Manuscript submitted to *The Econometrics Journal*, pp. 1–26.

The spread of COVID-19 and the BCG vaccine: A natural experiment in reunified Germany

Richard Bluhm*,† and Maxim Pinkovskiy‡

*Leibniz University Hannover, Institute of Macroeconomics, 30167 Hannover, Germany
†University of California San Diego, Department of Political Science, La Jolla, CA 92093
E-mail: bluhm@mak.uni-hannover.de
‡Federal Reserve Bank of New York, Microeconomic Studies Function, New York, NY 10045
E-mail: maxim.pinkovskiy@ny.frb.org

Summary The "BCG hypothesis" suggests that the Bacillus Calmette-Guérin (BCG) vaccine against tuberculosis limits the severity of COVID-19. We exploit the differential vaccination practices of East Germany and West Germany prior to reunification to test this hypothesis. Using a differences in regression discontinuities (RD-DD) design centred on the end of universal vaccination in the West, we find that differences in COVID-19 severity across cohorts in the East and West are insignificant or have the wrong sign. We document a sharp cross-sectional discontinuity in severity of the disease, which we attribute to limited mobility across the long-gone border and which disappears when we control for social connectedness. Case and death data after the end of the first lockdown on April 26 does not display a discontinuity at the former border, suggesting that mobility (as opposed to BCG vaccination) played a major role during the initial outbreak.

Keywords: *COVID-19, BCG vaccine, Germany, mobility, SIR model with commuting flows*

1. INTRODUCTION

Since December 2019, the disease caused by the novel coronavirus (COVID-19) has infected over 72 million people worldwide, of whom over 1.6 million have died.[1] The pandemic has produced an unprecedented decline in global economic activity as countries repeatedly implement social distancing measures to contain the spread of the virus.

The development of targeted vaccines to combat COVID-19 is occurring at record speed. While first vaccines, which appear to be safe and effective,[2] are now available to risk groups in some countries, vaccinating large parts of the population in developed and developing countries is expected to take at least 1-2 years. This lack of a widely available vaccine has sparked considerable interest in whether some types of vaccines that are already known to be safe may have positive indirect effects on the spread and severity of COVID-19 infections. Specifically, there is now a lively controversy over whether the Bacillus Calmette-Guérin (BCG) vaccine against tuberculosis also partially protects individuals against COVID-19. Several studies point out that countries with mandatory BCG vaccination tend to have substantially fewer coronavirus cases and deaths per capita than countries without mandatory vaccination, and that the intensity of the epidemic is lower for countries that began vaccinating earlier (Berg et al., 2020; Escobar et al.,

[1] As of December 14 2020, based on data from coronavirus.jhu.edu.

[2] The US Food and Drug Administration (FDA) licensed the first COVID-19 vaccine that has completed phase 3 trials on December 11, 2020.

2020; Gursel and Gursel, 2020; Hauer et al., 2020; Sharma et al., 2020). Non-specific or off-target effects of live vaccines are not uncommon and have been documented in a variety of settings (Kleinnijenhuis et al., 2015; Chumakov et al., 2020). The BCG vaccine appears to protect against its target, some forms of tuberculosis, for up to 60 years (Aronson et al., 2004) but has also been associated with long-term reductions in all-cause mortality and mortality from respiratory diseases (Rieckmann et al., 2016). Live vaccines appear to elicit a strong response of the immune system, which subsequently offers broad protection against unrelated pathogens (Chumakov et al., 2020). Similar 'trained immunity' responses against COVID-19 in BCG vaccinated individuals are therefore plausible (O'Neill and Netea, 2020), but there is a lack of rigorous evidence investigating these non-specific effects. Clinical trials are taking place across the globe which test the effectiveness of the BCG vaccine against COVID-19.[3] These trials are likely to take at least a year while the virus continues to spread at a rapid pace. The WHO cautions that there is currently no evidence that the BCG vaccine protects against the novel coronavirus (Curtis et al., 2020).

In the absence of experimental results, we propose to use the tools of modern applied econometrics to test the hypothesis that BCG may offer long-run protection against COVID-19. We use a regression discontinuity differences-in-differences (RD-DD) analysis of severe coronavirus cases to exploit a natural experiment along the former border between East Germany and West Germany. This border separated the two sides from 1949 until Germany was reunified in October 1990. West Germany phased out a policy of *de facto* universal BCG vaccination (which began in the 1950s) for the general population starting in 1974, while East Germany strictly enforced a policy of mandatory BCG neonatal vaccination from 1953 until 1990. Moreover, neonatal vaccination was completely interrupted in West Germany in 1975 and 1976 when the only licensed vaccine caused unexpected side effects and had to be withdrawn from the market (Genz, 1977). Residual BCG vaccination of risk groups ended permanently in reunified Germany in 1998. This gives rise to a natural experiment that creates variation over time and space. Individuals born shortly after 1974 just to the west of the former border experienced a sharp drop in BCG coverage relative to those born shortly before 1974, while their peers just to the east were consistently vaccinated throughout (at rates exceeding 95%).

Our analysis takes advantage of several important features that are present in the German context. First, there is an internal border sharply dividing different vaccination regimes for a limited period of time. Second, the German health ministries of each state report detailed records to the Robert Koch Institute (RKI), which disseminates standardized data on COVID-19 cases by geographic location and by single year of age. The latter allows us to compare individuals in close age cohorts who are also close to each other in physical space, and therefore differ less on unobservable characteristics. This alleviates the concerns raised by Becker et al. (2020) about cross-sectional comparisons and discontinuity designs between East and West Germany. While differences in BCG vaccination rates across space (Hauer et al., 2020) or cohorts have been used in related studies (de Chaisemartin and de Chaisemartin, 2020; Hamiel et al., 2020), to our knowledge this is the first paper combining geographic *and* age variation in BCG vaccination

[3]For example, the BRACE trial in Australia (`NCT04327206`), the BADAS trial in the United States (`NCT04348370`), or the BCG-CORONA trial in the Netherlands (`NCT04328441`). Clinicaltrials.gov identifiers are provided in parentheses. However, most of these trials are small scale and may lack sufficient power to provide a definitive answer.

status for reliable causal inference. A third feature of the German reporting system is that every test result submitted to the RKI indicates a case definition, allowing us to distinguish cases with acute respiratory symptoms (including pneumonia), or patients who have died from COVID-19, from benign cases without acute symptoms. Fourth, areas of Germany on both sides of the border have been subject to the same state response to the COVID-19 pandemic and have comparable access to medical services, creating a high level of homogeneity in pandemic policy. Finally, Germany publishes detailed data on county-by-county commuter flows, allowing us to investigate a potentially important factor in the transmission of COVID-19.

Our RD-DD results contradict the BCG hypothesis. We find that individuals born just to the east of the former border shortly *after* 1974 are, if anything, more likely to have a reported case of COVID-19 relative to their peers in the west than individuals born just to the east of the border shortly *before* 1974. Our estimates for the potential "effect" of the BCG vaccine consistently have the wrong sign and typically cannot be distinguished from zero. We perform covariate balancing tests to rule out that other outcomes, such as mortality and respiratory hospitalizations, change differentially between East and West Germany for cohorts born after rather than before 1974. Performing a less restrictive differences-in-differences (DD) analysis between all of East and West Germany, as well as between cohorts born before and after 1974, yields similar results. We also obtain comparable results when we focus on severe cases only. Analogous RD-DD and DD analyses in 1990—the other critical date in the history of the BCG vaccine in Germany—deliver even stronger estimates of the differential effect that run counter to the BCG hypothesis, although their covariate balancing tests are not as stable. Taken together, these findings cast serious doubt that the correlations adduced by the literature supporting the BCG hypothesis capture a causal relationship (Berg et al., 2020; Escobar et al., 2020; Gursel and Gursel, 2020; Hauer et al., 2020; Sharma et al., 2020) and, instead, support the findings of studies using temporal differences in BCG vaccination in the same country (de Chaisemartin and de Chaisemartin, 2020; Hamiel et al., 2020).

Our analysis reveals a puzzle. If we ignore the variation over time, we observe a sharp discontinuity in COVID-19 cases at the former border separating West and East Germany. There are considerably fewer cases and deaths in the former East. If it is not the BCG vaccine, then what explains this jump?[4] Our solution to this puzzle suggests that while the virus does not stop at the long-gone border, people who carry the virus still do. We find that controlling for social connectedness as proxied by cross-county Facebook connections significantly decreases the discontinuities in COVID-19 cases at the border and makes them statistically insignificant. This echoes the findings of a recent literature on border controls and travel restrictions (e.g., Chinazzi et al., 2020; Eckardt et al., 2020). Going further, we document that long-range commuter flows in Germany are sharply discontinuous at the former border, which reflects a continued lack of connectedness between West German and East German counties. We demonstrate that discontinuities in commuter flows can generate discontinuities in COVID-19 cases at the former border without any reference to the BCG hypothesis. We simulate a spatial SIR model of viral

[4]Prominent newspapers in Germany have noticed that there is much lower COVID-19 prevalence in the former East than in the West but offer only suggestive explanations (e.g., *Die Zeit*, a German weekly, `www.zeit.de/2020/13/coronavirus-ausbreitung-osten-westen-faktoren`, or *Der Tagesspiegel*, a Berlin-based German daily, `www.tagesspiegel.de/politik/mehr-flaeche-mehr-alte-warum-der-osten-weniger-unter-corona-leidet/25796940.html`). Moreover, low COVID-19 mortality in Germany as a whole has been the subject of media interest.

spread across German counties where commuter flows act as a transmission channel (as in Wesolowski et al., 2017). We show that even within this very stylized framework, we can obtain a sizable discontinuity in COVID-19 intensity without incorporating anything relating to vaccination into the model. Moreover, we show that the discontinuities in cases and severe cases weaken over time. In fact, we do not observe significant discontinuities in new cases occurring after the initial outbreak in spring. We interpret this as additional evidence in favor of mobility playing a key role in "seeding" the early distribution of the outbreak, as opposed to the spread being contingent on innate characteristics of the population. Finally, we consider evidence for a discontinuity in COVID-19 deaths as a share of the underlying population. We find no statistically significant evidence that the pandemic produced a discontinuity in the logarithm of deaths at the former border, no matter if we consider data until the end of the first wave or until mid-December 2020.

The remainder of the paper is organized as follows. Section 2 describes the county-level data on cases, covariates, and commuter flows. Section 3 outlines our empirical strategy. Section 4 presents the balancing tests and main RD-DD results for cases by age-groups. Section 5 explores mobility as an alternative explanation for the sharp discontinuity in overall cases, presents results from placebo tests with cases simulated from a county-level SIR model with mobility, and examines the evidence on deaths rather than cases. Section 6 concludes.

2. DATA

The COVID-19 pandemic in Germany can be characterized by three distinct phases: *i)* an initial outbreak in spring which began in late February and lasted about four weeks into the lockdown on 23 March 2020 until about April 26 2020, *ii)* a slow resurgence of coronavirus cases as travel restrictions were lifted over the summer, and *iii)* a second wave of infections and mobility restrictions in the winter.[5] For most of the analysis, we focus on the first period between the start of the pandemic in Germany over the peak with more than 6,000 daily cases until new infections dropped again to less than 2,000 cases per day, but we also investigate new cases and deaths from April 26 2020 until December 13 2020, the day before Germany imposed another lockdown to confront the autumn wave of the pandemic.

Our main dependent variable is the logarithm of one plus the number of cumulative reported COVID-19 cases per million people in a German county (*Kreis*) as of April 26, 2020. Germany reported a total of 158,047 COVID-19 cases and 8,122 deaths by this date. We obtain counts of cumulative and new COVID-19 cases and deaths by German county for every date since January 27 2020 (the start of week 5) from the RKI's *NPGEO Corona Hub 2020*. Figure 1 shows a map of cumulative cases by county as of April 26 2020 (the end of week 17). Given the difficulties in recording asymptomatic cases of COVID-19, it is all but certain that the case counts we employ in this paper are a lower bound for COVID-19 cases in Germany. In the period of observation, testing capacity was high but testing was limited to individuals displaying symptoms, returnees from high risk countries, and those who have been in contact with confirmed cases. We proceed on the necessary assumption that undercounting errors do not vary systematically across age groups and locations.

The RKI also separately reports case data by single-year age groups and week via the

[5]Figure S.3. provides a time-series of daily cases.

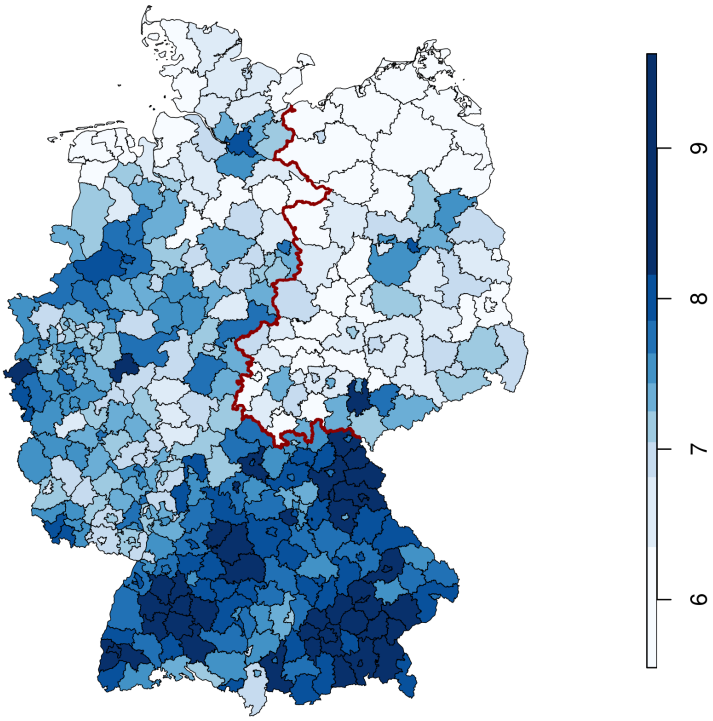


Figure 1. Spatial distribution of COVID-19 cases in Germany as of April 26 2020. The map shows the log(1 + cases/million people) in each county. The former border is marked in red. Darker (lighter) shades of blue indicate more (fewer) COVID-19 cases.

platform *SurvStat@RKI 2.0*. By federal law, all cases reported to the RKI have to follow a case definition (*Falldefinition*).[6] This categorization allows us to distinguish *severe* cases which have been confirmed by a PCR test and display acute symptoms (about 63% of all cases)—ranging from respiratory distress over pneumonia to death—from cases which do not display any clinical symptoms or are entirely asymptomatic (26%), and from cases whose clinical status is unknown (11%).[7] The high share of severe cases in overall cases is consistent with the notion that the cumulative case count by and large indicates COVID-19 severity rather than just incidence. We use both overall and severe cases throughout the analysis. Deaths by single-year age groups and week are not reported by the RKI.

In addition to COVID-19 cases, we compile several other data sources at the county-by-age level to use in balance tests. We collect data on overall mortality, mortality from selected infectious diseases, mortality from respiratory diseases, and similar statistics for hospitalizations in 2016 from the German federal statistical office (via the *GENESIS* regional database). We also obtain data on labour force participation, unemployment, full-time employment, public employment and commuting to work from the full 2017

[6]For a full description (in German), see `www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Falldefinition.pdf?__blob=publicationFile`.

[7]Figure S.4. includes a figure of the age distribution of cases and severe cases for the territories of the former West and East.

Microcensus (a 1% sample of the German population) at the county-by-age level.[8] We also use the latest available data on commuting flows published by the Federal Employment Agency from December 2019. The agency regularly releases an origin-destination matrix of commuting flows across German counties. These flows capture about 33 million jobs. Approximately 13 million of the jobs are in a different county than the primary tax residence of the employee. To focus on long-run commuting patterns, we eliminate origin-destination pairs whose centroids are less than 50 km apart, and for which the rail and highway network plays a smaller role. Finally, we use data on Facebook connections between individuals living in different counties, which has been made publicly available by Facebook for researchers interested in studying the COVID-19 pandemic. The provided scores can be interpreted as scaled probabilities that two randomly selected Facebook users from the different counties are Facebook friends (Bailey et al., 2018).

3. EMPIRICAL STRATEGY

Our study exploits discontinuous changes in vaccination policies across the former border dividing East and West Germany from 1949 until 1990, with a focus on the cessation of widely recommended BCG vaccination after 1974 in the West.

3.1. Identifying variation

Even though tuberculosis was widespread among the war-ravaged population, Germany had a non-vaccination policy until the end of World War II and did not join Red Cross-led vaccination campaigns in the early post-war years. This decision was in part due to the "Lübeck vaccination disaster" in which 251 infants were vaccinated with a BCG vaccine contaminated with live tuberculosis bacteria. Almost all children fell ill with tuberculosis and 72 died, leading the Interior Ministry to reject BCG vaccination as unsafe in 1930 (Loddenkemper and Konietzko, 2018).

BCG policies then diverged quickly when the country was divided. In 1953, the German Democratic Republic (GDR) introduced mandatory vaccination for a variety of diseases, including the BCG vaccine against tuberculosis. Enforcement in the GDR was strict. "From 1954 on, school children who had not yet been vaccinated had to present a letter of exemption not only from their parents but also from a physician" (Harsch, 2012, p. 420). Vaccinations substantially outstripped newborns in the early 1950s, suggesting that young adults born before the GDR existed were also vaccinated ex post (Kreuser, 1967). The policy lasted until the collapse of the GDR in 1990.

The Federal Republic of Germany (FRG) only required mandatory vaccination for smallpox from 1949 until the end of 1975. The BCG vaccine was highly recommended but administered on a voluntary basis. In practice, vaccination of newborns was near universal by the mid-1960s. Due to the decentralized nature of the West German health care system, the initial roll out of the vaccination policy varied by state in the 1950s. By 1964, practically all newborns in West Germany were BCG vaccinated shortly after birth (Kreuser, 1967). In 1974, the policy was changed to vaccinate only children in risk groups and, in May 1975, the BCG vaccine was temporarily withdrawn from the market in the West, when a new vaccine compatible with WHO guidelines was discovered to have unintended side-effects. Neo-natal vaccination practically ceased for two years and

[8]The Microcensus data is made available for remote or on-site processing by the Research Data Centres (*Forschungsdatenzentrum*) of the Federal Statistical Office and the Statistical Offices of the federal states.

Table 1. Timeline of vaccination policies in both parts of Germany, 1949 until today

Year	West Germany (FRG)	East Germany (GDR)
1949		First BCG vaccinations
1951-52		Extended program with GDR manufactured BCG vaccine
1953	BCG vaccine is licensed	Mandatory vaccination (with refresher), target rate at least 95%
1955	Recommendation to vaccinate all newborn children	
Mid 1960s	Near universal vaccination of newborns	Near universal vaccination of newborns
1974	National recommendation to only vaccinate children in risk groups, some states continue to recommend universal vaccination of newborns	
1975	BCG vaccination temporarily halted for two years	
1983	Further restriction to only those children that have TB in the family	
1988	Vaccine recommended only for children that tested negative for tuberculin and are risk groups	
1990	Reunification, policies of FRG continue	Reunification, policies of FRG apply
1998	Vaccination no longer recommended, risk groups are no longer vaccinated, last states drop recommendation to vaccinate	

Note: Based on Klein et al. (2012), Klein (2013), and various sources cited in the text.

tuberculosis incidence among newborns doubled (Genz, 1977). Voluntary vaccination of risk groups continued thereafter until 1998 (Robert Koch-Institut, 1976, 1998) but vaccination was no longer universal in West Germany from 1975 onward or in reunified Germany after 1990. Some states continued to recommend universal vaccination until 1998 but with low compliance and confusion among parents about which recommendations apply (Danner and Qast, 1995). Currently, no BCG vaccine is licensed in Germany. We summarize these changes in Table 1.

3.2. Regression discontinuity design

Ignoring variation across cohorts, we can estimate the discontinuity between East and West Germany in COVID-19 intensity, y_c, by running a standard geographic regression discontinuity (RD) design (see e.g. Lalive, 2008; Dell, 2010; Keele and Titiunik, 2015):

$$y_c = \alpha + \beta \textsc{East}_c + \delta_1 d_c + \delta_2 \left(d_c \times \textsc{East}_c\right) + \lambda_{s(c)} + \varepsilon_c \quad \text{for} \quad |d_c| < b \tag{3.1}$$

where c indexes counties (*Kreise*), EAST_c indicates whether the county was part of East Germany before reunification, d_c is the distance of county c from the former East German border (it is negative if $\text{EAST}_c = 0$ and positive if $\text{EAST}_c = 1$), and $\lambda_{s(c)}$ is a fixed effect for the border segment associated with county c. Border segments are defined by breaking up the former border into five consecutive segments of about 265 km each.[9] We drop Berlin throughout the analysis, as it was an enclave with its own internal border between East and West, and focus on the border which used to separate the two larger countries.

The coefficient of interest in this design, β, does *not* identify the effect of the BCG vaccine on COVID-19 severity. Instead, it captures a compound treatment effect of discontinuous changes in other variables that are related to being born in East Germany and directly affect COVID-19 transmission. Put simply, East Germany differed from West Germany in many more ways than BCG vaccination (see e.g., Becker et al., 2020). We only use β as a reference to gauge the size of the baseline discontinuity for all ages or for specific cohorts.

To isolate the effect of BCG vaccination on COVID-19 severity, we estimate the *difference in discontinuities* at the border for cohorts born just before and just after West Germany suspended widespread recommendation of the BCG vaccine in 1974. We call this a regression discontinuity differences-in-differences (RD-DD) design (see e.g., Deshpande, 2016), which we estimate using the following specification:

$$\begin{aligned} y_{c,a} = \; & \alpha_a + \beta \text{EAST}_c + \gamma \text{EAST}_c \times \text{TREATED}_a + \delta_1 d_c + \delta_2 d_c \times \text{EAST}_c + \\ & \delta_3 d_c \times \text{TREATED}_a + \delta_4 d_c \times \text{EAST}_c \times \text{TREATED}_a + \lambda_{s(c),a} + \varepsilon_{c,a} \quad \text{for} \quad |d_c| < b \end{aligned} \tag{3.2}$$

where a indexes age groups (pre- and post-vaccination change), TREATED_a is an indicator for the age group for which the policy experiment created no differential in vaccination status across the border (here it is the age group born before the vaccination policy change), and the intercept and the border segment fixed effects are allowed to vary by age group. The rest of the notation is the same as before. The coefficient of interest, γ, delivers an estimate of the difference in discontinuities across cohorts.

The assumption needed for this specification to recover a causal effect of BCG vaccination on the outcome variable is that any discontinuities across the former border—other than the effect of the vaccine—are constant across cohorts born shortly before and shortly after the change in vaccination regime. As it is not apparent that anything else happened in 1974 that would affect newborns differently in East and West Germany (e.g. no other vaccines were introduced or withdrawn) we view this as a plausible assumption.[10] Moreover, we can test this assumption by looking at other outcomes for successive cohorts around 1974.

Equation (3.2) can also be viewed as nesting a standard differences-in-differences specification, in which we compare outcomes in all of East Germany with those in all of West Germany, and for cohorts born just before or just after 1974. In the framework above, such a specification would set the bandwidth b equal to infinity, and the coefficients on the distance terms and the border segment fixed effects to zero.

[9] Another way to define border segments would have been by using pairs of bordering states. However, this division produces unevenly sized clusters, with some states sharing a border at only one pair of counties.

[10] The closest policy change is the near synchronous end of universal smallpox vaccination in 1983 in West Germany and 1982 in East Germany (Klein et al., 2012; Klein, 2013).

3.3. Optimal bandwidth selection and bias correction

An integral part of regression discontinuity analysis is choosing an appropriate bandwidth for estimation and inference. For the RD specifications, we follow Calonico et al. (2020) and present our baseline estimates using inference-optimal bandwidth choices that minimize the coverage error rate (CER) of the robust bias-corrected confidence interval. We also report estimates using bandwidths that minimize the mean squared error (MSE) of the RD point estimate (Calonico et al., 2014) which have become the *de facto* standard in the literature. Additionally, we present results for two fixed bandwidths (30 km and 60 km), which are easy to interpret and bracket most of the optimal bandwidths selected by the two data-driven procedures.

Calonico et al. (2014) show that all RD estimates have finite-sample bias that may be separately estimated (using an undersmoothed bandwidth) and corrected for, with corresponding corrections to asymptotic standard errors to take into account additional variability from the bias correction. We bias-correct all RD estimates and present robust confidence intervals throughout our RD analysis. Rules for optimal bandwidth selection and bias correction have not yet been developed for RD-DD designs. We proceed by calculating CER-minimizing bandwidths for each group in the RD-DD and then use these bandwidths directly in equation (3.2). The CER-bandwidths are undersmoothed relative to their MSE-optimal counterpart, allowing us to use conventional confidence intervals without bias-correction in the RD-DD design.

3.4. Bootstrap inference for few clusters

Germany is a federal country in which most public health policy, including BCG vaccination, COVID-19 containment measures, and the reporting of cases and deaths through local health authorities, is the responsibility of the states (*Bundesländer*). While this is the natural level at which the errors should be clustered, there are only 16 states in Germany (and only six in former East Germany). Relatively narrow bandwidths will then select a subset of those states, leading to the concern that asymptotic approximations for standard errors clustered on states may be inaccurate.

We address this concern by estimating wild cluster bootstrap (WCB) confidence intervals (Cameron et al., 2008). MacKinnon and Webb (2017) show that the WCB can fail in situations where there are only few treated clusters. Our setting, however, is one where the WCB typically performs well (as the number of clusters is not too small and about half of the units are treated). For the RD-DD design, we report conventional 95% confidence intervals obtained from bootstrap DGPs where the null of no difference in discontinuities ($\gamma = 0$) has been imposed.

For the RD results, we use a variant of the wild bootstrap for robust bias-corrected confidence intervals proposed by Bartalotti et al. (2017), which uses second-order local polynomials to estimate the bias of the RD estimator and confidence interval of the bias-corrected RD estimator. Our version accommodates clustering (as in He and Bartalotti, 2020), additional covariates (Calonico et al., 2019), Webb's six-point distribution[11] as wild weights (Webb, 2014), and parallel computation. We present WCB versions of the

[11] Wild bootstrap schemes typically use 2-point distributions as weights for the original residuals. This implies that there are only 2^G unique samples, where G is the number of groups (clusters). Some of our specifications are based on as few as 8 clusters or 256 unique samples. A 6-point distribution expands the universe of unique samples and performs well in simulations with few clusters (Webb, 2014).

95% bias-corrected confidence intervals. As with the RD-DD design, the procedure takes the bandwidth choices as given, which we either fix or calculate prior to the bootstrap using asymptotic data-driven methods for the same specification and sampling scheme.

4. RESULTS

4.1. Balancing tests

We start by presenting evidence that several other outcomes for cohorts born just before and just after 1974 are similar, suggesting that the RD-DD design is not capturing the effects of differences between these cohorts unrelated to BCG vaccination. Table 3 presents these balancing tests. Unfortunately, we do not have data on alternative outcomes by single year of age. However, we have mortality and hospitalization rates in 5-year age bins derived from administrative data in 2016 and we construct labour market indicators, also in 5-year age bins, from the 2017 Microcensus.[12]

Table 2 presents the coefficient γ from equation 3.2 when the logarithms of unemployment rates, labor force participation rates or other labor market indicators are used as dependent variables in the RD-DD design. Since we observe the reported birth year of each individual in the 2017 Microcensus, we construct exact 5- and 10-year age bins around birth years in 1974 and 1990, for which we perform the balancing tests. Columns 1 and 2 present balancing tests for the 1974 policy change, whereas columns 3 and 4 present these tests for reunification in 1990. Table 3 presents the coefficient γ from equation 3.2 when the logarithms of mortality or hospitalization rates are used as dependent variables in the RD-DD design. Unlike the Microcensus data, the mortality and hospitalization data comes in predefined 5-year age bins, preventing us from building these bins around the policy change years. Instead, we omit the 40–44 year old group (since they were born between 1972 and 1976) as we cannot classify them neatly as treated or untreated relative to the 1974 vaccination suspension in the West. In column 1 we take the 5-year age group immediately older than the omitted group as the "treated" group and the 5-year age group immediately younger than the omitted group as the "untreated" group. For example, in the case of mortality this compares 45–49 year olds to those that are 35–39. In column 2 we take the two 5-year age groups immediately older than the omitted group as the "treated group" and construct the "untreated group" symmetrically (e.g., 45–54 year olds versus 30–39 year olds). All other cohorts are omitted from the analysis. Columns 3 and 4 present analogous balance tests for the cohorts born around 1990, where the age cohorts for analysis are chosen similarly.

For each balance test, we present the RD-DD estimate, its clustered standard error based on standard asymptotics (in parentheses), a 95% confidence interval based on the cluster wild bootstrap, and the bandwidths used for the treated and untreated groups, respectively. We conduct the RD-DD analysis using CER-optimal bandwidths obtained separately for the treated and untreated age groups. Across both tables 2 and 3, all labour market and health outcomes appear to be balanced for the cohorts born around 1974, regardless of whether we consider standard asymptotic inference or wild bootstrap-based confidence intervals. For example, Column 1 of Panel A in Table 2 shows that the labour force participation rate is 0.03 log points higher for those just to the East of the

[12]While it is technically possible to construct smaller age groups for every county using the Microcensus, the resulting summary statistics are based on too few observations to pass the confidentiality tests required by Research Data Centres.

Table 2. Balance tests: Microcensus 2017

	Policy change			
	1974 West ends universal		1990 East ends mandatory	
	Age interval around policy change			
	5 years	10 years	5 years	10 years
	(1)	(2)	(3)	(4)
Panel A. Log labour force participation				
East × Treated	0.03 (0.03)	0.03 (0.02)	-0.33** (0.12)	-0.14 (0.14)
Wild 95% CI	[-0.09, 0.13]	[-0.04, 0.12]	[-0.56, 0.46]	[-0.55, 0.90]
BWs (h_{NT}, h_T)	56.2, 32.5	63.1, 36.2	27.2, 42.6	37.4, 33.6
Panel B. Log unemployed				
East × Treated	0.13 (0.29)	-0.17 (0.19)	-0.58** (0.19)	0.61 (0.53)
Wild 95% CI	[-0.33, 1.42]	[-1.08, 0.49]	[-1.12, 0.14]	[-1.23, 2.45]
BWs (h_{NT}, h_T)	40.0, 34.0	61.1, 35.2	67.8, 42.1	37.6, 42.4
Panel C. Log full time employment				
East × Treated	0.03 (0.08)	-0.01 (0.05)	-0.01 (0.05)	0.05 (0.07)
Wild 95% CI	[-0.43, 0.45]	[-0.28, 0.22]	[-0.19, 0.23]	[-0.21, 0.29]
BWs (h_{NT}, h_T)	32.3, 67.3	31.0, 44.6	57.1, 46.4	55.7, 45.4
Panel D. Log public employment				
East × Treated	-0.16 (0.27)	-0.18 (0.17)	-0.91* (0.46)	-0.78** (0.26)
Wild 95% CI	[-1.72, 0.82]	[-0.57, 0.80]	[-2.58, 1.22]	[-2.16, 0.58]
BWs (h_{NT}, h_T)	30.9, 38.5	32.1, 48.0	32.4, 30.9	32.7, 48.5
Panel E. Log commuters				
East × Treated	-0.02 (0.14)	-0.12 (0.15)	-0.54** (0.21)	-0.36 (0.24)
Wild 95% CI	[-0.86, 0.96]	[-0.85, 0.99]	[-1.32, 0.71]	[-1.14, 0.73]
BWs (h_{NT}, h_T)	49.6, 28.7	45.1, 29.9	34.6, 45.4	44.0, 39.0

Note: All coefficient estimates are based on a RD-DD specification with a CER-optimal bandwidth (Calonico et al., 2020) computed separately for the treated (h_T) and untreated cohorts (h_{NT}). Pilot bandwidths for the bias correction are not reported. The number of observations differs per cohort, depending on the bandwidth. Conventional standard errors clustered on states level are reported in parentheses. The wild cluster bootstrap confidence intervals (Cameron et al., 2008) are based on 99,999 replications where the null of no difference in discontinuities has been imposed.

former border compared to those located just to the West of the former border in the cohort born 5 years before 1974, relative to the same East-West differential for cohorts born after 1974. The difference-in-discontinuities is insignificantly different from zero using either asymptotic or finite-sample inference. The bounds on this difference under finite-sample inference range from -0.09 to 0.12 log points, suggesting that this difference-in-discontinuities is estimated as a reasonably precise zero. Some of the estimates tend to be large in magnitude and be accompanied by even larger standard errors. However, the fact that they are all insignificant under exact finite-sample inference reassures us that we are not getting covariate balance because the asymptotic formulas for the standard errors are invalid. For cohorts born around 1990, the labour market covariates appear to

Table 3. Balance tests: Health status 2016

	Policy change			
	1974 West ends universal		1990 East ends mandatory	
	Age interval around policy change			
	5 years	10 years	5 years	10 years
	(1)	(2)	(3)	(4)
Panel A. Log all-cause mortality				
East × Treated	-0.22 (0.15)	0.03 (0.19)	-0.61 (0.95)	-1.70** (0.54)
Wild 95% CI	[-0.80, 0.40]	[-0.93, 0.80]	[-2.68, 5.69]	[-2.68, 0.85]
BWs (h_{NT}, h_T)	53.0, 40.4	31.2, 35.5	53.9, 17.9	31.2, 31.2
Panel B. Log mortality from infectious diseases				
East × Treated	-1.00 (1.21)	-0.57 (1.05)	–	–
Wild 95% CI	[-6.85, 1.38]	[-6.28, 2.51]		
BWs (h_{NT}, h_T)	43.3, 40.8	34.6, 47.3		
Panel C. Log mortality from respiratory diseases				
East × Treated	-1.20 (1.77)	0.47 (1.54)	1.02 (1.06)	2.11 (1.32)
Wild 95% CI	[-9.19, 2.06]	[-8.33, 5.60]	[-0.35, 7.40]	[0.35, 10.32]
BWs (h_{NT}, h_T)	47.6, 47.0	50.9, 35.2	39.3, 71.0	37.9, 50.9
Panel D. Log hospitalizations				
East × Treated	0.07 (0.08)	0.08 (0.08)	-0.18*** (0.04)	-0.18*** (0.04)
Wild 95% CI	[-0.59, 0.17]	[-0.56, 0.20]	[-0.38, -0.03]	[-0.36, -0.12]
BWs (h_{NT}, h_T)	46.7, 30.3	41.4, 28.7	44.9, 41.5	40.3, 41.4
Panel E. Log hospitalizations for infectious diseases				
East × Treated	-0.12 (0.12)	0.04 (0.04)	0.17 (0.10)	-0.23* (0.12)
Wild 95% CI	[-0.67, 0.61]	[-0.01, 0.22]	[-0.21, 0.43]	[-1.05, -0.06]
BWs (h_{NT}, h_T)	28.8, 34.5	34.8, 41.9	42.3, 44.5	47.2, 34.8
Panel F. Log hospitalizations for respiratory diseases				
East × Treated	0.17 (0.12)	0.15 (0.09)	-0.24** (0.09)	-0.06 (0.05)
Wild 95% CI	[-0.50, 0.89]	[-0.03, 0.85]	[-0.58, 0.44]	[-0.26, 0.22]
BWs (h_{NT}, h_T)	23.2, 41.3	26.5, 40.1	37.4, 28.5	48.0, 26.5

Note: All coefficient estimates are based on a RD-DD specification with a CER-optimal bandwidth (Calonico et al., 2020) computed separately for the treated (h_T) and untreated cohorts (h_{NT}). Pilot bandwidths for the bias correction are not reported. The number of observations differs per cohort, depending on the bandwidth. Conventional standard errors clustered on states level are reported in parentheses. The wild cluster bootstrap confidence intervals (Cameron et al., 2008) are based on 99,999 replications where the null of no difference in discontinuities has been imposed.

be balanced when finite sample inference is used (although four out of ten tests reject at 5% when asymptotic inference is used).[13] However the health covariates are not balanced,

[13]Tables S.1. and S.2. in the Online Supplement show that the results are qualitatively similar when we use MSE-optimal bandwidths for the placebo checks.

with 4 out of 10 tests rejecting at 5% using the wild bootstrap inference (and a different set of 4 out of 10 rejecting when asymptotic formulas are used for inference). This may be because we are dealing with a much younger population in which mortality and hospitalization events are even rarer or because political changes in the 1990s had long-run consequences as the fetal origins hypothesis may suggest (Almond and Currie, 2011). For this reason, we focus on the 1974 experiment and present results using the end of mandatory BCG vaccination in the East in 1990 only as a robustness check.

4.2. Main results

We start by examining the reference RD specification for overall cases and symptomatic cases, across all ages and for the constituent cohorts of the RD-DD design.

Figure 2*a)* presents regression discontinuity plots of log(1+cases/million) by distance to the border of the former GDR, with positive (negative) distances indicating locations in former East Germany (West Germany). We use the methods of Calonico et al. (2014) to compute an optimal number of quantile-spaced bins and then estimate a local polynomial on each side using a fixed bandwidth of 30 km. We find that the local linear estimates are discontinuous at zero, falling from west to east. Otherwise, the conditional expectation function exhibits no apparent discontinuities. We observe a drop in cases per million of a little under 1 log point as one crosses the border from west to east. Hence, there are more than half as many cases per capita in a former East German county relative to a West German county just across the border. Figure 2*b)* presents similar local linear estimates for the discontinuity in log symptomatic cases. Here too, crossing the border from west to east entails a nearly 1 log point decrease in the number of severe cases per million residents. These correlations mirror those documented in the recent literature which typically finds a strong association between BCG status and COVID-19 severity in cross-sectional analyses. While remarkable, these apparent jumps are not causal, as many other confounders also change discontinuously at the former border.

Panels *c)* to *f)* of Figure 2 present the corresponding local linear estimates for the 35–44 age group and the 46–55 age group. These correspond to the cohorts born in the 10 years following and the 10 years preceding the 1974 discontinuation of the recommendation to give the BCG vaccine to most newborns. The jump at the border is present for the older group and is quantitatively similar to the discontinuity for the younger group. As the BCG hypothesis would have predicted little if any discontinuity for the older group (as members of that cohort in both the East and the West received the vaccine), this exercise already provides evidence against the BCG hypothesis. Importantly, using symptomatic cases instead of all reported cases as the COVID-19 intensity measure hardly changes the qualitative conclusion that the change at the border is similar across the two age groups.

Table 4 formalizes the intuition of the cohort-wise figures and presents estimates of the coefficient γ from specification 3.2 for various definitions of the "treated" and "untreated" groups and outcomes. We start with the 10-year cohorts underlying Figure 2, then present 5-year cohorts following Table 3 and then consider a narrow age group of two years before and after the 1974 policy experiment (46–47 versus 43–44 year olds). We present our baseline results in the first column by computing difference-in-difference regressions within samples defined by CER-optimal bandwidths estimated separately for the treated

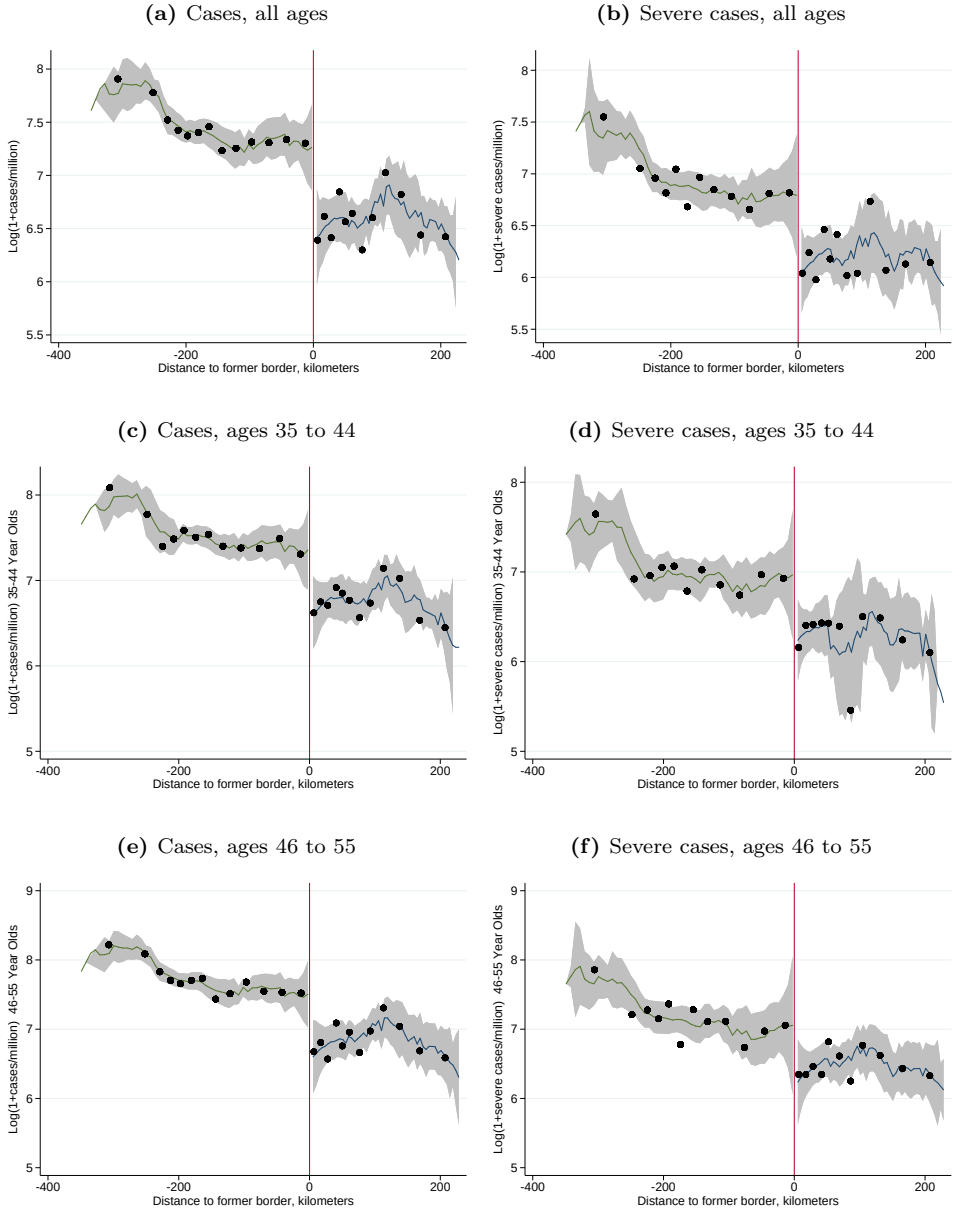


Figure 2. Discontinuities in log(1+cases/million) at the former border estimated using an optimal number of quantile-spaced bins and a local polynomial on each side with a fixed bandwidth of 30 km.

and the untreated group.[14] If the discontinuity in COVID-19 cases is caused by the direct long-term effect of BCG vaccination, then we would expect that the discontinuity among individuals born just before 1974 should be smaller than the discontinuity among people born just after 1974. The end of the recommendation to vaccinate in the West together with the temporary cessation would have lowered vaccinations, whereas the East continued mandatory vaccination with no change in 1974. As the sign of the discontinuity in COVID-19 intensity across the border is negative, we would expect the coefficient γ, the additional effect on the treated group (those born before 1974), to be positive as it should cancel out the average discontinuity for the population as a whole, or at least counter some portion of the compound treatment effect at the former border.

Contrary to the BCG hypothesis, the coefficient γ for overall cases is typically negative. In two instances, it is positive but not different from zero under either standard asymptotic or wild bootstrap inference (Panel A of Table 4). For example, Column 2 of Panel A suggests that the cohorts in the East born in the 8 years before 1974 had -0.49 log points (-39%) fewer COVID-19 cases than their counterparts in the West, when the differential between the East and West cohorts born in the 8 years after 1974 is used as a comparison. When γ is positive (as in column 4 of Panel A), it is slightly more than half the magnitude of the baseline discontinuity shown in Figure 2*a)* and statistically insignificant, although both the asymptotic and the wild bootstrap confidence intervals contain the large positive estimates that one would expect if the BCG hypothesis were true. This appears to be a function of the size of the age bins (and noise in county-by-age case data). With smaller age groups and fewer underlying cases, the confidence intervals on the coefficient γ with age windows of 4 and 2 years (columns 4 and 5) become large and contain the magnitude of the population discontinuity estimate. However, several estimates for wider age groups have modest standard errors that comfortably exclude an estimate of equal or greater magnitude and opposite sign from the population discontinuity. We also present the simple differences-in-differences estimates for the 1974 experiment, in which the whole of West Germany is used as a control for the whole of East Germany (Panel B). We now obtain nonpositive estimates of the coefficient γ for four of the five ways of structuring the treated group, while the lone positive estimate (in column 4) has a large standard error. For all but the narrowest age group definition, we can reject that the policy experiment fully offsets the population discontinuity estimate, while several estimates are centered relatively narrowly around zero (columns 3 and 4). In Panels C and D of Table 4 we exploit the unique data reported by the RKI and repeat our analysis using symptomatic cases only. Once again, our estimates of γ are negative, or positive but insignificantly different from zero in all cases when wild bootstrap confidence intervals are used, and in all but one case when standard inference is used, inconsistent with the BCG hypothesis. However, our confidence intervals are wide and typically include the population RD estimate.

As a robustness check we examine a second change to the BCG vaccination regime in Germany: the end of mandatory vaccination in the East following reunification in 1990. Now, the "treated" group is the post-1990 cohort because the end of mandatory vaccination in the East reduced the differential in vaccination rates between the East and the West for individuals born just after 1990 relative to individuals born just before. Panels A through D of Table 5 show that all the corresponding estimates but one are

[14]Table S.3. in the Online Supplement shows that we obtain similar results when MSE-optimal bandwidths are used.

Table 4. RD-DD results: 1974 experiment

	Age interval around policy change				
	10 years	8 years	6 years	4 years	2 years
	(1)	(2)	(3)	(4)	(5)
Panel A. 1974 baseline (cases/million)					
East × Treated	0.08 (0.11)	-0.49 (0.36)	-1.12** (0.44)	0.39 (0.60)	-2.17* (1.06)
WCR 95% CI	[-0.46, 0.42]	[-2.22, 1.01]	[-3.05, 0.26]	[-2.11, 1.82]	[-3.86, 4.85]
BWs (h_{NT}, h_T)	49.26, 50.56	45.94, 36.76	63.94, 37.74	50.08, 39.60	46.82, 25.79
Panel B. 1974 differences-in-differences (cases/million)					
East × Treated	-0.09** (0.04)	-0.02 (0.11)	0.00 (0.15)	0.22 (0.20)	-0.16 (0.54)
WCR 95% CI	[-0.19, 0.03]	[-0.27, 0.24]	[-0.36, 0.35]	[-0.50, 0.68]	[-1.48, 1.14]
BWs (h_{NT}, h_T)	348.89, 348.89	348.89, 348.89	348.89, 348.89	348.89, 348.89	348.89, 348.89
Panel C. 1974 severity measure (symptomatic cases/million)					
East × Treated	0.38 (0.49)	-0.53** (0.19)	-1.48** (0.55)	1.77*** (0.50)	-1.10 (1.16)
WCR 95% CI	[-1.66, 3.02]	[-1.26, 0.32]	[-4.01, -0.69]	[-0.83, 4.88]	[-4.63, 5.85]
BWs (h_{NT}, h_T)	31.68, 47.54	41.69, 38.76	55.24, 40.81	32.18, 32.83	40.65, 23.75
Panel D. 1974 differences-in-differences (symptomatic cases/million)					
East × Treated	0.04 (0.08)	0.22 (0.18)	0.17 (0.24)	0.33 (0.28)	-0.04 (0.56)
WCR 95% CI	[-0.10, 0.30]	[-0.23, 0.63]	[-0.37, 0.76]	[-0.34, 1.16]	[-1.31, 1.33]
BWs (h_{NT}, h_T)	348.89, 348.89	348.89, 348.89	348.89, 348.89	348.89, 348.89	348.89, 348.89

Note: All coefficient estimates are based on a RD-DD specification with a CER-optimal bandwidth (Calonico et al., 2020) computed separately for the treated (h_T) and untreated cohorts (h_{NT}). Pilot bandwidths for the bias correction are not reported. The number of observations differs per cohort, depending on the bandwidth. Conventional standard errors clustered on states level are reported in parentheses. The wild cluster bootstrap confidence intervals (Cameron et al., 2008) are based on 99,999 replications where the null of no difference in discontinuities has been imposed.

Table 5. RD-DD results: 1990 experiment

	Age interval around policy change				
	10 years	8 years	6 years	4 years	2 years
	(1)	(2)	(3)	(4)	(5)
Panel A. 1990 East Germany ends mandatory vaccination (cases/million)					
East × Treated	-0.21 (0.46)	-0.67 (0.60)	-2.25** (0.88)	-2.00** (0.84)	-0.39 (1.28)
WCR 95% CI	[-1.49, 1.75]	[-1.67, 2.07]	[-3.60, 1.38]	[-3.19, 1.99]	[-2.90, 5.35]
BWs (h_{NT}, h_T)	39.66, 82.03	38.79, 77.98	59.77, 65.66	45.32, 57.87	43.07, 70.15
Panel B. 1990 differences-in-differences (cases/million)					
East × Treated	-0.05 (0.11)	-0.14 (0.11)	-0.23 (0.21)	-0.61 (0.42)	-1.00** (0.40)
WCR 95% CI	[-0.34, 0.25]	[-0.38, 0.18]	[-0.80, 0.20]	[-1.73, 0.34]	[-2.00, -0.06]
BWs (h_{NT}, h_T)	348.89, 348.89	348.89, 348.89	348.89, 348.89	348.89, 348.89	348.89, 348.89
Panel C. 1990 severity measure (symptomatic cases/million)					
East × Treated	-0.16 (0.34)	-0.85 (0.49)	-1.37** (0.54)	-2.78*** (0.81)	0.07 (1.20)
WCR 95% CI	[-1.17, 1.05]	[-1.66, 0.87]	[-3.39, 1.49]	[-6.17, 0.43]	[-2.40, 5.15]
BWs (h_{NT}, h_T)	50.37, 57.85	65.65, 66.96	42.95, 38.45	55.05, 43.55	47.80, 54.65
Panel D. 1990 differences-in-differences (symptomatic cases/million)					
East × Treated	-0.02 (0.17)	-0.16 (0.14)	-0.28 (0.29)	-0.56 (0.54)	-1.22** (0.54)
WCR 95% CI	[-0.38, 0.45]	[-0.44, 0.36]	[-0.87, 0.76]	[-1.79, 0.79]	[-2.67, -0.03]
BWs (h_{NT}, h_T)	348.89, 348.89	348.89, 348.89	348.89, 348.89	348.89, 348.89	348.89, 348.89

Note: All coefficient estimates are based on a RD-DD specification with a CER-optimal bandwidth (Calonico et al., 2020) computed separately for the treated (h_T) and untreated cohorts (h_{NT}). Pilot bandwidths for the bias correction are not reported. The number of observations differs per cohort, depending on the bandwidth. Conventional standard errors clustered on states level are reported in parentheses. The wild cluster bootstrap confidence intervals (Cameron et al., 2008) are based on 99,999 replications where the null of no difference in discontinuities has been imposed.

consistently negative, and none is positive and significant under either asymptotic or finite sample inference. If anything, these estimates disagree even more with the BCG hypothesis than the 1974 experiment but have to be interpreted with caution given the covariate imbalance documented earlier.

5. ALTERNATIVE EXPLANATIONS FOR THE DISCONTINUITY

5.1. Social connectedness

If the BCG vaccine does not explain the East-West differential in COVID-19 cases, then what does? We investigate a broad set of additional variables to assess whether they may explain the discontinuity in overall COVID-19 cases and severity across the border. Regardless of the bandwidth used, log population density, log disposable income, the share of the population aged 45–64 and the share older than 64, the date that the first COVID-19 case was recorded, and age-adjusted mortality from all causes, infectious diseases and respiratory diseases all show discontinuities at the former border.[15] This echoes the pre-division East-West differences documented elsewhere (e.g. Becker et al., 2020) and is precisely why we do not consider the overall discontinuity as a causal estimate, since it may be driven by some of these variables. It is well known that counties just to the east of the border have lower population density, lower consumption, and an older population. We also find that they recorded their first COVID-19 case later and have higher age-adjusted mortality rates than counties just to the west of the border. Nevertheless, these explanations cannot fully account for the East-West difference in COVID-19 intensity. The discontinuity in cases per capita remains after including these variables. The raw correlations within the East and West further suggest that the geography of the early outbreak in Germany was very particular. The virus first spread among affluent and less vulnerable populations.[16]

In Table 6 we show how another variable—social connectedness to East Germany, as proxied by the ratio of Facebook connections to East Germany to overall Facebook connections—does explain the discontinuity in COVID-19 cases across the former border.[17] Panels A and B reproduce the benchmark RD specification for the CER-optimal bandwidth. We also include results for an MSE-optimal bandwidth and for fixed bandwidths of 30 km and 60 km, which are close to the largest and smallest among the CER-optimal and MSE-optimal bandwidths. As before, we also present wild cluster bootstrap confidence intervals (now following Bartalotti et al., 2017). Regardless of the choice of bandwidth and of the use of standard or finite-sample inference procedures, there is a strong and statistically significant decline in both total and symptomatic cases per million as one crosses the former border from west to east, typically on the order of 1 log point. Panel C shows that the fraction of Facebook connections of individuals in the county to individuals in East Germany discontinuously increases at the old East Germany border by at least 0.3, or 30 percentage points. Finally, Panel D shows that, once the fraction of

[15] Tables S.4. and S.5. in the Online Supplement report the corresponding results.

[16] See RD estimates with controls in Supplementary Table S.6. and the bivariate OLS regressions in Table S.7.

[17] The precise measure that we use for each county is the ratio of the sum of the Social Connectedness Index between that county and all East German counties and the sum of the Social Connectedness Index between that county and all other German counties. Connections within the county itself are excluded from the computation.

Facebook connections to East Germany is controlled for, the discontinuity in log severe cases per million at the border ceases to be statistically significant and declines in magnitude substantially. For example, using the CER-optimal bandwidth, the discontinuity declines from a statistically significant -0.97 log points to a statistically insignificant -0.23 log points. Therefore, the discontinuity in COVID-19 cases at the former border is statistically explained by social connectedness.

5.2. Long-distance commuting

Mobility is a key driver of the spread of COVID-19 in Germany and across the world (Dehning et al., 2020; Hsiang et al., 2020). Hence, a likely explanation for the discontinuity at the former border are Germany's regional commuting patterns. Next, we show that long-distance commuter networks (between counties that are more than 50 km apart) are also discontinuous across the former border, and that this discontinuity alone can give rise to a discontinuous distribution in COVID-19 cases.

We examine the role played by mobility in Table 6. If flows from the West usually terminate in the West and flows from the East usually terminate in the East, then cross-border transmission of the virus could be relatively slow.[18] As the epidemic started in the West, it will have had a harder time spreading eastward. The eastward spread was then further interrupted by the nation-wide lock-down on March 22 2020. Panel A shows that there is a stark and statistically significant discontinuity in long-distance commuter flows across the former border, which is robust to different bandwidth selection procedures. In particular, border counties on the eastern side are considerably less likely to receive commuter flows from a West German county than counties on the western side. Although commuting over long distances is very common in Germany, decades of partition meant that its infrastructure was re-oriented to connect counties within the West or East (Santamaria, 2020), with lasting effects on the spatial equilibrium in Germany.

We now simulate the epidemic in each county using a canonical SIR model with mobility flows (Bjørnstad and Grenfell, 2008; Wesolowski et al., 2017). This allows us to demonstrate that mobility patterns and the geography of the initial outbreaks can create a counterfactual discontinuity just like the one we observe in the data. In the model, we allow infections to spread along commuting patterns starting from the distribution of COVID-19 cases on February 29 2020 and use the approximate epidemiological characteristics of the outbreak in Germany (e.g., an R_0 of 2.5). We use the observed commuting flows from December 2019 together with county population data to proxy for actual mobility around the time of the outbreak. We simulate the model for 60 periods (days) but stop all commuting flows after 22 days to reflect the nation-wide shutdown. We do not explicitly model social distancing or other local containment measures, apart from the lack of commuting, which implies that the simulation overpredicts total cases.[19]

Panel B in Table 7 shows that the number of cases per million also discontinuously declines in the simulated data as one crosses from west to east over the former border. With about -0.82 log points at the CER-optimal bandwidth, the discontinuity is a little over 70% of the discontinuity for observed cases (in Panel A of Table 6). The dis-

[18] Figure S.1. shows that few counties have flows across the former East-West border of more than one thousand people. The only major destination in former East Germany is Berlin.

[19] Details of the simulation can be found in Online Appendix S1.

Table 6. Regression discontinuities in cases and social connectedness

	The bandwidth is			
	CER opt.	MSE opt.	30 km	60 km
Panel A. Log(1+Cases/Million)				
East	-1.11*** (0.36)	-0.76*** (0.25)	-1.74*** (0.29)	-1.27*** (0.29)
Wild 95% CI	[-1.6, -0.63]	[-1.19, -0.35]	[-2.23, -1.3]	[-1.75, -0.79]
BWs (h, b)	44.4, 106.4	50.9, 106.4	30, 30	60, 60
Panel B. Log(1+Severe Cases/Million)				
East	-0.97*** (0.28)	-1.04*** (0.38)	-1.12*** (0.28)	-1.07*** (0.37)
Wild 95% CI	[-1.44, -0.5]	[-1.64, -0.44]	[-1.59, -0.66]	[-1.69, -0.47]
BWs (h, b)	36.6, 95.2	41.9, 95.2	30, 30	60, 60
Panel C. Fraction of Facebook Connections with East Germany (FFC)				
East	0.33*** (0.07)	0.31*** (0.06)	0.31*** (0.11)	0.3*** (0.04)
Wild 95% CI	[0.22, 0.44]	[0.21, 0.4]	[0.13, 0.49]	[0.24, 0.36]
BWs (h, b)	18, 56.2	20.7, 56.2	30, 30	60, 60
Panel D. Log(1+Severe Cases/Million), Controlling for FFC				
East	-0.23 (0.3)	-0.24 (0.26)	-0.38 (0.4)	-0.31 (0.38)
Wild 95% CI	[-0.71, 0.26]	[-0.64, 0.22]	[-1.01, 0.24]	[-0.92, 0.29]
BWs (h, b)	24.9, 82.6	28.5, 82.6	30, 30	60, 60

Note: All coefficient estimates are bias-corrected (Calonico et al., 2014). The RD bandwidths for the point estimate (h) and bias corrections (b) are indicated in each column. CER and MSE optimal bandwidths use the selection rules developed in Calonico et al. (2020). Robust standard errors clustered on states are reported in parentheses. The Wild 95% confidence intervals use a cluster variant of the iterative bootstrap proposed by Bartalotti et al. (2017) with 99,999 replications for each bias correction and 1,000 replications to obtain the empirical distribution of the bias-corrected estimator.

continuity in simulated cases is statistically significantly different from zero under both asymptotic and finite-sample inference. Our simulation constructs a situation that shares some essential features of the data, and that explains the discontinuously lower COVID-19 prevalence across the border from East to West without any reference to the BCG hypothesis. On the other hand, our approach cannot exclude other alternative explanations, and officially registered commuter flows likely do not represent person-to-person movement across Germany perfectly.

5.3. The discontinuity after the initial outbreak

An implication of the hypothesis that discontinuities in mobility rather than of BCG vaccination explain the jump in COVID-19 intensity is that the discontinuity in cases should weaken over time. If BCG vaccination has a protective effect, this effect should manifest itself in both early and later stages of the pandemic. If mobility matters, once the virus was introduced everywhere, individuals in East Germany should transmit the virus to others in the East at similar rates to those in the West (especially after the lockdown was relaxed in May and June). In Panels C and D of Table 7 we present estimates of the discontinuity at the former border for log total reported cases and log

Table 7. Regression discontinuities in commuting, simulated cases, cases and deaths

	The bandwidth is			
	CER opt.	MSE opt.	30 km	60 km
Panel A. Fraction of Incoming Flows from West Germany (WGF)				
East	-0.42*** (0.06)	-0.38*** (0.07)	-0.55*** (0.17)	-0.45*** (0.08)
Wild 95% CI	[-0.52, -0.32]	[-0.48, -0.27]	[-0.81, -0.27]	[-0.59, -0.3]
BWs (h, b)	44.6, 101.4	51.1, 101.4	30, 30	60, 60
Panel B. Log(1+Simulated Cases/Million)				
East	-0.82*** (0.22)	-0.78*** (0.16)	-2.7** (1.26)	-1.38*** (0.5)
Wild 95% CI	[-1.16, -0.49]	[-1.03, -0.5]	[-4.6, -0.81]	[-2.15, -0.62]
BWs (h, b)	55.9, 123.3	64, 123.3	30, 30	60, 60
Panel C. Log(1+New Cases/Million)				
East	0 (0.16)	0.04 (0.13)	0.2 (0.23)	0.15 (0.22)
Wild 95% CI	[-0.31, 0.29]	[-0.19, 0.25]	[-0.16, 0.55]	[-0.23, 0.53]
BWs (h, b)	75.8, 156.8	86.7, 156.8	30, 30	60, 60
Panel D. Log(1+New Severe Cases/Million)				
East	-0.09 (0.2)	-0.08 (0.25)	-0.61 (0.7)	-0.23 (0.26)
Wild 95% CI	[-0.4, 0.28]	[-0.51, 0.37]	[-1.75, 0.56]	[-0.67, 0.21]
BWs (h, b)	39.9, 87.2	45.7, 87.2	30, 30	60, 60
Panel E. Log(1+Deaths/Million)				
East	-1.07 (0.95)	-0.51 (0.65)	0.47 (1.28)	-0.48 (1.04)
Wild 95% CI	[-2.68, 0.55]	[-1.59, 0.55]	[-1.65, 2.53]	[-2.15, 1.17]
BWs (h, b)	26.9, 91.5	30.8, 91.5	30, 30	60, 60
Panel F. Log(1+New Deaths/Million)				
East	0.01 (0.15)	0.21 (0.19)	1.09 (0.83)	0.37 (0.43)
Wild 95% CI	[-0.24, 0.27]	[-0.12, 0.58]	[-0.21, 2.41]	[-0.34, 1.03]
BWs (h, b)	37.9, 113.7	43.4, 113.7	30, 30	60, 60

Note: All coefficient estimates are bias-corrected (Calonico et al., 2014). The RD bandwidths for the point estimate (h) and bias corrections (b) are indicated in each column. CER and MSE optimal bandwidths use the selection rules developed in Calonico et al. (2020). Robust standard errors clustered on states are reported in parentheses. The Wild 95% confidence intervals use a cluster variant of the iterative bootstrap proposed by Bartalotti et al. (2017) with 99,999 replications for each bias correction and 1,000 replications to obtain the empirical distribution of the bias-corrected estimator.

total reported systematic cases per million between April 27 and December 13 (shortly before Germany announced a new lockdown in response to the autumn wave of COVID-19 in Europe). The RKI reported 1,178,997 new cases over this period. Estimates of the discontinuity in new cases are positive and statistically insignificant, with the wild bootstrap 95% confidence intervals rejecting estimates less than -0.31 regardless of bandwidth choice. For new symptomatic cases, the discontinuity estimates are negative but small in magnitude at the CER-optimal bandwidth, and statistically insignificant regardless of the inference procedure for all bandwidth choices. Whatever was responsible for the initial discontinuous drop in COVID-19 cases in East relative to West Germany appears

to have been particular to the first wave of COVID-19 infections—another finding which is inconsistent with the BCG hypothesis.

5.4. Evidence on death rates

So far we have relied on cases or symptomatic cases, rather than deaths, as our main measures of COVID-19 intensity. This choice is in part due to these measures being available by single year of age for each county. COVID-19 deaths are available only in large, aggregated categories at the county level, which are not amenable to the RD-DD analysis we carry out in this article. Nevertheless, it could be argued that BCG vaccination only helps to avoid the worst outcomes associated with contracting the novel coronavirus.

We consider discontinuities in COVID-19 related deaths per million in panel E of Table 7. We do not find statistically significant discontinuities in deaths per million as of April 26 2020 at any of the optimal or fixed bandwidths. The point estimates tend to be negative but have large asymptotic and finite-sample confidence intervals. Panel F presents the corresponding estimates for the period between April 27 and December 13. Now the estimated discontinuities are positive and insignificant. In fact, at the CER-optimal bandwidth, we obtain an estimate very close to zero with a relatively tight 95% wild bootstrap confidence interval, ruling out more substantial East-West discontinuities below -21%. Therefore, it does not appear that COVID-19 related deaths as a fraction of the population are discontinuously lower in the East than in the West, which also runs counter to the BCG hypothesis.

To summarize, our RD evidence shows that the discontinuity in (severe) cases per capita is fragile to inclusion of connectivity metrics like Facebook connectedness, that simulated case data with no reference to the border exhibits a similar discontinuity, and that case and death data after the end of the first lockdown on April 26 does not display a discontinuity at the former border. These facts along with our analysis of discontinuities in cases by cohort leads us to conclude that differential BCG vaccine coverage does not play an important role in explaining the geography of the outbreak in Germany.

6. CONCLUSION

We use variation in vaccination policy across the former East and West Germany to test whether the BCG vaccine offers protection against COVID-19. We identify patterns in the data that are inconsistent with the hypothesis that the BCG vaccine limits the spread or the severity of COVID-19. Instead, a more plausible explanation for the stark discontinuity in COVID-19 cases observed at the border is the continued presence of limits to mobility and interpersonal connectedness between the former East and West. These limits, coupled with the epidemic beginning in the West, decreased early COVID-19 exposure in the East.

An important limitation that our paper shares with the nonexperimental literature on the BCG hypothesis is that it looks only at whether or not there is a long-run effect of the BCG vaccine (decades after it was administered). We consider this broad version of the hypothesis to be of first-order importance. However, well-documented protective effects of the BCG vaccine regarding other viral infections, such as yellow fever (Arts et al., 2018), arise from a trained response of the innate immune system which typically occurs within one to twelve months after the vaccine has been administered (Kleinnijenhuis et al., 2015;

Covián et al., 2019; Chumakov et al., 2020). Hence, we cannot rule out that the vaccine might have a short-run effect which could offer some protection to risk groups.

Our results may be of interest as decisions are made on allocating resources to various ways of fighting COVID-19. The BCG vaccine is already in low supply (Guallar-Garrido and Julián, 2020) and is an important tool in the fight against tuberculosis—a disease which killed 1.5 million people in 2018 alone. Efforts to combat COVID-19 are already interrupting routine vaccination and detection efforts, which is projected to lead to a steep rise in fatalities from tuberculosis and other infectious diseases (Nature, 2020).

ACKNOWLEDGEMENTS

The authors are grateful to Gerda Asmus, Eli Berman, Victor Chernozhukov, Hunter Clark, Gordon Dahl, Beverly Hirtle, Melanie Krause, Keith Meyers, Linda Oz, Devesh Rustagi and the participants at the UCSD applied micro and omni-methods group seminars for valuable comments. We thank Colin Schmidt and Lisa-Marie Jannaschk from the German Research Data Centre of the Federal and State Statistical Offices for help with remotely processing the Microcensus data. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

REFERENCES

Almond, D. and J. Currie (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives 25*(3), 153–72.

Aronson, N. E., M. Santosham, G. W. Comstock, R. S. Howard, L. H. Moulton, E. R. Rhoades, and L. H. Harrison (2004, 05). Long-term Efficacy of BCG Vaccine in American Indians and Alaska Natives: A 60-Year Follow-up Study. *JAMA 291*(17), 2086–2091.

Arts, R. J., S. J. Moorlag, B. Novakovic, Y. Li, S.-Y. Wang, M. Oosting, V. Kumar, R. J. Xavier, C. Wijmenga, L. A. Joosten, et al. (2018). BCG vaccination protects against experimental viral infection in humans through the induction of cytokines associated with trained immunity. *Cell Host & Microbe 23*(1), 89–100.

Bailey, M., R. Cao, T. Kuchler, J. Stroebel, and A. Wong (2018, August). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives 32*(3), 259–80.

Bartalotti, O., G. Calhoun, and Y. He (2017, July). Bootstrap Confidence Intervals for Sharp Regression Discontinuity Designs: . In M. D. Cattaneo and J. C. Escanciano (Eds.), *Regression Discontinuity Designs*, Volume 38 of *Advances in Econometrics*, pp. 421–453. Emerald Publishing Ltd.

Becker, S. O., L. Mergele, and L. Woessmann (2020). The separation and reunification of Germany: Rethinking a natural experiment interpretation of the enduring effects of communism. *Journal of Economic Perspectives 34*(2), 71–143.

Berg, M. K., Q. Yu, C. E. Salvador, I. Melani, and S. Kitayama (2020). Mandated Bacillus Calmette-Guérin (BCG) vaccination predicts flattened curves for the spread of COVID-19. *Science Advances 6*(32).

Bjørnstad, O. N. and B. T. Grenfell (2008). Hazards, spatial transmission and timing of outbreaks in epidemic metapopulations. *Environmental and Ecological Statistics 15*(3), 265–277.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *Econometrics Journal 23*(2), 192–210.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2019). Regression discontinuity designs using covariates. *The Review of Economics and Statistics 101*(3), 442–451.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Chinazzi, M., J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, and A. Vespignani (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science 368*(6489), 395–400.

Chumakov, K., C. S. Benn, P. Aaby, S. Kottilil, and R. Gallo (2020). Can existing live vaccines prevent COVID-19? *Science 368*(6496), 1187–1188.

Covián, C., A. Fernández-Fierro, A. Retamal-Díaz, F. E. Díaz, A. E. Vasquez, M. K.-L. Lay, C. A. Riedel, P. A. González, S. M. Bueno, and A. M. Kalergis (2019). BCG-induced cross-protection and development of trained immunity: Implication for vaccine design. *Frontiers in Immunology 10*, 1–14.

Curtis, N., A. Sparrow, T. A. Ghebreyesus, and M. G. Netea (2020). Considering BCG vaccination to reduce the impact of COVID-19. *The Lancet 395*, 1545–1546.

Danner, K. and U. Qast (1995). Neuere Impfempfehlungen — Impfschemata. In H. Schneemann, G. Wurm, R. Batty, A. Berg, J. Cope, K. Danner, S. Dhillon, P. Elias, W. Feldheim, R. Großklaus, R. Grüttner, G. Gündermann, H. Haindl, H.-J. Hapke, H. Hehenberger, P. E. Heide, G. Heil, J. Keul, R. Kilian, U. Kirschner, A. Klaus, F. Klingauf, A. Kostrewski, I. Krämer, A. Liersch, N. P. Lüpke, G. Mould, H. Müller, A. Obermayer, D. Paar, U. Quast, A. Quilling, A. Rabitz, F. v. Rheinbaben, R. S. Roß, J. E. Schmitz, H. Schütz, E. Telser, E. J. Verspohl, J. Wachsmuth, U. Wahrburg, C. Ward, E. Wisker, and G. Wurm (Eds.), *Hagers Handbuch der Pharmazeutischen Praxis: Waren und Dienste Folgeband 1*, pp. 515–529. Berlin, Heidelberg: Springer.

de Chaisemartin, C. and L. de Chaisemartin (2020, 08). Bacille Calmette-Guérin Vaccination in Infancy Does Not Protect Against Coronavirus Disease 2019 (COVID-19): Evidence From a Natural Experiment in Sweden. *Clinical Infectious Diseases*. ciaa1223.

Dehning, J., J. Zierenberg, F. P. Spitzner, M. Wibral, J. P. Neto, M. Wilczek, and V. Priesemann (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science 369*(6500).

Dell, M. (2010). The persistent effects of Peru's mining Mita. *Econometrica 78*(6), 1863–1903.

Deshpande, M. (2016). Does welfare inhibit success? The long-term effects of removing low-income youth from the disability rolls. *American Economic Review 106*(11), 3300–3330.

Eckardt, M., K. Kappner, and N. Wolf (2020). COVID-19 across European regions: the role of border controls. CEPR Discussion Paper No. DP15178, Centre for Economic Policy Research.

Escobar, L. E., A. Molina-Cruz, and C. Barillas-Mury (2020). BCG vaccine protection from severe coronavirus disease 2019 (COVID-19). *Proceedings of the National Academy of Sciences 117*(30), 17720–17726.

Genz, H. (1977). Entwicklung der Säuglingstuberkulose in Deutschland im ersten Jahr nach Aussetzen der ungezielten BCG-Impfung. *Deutsche Medizinische Wochenschrift 102*(36), 1271–1273.

Guallar-Garrido, S. and E. Julián (2020). Bacillus Calmette-Guérin (BCG) therapy for bladder cancer: an update. *ImmunoTargets and Therapy 9*, 1.

Gursel, M. and I. Gursel (2020). Is global BCG vaccination-induced trained immunity relevant to the progression of SARS-CoV-2 pandemic? *Allergy 75*, 1815–1819.

Hamiel, U., E. Kozer, and I. Youngster (2020). SARS-CoV-2 Rates in BCG-Vaccinated and Unvaccinated Young Adults. *JAMA 323*(22), 2340–2341.

Harsch, D. (2012). Medicalized social hygiene? Tuberculosis policy in the German Democratic Republic. *Bulletin of the History of Medicine 86*(3), 394–423.

Hauer, J., U. Fischer, F. Auer, and A. Borkhardt (2020). Regional BCG vaccination policy in former East-and West Germany may impact on both severity of SARS-CoV-2 and incidence of childhood leukemia. *Leukemia 34*, 2217–2219.

He, Y. and O. Bartalotti (2020). Wild bootstrap for fuzzy regression discontinuity designs: obtaining robust bias-corrected confidence intervals. *Econometrics Journal 23*(2), 211–231.

Hsiang, S., D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, L. Y. Huang, A. Hultgren, E. Krasovich, P. Lau, J. Lee, E. Rolf, J. Tseng, and T. Wu (2020). The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*, 1–9.

Keele, L. J. and R. Titiunik (2015). Geographic boundaries as regression discontinuities. *Political Analysis 23*(1), 127–155.

Klein, S. (2013). *Zusammenhang zwischen Impfungen und Inzidenz und Mortalität von Infektionskrankheiten: Zeitreihenanalysen mit Meldedaten zu Diphtherie, Pertussis, Poliomyelitis und Tetanus von 1892 bis 2011 in Deutschland*. Ph. D. thesis, Freie Universität Berlin.

Klein, S., I. Schöneberg, and G. Krause (2012). Vom Zwang zur Pockenschutzimpfung zum Nationalen Impfplan. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz 55*(11-12), 1512–1523.

Kleinnijenhuis, J., R. van Crevel, and M. G. Netea (2015). Trained immunity: consequences for the heterologous effects of BCG vaccination. *Transactions of The Royal Society of Tropical Medicine and Hygiene 109*(1), 29–35.

Kreuser, F. (1967). Stand der Tuberkulose-Bekämpfung im Bundesgebiet, in West-Berlin und in Mitteldeutschland. In F. Kreuser (Ed.), *Tuberkulose-Jahrbuch 1964/65 — Band 14*, pp. 33–147. Berlin, Heidelberg: Springer.

Lalive, R. (2008). How do extended benefits affect unemployment duration? A regression discontinuity approach. *Journal of Econometrics 142*(2), 785–806.

Loddenkemper, R. and N. Konietzko (2018). Tuberculosis in Germany before, during and after World War II. In *Tuberculosis and War*, Volume 43, pp. 64–85. Karger Publishers.

MacKinnon, J. G. and M. D. Webb (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics 32*(2), 233–254.

Nature (2020). How to stop COVID-19 fuelling a resurgence of AIDS, malaria and tuberculosis. *Nature 584*(7820), 169. Editorial.

O'Neill, L. A. and M. G. Netea (2020). BCG-induced trained immunity: can it offer protection against COVID-19? *Nature Reviews Immunology 20*(6), 335–337.

Rieckmann, A., M. Villumsen, S. Sørup, L. K. Haugaard, H. Ravn, A. Roth, J. L. Baker, C. S. Benn, and P. Aaby (2016). Vaccinations against smallpox and tuberculosis are associated with better long-term survival: a Danish case-cohort study 1971–2010. *International Journal of Epidemiology 46*(2), 695–705.

Robert Koch-Institut (1976). Impfempfehlungen der Ständigen Impfkommission. *Bundesgesundheitsblatt 19*, 270–273.

Robert Koch-Institut (1998). Impfempfehlungen der Ständigen Impfkommission. *Epidemiologisches Bulletin 15*, 101–114.

Santamaria, M. (2020, February). Reshaping Infrastructure: Evidence from the division of Germany. The Warwick Economics Research Paper Series (TWERPS) 1244, University of Warwick, Department of Economics.

Sharma, A., S. K. Sharma, Y. Shi, E. Bucci, E. Carafoli, G. Melino, A. Bhattacherjee, and G. Das (2020). BCG vaccination policy and preventive chloroquine usage: do they have an impact on COVID-19 pandemic? *Cell Death & Disease 11*(7), 1–10.

Webb, M. D. (2014, November). Reworking Wild Bootstrap Based Inference For Clustered Errors. Working Paper 1315, Economics Department, Queen's University.

Wesolowski, A., E. zu Erbach-Schoenberg, A. J. Tatem, C. Lourenço, C. Viboud, V. Charu, N. Eagle, K. Engø-Monsen, T. Qureshi, C. O. Buckee, and C. J. E. Metcalf (2017). Multinational patterns of seasonal asymmetry in human movement influence infectious disease dynamics. *Nature Communications 8*(1), 1–9.

Manuscript submitted to *The Econometrics Journal*, pp. S1–S13.

The spread of COVID-19 and the BCG vaccine: Online supplement

Richard Bluhm[*,†] and Maxim Pinkovskiy[‡]

[*] *Leibniz University Hannover, Institute of Macroeconomics, 30167 Hannover, Germany*
[†] *University of California San Diego, Department of Political Science, La Jolla, CA 92093*
E-mail: bluhm@mak.uni-hannover.de
[‡] *Federal Reserve Bank of New York, Microeconomic Studies Function, New York, NY 10045*
E-mail: maxim.pinkovskiy@ny.frb.org

S1. SIR MODEL WITH COMMUTER FLOWS

We simulate a SIR model with multiple locations and exogenous migration flows between locations (Bjørnstad and Grenfell, 2008; Wesolowski et al., 2017). Let

$$\tilde{I}_{i,t} = I_{i,t} + \frac{N_i \sum_j m_{j,i} \frac{I_{j,t}}{N_j}}{N_i + \sum_j m_{j,i}} \tag{S.1}$$

$$S_{i,t+1} = S_{i,t} - \beta S_{i,t} \frac{\tilde{I}_{i,t}}{N_i} \tag{S.2}$$

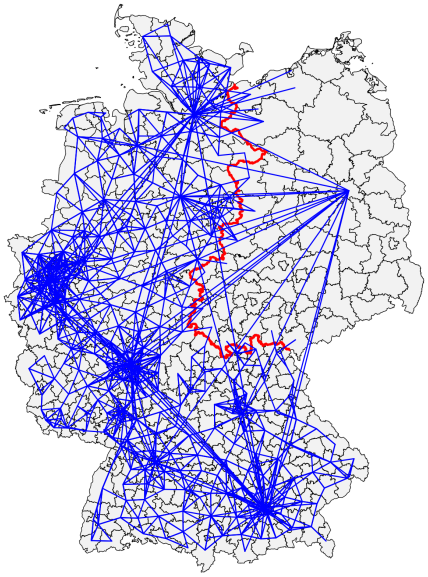
$$I_{i,t+1} = I_{i,t} + \beta S_{i,t} \frac{\tilde{I}_{i,t}}{N_i} - \gamma I_{i,t} \tag{S.3}$$

$$R_{i,t+1} = R_{i,t} + \gamma I_{i,t} \tag{S.4}$$

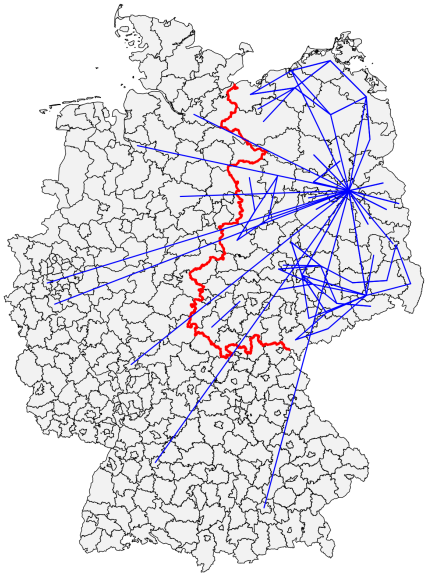
where $m_{j,i}$ is the number of commuters going from location j to location i each period and all other variables are as in the classical SIR model (Kermack and McKendrick, 1927). We take German counties as the locations in our models. We assume $\gamma = 1/7$ (because the incubation period is 7 days on average, and much of the transmission is pre-symptomatic) and $R_0 = \beta/\gamma = 2.5$. We assume the initial counts of infected to correspond to the reported cases by county on February 29 2020. We simulate the model for 60 time periods, assuming that after time period 22, all cross-county commuting flows are shut down to simulate measures taken by the German government.

We have tried other parametrizations of the SIR model and we get similar results provided that the epidemic is not allowed to evolve too close to the long-run equilibrium (which, when migration flows are eventually shut down, is the same for each county and hence, would not generate a discontinuity). The continued growth in cases over the summer of 2020 suggests that assuming that the epidemic did not attain long-run equilibrium is reasonable. The magnitude of the case counts resulting from the epidemic vary widely between parametrizations. We view this exercise not as an attempt to model the COVID-19 epidemic in Germany but to provide an illustration that mobility patterns can generate discontinuities in the spread of an epidemic without there being essential discontinuities in the underlying resistance of the population.

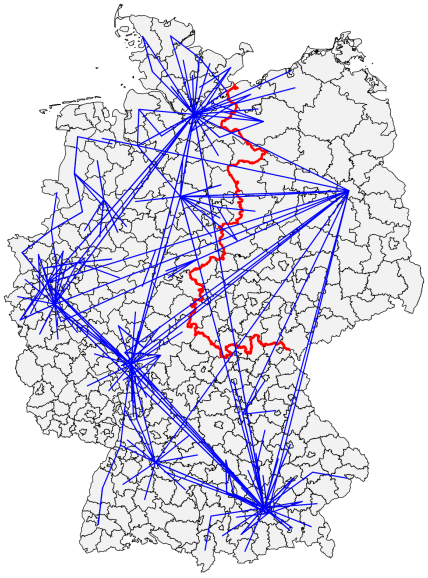
(a) Originating in the West



(b) Originating in the East



(c) Originating in the West & $>$ 50 km



(d) Originating in the East & $>$ 50 km

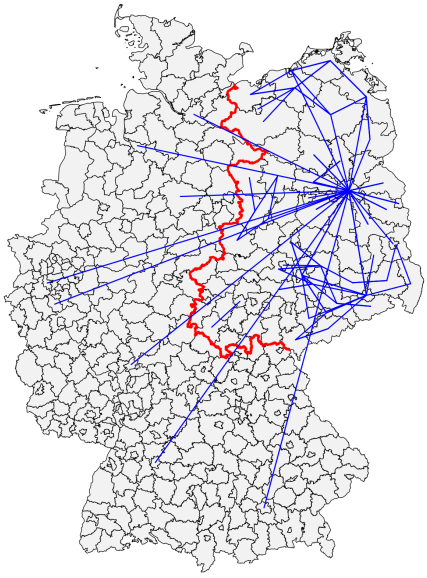


Figure S.1. Major commuting flows (at least 1,000 people) by origin

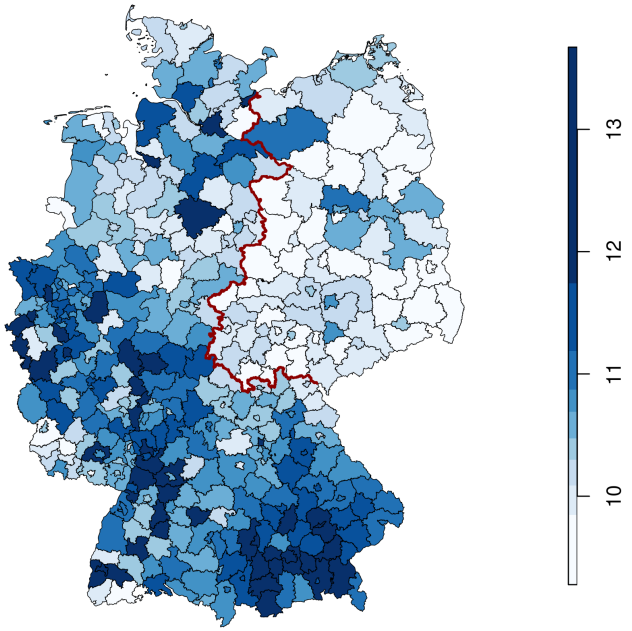


Figure S.2. Spatial distribution of simulated COVID-19 cases in Germany

S2. ADDITIONAL FIGURES

Figure S.3 illustrates the time-series of daily cases in Germany from Jan 27 2020 (week 5) until December 13 2020 (week 50). Week 17 is the last week during which reported cases exceeded 2,000 on at least one day Coronavirus-related travel restrictions were progressively relaxed starting in May and then re-instituted in the fall. Week 37 marks the first week in the fall wave during which reported cases exceeded 2,000 on at least one day. Our data ends three days before the second nationwide "hard" lockdown on December 16 2020.

Figure S.4 shows the age distribution of cases in the territory of former West and East Germany on April 26, 2020. Cases are all positive COVID-19 tests reported by the RKI. Severe cases display acute respiratory symptoms (including pneumonia) or have died from COVID-19. The distribution is truncated after 79 years, as the RKI reports cases for those aged 80 and above in a single combined category which we omit for display purposes.

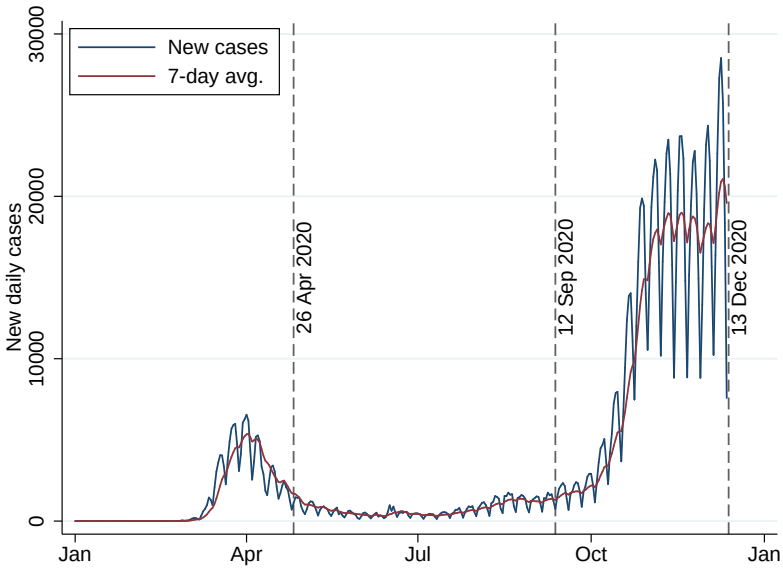


Figure S.3. Time-series of daily cases

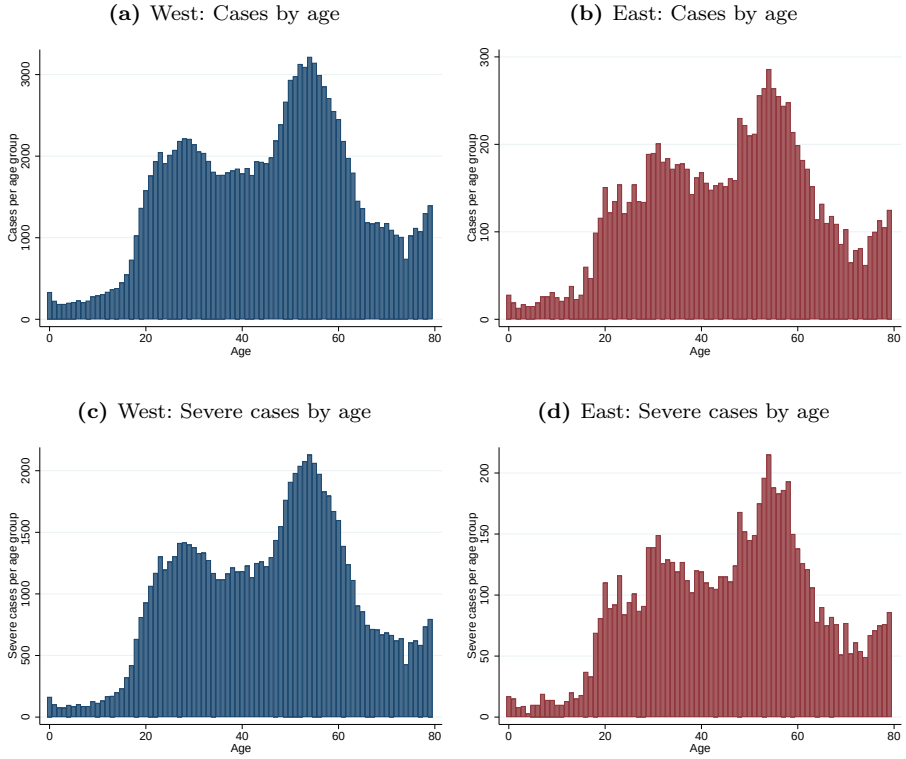


Figure S.4. Distribution of (severe) cases by age on both sides of the former border

S3. ADDITIONAL REGRESSION RESULTS

Table S.1. Balance tests: Microcensus 2017, MSE-optimal bandwidth

	Policy change			
	1974 West ends universal		1990 East ends mandatory	
	Age interval around policy change			
	5 years	10 years	5 years	10 years
	(1)	(2)	(3)	(4)
Panel A. Log labour force participation				
East × Treated	0.01 (0.01)	0.05*** (0.01)	-0.22 (0.12)	-0.09 (0.14)
WCR 95% CI	[-0.03, 0.04]	[0.03, 0.10]	[-0.40, 0.50]	[-0.34, 0.74]
BWs (h_{NT}, h_T)	64.3, 37.2	72.3, 41.4	31.2, 48.7	42.8, 38.5
Panel B. Log unemployed				
East × Treated	0.02 (0.18)	-0.43 (0.25)	-0.42 (0.28)	0.89* (0.40)
WCR 95% CI	[-0.71, 0.55]	[-1.30, 0.01]	[-1.25, 0.54]	[-0.47, 2.42]
BWs (h_{NT}, h_T)	45.8, 38.9	70.0, 40.3	77.7, 48.2	43.1, 48.6
Panel C. Log full time employment				
East × Treated	0.04 (0.07)	0.00 (0.04)	-0.02 (0.06)	0.01 (0.06)
WCR 95% CI	[-0.39, 0.28]	[-0.23, 0.09]	[-0.23, 0.18]	[-0.23, 0.20]
BWs (h_{NT}, h_T)	37.0, 77.0	35.5, 51.1	65.4, 53.1	63.7, 51.9
Panel D. Log public employment				
East × Treated	-0.19 (0.18)	-0.14 (0.11)	-0.24 (0.26)	-0.12 (0.39)
WCR 95% CI	[-1.02, 0.55]	[-0.59, 0.59]	[-1.62, 0.58]	[-0.65, 1.69]
BWs (h_{NT}, h_T)	35.4, 44.1	36.8, 55.0	37.1, 35.4	37.4, 55.5
Panel E. Log commuters				
East × Treated	-0.04 (0.22)	-0.17 (0.19)	-0.41 (0.24)	-0.21*** (0.06)
WCR 95% CI	[-1.42, 0.99]	[-1.38, 0.57]	[-1.30, 0.57]	[-0.48, -0.04]
BWs (h_{NT}, h_T)	56.8, 32.9	51.6, 34.2	39.7, 52.0	50.4, 44.6

Note: All coefficient estimates are based on a RD-DD specification with a MSE-optimal bandwidth (Calonico et al., 2020) computed separately for the treated (h_T) and untreated cohorts (h_{NT}). Pilot bandwidths for the bias correction are not reported. The number of observations differs per cohort, depending on the bandwidth. Conventional standard errors clustered on states level are reported in parentheses. The wild cluster bootstrap confidence intervals (Cameron et al., 2008) are based on 99,999 replications where the null of no difference in discontinuities has been imposed.

Table S.2. Balance tests: Health status 2016, MSE-optimal bandwidth

	Policy change			
	1974 West ends universal		1990 East ends mandatory	
	Age interval around policy change			
	5 years	10 years	5 years	10 years
	(1)	(2)	(3)	(4)
Panel A. Log all-cause mortality				
East × Treated	-0.18 (0.10)	-0.05 (0.15)	-0.25 (0.68)	-1.73** (0.66)
WCR 95% CI	[-0.63, 0.04]	[-0.64, 0.75]	[-2.51, 3.33]	[-2.89, 1.20]
BWs (h_{NT}, h_T)	60.6, 46.3	35.7, 40.7	61.7, 20.5	35.7, 35.7
Panel B. Log mortality from infectious diseases				
East × Treated	-0.74 (1.27)	0.56 (1.11)	–	–
WCR 95% CI	[-7.19, 1.15]	[-5.30, 3.33]		
BWs (h_{NT}, h_T)	49.6, 46.7	39.6, 54.1		
Panel C. Log mortality from respiratory diseases				
East × Treated	-1.17 (1.54)	0.03 (1.60)	0.83 (0.92)	2.11* (1.00)
WCR 95% CI	[-8.22, 1.38]	[-7.19, 4.00]	[-0.58, 5.50]	[0.74, 7.27]
BWs (h_{NT}, h_T)	54.5, 53.8	58.2, 40.3	45.0, 81.3	43.3, 58.2
Panel D. Log hospitalizations				
East × Treated	0.08 (0.05)	0.05 (0.07)	-0.20*** (0.04)	-0.11* (0.05)
WCR 95% CI	[-0.27, 0.14]	[-0.37, 0.17]	[-0.34, -0.06]	[-0.37, 0.02]
BWs (h_{NT}, h_T)	53.5, 34.7	47.4, 32.9	51.5, 47.5	46.2, 47.4
Panel E. Log hospitalizations for infectious diseases				
East × Treated	-0.02 (0.14)	-0.06 (0.08)	0.20** (0.07)	-0.13 (0.12)
WCR 95% CI	[-0.26, 0.76]	[-0.39, 0.32]	[-0.12, 0.46]	[-0.98, 0.32]
BWs (h_{NT}, h_T)	33.0, 39.5	39.9, 48.0	48.4, 51.0	54.0, 39.9
Panel F. Log hospitalizations for respiratory diseases				
East × Treated	0.05 (0.05)	0.13 (0.10)	-0.14 (0.12)	-0.05 (0.06)
WCR 95% CI	[-0.10, 0.35]	[-0.20, 0.84]	[-0.46, 0.58]	[-0.26, 0.29]
BWs (h_{NT}, h_T)	26.6, 47.3	30.4, 45.9	42.9, 32.6	55.0, 30.4

Note: All coefficient estimates are based on a RD-DD specification with a MSE-optimal bandwidth (Calonico et al., 2020) computed separately for the treated (h_T) and untreated cohorts (h_{NT}). Pilot bandwidths for the bias correction are not reported. The number of observations differs per cohort, depending on the bandwidth. Conventional standard errors clustered on states level are reported in parentheses. The wild cluster bootstrap confidence intervals (Cameron et al., 2008) are based on 99,999 replications where the null of no difference in discontinuities has been imposed.

Table S.3. RD-DD results: MSE-optimal bandwidth

	Age interval around policy change				
	10 years	8 years	6 years	4 years	2 years
	(1)	(2)	(3)	(4)	(5)
Panel A. 1974 baseline (cases/million)					
East × Treated	-0.11 (0.17)	-0.89** (0.32)	-1.22** (0.49)	0.00 (0.81)	-3.29*** (1.00)
Wild 95% CI	[-0.97, 0.17]	[-2.14, -0.23]	[-3.31, -0.42]	[-3.48, 1.20]	[-9.44, 0.92]
BWs (h_{NT}, h_T)	56.41, 57.89	52.60, 42.09	73.21, 43.21	57.34, 45.34	53.61, 29.53
Panel B. 1974 severity measure (symptomatic cases/million)					
East × Treated	0.12 (0.39)	-0.91** (0.38)	-1.37** (0.55)	0.82* (0.39)	-0.73 (1.35)
Wild 95% CI	[-1.35, 1.82]	[-2.81, 1.03]	[-3.65, -0.28]	[-0.06, 3.41]	[-4.97, 8.60]
BWs (h_{NT}, h_T)	36.27, 54.43	47.74, 44.38	63.25, 46.73	36.85, 37.59	46.55, 27.19
Panel C. 1990 East Germany ends mandatory vaccination (cases/million)					
East × Treated	-0.37 (0.46)	-0.67 (0.51)	-1.88* (0.87)	-2.83*** (0.32)	-1.00 (0.81)
Wild 95% CI	[-2.19, 1.18]	[-1.84, 1.18]	[-3.33, 1.20]	[-3.59, -1.46]	[-4.34, 2.39]
BWs (h_{NT}, h_T)	45.41, 93.93	44.41, 89.29	68.43, 75.18	51.90, 66.26	49.32, 80.32
Panel D. 1990 severity measure (symptomatic cases/million)					
East × Treated	-0.27 (0.39)	-0.87 (0.49)	-2.08*** (0.64)	-1.86* (0.92)	-1.65* (0.84)
Wild 95% CI	[-1.62, 0.99]	[-1.77, 0.83]	[-4.31, 1.30]	[-5.64, 1.93]	[-4.38, 1.95]
BWs (h_{NT}, h_T)	57.68, 66.24	75.16, 76.67	49.17, 44.03	63.04, 49.87	54.74, 62.57

Note: All coefficient estimates are based on a RD-DD specification with a MSE-optimal bandwidth (Calonico et al., 2020) computed separately for the treated (h_T) and untreated cohorts (h_{NT}). Pilot bandwidths for the bias correction are not reported. The number of observations differs per cohort, depending on the bandwidth. Conventional standard errors clustered on states level are reported in parentheses. The wild cluster bootstrap confidence intervals (Cameron et al., 2008) are based on 99,999 replications where the null of no difference in discontinuities has been imposed.

Table S.4. Discontinuities in other variables, part I

	The bandwidth is			
	CER opt.	MSE opt.	30 km	60 km
Panel A. Disposable income per capita				
East	-0.11*** (0.03)	-0.09*** (0.03)	-0.15 (0.1)	-0.07 (0.07)
Wild 95% CI	[-0.16, -0.05]	[-0.15, -0.04]	[-0.29, -0.01]	[-0.18, 0.04]
BWs (h, b)	33.3, 92.3	38.1, 92.3	30, 30	60, 60
Panel B. Population density				
East	-0.41 (0.44)	-0.49 (0.47)	-1.59** (0.75)	-0.31 (0.64)
Wild 95% CI	[-1.19, 0.36]	[-1.33, 0.35]	[-2.85, -0.34]	[-1.31, 0.63]
BWs (h, b)	38.4, 104.1	44, 104.1	30, 30	60, 60
Panel C. Area				
East	0.54 (0.37)	0.73*** (0.23)	-0.6 (0.62)	0.56 (0.66)
Wild 95% CI	[-0.09, 1.19]	[0.37, 1.1]	[-1.57, 0.44]	[-0.54, 1.66]
BWs (h, b)	41.8, 100.3	47.9, 100.3	30, 30	60, 60
Panel D. Percent older than 60				
East	2.81*** (0.74)	2.61*** (0.75)	4.68*** (1.38)	2.05 (1.25)
Wild 95% CI	[1.65, 4.03]	[1.3, 3.84]	[2.43, 6.95]	[0, 4.13]
BWs (h, b)	36.8, 86.3	42.1, 86.3	30, 30	60, 60
Panel E. Percent younger than 35				
East	-4.58*** (1.23)	-5.43*** (1.36)	-5.08*** (1.33)	-3.93** (1.94)
Wild 95% CI	[-6.72, -2.4]	[-7.78, -3.03]	[-7.48, -2.68]	[-7.08, -0.72]
BWs (h, b)	41.8, 103.5	47.8, 103.5	30, 30	60, 60
Panel F. Days since first case				
East	-1.38 (2.99)	-1.87 (2.92)	-21.18*** (4.49)	-1.14 (2.88)
Wild 95% CI	[-6.53, 3.5]	[-6.79, 3.1]	[-28.45, -13.92]	[-5.99, 3.7]
BWs (h, b)	33.8, 65.2	38.7, 65.2	30, 30	60, 60

Note: All coefficient estimates are bias-corrected (Calonico et al., 2014). The RD bandwidths for the point estimate (h) and bias corrections (b) are indicated in each column. CER and MSE optimal bandwidths use the selection rules developed in Calonico et al. (2020). Robust standard errors clustered on states are reported in parentheses. The Wild 95% confidence intervals use a cluster variant of the iterative bootstrap proposed by Bartalotti et al. (2017) with 99,999 replications for each bias correction and 1,000 replications to obtain the empirical distribution of the bias-corrected estimator.

Table S.5. Discontinuities in other variables, part II

	The bandwidth is			
	CER opt.	MSE opt.	30 km	60 km
Panel A. Age-adjusted overall death rate per million				
East	0 (0.02)	0 (0.02)	0 (0.03)	-0.03** (0.01)
Wild 95% CI	[-0.03, 0.02]	[-0.04, 0.03]	[-0.05, 0.05]	[-0.05, -0.01]
BWs (h, b)	39.7, 95.9	45.4, 95.9	30, 30	60, 60
Panel B. Age-adjusted infectious diseases death rate per million				
East	4.4*** (1.13)	3.86*** (0.88)	5.08*** (1.57)	3.84*** (1.03)
Wild 95% CI	[2.69, 6.04]	[2.48, 5.32]	[2.4, 7.69]	[2.16, 5.57]
BWs (h, b)	23.1, 56.5	26.5, 56.5	30, 30	60, 60
Panel C. Age-adjusted respiratory diseases death rate per million				
East	5.84*** (1.49)	5.17*** (1.16)	5.92*** (2.21)	4.94*** (1.36)
Wild 95% CI	[3.66, 8.02]	[3.25, 6.93]	[1.99, 9.75]	[2.69, 7.17]
BWs (h, b)	23.5, 53.9	26.9, 53.9	30, 30	60, 60
Panel D. Age-adjusted overall hospitalization rate per million				
East	0.1** (0.05)	0.02 (0.07)	0.25*** (0.1)	0.04 (0.03)
Wild 95% CI	[0.02, 0.17]	[-0.09, 0.13]	[0.1, 0.4]	[-0.01, 0.08]
BWs (h, b)	25.2, 70	28.9, 70	30, 30	60, 60
Panel E. Age-adjusted infectious diseases hospitalization rate per million				
East	0.05 (0.06)	0.1 (0.08)	0.12 (0.21)	-0.11 (0.08)
Wild 95% CI	[-0.06, 0.16]	[-0.04, 0.23]	[-0.22, 0.46]	[-0.25, 0.03]
BWs (h, b)	53.1, 135.4	60.8, 135.4	30, 30	60, 60
Panel F. Age-adjusted respiratory diseases hospitalization rate per million				
East	-0.01 (0.06)	-0.01 (0.05)	0.12 (0.14)	0.02 (0.04)
Wild 95% CI	[-0.1, 0.09]	[-0.1, 0.07]	[-0.09, 0.34]	[-0.04, 0.08]
BWs (h, b)	27.9, 65.1	31.9, 65.1	30, 30	60, 60

Note: All coefficient estimates are bias-corrected (Calonico et al., 2014). The RD bandwidths for the point estimate (h) and bias corrections (b) are indicated in each column. CER and MSE optimal bandwidths use the selection rules developed in Calonico et al. (2020). Robust standard errors clustered on states are reported in parentheses. The Wild 95% confidence intervals use a cluster variant of the iterative bootstrap proposed by Bartalotti et al. (2017) with 99,999 replications for each bias correction and 1,000 replications to obtain the empirical distribution of the bias-corrected estimator.

Table S.6. Discontinuity in cases with controls

	Dependent variable: Log(1+Cases/Million)			
	The bandwidth is			
	CER opt.	MSE opt.	30 km	60 km
Panel A. No controls				
East	-1.11*** (0.36)	-0.76*** (0.25)	-1.74*** (0.29)	-1.27*** (0.29)
Wild 95% CI	[-1.6, -0.63]	[-1.19, -0.35]	[-2.23, -1.3]	[-1.75, -0.79]
BWs (h, b)	44.4, 106.4	50.9, 106.4	30, 30	60, 60
Panel B. Population density				
East	-1.18** (0.46)	-0.96** (0.42)	-1.66*** (0.31)	-1.19*** (0.34)
Wild 95% CI	[-1.82, -0.56]	[-1.53, -0.4]	[-2.14, -1.14]	[-1.73, -0.65]
BWs (h, b)	38, 94.4	43.6, 94.4	30, 30	60, 60
Panel C. Disposable income p.c.				
East	-1.2** (0.47)	-1.07** (0.42)	-1.66*** (0.29)	-1.26*** (0.29)
Wild 95% CI	[-1.86, -0.55]	[-1.62, -0.51]	[-2.1, -1.21]	[-1.74, -0.78]
BWs (h, b)	38.3, 98.8	43.9, 98.8	30, 30	60, 60
Panel D. Population density and disposable income p.c.				
East	-1.18** (0.46)	-0.96** (0.42)	-1.66*** (0.31)	-1.19*** (0.34)
Wild 95% CI	[-1.83, -0.51]	[-1.54, -0.37]	[-2.18, -1.13]	[-1.73, -0.64]
BWs (h, b)	38, 94.4	43.6, 94.4	30, 30	60, 60
Panel E. Percent younger 35 and percent older than 60				
East	-0.99*** (0.38)	-0.73** (0.29)	-1.58*** (0.3)	-1.25*** (0.29)
Wild 95% CI	[-1.51, -0.5]	[-1.21, -0.24]	[-2.04, -1.1]	[-1.75, -0.75]
BWs (h, b)	42.5, 101.6	48.6, 101.6	30, 30	60, 60
Panel F. Days since first case				
East	-1.2** (0.52)	-1.08** (0.46)	-1.16*** (0.3)	-1.26*** (0.31)
Wild 95% CI	[-1.92, -0.49]	[-1.67, -0.46]	[-1.65, -0.66]	[-1.79, -0.75]
BWs (h, b)	39.6, 117.6	45.3, 117.6	30, 30	60, 60
Panel G. All of the above				
East	-1.29*** (0.42)	-1.21*** (0.39)	-1.26*** (0.3)	-1.17*** (0.35)
Wild 95% CI	[-1.89, -0.66]	[-1.79, -0.65]	[-1.76, -0.77]	[-1.72, -0.59]
BWs (h, b)	38.1, 85.6	43.6, 85.6	30, 30	60, 60

Note: All coefficient estimates are bias-corrected (Calonico et al., 2014). The RD bandwidths for the point estimate (h) and bias corrections (b) are indicated in each column. CER and MSE optimal bandwidths use the selection rules developed in Calonico et al. (2020). Robust standard errors clustered on states are reported in parentheses. The Wild 95% confidence intervals use a cluster variant of the iterative bootstrap proposed by Bartalotti et al. (2017) with 99,999 replications for each bias correction and 1,000 replications to obtain the empirical distribution of the bias-corrected estimator.

Table S.7. Bivariate OLS regressions

	Dependent variable: Log(1+Cases/Million)		
	West (1)	East (2)	All (3)
Panel A. Disposable income p.c.			
Coefficient (Std. Err.)	2.36** (0.74)	3.52 (2.02)	3.38*** (0.61)
Wild 95% CI	[0.76, 4.98]	[-1.40, 7.13]	[1.76, 4.92]
Panel B. Population density			
Coefficient (Std. Err.)	-0.03 (0.08)	0.16*** (0.02)	0.09 (0.07)
Wild 95% CI	[-0.15, 0.26]	[0.09, 0.21]	[-0.07, 0.29]
Panel C. Percent older than 60			
Coefficient (Std. Err.)	-0.06** (0.02)	-0.03 (0.04)	-0.09*** (0.02)
Wild 95% CI	[-0.12, -0.01]	[-0.16, 0.23]	[-0.13, -0.05]
Panel D. Percent younger than 35			
Coefficient (Std. Err.)	0.03 (0.02)	0.03 (0.02)	0.07*** (0.01)
Wild 95% CI	[-0.01, 0.08]	[-0.01, 0.11]	[0.04, 0.10]
Panel E. Age-adjusted overall death rate per million			
Coefficient (Std. Err.)	-3.17** (1.30)	-3.51** (0.97)	-4.02*** (1.02)
Wild 95% CI	[-6.85, -0.65]	[-7.35, -1.32]	[-6.58, -1.86]
Panel G. Age-adjusted infectious diseases death rate per million			
Coefficient (Std. Err.)	-0.12*** (0.03)	-0.08 (0.25)	-0.15*** (0.03)
Wild 95% CI	[-0.42, 0.17]	[-0.91, 0.46]	[-0.47, 0.26]
Panel H. Age-adjusted respiratory diseases death rate per million			
Coefficient (Std. Err.)	-0.10*** (0.03)	-0.22 (0.42)	-0.13*** (0.02)
Wild 95% CI	[-0.32, 0.09]	[-2.25, 1.47]	[-0.35, 0.11]
Panel I. Age-adjusted hospitalization rate per million			
Coefficient (Std. Err.)	-0.62 (0.64)	-2.00* (0.81)	-1.49* (0.72)
Wild 95% CI	[-2.06, 1.26]	[-4.31, 1.65]	[-3.30, 0.22]
Panel J. Age-adjusted infectious diseases hospitalization rate per million			
Coefficient (Std. Err.)	-0.01 (0.03)	-1.33** (0.30)	-0.05 (0.04)
Wild 95% CI	[-2.55, 1.75]	[-2.25, 0.13]	[-1.87, 1.78]
Panel K. Age-adjusted respiratory diseases hospitalization rate per million			
Coefficient (Std. Err.)	-0.03 (0.02)	-1.37*** (0.13)	-0.06* (0.03)
Wild 95% CI	[-2.10, 1.71]	[-1.59, -0.80]	[-2.26, 2.08]
Observations	324	76	400

Note: The table reports results from bivariate ordinary least squares regressions for the samples indicated in the column headers. Conventional standard errors clustered on states level are reported in parentheses. The wild cluster bootstrap confidence intervals (Cameron et al., 2008) are based on 99,999 replications where the null of no correlation between the two variables has been imposed.

REFERENCES

Bartalotti, O., G. Calhoun, and Y. He (2017, July). Bootstrap Confidence Intervals for Sharp Regression Discontinuity Designs: . In M. D. Cattaneo and J. C. Escanciano (Eds.), *Regression Discontinuity Designs*, Volume 38 of *Advances in Econometrics*, pp. 421–453. Emerald Publishing Ltd.

Bjørnstad, O. N. and B. T. Grenfell (2008). Hazards, spatial transmission and timing of outbreaks in epidemic metapopulations. *Environmental and Ecological Statistics 15*(3), 265–277.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *Econometrics Journal 23*(2), 192–210.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Kermack, W. O. and A. G. McKendrick (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A 115*(772), 700–721.

Wesolowski, A., E. zu Erbach-Schoenberg, A. J. Tatem, C. Lourenço, C. Viboud, V. Charu, N. Eagle, K. Engø-Monsen, T. Qureshi, C. O. Buckee, and C. J. E. Metcalf (2017). Multinational patterns of seasonal asymmetry in human movement influence infectious disease dynamics. *Nature Communications 8*(1), 1–9.