference-in-difference tests. The reported trajectories of synthetic Jena in Fig. 3*A* 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 policy interventions show measurable effects (a detailed description is given in *SI Appendix, section F*). 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 2 wk after the introduction of face masks on 6 April. If we resort to the estimated incubation and reporting lag as shown in *SI Appendix, section 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.

Results for Other Regions. Jena may be a unique case. We therefore also study treatment effects for other individual regions that introduced face masks earlier than other regions. Further single-unit treatment analyses are shown in *SI Appendix, section D.1*. SCM estimation for multiple treated units can be undertaken as not only did some individual regions introduce face masks earlier than their federal states (see Fig. 5, below the time line) but also some federal states (Saxony and Saxony-Anhalt) before the remaining German federal states (*SI Appendix, Fig. S1*). To ensure a sufficiently long treatment period, we consider all regions as treated which introduced face masks on or before 22 April. We estimate average treatment effects in a multiple treatment SCM approach for 1) all of these regions and 2) a subset consisting of larger cities (*kreisfreie Städte*) only. In the former case we have a total of 32 treated units, and in the latter there are 8 treated units. The donor pool of control regions (all regions in Germany and only larger cities, respectively) is specified such that the minimum time lag in the introduction of face masks between treated and control regions ranges between 5 and 13 d.

The results, visible in Fig. 4, point to a significant face mask effect in the reduction of SARS-CoV-2 infections over a period of 20 d 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 Fig. 4 *A* and *B*, respectively. The reported 90% confidence intervals in Fig. 4 *C* and *D* calculated on the basis of adjusted *P* values shown in Fig. 4 *E* and *F* indicate that the estimated treatment effects are not random for both samples. While treatment effects of face masks turn significant after roughly 1 wk 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 20 April.

^{‡‡}Alternatively, we have also tested for pseudo-treatment effects in Jena over a period of 20 d before the introduction of face masks. This period is equally split into a pre- and pseudo posttreatment period. As *SI Appendix, Fig. S11B* shows, there is no strong deviation from the path of the synthetic control group.

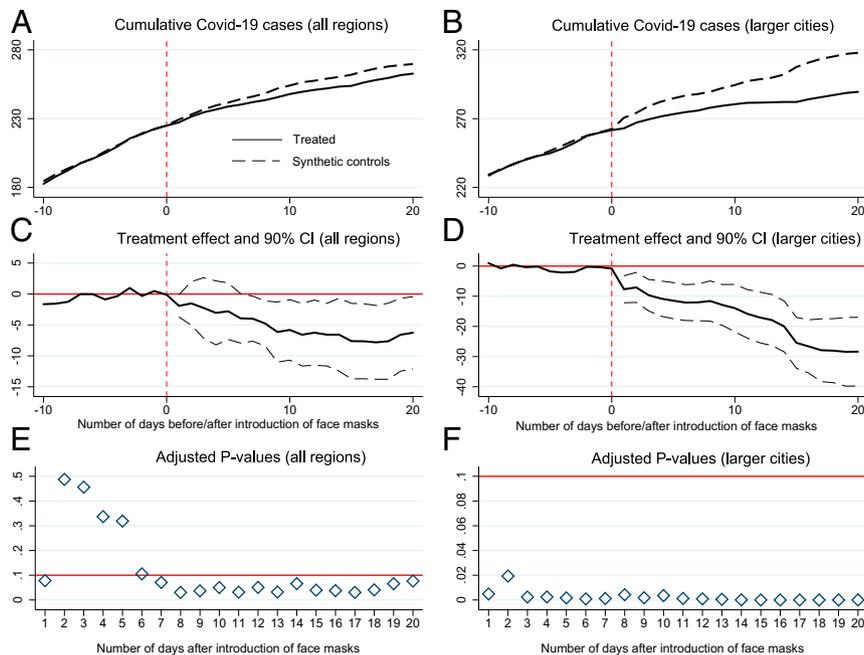


Fig. 4. Average treatment effects for the introduction of face masks with multiple treated units. The figure shows the average development of the cumulative number of COVID-19 cases for treated units (defined as all regions that introduced face masks on or before 22 April), for their synthetic control groups and estimated treatment effects. **A** plots the average number of cumulative COVID-19 cases in treated regions (black solid line) and synthetic controls (black dashed line) for the full sample of German regions in the donor pool (see *Method and Data* for details on the specification of the donor pool). The horizontal axis plots the number of days before and after the start of the treatment (mandatory introduction of face masks). The vertical dashed line (red color) indicates the start of the treatment period, which is allowed to vary by treated regions. In **B**, the number of treated and the donor pool for control regions is limited to larger cities (*kreisfreie Städte*). **C** and **D** plot the estimated average treatment effects, that is, average reduction in the cumulative number of COVID-19 cases (black solid lines) over time joint with 90% confidence intervals (gray dashed lines) for the two samples. Confidence intervals are constructed on the basis of pseudo *P* values as shown in **E** and **F** for the first 20 d after the start of the treatment. These (one-sided) *P* values are adjusted for the pretreatment match quality (see *Method and Data* for details). Inference has been conducted on the basis of a randomly drawn sample of 1 million placebo averages. The red horizontal line in **E** and **F** indicates a threshold *P* value of 0.1.

Importantly, however, the trend development for larger cities as shown in Fig. 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 reopened. As Fig. 4B highlights, such an avoidance effect may be particularly important in larger cities with higher population density and accordingly higher intensity of social interaction^{§§}

Taken together, over a period of 20 d, 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 11 May 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 cases. 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 *SI Appendix, section D.2* for an overview of all measures and *Method and Data* 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.

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 6 April reduced the number of newly registered COVID-19 cases over the next 20 d 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 implicitly assume that compliance to rules in Germany is sufficiently homogenous. Some field observations in this respect would be very useful, especially across federal states in Germany and worldwide. Ref. 19 reports that compliance for distancing rules rises when masks are worn. As a first guess and assuming a compliance of 100% in our treated regions, one would expect that a reduction in compliance by *x*% of the population leads to a reduction of the effects of masks by *x*%.

^{§§}This is perfectly in line with ref. 7 given the reduction in aerosols and droplets via using masks.

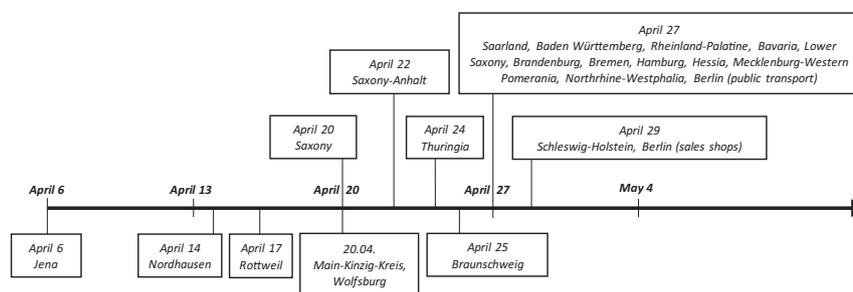


Fig. 5. The timing of mandatory face mask wearing in German federal states (*Top*) and individual regions (*Bottom*). The figure shows the regional variation in the introduction of face masks in public transport and shops over time. While text boxes above the timeline on the horizontal axis indicate the timing when the wearing of face masks became compulsory in the respective federal states (NUTS1 level), text boxes below the timeline identify individual NUTS3 regions that have anteceded the general introduction of face masks at the federal state level. The first NUTS3 region that introduced mandatory face masks in Germany was Jena on 6 April. By 29 April face masks had become mandatory in all German regions. (See *SI Appendix*, section A.1 for more background.)

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 posttreatment 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 in *Method and Data*, Jena is a fairly representative region of Germany in terms of COVID-19 cases. Third, the smaller treatment effects observed in 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 at home more, etc.) more seriously.

Concerning the point in time (or better, the point in 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 to 3%. These are low growth rates compared to the early days of the epidemic in Germany, where daily growth rates lay above 50% (20). 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 6 April, the general trend in 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 (which relate to compliance), 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. Fourth, there are various types of face masks. We cannot identify differential effects since mask regulations in German regions do not require a certain type. Finally, economic costs should be taken into account. When we compare masks with other common measures,^{***} the implied economic costs for community masks seem comparatively low. This applies to disposable masks and reusable non-medical masks. Yet, the cost of information campaigns should be taken into account. While a detailed cost–benefit analysis is needed, we would expect that a comparison with other policy actions would speak in favor of face masks. [See *SI Appendix*, section E.3 for a brief introduction to the literature. We estimate that costs (of households only) for face masks amount to 1.4 to 2.5% of disposable income.]

Method and Data

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

Statistical Method. We estimate treatment effects from introducing face masks by means of the SCM for single and multiple treated units. (All analyses are undertaken in Stata. The corresponding files plus data are available in *SI Appendix*.) The SCM has become increasingly popular for policy evaluations that rely on comparative case studies. The intuition of SCM estimation can briefly be described as follows (see ref. 2–5 and *SI Appendix*, section B for more background).

First, the method can be applied to policy interventions (thereafter treatment) which only target a small number of treated units (in our case, one or a few regions). Treatment effects are identified by comparing the development of outcomes in treated and control regions during the treatment period. For causal inference, the proper selection and weighting of control regions is crucial to credibly estimate what outcome would have been observed in the treated region in the absence of the treatment. To establish this counterfactual, the SCM approach constructs a synthetic control group as weighted average of regions in the donor pool of controls. Weights for individual control regions vary between 0 and 1, sum up to 1 over all control regions, and are determined on the basis of a minimum distance approach. The latter involves a set of predictor variables chosen by the researcher as to closely match the outcome of the treated region prior to the treatment (predictors used here are listed in *SI Appendix*, Table S4 and selected weights for control regions are displayed in *SI Appendix*, Table S5).

Second, statistical significance of the estimated treatment effect is based on permutation. The SCM estimates a series of placebo treatment effects for all regions in the donor pool, that is, each region in the donor pool is treated

^{***}Common measures can be grouped inter alia into closures (of, e.g., restaurants or hotels, educational institutions or clubs, and leisure facilities), contact bans (for individuals, faith groups, or visits to hospitals or retirement homes), and contact rules (social distancing or quarantines after traveling).

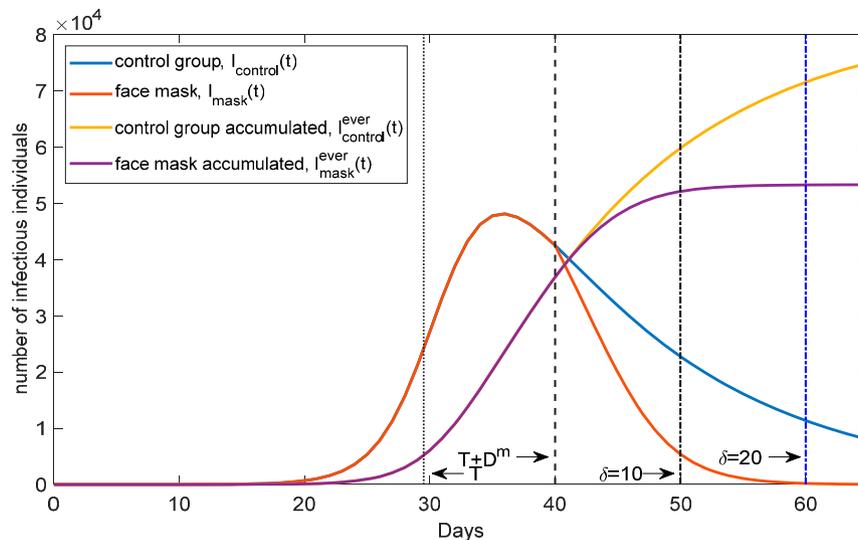


Fig. 6. Theoretical effects of face masks on the number of infectious individuals $I(t)$ and on the accumulated number of infectious individuals $I^{ever}(t)$. The horizontal axis plots time and the vertical axis the number of infectious individuals, both new cases (blue and red curves) and accumulated cases (yellow and purple curves). The curves show time paths following from a standard SIR model. We let an intervention take place on day $T = 29$. After a delay of D^m days, where D^m is the median of the delay between infection and reporting, the effect of the intervention is visible. Waiting 10 or 20 d then allows us to quantify the effect of the intervention.

as if it had been treated. The distribution of placebo effects is then compared with the treatment effect for the treated region. If the magnitude of the latter effect is large relative to the distribution of the placebo effects, the treatment effect is considered not to be observed by chance, that is, it is deemed to be significant.

Third, the effective use of SCM relies on contextual requirements: 1) A donor pool of controls is available, that is, not all regions receive the treatment during the period of the study; 2) predictor values of the treated region are not extreme relative to those of controls, that is, the treated region lies in the convex hull of control regions; 3) data are available for a sufficiently long time horizon before and after the start of the treatment; 4) spillover effects of the treatment on controls are absent; and 5) there are no early anticipation effects. Implications of these requirements for our analysis are discussed in *SI Appendix, section B* and we apply a series of robustness tests to check if the requirements hold in our data settings (reported in *Robustness Checks*).

Finally, as we also employ DiD estimation, we briefly touch upon the relationship between these two approaches. Similar to SCM, the DiD approach estimates treatment effects by contrasting changes in outcomes between a pretreatment and treatment period for treated and nontreated (control) regions. An attractive feature of the DiD approach is its flexibility in estimating dynamic treatment effects, that is, those that build up over time potentially determined by (unobserved) interventions starting at different time periods. (This is the reason why we employed DiD above.) By contrast, DiD has its limits when only one or a few treated regions are available as heteroscedastic errors might occur (see ref. 21). DiD estimation also relies strongly on the validity of the parallel trends assumption: It requires that treated and control regions would have followed parallel trajectories over time if treatment had not occurred. Given the highly dynamic development of regional COVID-19 cases and the likely presence of macroregional trends, the validity of the common trend assumption is questionable in our data settings. For SCM estimation, we do not need to impose this assumption as the presence of common trends between treated and control regions is in itself a favorable factor for finding an appropriate counterfactual trajectory (2, 3, 22).

Data. We use the official German statistics on reported COVID-19 cases from the Robert Koch Institute (23). We build a balanced panel for 401 NUTS level 3 regions and 105 d spanning the period from 28 January to 11 May 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 variables. We estimate overall effects for these variables together with disaggregated effects by age groups (persons aged 15 to 34 y, 35 to 59 y, and 60+ y). We also employ regional

data to inter alia identify control regions. *SI Appendix, Table S4* 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 needs 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 *SI Appendix, sections A.1 and A.2* for details. The database indicates that our results indeed capture the effects of face masks and not of other public health measures.

Conceptual Background. To facilitate the interpretation of our findings, we employ a standard SIR model with three states: susceptible, infectious, and removed (see *SI Appendix, section A.4* for more details.) Imagine we study a region where face masks are not mandatory. The time path $I(t)$ of infectious individuals in this (synthetic) control group is displayed in Fig. 6 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 Fig. 6).^{†††} 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 d in *SI Appendix, section 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 purple curve here and the corresponding curves in Figs. 1A and 4A and B.

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 $T + D^m + \delta$, that is, δ days after face masks became effective. The difference between the control region and the face-mask region is given by $I_{control}^{ever}(T + D^m + \delta) - I_{mask}^{ever}(T + D^m + \delta)$. Hence, the introduction of face masks reduced the number of COVID-19 cases by

^{†††}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 had already been declining when face masks became mandatory. Numerical solutions are computed in MATLAB. The code is available in *SI Appendix*.

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

[1]

This equation produces the numbers we report to quantify the effects of face masks. *SI Appendix, section D.2* describes our measures based on daily growth rates.

Data Availability. Public health data have been deposited in FigShare (<https://doi.org/10.6084/m9.figshare.13065920>). All study data are included in the paper and *SI Appendix*.

- G. Seminara *et al.*, Biological fluid dynamics of airborne COVID-19 infection. *Rend. Lincei Sci. Fis. Nat.*, 10.1007/s12210-020-00938-2 (2020).
- A. Abadie, J. Gardeazabal, The economic costs of conflict: A case study of the Basque Country. *Am. Econ. Rev.* **93**, 113–132 (2003).
- A. Abadie, A. Diamond, J. Hainmueller, Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J. Am. Stat. Assoc.* **105**, 493–505 (2010).
- A. Abadie, Using synthetic controls: Feasibility, data requirements, and methodological aspects. *J. Econ. Lit.*, 10.1257/jel.20191450.
- C. A. W. Bruhn *et al.*, Estimating the population-level impact of vaccines using synthetic controls. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1524–1529 (2017).
- C. J. Kähler, R. Hain, Fundamental protective mechanisms of face masks against droplet infections. *J. Aerosol Sci.* **148**, 105617 (2020).
- K. A. Prather, C. C. Wang, R. T. Schooley, Reducing transmission of SARS-CoV-2. *Science* **368**, 1422–1424 (2020).
- J. W. Tang, T. J. Liebner, B. A. Craven, G. S. Settles, A schlieren optical study of the human cough with and without wearing masks for aerosol infection control. *J. R. Soc. Interface* **6** (suppl. 6), S727–S736 (2009).
- I. M. Viola *et al.*, Face coverings, aerosol dispersion and mitigation of virus transmission risk. arXiv:2005.10720 (19 May 2020).
- D. K. Chu *et al.*; COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors, 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. *Lancet* **395**, 1973–1987 (2020).
- J. Howard *et al.*, Face masks against COVID-19: An evidence review. 10.20944/preprints202004.0203.v2 (30 October 2020).
- N. H. L. Leung *et al.*, Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat. Med.* **26**, 676–680 (2020).
- 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. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 14857–14863 (2020).
- Tagesschau, Maskenpflicht in allen Bundesländern. <https://www.tagesschau.de/inland/corona-maskenpflicht-103.html>. Accessed 14 November 2020.
- Jenaer Nachrichten, Jena zeigt Maske: Kampagne für Mundschutz startet. <https://www.jenaer-nachrichten.de/stadtleben/13069-jena-zeigt-maske-kampagne-f%C3%BCr-mundschutz-startet>. Accessed 14 November 2020.
- D. G. Altman, J. M. Bland, How to obtain the confidence interval from a P value. *BMJ* **343**, d2090 (2011).
- A. Abadie, A. Diamond, J. Hainmueller, Comparative politics and the synthetic control method. *Am. J. Pol. Sci.* **59**, 495–510 (2015).
- S. Galiani, B. Quistorff, The Synth_Runner package: Utilities to automate synthetic control estimation using synth. *Stata J.* **17**, 834–849 (2017).
- Y. Liu, Face masks increase compliance with physical distancing recommendations during the COVID-19 pandemic. <https://osf.io/db8sj/>. Accessed 4 June 2020.
- K. Wälde, "Einschätzung vom Freitag, 20. März 2020." *Der Corona-Blog* (2020). <https://www.macro.economics.uni-mainz.de/corona-blog/>. Accessed 4 June 2020.
- B. Ferman, C. Pinto, Inference in differences-in-differences with few treated groups and heteroskedasticity. *Rev. Econ. Stat.* **101**, 452–467 (2019).
- N. Kreif *et al.*, Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health Econ.* **25**, 1514–1528 (2016).
- Robert Koch Institute, Covid-19 Infektionen, NPGEO Corona Hub 2020. <https://npgeo-corona-npgeo-de.hub.arcgis.com/>. Accessed 4 June 2020.

Supplementary Information for

Face Masks Considerably Reduce Covid-19 Cases in Germany

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

November 13, 2020

Corresponding Author: Klaus Wälde

Email: waelde@uni-mainz.de

This PDF file includes:

Supplementary text
Figures S1 to S16
Tables S1 to S14
SI References

Other supplementary materials for this manuscript include the following:

STATA and Matlab codes
Data for replication purposes
Excel file delivering results in Table S13

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.

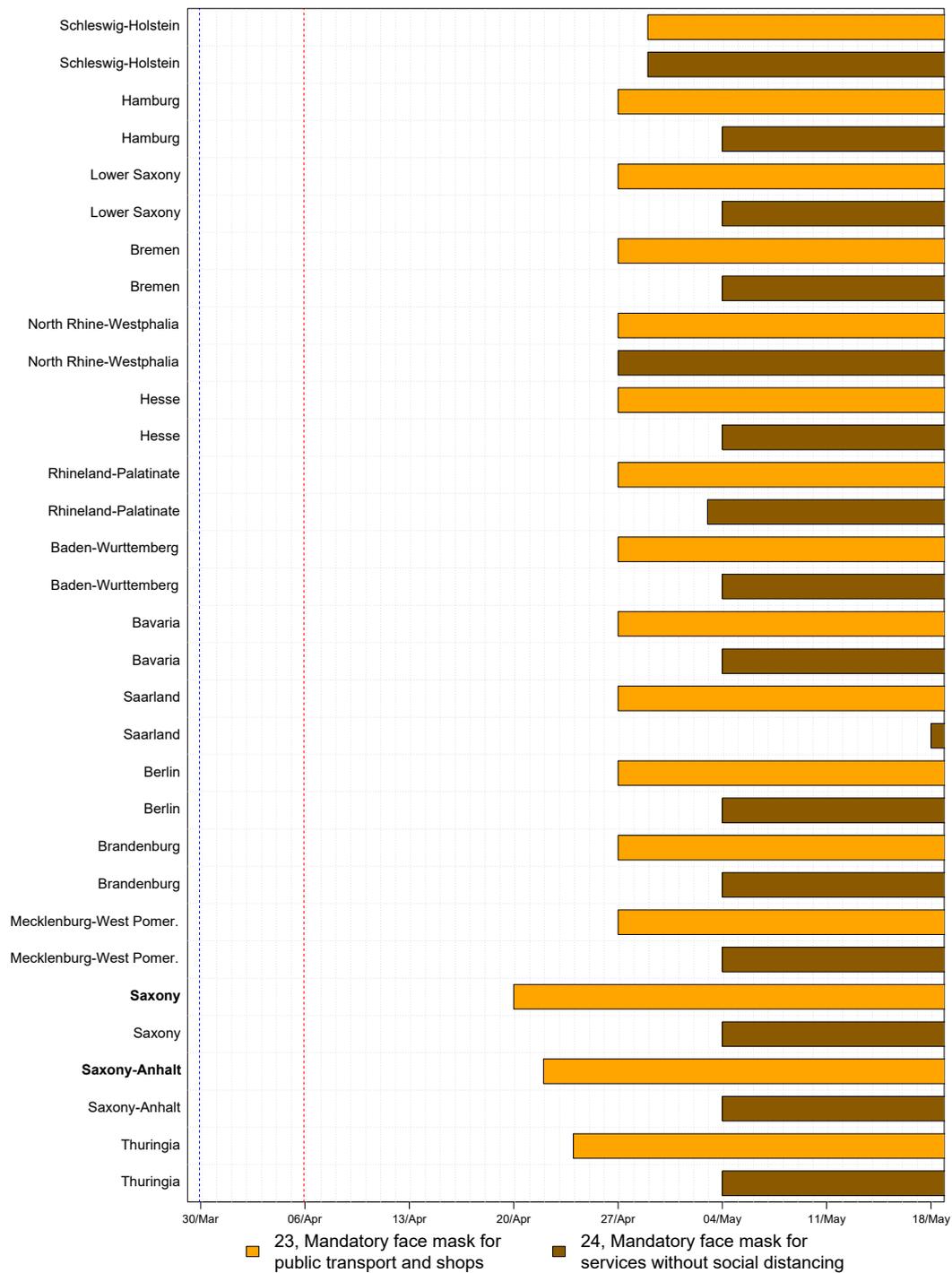


Figure S1: Time line of making face masks mandatory across federal states

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 transport

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, which introduced face masks earlier than the federal state to which they belong.

Table S1: 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-Wuerttemberg	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 S2 shows the points in time when measures entered into force in Jena. All measures for Thuringia are also binding for Jena.¹⁷ As Jena introduced three regulations concerning face masks, they became mandatory in three steps. April 1 saw the introduction of face masks for services where 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

¹⁷ We are grateful to Jan Franke for many explanations related to public health measures in Jena and Thuringia.

10, masks also became mandatory at work and in public buildings, assuming a distance of 1.5 m could not be maintained. (See also the box on the next page.) Measures of April 1 and 6 are measures also employed by federal states subsequently (see SI Appendix, Section A.1). The measure of April 10 was employed only by Jena (at least in this wording).

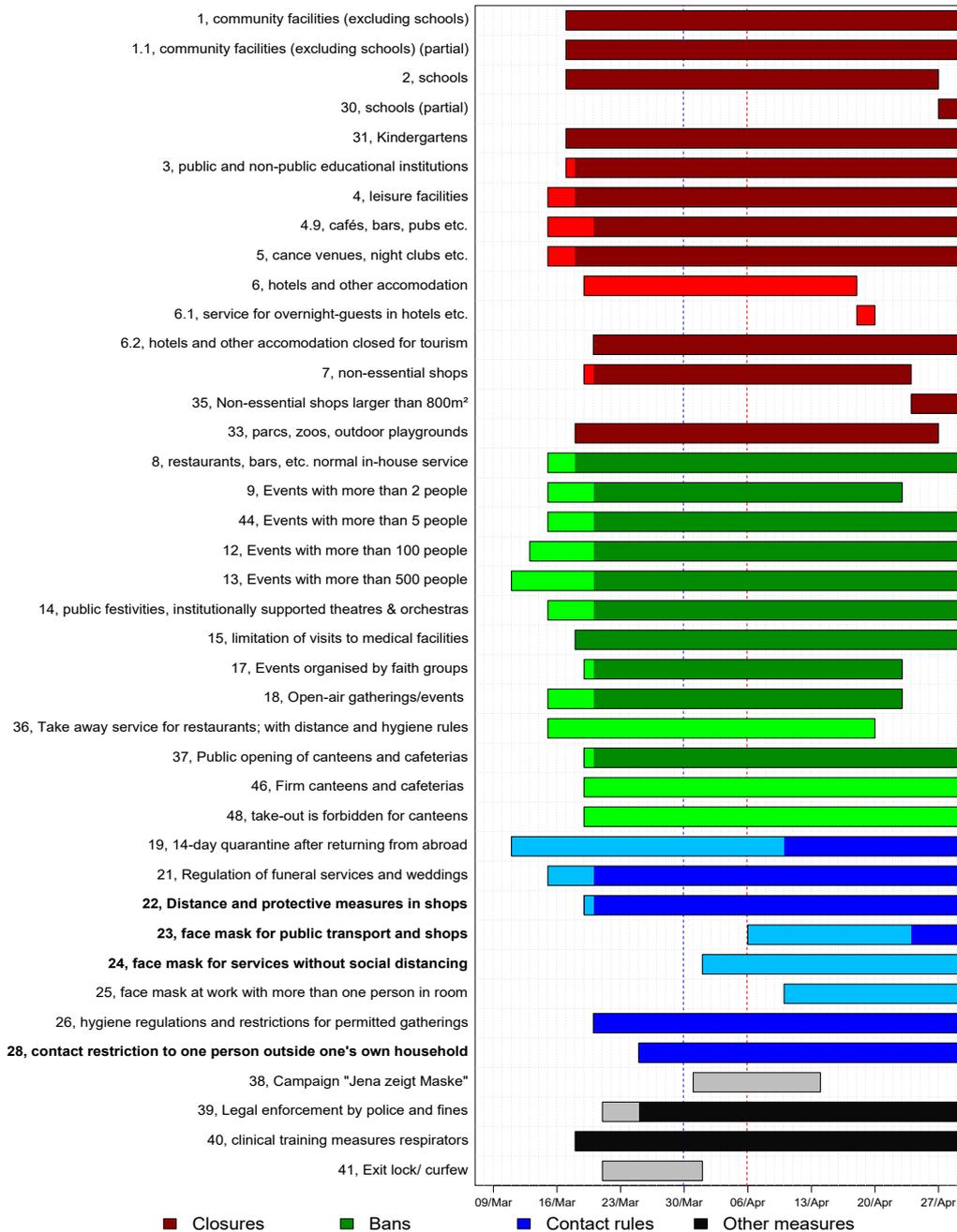


Figure S2: Time line of public health measures in Jena. Light bars indicate measures in force only in Jena, dark bars indicate measures in force in Thuringia (and thereby also in Jena)

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.¹⁸

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 were introduced. We therefore conclude that it was indeed face masks whose effect we measured 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 31, 2020 and enters in force on April 1, 2020.

Box S1: 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 general practitioner in order to obtain a diagnosis. With typical symptoms, a test is undertaken, and the result needs to be reported to the 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

¹⁸ 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.

properties of the 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 119,917 observations.

Findings for incubation. (1) and (2) 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

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

(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 $\sigma=0.4149$. The scale parameter is given by $\mu = \ln 5.1 = 1.63$.

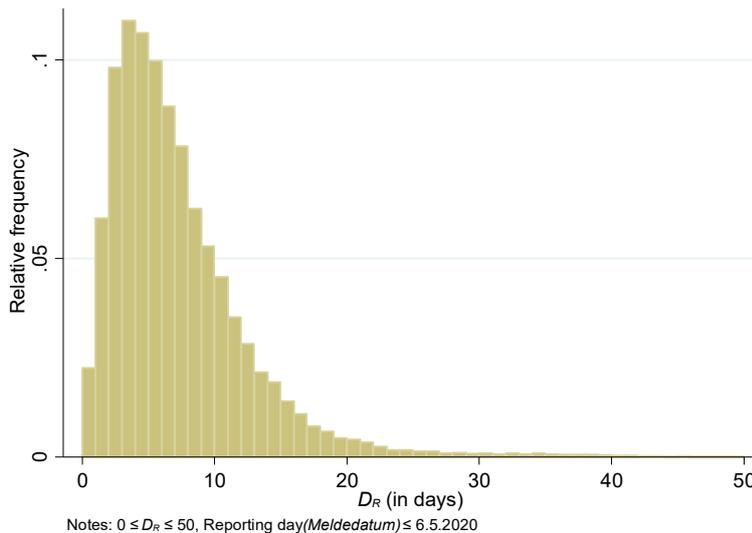


Figure S3: Histogram of delay between first symptoms and reporting

Findings for reporting. For illustration purposes, we plot a histogram of realizations of D_R in Figure S3. The mean, median (50% percentile), variance and standard deviation of D_R are reported in the next table.

Table S2: 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 (downloaded on June 7, 2020), there are 119,917 observations with information on day of infection (until reporting day May 6, 2020). We focus on 118,618 with $D_R \geq 0$.

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 $\text{Var}D = \text{Var}D_I + \text{Var}D_R$ 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. $F_D(\delta) = \text{Prob}(D \leq \delta)$. 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

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

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 $\delta - \delta_I$ 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.

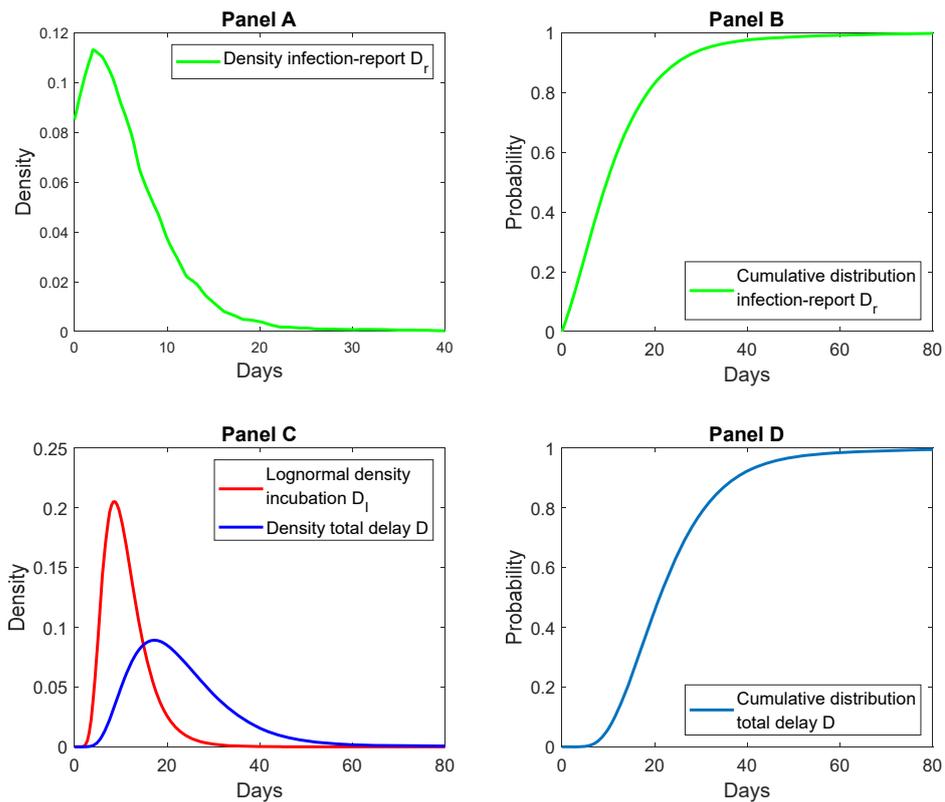


Figure S4: Density of the total delay D

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 S3 above, we can easily compute the density numerically. Figure S4 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., ref. 4) that need to assume a certain delay between infection and visibility in the data.¹⁹ 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 S3: 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

Conceptual 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 $R(t)$.²⁰ 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 $I^{ever}(t) = \int_0^t rI(x)S(x)dx$. This number is the theoretical counterpart to the number of Covid-19 cases reported by health authorities worldwide. This model is used for our conceptual 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.²¹ 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).

¹⁹ We are grateful to Christof Kuhbandner for discussions of this point.

²⁰ More elaborate models designed for Covid-19 exist (e.g., ref. 5–7). The simple model employed here is, however, sufficient for our interpretation purposes.

²¹ 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 Synthetic control method: Design, implementation and inference

This SI Appendix section provides further information on the design and implementation of the synthetic control method (SCM) to estimate the effect of face masks on the spread of Covid-19. We also explain how robustness tests can be used to conduct statistical inference.

Design. SCM is designed to run comparative case studies based on a data driven process. We have chosen SCM as the main vehicle for our empirical analysis here as the approach offers several attractive features relevant for our data settings. First, the scope 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 used to identify treatment effects of a policy intervention (thereafter treatment), such as difference-in-difference (DiD) estimation, SCM is performs well when only one or very few units receive the treatment in focus (11, 12). Second, the method is flexible, transparent in terms of presenting the model fit and counterfactual of interest; it relies on relatively few requirements and has thus become a widely utilized tool in the policy evaluation literature (13) and for causal analyses in related disciplines (see, e.g., ref. 14, for an overview of SCM in health economics, 15, for a biomedical application).²²

This section shall not serve the purpose of providing a formal description of the SCM approach (see ref. 10, 11, 17 for methodical details). The key *idea* of SCM is to establish a counterfactual that mimics a situation in which the treatment *ceteris paribus* would not have taken place. This is obtained by creating a synthetic control group out of the donor pool of control units (which have not received the treatment) and by comparing the outcomes of treated units and the synthetic control after the start of the treatment. Using a minimum distance approach, SCM selects weights for control units in the synthetic control group by relating treated and control units through a set of predictor variables measured prior to the start of the treatment. Weights for control regions are restricted to non-negative numbers and sum up to one in order to preclude extrapolation outside the support of the data.

To construct a suitable synthetic control group for our comparative case study of Jena, we need to find structurally similar regions in terms of their Covid-19 development before mandatory masks were introduced in Jena. Moreover, we need to make sure that face masks did not become mandatory in those regions during the treatment period. Such a control group would then most likely have had the same behavior as Jena in the absence of the mask obligation. We can then use this group to ‘synthesize’ Jena and conduct causal inference on the treatment effects of introducing face masks.

Synthetization means constructing a synthetic control group as weighted average of all regions in the donor pool of controls in which masks did not become compulsory earlier on. Regional weights are chosen in order to minimize a pre-treatment prediction error function for a set of predictor variables observed for Jena and the donor pool of control regions. Historical realizations of the outcome variable (cumulative number of Covid-19 cases; cumulative incidence rate) prior to the start of the treatment and several other regional characteristics (demography,

²² SCM is employed by (16) 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.

health care system) comprise the set of predictors. The implementation of the SCM approach makes sure that the essential requirements are met.

Requirements and implementation. The effective use of SCM relies on certain contextual requirements (see ref. 10, 18, 19). These requirements, also listed in the Method and Data section in the main text, include that

- (i) a donor pool of controls is available, i.e. not all regions receive the treatment during the period of the study,
- (ii) data is available for a sufficiently long time period before and after the start of the treatment,
- (iii) predictor values of the treated region are not extreme relative to those of controls, i.e. the treated region lies in the convex hull of control regions,
- (iv) spillover effects of the treatment on controls are absent and
- (v) there are no early anticipation effects, possibly related to concurrent policy interventions in the treated region.

In the empirical implementation of the SCM analysis, we ensure that these requirements are met as well as possible given the data at hand. The implementation is organized as follows. As baseline analysis, we focus on the single treatment case for the city of Jena. Our choice of Jena as essential comparative case study for the identification of treatment effects of mandatory face masks was made for several reasons, which can be directly linked to the aforementioned requirements. First, as shown in Figure 5 in the main text, Jena was the pioneer region for introducing 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. Referring to requirement i), a sufficiently long lag between the start of the treatment in Jena vis-à-vis control regions is important for effect identification as it ensures that a donor pool of controls without policy intervention is available.

Concerning the second requirement, our data sample for the daily development of (cumulative) Covid-19 cases throughout the period January 28 and May 11, 2020 allows us to operate with a sufficiently long pre- and treatment period. While we set the treatment period to the first 20 days after the introduction of face masks, the pre-treatment period covers the 14 days prior to the start of the treatment. The latter time window ensures that the highly dynamic nature of regional Covid-19 trajectories is properly captured for the construction of the synthetic control group. This helps to increase the fit between Jena and its synthetic control group in the pre-treatment period.

The synthetic control group is constructed by using 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. As outlined above, the chosen lag structure will 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. A list of variables together with summary statistics is given in SI Appendix, Section C.

With regard to the choice of predictor variables, SCM requires that values of predictor variables for the treated region are not extreme relative to those of controls, i.e. the treated region “lies in the convex hull” of control regions. As we show in greater detail in SI Appendix, Section C, 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 S6 in SI Appendix, Section C.2).

In order to preclude direct spillover effects (see requirement (iv)) from the treatment on non-treated control regions, we eliminate the immediate geographical neighbors of Jena from the donor pool. We also exclude those regions for which anticipation effects may have taken place because face masks became compulsory in quick succession to Jena (see Figure 5).

A further requirement for the effective use of SCM is that the timing of the introduction of face masks in Jena is not affected by concurrent public health intervention related to the Covid-19 spread. To support this requirement (v) from above, we have looked at all regulations (totaling almost 50) that were implemented in Jena between the beginning of March 2020 and end of April.²³ We have also looked at all regulations in Thuringia as these become automatically binding in Jena. A graphical illustration of the timing of the various measures and related discussion can be found in SI Appendix, Section 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.

In the implementation of our SCM analysis, we also run a series of sensitivity checks and placebo tests that will help to investigate the robustness of the results and check to what extent the underlying requirements of SCM are met. The results are reported in the main text. Important sensitivity checks are:

1. 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.²⁴
2. 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.
3. Significant public health measures that were introduced in Jena but not (or only slightly delayed) in the federal state of Thuringia or other German regions will be tested for their

²³ 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 (20) for more background.

²⁴ 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 by the local administration in Jena was published on March 31.

intervening effect on the introduction of face masks by means of *placebo-in-time tests* and complementary *difference-in-difference regressions*.

We also run SCM analysis for a different outcome variable (cumulative incidence rate) and by disaggregating the data into specific age groups. A final important robustness test, which can also be used to conduct statistical inference in the SCM framework are *placebo-in-space tests*. We conduct such tests by estimating placebo treatment effects for each control region in which masks did not become compulsory early on. These placebo treatment effects treat each region in the donor pool of controls as treated region for the treatment period that applies for Jena. Using SCM, a synthetic control group is constructed for each control region (based on the same set of predictor variables as for Jena), and outcome differences between the region and its synthetic control measure the size of the placebo treatment effect. How these tests are employed to calculate p-values for assessing the significance of the estimated treatment effects is described below.

Inference. The implementation of comprehensive *placebo-in-space tests* allows us to conduct statistical inference on the basis of permutation tests as suggested by (18) and applied, for example, by (21) or (22). Statistical significance is established by comparing the difference of outcomes between the treated region and its synthetic control group and the differences among donor regions and their respective synthetic control groups. For a significant treatment effect, we expect that the estimated placebo treatments for control regions are small (or even reversed) relative to those for the treated region. We calculate time-specific significance levels for the test of the hypothesis that mandatory face masks did not significantly reduce the number of reported Covid-19 cases for each day after the treatment.

The (pseudo) p -values reported in the main text are derived from a ranking of the actual treatment effect within the distribution of placebo treatment. We follow the suggestion in (23) and compute adjusted p -values taking the pre-treatment match quality of the placebo treatments into account. The refined procedure draws inference only from donors with a good fit in the pre-treatment period. Specifically, we do not include placebo effects in the pool for inference if the match quality of the control region, measured in terms of the pre-treatment root mean squared prediction error (RMSPE), is greater than 20 times the match quality of the treated unit. Further, we adjust p -values by dividing all underlying effects with the corresponding pre-treatment match quality.²⁵ We finally use the set of p -values to compute confidence intervals for treatment effects to visualize the significance and precision of the estimated effects (25).

Multiple treatment analysis. 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. This may prevent a straightforward generalization of estimated effects. We therefore also test for treatment effects in regions that introduced face masks after Jena but still before they became compulsory all across Germany. To do so, 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, of which 8 are larger cities (*kreisfreie Städte*).

All other regions apart from 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

²⁵ We conduct all estimations in STATA using “Synth” and “Synth Runner” packages (23, 24). Data and estimation files are available on the journal’s web page.

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 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 ref. 1, 2) and a similar reporting lag. This overall time lag between the infection with SARS-CoV-2 and registration in the data is also crucial for the interpretation of our results and we discuss it in detail in SI Appendix, Section A.3.

SCM and difference-in-difference estimation. We use SCM and DiD regressions as complementary estimation approaches to identify the treatment effects of face masks. The choice of SCM as our main vehicle of analysis results from the different requirements of both estimators. Regression approaches such as DiD usually perform poorly when the treatment group is very small. Moreover, the identification of treatment effects in the DiD approach strongly relies on the validity of the parallel trend assumption. A parallel trend in the outcome variable between the treated and control units ascertained in the pre-treatment period is assumed to hold in the treatment period in the counterfactual case without the intervention. For establishing the treatment effect by SCM, the parallel trend assumption need not to imposed (11, 14, 17). This is due to the construction of the SCM, in which common trends between the treated and similar control regions are favorable for finding an appropriate counterfactual trajectory. While DiD accordingly associates equal weights to all control regions, SCM replaces equal weights by optimal weights. Optimal weights minimize the distance between treated and control regions in terms of pre-intervention characteristics including lagged outcome values.

However, DiD estimation also has certain advantages over SCM. These particularly relate to the estimation of dynamic treatment effects. We accordingly use DiD regressions in order to check the robustness of our empirical results against the potential presence of effects stemming from latent policy interventions or other events that may affect the outcome variable (see SI Appendix, Section F for a detailed description of these robustness tests).

C Data description and additional SCM estimation results for Jena

C.1 Summary statistics for outcome and predictor variables

The Robert Koch Institute (RKI) collects data on registered Covid-19 cases from local health authorities in the individual municipal districts (NUTS3 regions) in Germany and provides updates to this database on a daily basis (available via API). We use the cumulative number of registered Covid-19 cases in each municipal district as main outcome variable.²⁶ As an alternative outcome variable, we also employ the cumulative incidence rate, i.e. the number of cumulative Covid-19 cases per 100,000 inhabitants. Summary statistics for these two outcome variables together with information on the daily number of newly registered Covid-19 cases are shown in Panel A of Table S4.

²⁶ 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 S4: 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 aged 15-64 years)	34.34	5.46	22.40	53.98
Young-age dependency ratio (persons aged 14 years and below per 100 of population aged 15-64 years)	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 (categorical variable [§])	2.59	1.04	1	4

Notes: * = International Standard Classification of Education (ISCED) Level 6 and above; § = 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*).

The variables presented in Table S4 are also used as predictor variables in the SCM approach. Selection of the latter has been guided by their ability to describe the regional number and dynamics of reported Covid-19 cases in the treatment period. Obviously, past values of (newly) registered Covid-19 cases are important for predicting 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. These latter regional predictor variables are only available as annual averages and are obtained from the INKAR online database of the Federal Institute for Research on Building, Urban Affairs and Spatial Development (26). We use the latest year available in the database, which is 2017. Despite this time lag, we argue that regional demographic structures only gradually vary over time such that they can be used to proxy regional differences during the spread of Covid-19 in early 2020.

C.2 Trajectories of cumulative Covid-19 cases and box plots for predictor variables

As outlined in the main text and in SI Appendix, Section B, one requirement of SCM is that predictor values of the treated region are not extreme relative to those of controls. This SI Appendix section goes beyond descriptive statistics by visualizing the overlap in variable values for Jena and the donor pool. We do so for both outcome and predictor variables. First, Figure S5 plots the cumulative number of Covid-19 cases for Jena and the (daily) 2nd to 4th quintile of

the control regions. As Jena lies well within the control regions, its infection dynamics was not extreme relative to the donor pool in the pre-treatment period.

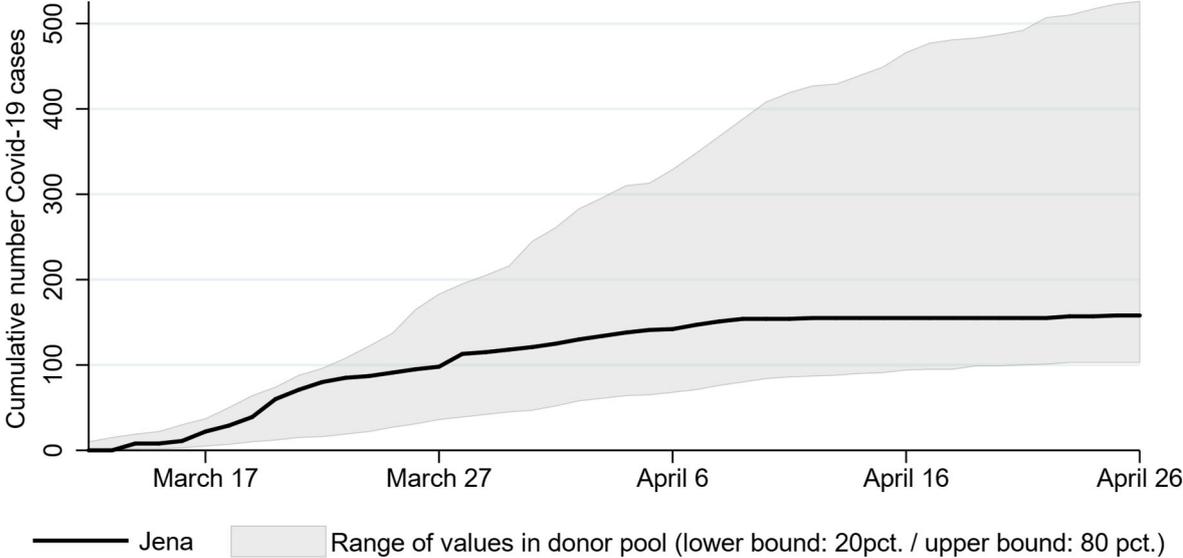


Figure S5: Overlap of cumulative Covid-19 cases between Jena and the donor pool of control regions

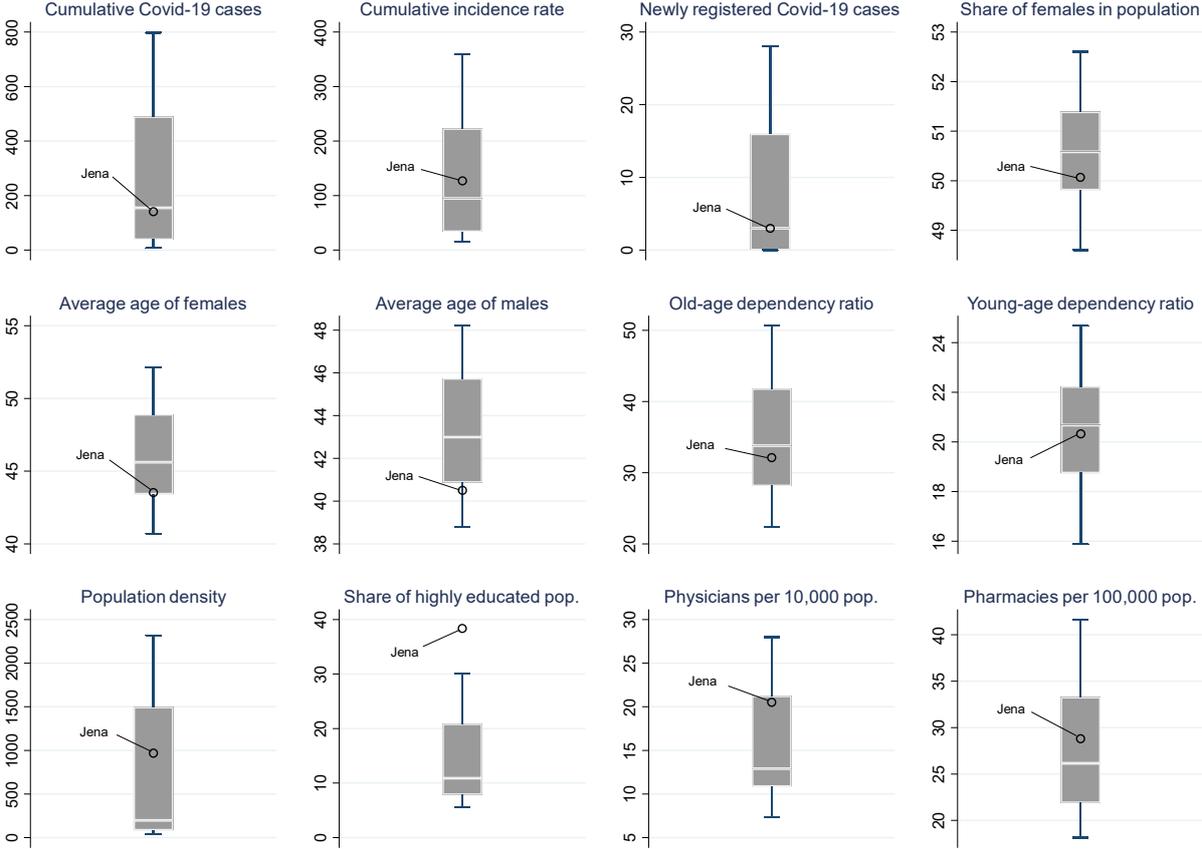


Figure S6: Box plots for the regional distribution of predictor variables (donor pool of controls and Jena).

Notes: Box plots are drawn to cover 90% of values for regional distribution across German regions; further outliers are ignored. For time-varying predictors, the regional distribution on April 5 is shown in the box plots.

Box plots in Figure S6 lend additional support for perceiving Jena as an “average region” in Germany. Infection level and dynamics prior to the treatment period on April 5 were very close to the median value of the donor pool (as shown for the cumulative number, the cumulative incidence rate and the number of new Covid-19 cases). Other predictor values (as summarized in Panel B of Table S4) also display a sufficient overlap between Jena and the control regions. Only the share of highly educated individuals is exceptional, probably reflecting the status of Jena as a university town with roughly 18,000 students out of approximately 108,000 inhabitants.

To test whether this exception drives our results, we have performed cross-validation tests that sequentially exclude individual variables from the set of predictors. As shown in SI Appendix, Section C.7, the results for the different SCM specifications do not show structural differences and identify similar treatment effects. Our results are therefore not driven by this one exception.

C.3 Control regions, SCM weights and pre-treatment predictor balance

This SI Appendix section provides details on the composition of the synthetic control group used to identify treatment effects of face masks in Figure 1 in the main text. 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 fit between Jena and its synthetic control group are also reported.

The donor pool used to construct the synthetic control group includes all other German NUTS3 regions except for the two immediate neighboring regions of Jena (Weimarer Land, Saale-Holzland-Kreis) and the regions Nordhausen and Rottweil. The latter two introduced face masks in rapid succession to Jena on April 14 and April 17, respectively.

Table S5: Composition of synthetic control group for Jena with associated SCM weights

Introduction of face masks (Panel A in Figure 1)		
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

Notes: Donor pools corresponds to SCM estimation in Panel A of Figure 1. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

Table S6: Pre-treatment predictor balance and RMSPE for SCM in Figure 1

	Treatment:		Introduction		Announcement	
	Jena	Synthetic	Jena	Synthetic	Jena	Synthetic
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 start of the treatment, average)	3.7	3.8	5	5.2		
Population density (Population/km ²)	968.1	1074.3	968.1	947.9		
Share of highly educated population (in %)	38.4	22.8	38.4	26.3		
Share of females 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 (categorical variable)	1	1.3	1	1.9		
RMSPE (pre-treatment)	3.145		4.796			

C.4 SCM results by age groups

We refined our Jena analysis also for different age groups. We looked at inhabitants aged 15 to 34, 35 to 60 and above 60. We again studied the treatment effect for the implementation date for masks on April 6 and for the announcement of masks on March 30.

Predictor variables are chosen as for the baseline specification shown in Figure 1 in the main text. Results visible in the next figure are also discussed in the main text. Regions in the synthetic control groups corresponding to the three age groups are listed in Table S7.

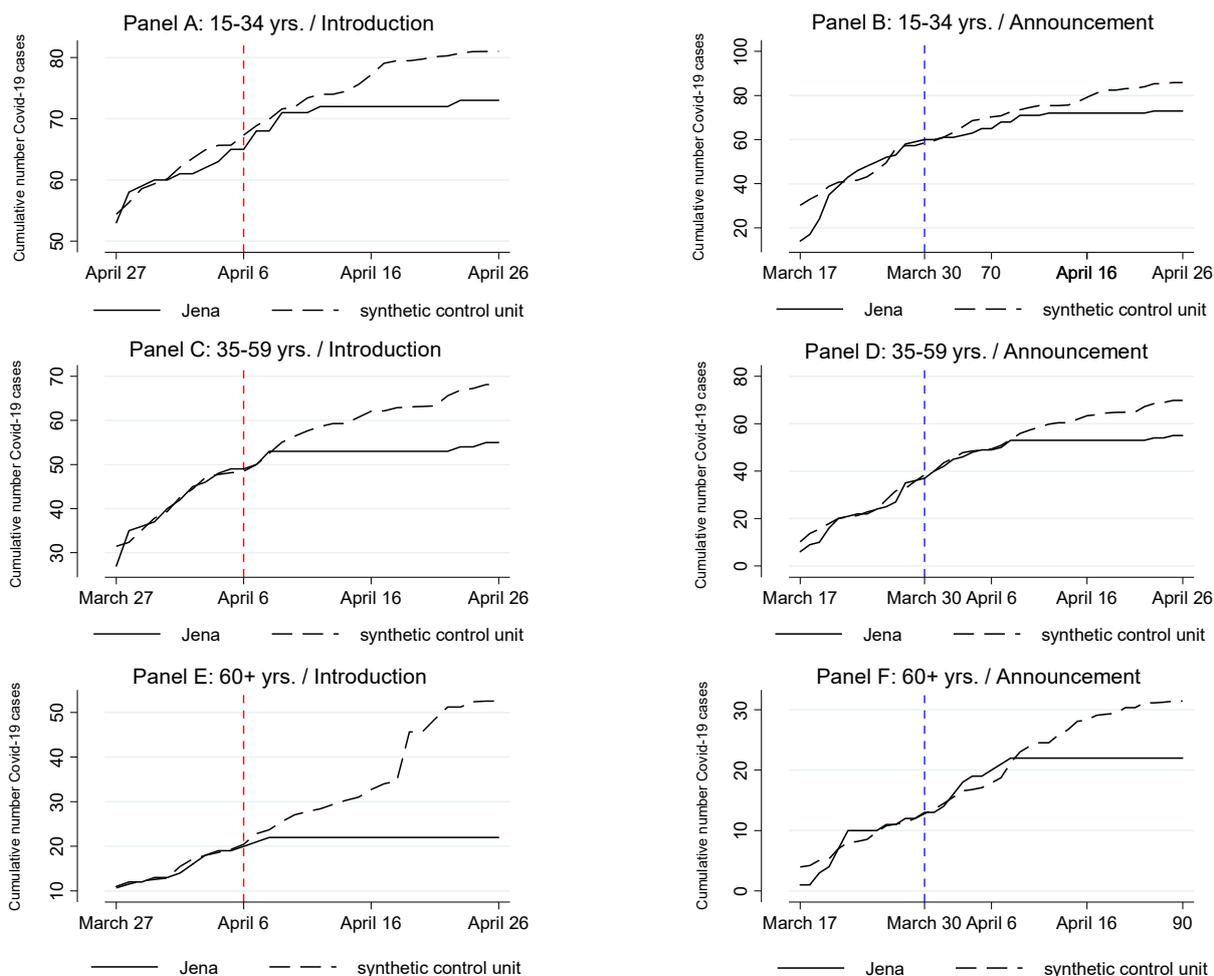


Figure S7: Treatment effects for introduction and announcement of face masks in Jena by age groups

Table S7: SCM weights for control regions used to construct synthetic Jena for three 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	Hochtaunuskreis	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

Notes: Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

C.5 Effects on cumulative number of infections per 100,000 inhabitants

One might be concerned that absolute infection numbers are not appropriate as regions differ in size measured by number of inhabitants. We checked our results by re-estimating effects for incidence, i.e. infections normalized by the size of the population. Our outcome variable is then given by cumulative reported infections divided by population size and multiplied by 100,000. Predictor variables are chosen as for the baseline specification shown in Figure 1.

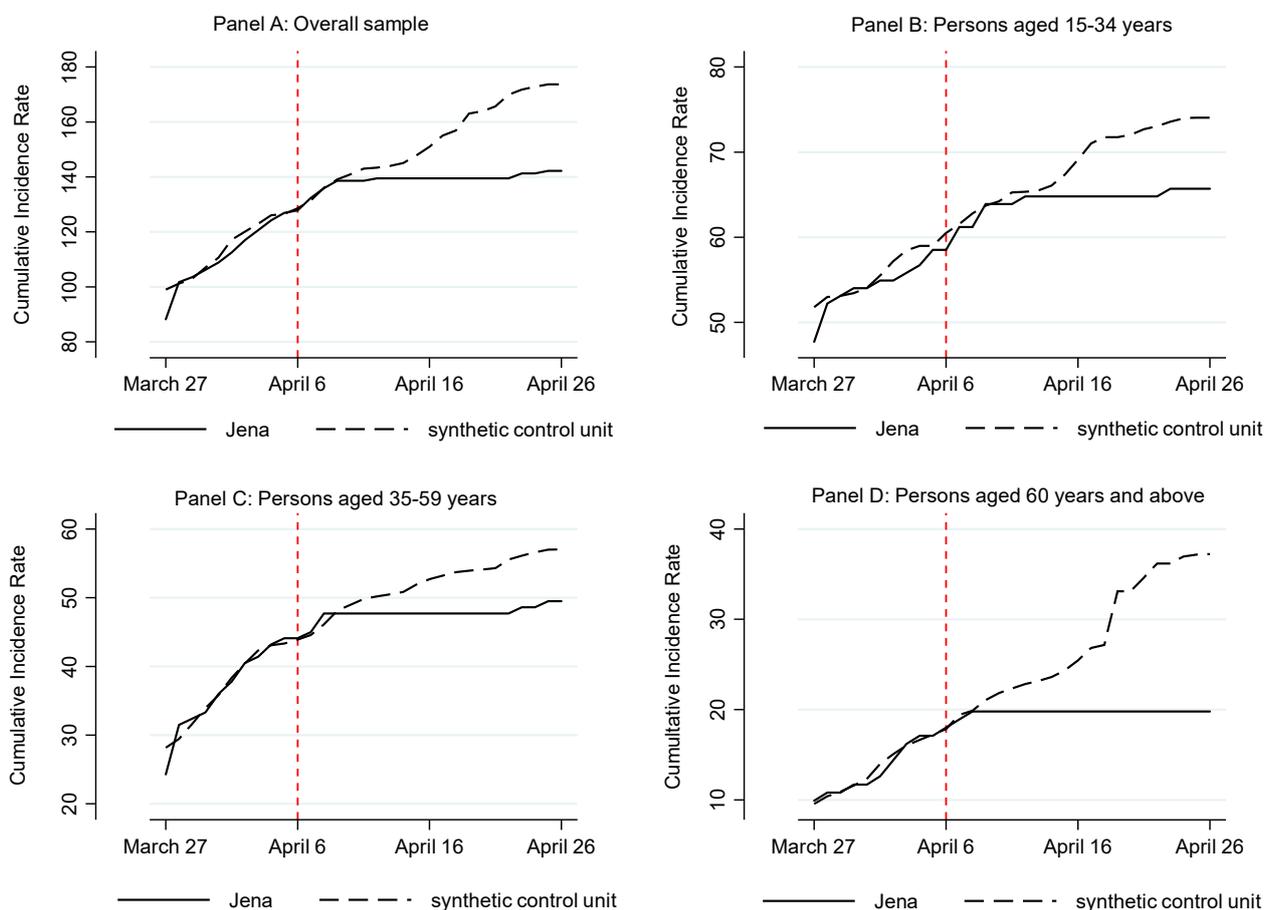


Figure S8: Treatment effects for introduction of face masks on cumulative incidence rate

Table S8: SCM weights for control regions used to construct 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: Synthetic control group corresponds to SCM estimation in Figure S8. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

Table S9: SCM weights for control regions used to construct 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

Notes: Donor pools corresponds to SCM estimations in Figure S8. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

C.6 Announcement and mobility

C.6.1 Google trends and announcement effects

To understand the role of a potential announcement effect or the effect of the information campaign in Jena for masks, we looked at search intensities in the internet. Time series are displayed in the next figure, results are discussed in the main text.

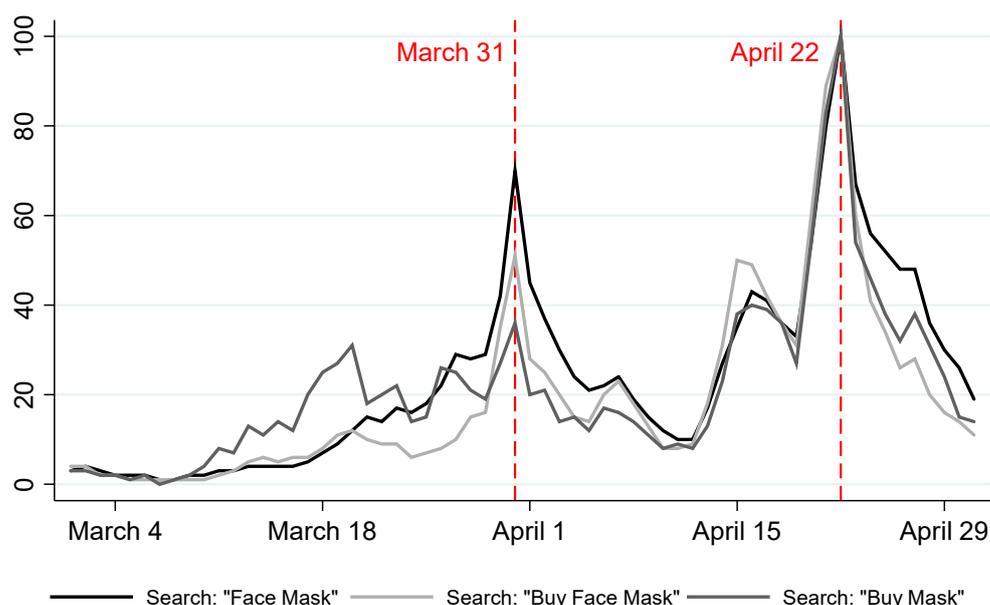


Figure S9: Online search for face masks and purchase options according to Google Trends

Notes: 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 (27).

C.6.2 Mobility trends across German federal states

Figure S10 shows overall mobility patterns across German federal states between Feb 17 and May 18, 2020 based on Google (28). 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 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 S10 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 ref. 20 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 us 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 (28).²⁷

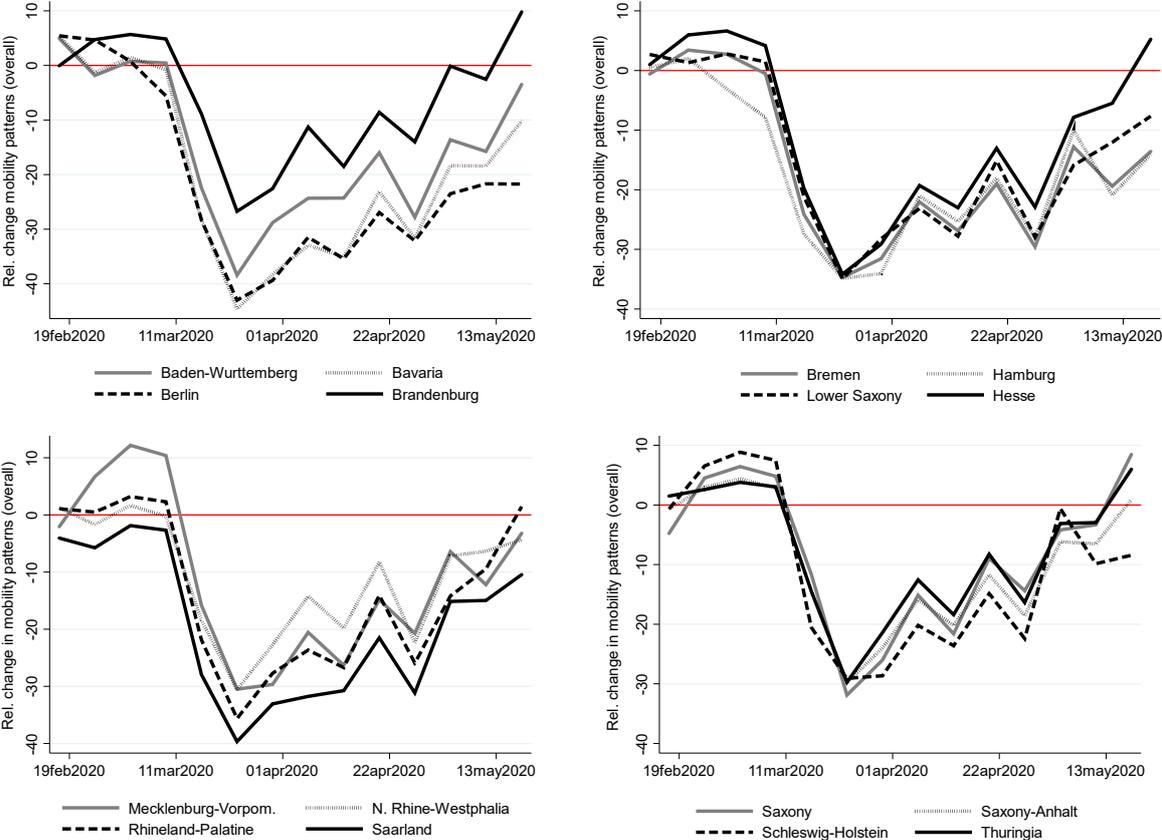


Figure S10: Trend in mobility patterns across German federal states (Feb 17 to May 18, 2020) from (28).

²⁷ For details see (29).

C.7 Cross validation and additional placebo-in-time test

The cross-validation and additional placebo-in-time tests reported in Figure S11 check the robustness of the estimated treatment effects for alterations to our benchmark SCM specification. The cross-validation tests in Panel A modify the lag structure of the included time-varying predictor variables. The additional placebo-in-time test in Panel B checks for anticipation effects 20 days before the start of the actual treatment on April 6 (equally split into a pre- and treatment period from March 26 onwards).

In Panel A the baseline specification for the synthetic control uses both 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 predictor variables. The alternative specifications lag both of these variables simultaneously by 1, 3 and 7 days.

Panel B displays pseudo-treatment effects for 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.

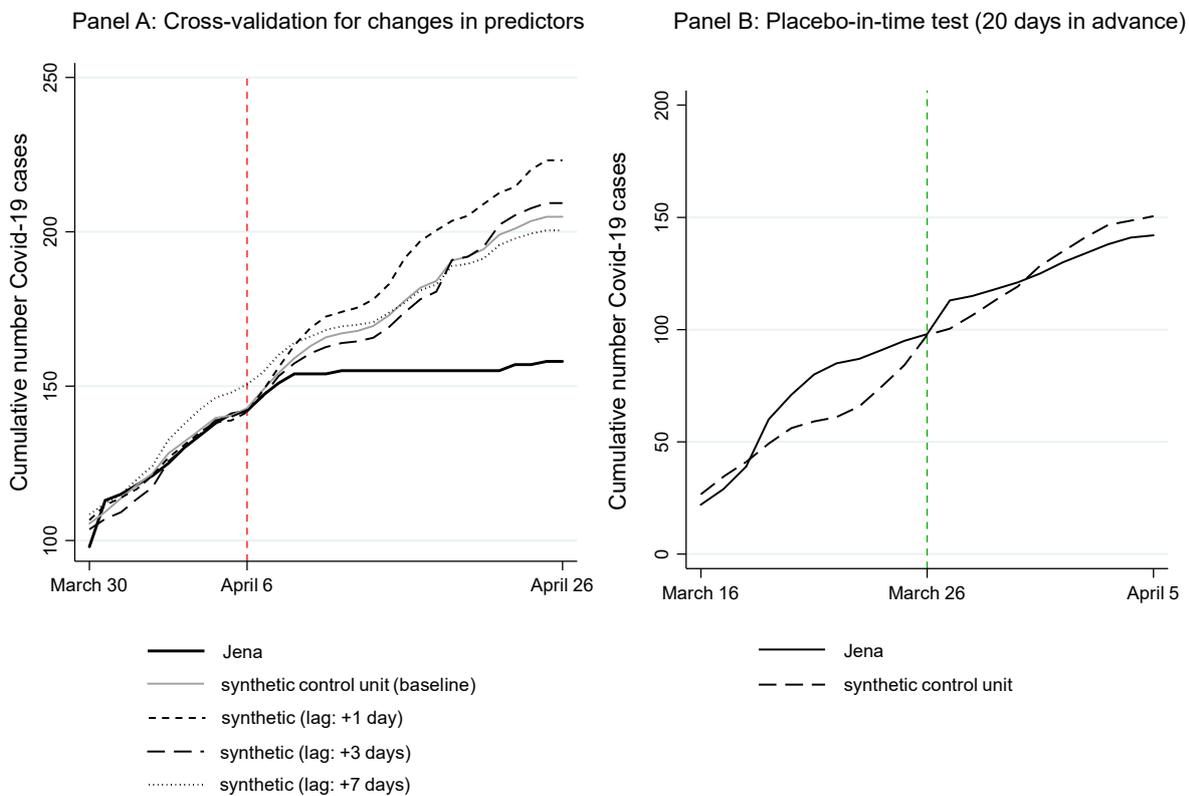


Figure S11: Cross-validation test for changes in time-varying predictor variables and placebo-in-time test

Figure S12 shows the results for an additional cross-validation test, which sequentially excludes time-constant predictors from the full set of predictor variables. The estimated trajectories for the respective synthetic control groups excluding a certain predictor follow a very similar trend and all identify a reduction in the number of cumulative Covid-19 cases in Jena vis-à-vis the synthetic Jena that widens over time. Excluding population density implies larger outcome difference than for other excluded variables. Most importantly, however, excluding the share of skilled individuals, the variable by which Jena seems exceptional according to Figure S6, does

not affect our main finding in any essential way. We therefore argue that using the full set of predictors is the most reasonable approach to identify reliable treatment effects.

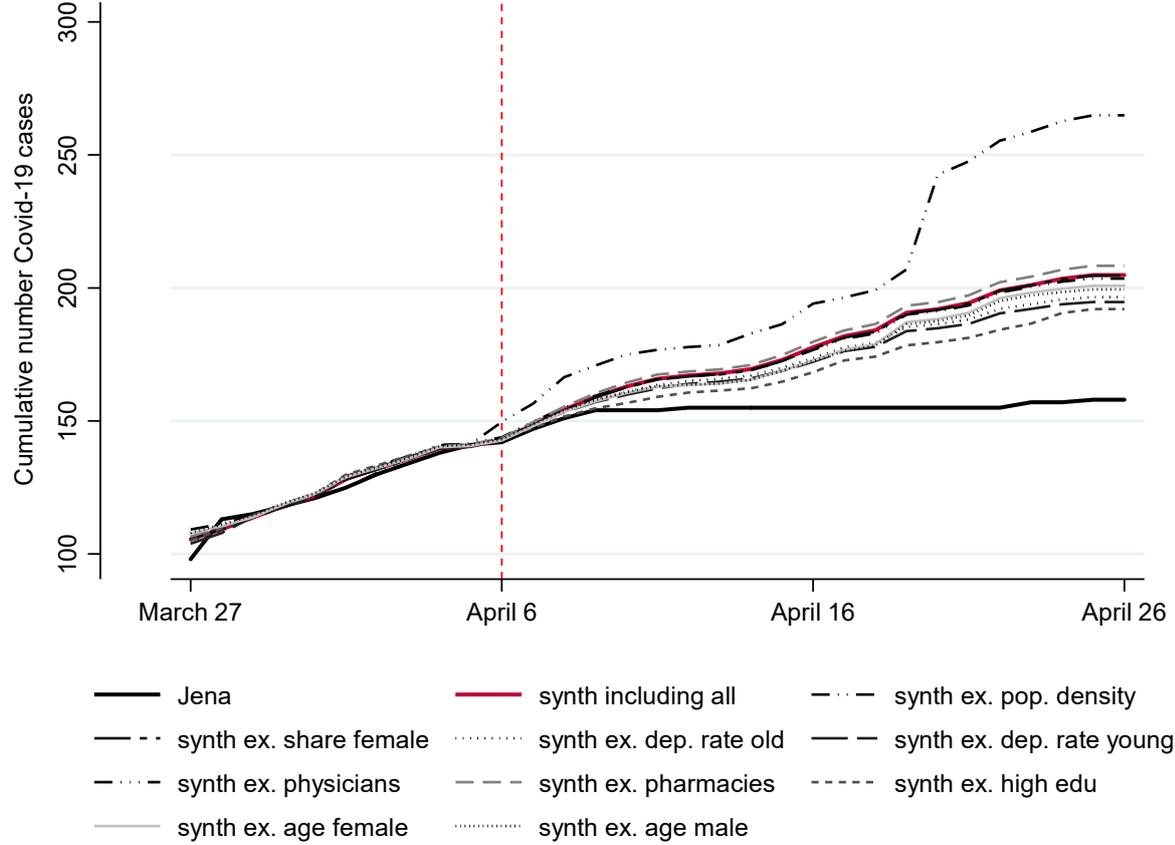


Figure S12: Cross-validation test for changes in the set of time-constant predictors

C.8 Changes in donor pool for synthetic Jena

A further robustness check changes the donor pool from which synthetic Jena is constructed (see main text for detailed definition). Infection dynamics in the various synthetic groups are shown in the next figure. Again, predictor variables are chosen as for baseline specification shown in Figure 1.

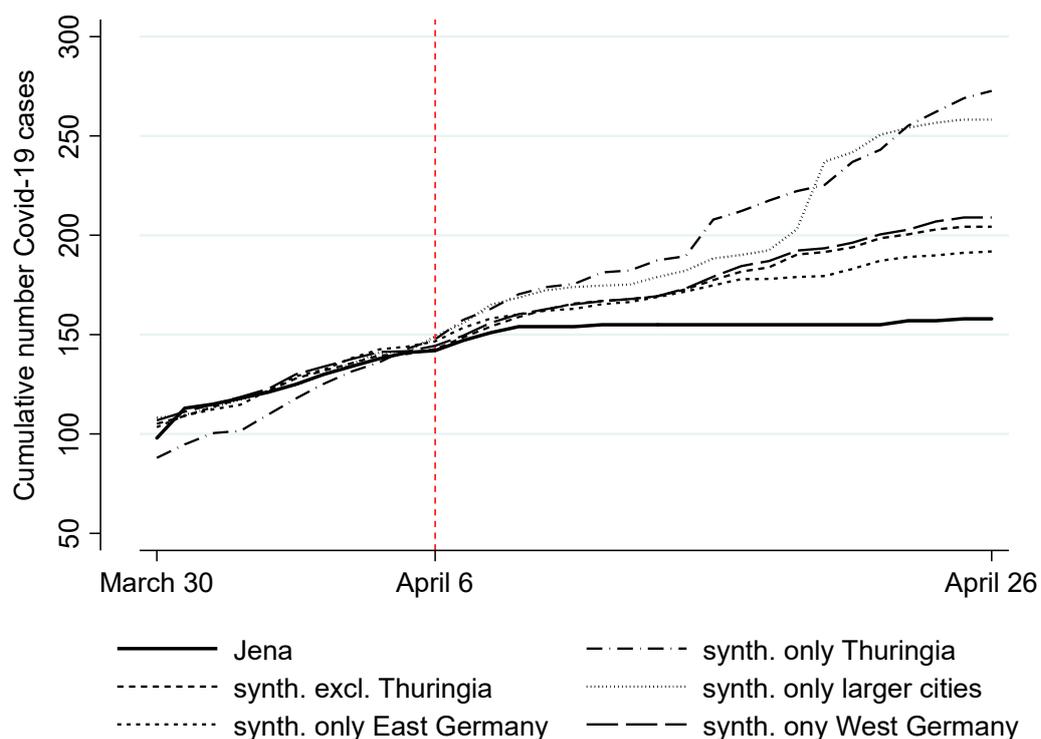


Figure S13: Treatment effects for changes in donor pool used to construct synthetic Jena

Table S10: Weights for control regions used to construct synthetic Jena for various donor pools

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 S13. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

C.9 Place-in-space tests for other major cities in Thuringia

The placebo-in-space tests for other cities and larger regions in Thuringia uses the same set of predictors as for Jena (see Figure 1). The reported regions cover all *kreisfreie Städte* plus Gotha (*Landkreis*). The cities Weimar, Suhl and Eisenach have been aggregated since the absolute number of reported Covid-19 is low in these cities, which made it hard to find a suitable control group. None of these regions shows a reduction in the number of Covid-19 cases as in Jena. This confirms that latent macro-effects, for example infection dynamics or other interdependencies in Thuringia as a whole, are not behind the measured treatment effects for Jena.

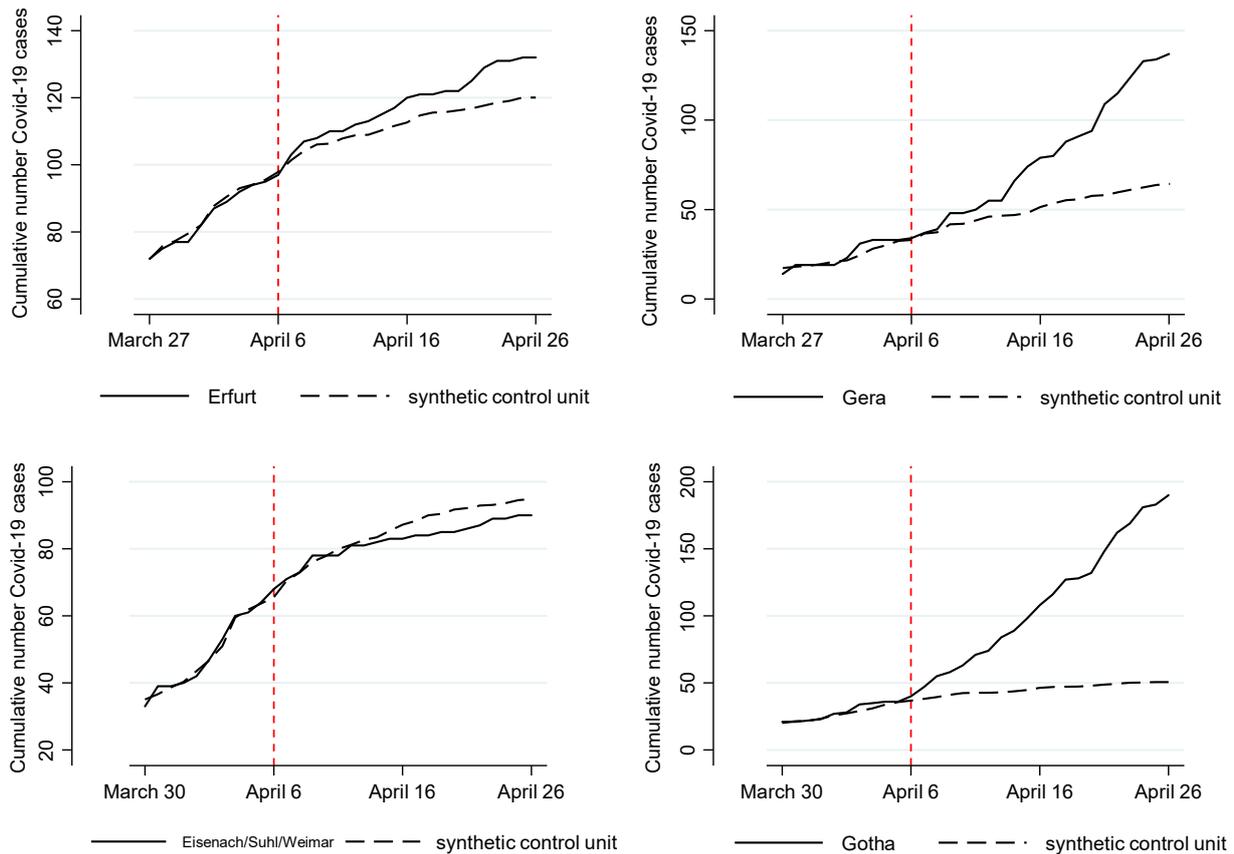


Figure S14: Placebo tests for the effect of face masks in other cities in Thuringia on April 6.

As a side note, the cumulative number of Covid-19 cases in Gotha relative to its synthetic control group stands out here. Local newspapers report²⁸ that there were outbreaks in two hospitals and three retirement homes. The fast rise in Gotha can therefore best be explained by a bad realization of an otherwise normal stochastic process. One could relate the outbreaks to issues with compliance but – to the best of our knowledge – data are not available in this respect.

²⁸ See e.g. (30).

Table S11: SCM weights for control regions in 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
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 S14. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

D Single and multiple treatment analyses

D.1 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).

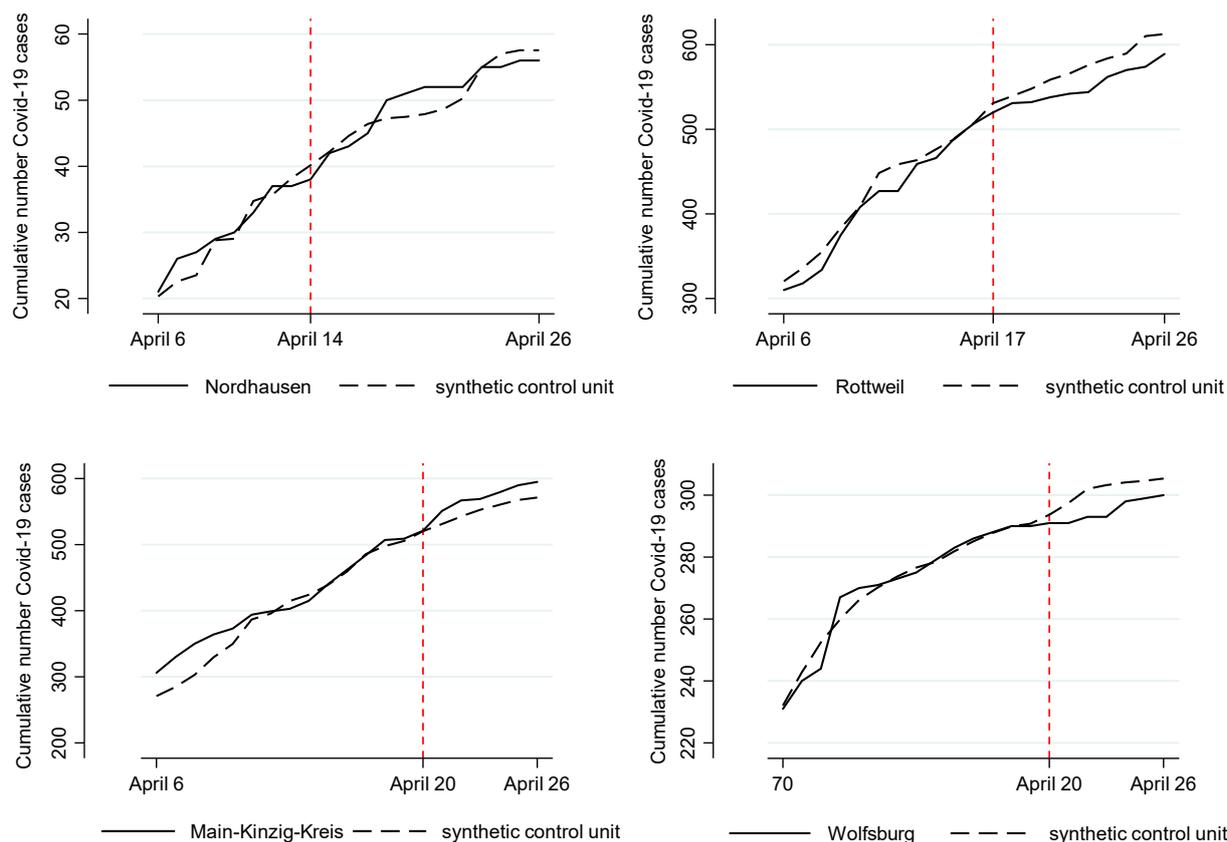


Figure S15: Treatment effects for introduction of face masks in other cities

Notes: Vertical dashed lines (in red color) indicate the introduction of mandatory face masks in each region.

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 S12: SCM weights for control regions in synthetic control groups (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			

Notes: Donor pools corresponds to SCM estimations in Figure S15. Sample weights are chosen to minimize the RMSPE ten days prior to the start of the treatment.

D.2 Growth rates and summary of single and multiple treatment effects

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 $142 [1 + \Delta x_{\text{Jena}}]^{20} = 158$. The daily growth rate in the control group is denoted by x_{control} and can be computed from $143 [1 + \Delta x_{\text{control}}]^{20} = 205$. Hence, the introduction of the face mask is associated with a decrease in the number of infections of $(\Delta x_{\text{control}} - \Delta x_{\text{Jena}})$ 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 S13: 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 as supporting information on the webpage of the journal.

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

E.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 is 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, 31–37). 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.²⁹

There are also many cross-country analyses, both in a structural SIR (susceptible, infectious and removed) sense (39) and with an econometric focus on forecasting the future development of the Covid-19 pandemic (40). Others draw parallels between earlier pandemics and Covid-19 (41). These studies do not explicitly take public health measures into account. Some studies discuss potential effects of public health measures and survey general findings (42–44) 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 were avoided by this measure during the first four weeks. (45) 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

²⁹ In a short note, (38) 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.

the actual dynamics in Sweden. Based 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.

E.2 Evidence for face masks

At present, more and more clinical evidence is being presented. While final conclusive results have not been reached (see ref. 46 for a review), research shows that face masks catch infectious particles that occur when speaking, coughing, or sneezing. This reduces the risk of infecting another person (47, 48). The effects of face masks have been systematically surveyed by (49) and (50). (50) mainly present evidence on the effect of face masks during non-Covid epidemics (influenza and SARS). (51) 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), (49) 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 (49) is one that analyzed *“exhaled breath and coughs of children and adults with acute respiratory illness”* (52, 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 (53), 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.³⁰

(54) 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

³⁰ 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 drove the estimated trend reversal.

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, (55) 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 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.

E.3 The economic costs of public health measures

We provide a short overview of studies that quantify economic costs of public health measures and a first rough estimate of the cost of face masks. For the case of school and child-care center closures, (56) estimate that 8.4 percent of total working hours will be lost, corresponding to 11.7 million employed persons in short-time-work. Beyond this short-term effect, intergenerational mobility and gender equality in the workplace is likely negatively affected. The macroeconomic impact of this public health measure is considered large.

Substantial GDP impacts of closures of stores, restaurants and other business outlets are found by (57) under various degrees of easing restrictions. Costs of face masks are neglected when assessing overall economic consequences of all measures. Economic effects of several interventions are examined by (58) for New Zealand. While a moderate decline in GDP of 0.7 percent is estimated for the ban of mass gatherings and closure of public venues, restricting domestic travel may induce a large reduction of up to 6 percent. Economic effects of social-distances policies are investigated by (59) and (60). The advantageousness of testing from the macroeconomic point of view is highlighted by (61) and (62). Finally, several papers examine macroeconomic effects of a general lockdown (cf. ref. 63–65).

We conduct a simple back-of-the-envelope calculation in order to understand the cost burden from wearing face masks for private household. Each person may need two face masks a day. In September 2020, two simple single use masks certainly cost less than 1 € in Germany. This translates to expenditures of less than 30 € per month. According to (66), the average monthly disposable income per one-person household was 2112€ in 2017 in Germany. Monthly per capita expenditures for face masks should therefore account for less than 1.4% of the average disposable income. We can do similar calculations for larger households. They have to spend a larger share of their disposable income.³¹ Of course, the monthly expenditures may be reduced if fabric masks are used. Washing and using them again is possible. We conclude that in relation to other public health measures and their consequences (among which part-time work or unemployment), the cost burden of face masks may be low for private households.

³¹ Two-person households had an average monthly disposable income of 3919 € in 2017 (66). They may need four masks per day, which translates to expenditures of less than 1.5% of the average disposable income for simple face masks. Expenditures for a four person household (average monthly disposable income of 5483 € in 2017, 8 masks may be needed per day) should account for less than 2.5% of the average disposable income.

F Difference-in-difference estimates for 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 given in Figure S2 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, accordingly, may 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 (67; see ref. 68 for a general discussion of the use of difference-in-difference estimation to identify causal effects of Covid-19 policies).

In order to test for the presence of anticipation effects stemming from (unobserved) public health measures taken prior to the introduction of face masks, we define a baseline treatment dummy, which takes a value of one for Jena from March 14 onwards and is zero before that day. This captures the start of public health measures taken in Jena (compare with Figure S2). 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 where each adds 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 sample as a whole ends on May 6.

Hence, 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 can identify additional effects, which relate to (unobserved) public measures introduced during the time interval. Again, as outlined in SI Appendix, Section A.3, for the correct interpretation of the obtained results, 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 a total of $m=(1,\dots,M)$ regressions 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 region-fixed effects at the level of municipal districts, $D_{Weekday}$ is a set of binary dummies for the different days of the week and $\Psi_{k(t)}$ are time-fixed effects for each with $k=1,\dots,K$ 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 estimate 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 include values of newly registered Covid-19 cases before March 14 as well as the set of structural regional characteristics (shown in Table S4 in SI Appendix, Section C.1) as regressors. Hence, in analogy to

our SCM approach, the main idea for this two-step approach is to give those control regions a larger sample weight that have similar characteristics to Jena before the baseline treatment starts (68). This may mitigate the problem of DiD estimation to result in a poor performance potentially linked to the problem of heteroscedasticity if there are very few (or even only one) treatment group (see ref. 69 for a general discussion of inference in DiD models with few treated units). The resulting two-step estimator is known as conditional difference-in-difference estimator (70). Estimated effects are shown in Figure S16.

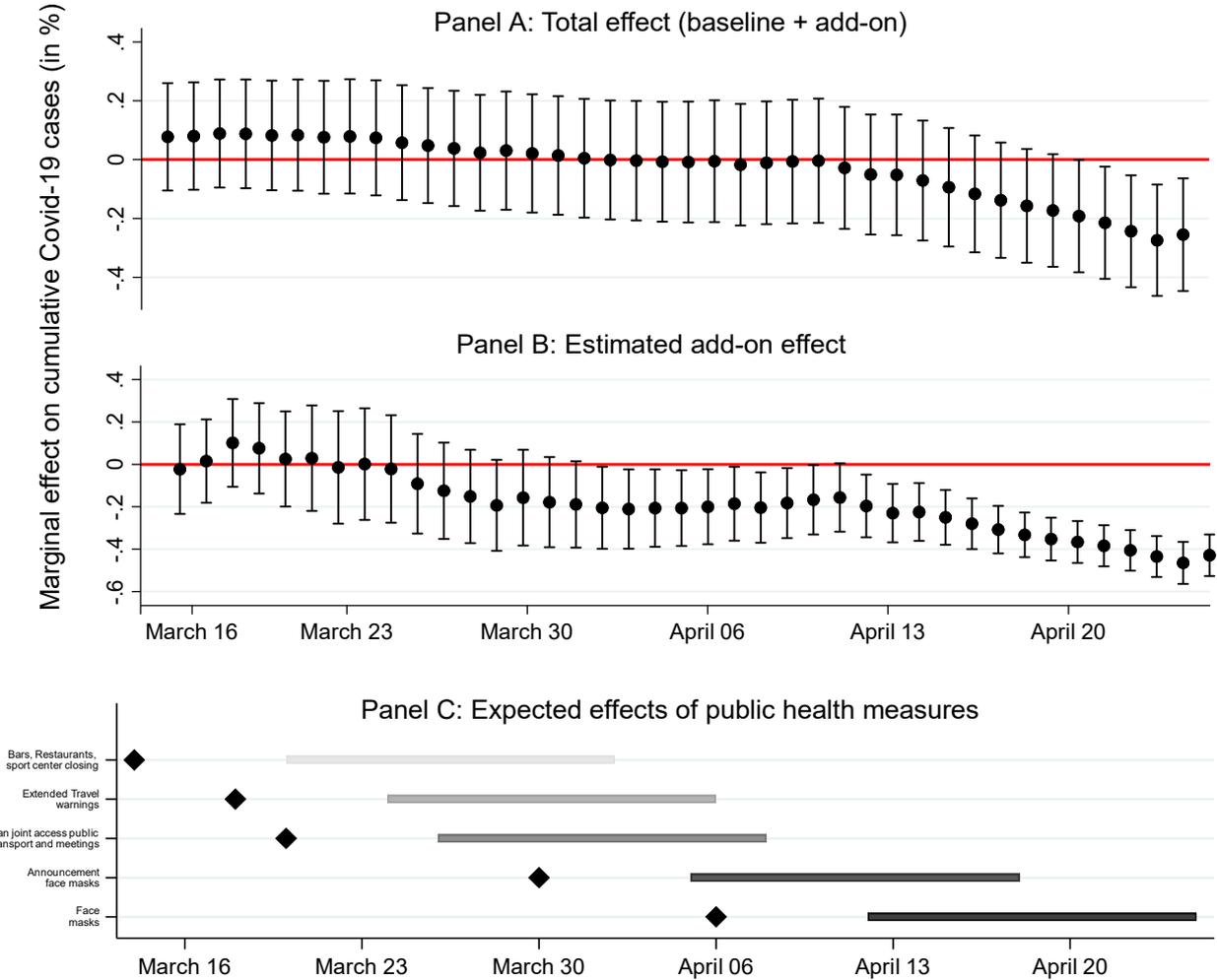


Figure S16: Estimated effects from incremental difference-in-difference (IDiD) model

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 lines 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.

The figure 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. As shown in Table S3 in SI Appendix, Section A.3, this covers more than 90% of cases associated with a specific date (i.e. timing of

public health measure). 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 S16 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 SI Appendix, Section 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/introduction of face masks.

Table S14: 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 SI Appendix section for details). In the FE-specification reported in Figure S16, 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 are in the zip-file on the Journal’s web page).

References for the SI 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/>, accessed on June 4, 2020.
4. U. Buchholz, S. Buda, K. Prahm, Abrupter Rückgang der Raten an Atemwegserkrankungen in der deutschen Bevölkerung. *Epidemiologisches Bulletin*. 10.25646/6636.2, 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* **66**, 115–133 (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 R_t from Coronavirus SARS-CoV-2 case data and implications for political measures based on prognostics. Available at <http://dx.doi.org/10.1101/2020.04.04.20053637> (deposited 2020).
8. T. Harko, F. S. 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. 1, 43–63. Available at: <https://cepr.org/sites/default/files/news/CovidEcon1%20final.pdf>.
10. A. Abadie, Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. [Forthcoming in] *Journal of Economic Literature*. 10.1257/jel.20191450.
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. IZA Discussion Paper No. 11255. <https://www.iza.org/publications/dp/11255>.
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. Available at: <https://www.iza.org/publications/dp/13160>.
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. E. Cavallo, S. Galiani, I. Noy, J. Pantano, Catastrophic Natural Disasters and Economic Growth. *Review of Economics and Statistics* **95**, 1549–1561 (2013).
19. N. F. Campos, F. Coricelli, L. Moretti, Institutional integration and economic growth in Europe. *Journal of Monetary Economics* **103**, 88–104 (2019).
20. K. Wälde, Corona-Blog, Einschätzung vom Freitag, 20. März 2020: <https://www.macro.economics.uni-mainz.de/corona-blog/>, accessed June 4, 2020.
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. INKAR, Indikatoren und Karten zur Raum- und Stadtentwicklung: Bundesinstitut für Bau-, Stadt- und Raumforschung (Federal Institute for Research on Building, Urban Affairs and Spatial Development). <https://www.inkar.de/>, accessed on June 4, 2020.
27. Google LLC, Google Trends. <https://trends.google.de/trends/>, accessed June 4, 2020 (deposited 2020).
28. Google LLC, COVID-19 Community Mobility reports: <https://www.google.com/covid19/mobility/>, accessed June 4, 2020 (deposited 2020).
29. Google LLC, COVID-19 Community Mobility Reports, Mobility Report CSV Documentation: https://www.google.com/covid19/mobility/data_documentation.html?hl=en, accessed on June 4, 2020.
30. Thüringer Allgemeine, Infektionsherde ließen Corona-Zahlen im Kreis Gotha steigen: <https://www.thueringer-allgemeine.de/regionen/gotha/infektionsherde-liessen-corona->

zahlen-im-kreis-gotha-steigen-id229070201.html, dated 8 May 2020, accessed September 9, 2020 (deposited 2020).

31. 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. Available at <http://dx.doi.org/10.1101/2020.03.25.20043109> (deposited 2020).
32. 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. Available at <http://dx.doi.org/10.1101/2020.04.18.20069955> (deposited 2020).
33. G. Bonaccorsi, F. Pierri, M. Cinelli, A. Flori, A. Galeazzi, F. Porcelli, A. L. Schmidt, C. M. Valensise, A. Scala, W. Quattrocioni, 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).
34. 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).
35. 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*. [10.1126/science.abb9789](https://doi.org/10.1126/science.abb9789) (2020).
36. C. Gros, R. Valenti, L. Schneider, K. Valenti, D. Gros, Containment efficiency and control strategies for the Corona pandemic costs. <http://arxiv.org/pdf/2004.00493v2> (deposited 2020).
37. T. Hartl, K. Wälde, E. Weber, Measuring the impact of the German public shutdown on the spread of Covid-19. *Covid Economics*. **1**, 25–32. Available at: <https://cepr.org/sites/default/files/news/CovidEcon1%20final.pdf>.
38. T. Hartl, E. Weber, Welche Maßnahmen brachten Corona unter Kontrolle? Available at: <https://www.oekonomenstimme.org/artikel/2020/05/welche-massnahmen-brachten-corona-unter-kontrolle/> (deposited 2020).
39. Chen, X. & Z. Qiu, Scenario analysis of non-pharmaceutical interventions on global Covid-19 transmissions. *Covid Economics*. **7**, 46–67. Available at: <https://cepr.org/sites/default/files/news/CovidEconomics7.pdf>.
40. A. Ritschl, Visualising and forecasting Covid-19. *Covid Economics*. **5**, 79–105. Available at: <https://cepr.org/sites/default/files/news/CovidEconomics5.pdf>.
41. 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. CESifo Working Paper No. 8166. Available at: <https://www.cesifo.org/node/53868> (deposited 2020).

42. 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).
43. 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).
44. 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. Medical Research Council (MRC). Available at <http://dx.doi.org/10.25561/77482> (deposited 2020).
45. Born, B., A. Dietrich & G. Müller, Do lockdowns work? A counterfactual for Sweden.: CEPR Discussion Paper No. 14744. Available at: https://cepr.org/active/publications/discussion_papers/dp.php?dpno=14744 (deposited 2020).
46. G. Seminara, B. Carli, G. Forni, S. Fuzzi, A. Mazzino, A. Rinaldo, Biological fluid dynamics of airborne COVID-19 infection. *Rendiconti Lincei. Scienze fisiche e naturali*. 10.1007/s12210-020-00938-2, 1–33 (2020).
47. 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).
48. K. A. Prather, C. C. Wang, R. T. Schooley, Reducing transmission of SARS-CoV-2. *Science*. 10.1126/science.abc6197 (2020).
49. 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. Preprints 2020040203. Available at <http://dx.doi.org/10.20944/preprints202004.0203.v1> (deposited 2020).
50. 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).
51. 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. Available at <http://dx.doi.org/10.21203/rs.3.rs-16701/v3> (deposited 2020).
52. 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).

53. 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).
54. 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).
55. 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).
56. Fuchs-Schündeln, Nicola and Kuhn, Moritz and Tertilt, Michèle, The Short-Run Macro Implications of School and Child-Care Closures. IZA Discussion Paper No. 13353, Available at: <https://www.iza.org/publications/dp/13353> (deposited 2020).
57. A. Basurto, H. Dawid, P. Harting, J. Hepp, D. Kohlweyer, Economic and epidemic implications of virus containment policies: insights from agent-based simulations. Universität Bielefeld Working Papers in Economics and Management; 05-2020. Available at <http://dx.doi.org/10.4119/unibi/2944282> (deposited 2020).
58. Stannard, T. and G. Steven and C. McDonald, Economic impacts of COVID-19 containment measures. Reserve Bank of New Zealand Analytical Notes AN2020/04, Reserve Bank of New Zealand. <https://www.rbnz.govt.nz/-/media/reservebank/files/publications/analytical%20notes/2020/an2020-04.pdf> (deposited 2020).
59. A. Strong, J. Welburn, An Estimation of the Economic Costs of Social-Distancing Policies. Santa Monica, CA: RAND Corporation, (2020). Available at <http://dx.doi.org/10.7249/RRA173-1> (deposited 2020).
60. J. Fernández-Villaverde, C. Jones, Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities. NBER Working Paper #27128. Available at <http://dx.doi.org/10.3386/w27128> (deposited 2020).
61. Brotherhood, Luiz and Kircher, Philipp and Santos, Cezar and Tertilt, Michèle, An Economic Model of the Covid-19 Epidemic: The Importance of Testing and Age-Specific Policies. IZA Discussion Paper No. 13265. Available at: <https://www.iza.org/publications/dp/13265> (deposited 2020).
62. M. Eichenbaum, S. Rebelo, M. Trabandt, The Macroeconomics of Testing and Quarantining. NBER Working Paper #27104. Available at <http://dx.doi.org/10.3386/w27104> (deposited 2020).
63. D. Acemoglu, V. Chernozhukov, I. Werning, M. Whinston, Optimal Targeted Lockdowns in a Multi-Group SIR Model. NBER Working Paper 27102, National Bureau of Economic Research. Available at <http://dx.doi.org/10.3386/w27102> (deposited 2020).

64. M. Eichenbaum, S. Rebelo, M. Trabandt, Epidemics in the Neoclassical and New Keynesian Models. NBER Working Paper #27430. Available at <http://dx.doi.org/10.3386/w27430> (deposited 2020).
65. M. Eichenbaum, S. Rebelo, M. Trabandt, The Macroeconomics of Epidemics. NBER Working Paper #26882. Available at <http://dx.doi.org/10.3386/w26882> (deposited 2020).
66. Statistisches Bundesamt, Table 63121-0001, Einkommen und Einnahmen sowie Ausgaben privater Haushalte (Laufende Wirtschaftsrechnungen): Deutschland, Jahre, Haushaltsgröße, Laufende Wirtschaftsrechnungen: Haushaltsbuch. Available at: <https://www-genesis.destatis.de/genesis/online>, accessed on September 30, 2020.
67. P. Dolton, C. R. Bondibene, J. Wadsworth, The UK National Minimum Wage in Retrospect. *Fiscal Studies* **31**, 509–534 (2010).
68. A. Goodman-Bacon, J. Marcus, Using Difference-in-Differences to Identify Causal Effects of COVID-19 Policies. *Survey Research Methods*. 10.18148/srm/2020.v14i2.7723, 153–158 (2020).
69. B. Ferman, C. Pinto, Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity. *Review of Economics and Statistics* **101**, 452–467 (2019).
70. J. Heckman, H. Ichimura, J. Smith, P. Todd, Characterizing Selection Bias Using Experimental Data. *Econometrica* **66**, 1017 (1998).

Content of SI Appendix

A	Timing of public health measures and visibility in data	2
A.1	Timing of the introduction of mandatory face masks	2
A.2	The timing of other public health measures	3
A.3	When are effects of public health measures visible in the data?	5
A.4	Visibility in data II – Conceptual background	8
B	Synthetic control method: Design, implementation and inference	9
C	Data description and additional SCM estimation results for Jena	13
C.1	Summary statistics for outcome and predictor variables	13
C.2	Trajectories of cumulative Covid-19 cases and box plots for predictor variables	14
C.3	Control regions, SCM weights and pre-treatment predictor balance	16
C.4	SCM results by age groups	17
C.5	Effects on cumulative number of infections per 100,000 inhabitants	19
C.6	Announcement and mobility	20
C.6.1	Google trends and announcement effects	20
C.6.2	Mobility trends across German federal states	20
C.7	Cross validation and additional placebo-in-time test	22
C.8	Changes in donor pool for synthetic Jena	23
C.9	Place-in-space tests for other major cities in Thuringia	25
D	Single and multiple treatment analyses	26
D.1	Single treatment analysis in other German cities and regions	26
D.2	Growth rates and summary of single and multiple treatment effects	28
E	A brief survey of research on public health measures against Covid-19	29
E.1	General overview	29
E.2	Evidence for face masks	30
E.3	The economic costs of public health measures	31
F	Difference-in-difference estimates for timing of treatment effects	32
	References for the SI Appendix	35

time index	reported infections				reported infections				reported infections			
	cumulative numbers		daily growth rates		cumulative numbers		daily growth rates		cumulative numbers		daily growth rates	
	Jena				Multiple Treatment all				Multiple Treatment Krsf. Städte			
	Treated	Control	Treated	Control	Treated	Control	Treated	Control	Treated	Control	Treated	Control
70	142	143			224,9375	225,17416			262	262		
71	147	149	3,5%	4,1%	227,3125	229,38291	1,1%	1,9%	263	271	0,5%	3,1%
72	151	154	2,7%	3,7%	231,4688	232,98772	1,8%	1,6%	267	274	1,6%	1,3%
73	154	159	2,0%	3,1%	234,5625	236,68691	1,3%	1,6%	270	279	0,9%	1,8%
74	154	163	0,0%	2,4%	236,6875	239,61191	0,9%	1,2%	272	282	0,8%	1,2%
75	154	166	0,0%	1,8%	238,8438	241,68959	0,9%	0,9%	273	285	0,6%	0,8%
76	155	167	0,6%	0,8%	240,1875	244,18825	0,6%	1,0%	275	287	0,5%	0,7%
77	155	168	0,0%	0,5%	242	246,08966	0,8%	0,8%	276	288	0,4%	0,4%
78	155	169	0,0%	1,0%	243,5	248,41103	0,6%	0,9%	278	290	0,8%	0,6%
79	155	173	0,0%	2,0%	245,5625	251,939	0,8%	1,4%	279	292	0,5%	0,9%
80	155	178	0,0%	2,7%	247,875	254,08616	0,9%	0,9%	281	294	0,4%	0,7%
81	155	182	0,0%	2,4%	249,5	256,43878	0,7%	0,9%	282	297	0,4%	1,0%
82	155	184	0,0%	1,2%	250,9063	257,57853	0,6%	0,4%	282	299	0,0%	0,4%
83	155	191	0,0%	3,6%	252	259,05481	0,4%	0,6%	282	300	0,1%	0,4%
84	155	192	0,0%	0,7%	253,2813	260,33541	0,5%	0,5%	282	302	0,1%	0,8%
85	155	194	0,0%	1,2%	253,8125	261,913	0,2%	0,6%	282	308	0,0%	1,8%
86	155	199	0,0%	2,4%	256,2813	264,42047	1,0%	1,0%	284	311	0,7%	1,0%
87	157	201	1,3%	1,0%	258,0938	266,44825	0,7%	0,8%	286	314	0,5%	0,9%
88	157	203	0,0%	1,2%	259,5938	267,93159	0,6%	0,6%	287	315	0,5%	0,5%
89	158	205	0,6%	0,7%	261,6563	268,90247	0,8%	0,4%	289	317	0,5%	0,6%
90	158	205	0,0%	0,0%	262,7188	269,70916	0,4%	0,3%	290	318	0,3%	0,3%
15	1,09	1,36	growth factor 14 days		1,13	1,16			1,08	1,17		
Growth_rate_15	0,59%	2,07%	1,49% daily growth rate		0,81%	1,01%	0,20%		0,50%	1,06%	0,56%	
		28,2%				79,83%				47,42%		
20	1,11	1,43	growth factor 20 days		1,17	1,20	growth factor 20 days		1,11	1,21	growth factor 20 days	
Growth_rate_20	0,54%	1,82%	1,28% daily growth rate		0,78%	0,91%	0,13%		0,51%	0,96%	0,46%	
		29,4%				86,0%				52,7%		

difference daily growth rate between treated and control
reduction of growth rate to this level

Difference between treated region(s) and synthetic control group(s)	multiple treatments	
	Jena	(all) (cities)
Absolute change in cumulative number of Covid-19 cases over 20 days	-46,9	-7,0
Percentage change in cumulative number of Covid-19 cases over 20 days	-22,9%	-2,6%
Percentage change in newly registered Covid-19 cases over 20 days	-75,6%	-15,7%
Difference in daily growth rates of Covid-19 cases in percentage points	-1,28%	-0,13%
Reduction in daily growth rates of Covid-19 cases (in percent)	70,6%	14,0%