Northwestern | INSTITUTE FOR POLICY RESEARCH

IPR Working Paper Series

WP-22-29

Economic vs. Epidemiological Approaches to Measuring the Human Capital Impacts of Infectious Disease Elimination

Caroline Chuard
University of St. Gallen

Hannes Schwandt
Northwestern University and IPR

Alexander Becker

Masahiko Haraguchi
Research Institute for Humanity and Nature

Version: July 5, 2022

DRAFT

Please do not quote or distribute without permission.

Abstract

A rich economic literature has examined the human capital impacts of disease-eliminating health interventions, such as the rollout of new vaccines. This literature is based on reduced-form approaches which exploit proxies for disease burden, such as mortality, instead of actual infection counts, which are difficult to measure. The researchers develop an epidemiological dynamic accounting model based on the susceptible-infected-recovered (SIR) framework to derive precise measles infection shares across U.S. cohorts born around the introduction of the measles vaccine. Measles is highly infectious and fully immunizing which makes the disease an ideal candidate for epidemiological modeling. The authors' epidemiological model is strongly predictive of future measles outbreaks, but the derived measles infection shares are not systematically related to cohorts' later educational, economic, or health outcomes. The reduced-form approach, on the other hand, shows that these long-term outcomes strongly improved among vaccinated cohorts in states with high pre-vaccine measles mortality. The researchers' results suggest that differences in disease severity are more relevant for long-term human capital impacts than raw differences in actual infection rates, supporting the reduced-form approach used in the economic literature.

The authors acknowledge support from the Value of Vaccination Research Network (Gates Foundation) in collaboration with Steve Elledge (Harvard University) and Michael Mina (eMed; Harvard University at project start). They are grateful to David Bloom, David Cutler, Atheendar Venkataramani, Casper Worm Hansen, and seminar participants at the Center for Disease Control (CDC), Harvard School of Public Health, Barcelona Summer Forum, and AEA Annual Meeting for comments. All of Alex Becker's work on this paper was carried out when he was employed at the Stanford Biology Department or the Johns Hopkins School of Public Health. This paper does not represent the views of his current employer (Merck).

1 Introduction

Epidemiological models of disease spread have become of increasing interest for economists analyzing the dynamics and societal impacts of the COVID-19 pandemic (Murray, 2020). But there has been limited use outside the current pandemic context, despite a rich and well-established literature in economics studying the impacts of disease spread on human capital accumulation and economic long-term outcomes.[1] Papers from this literature typically use reduced-form approaches that exploit public health interventions with a focus on the general severity of the pre-intervention disease environment, which is often proxied by pre-intervention mortality. But mortality might be a suboptimal proxy to measure a disease's long-term effects in the *surviving* population. Epidemiological models, in comparison, have the advantage that they can account for actual infection rates across cohorts. In this paper, we develop an epidemiological accounting model to compare the two approaches analyzing the long-term impacts of measles surrounding the measles vaccine introduction to the United States in the early 1960s.

Measles provide a useful context to assess the value of epidemiological modeling for the analysis of the economic consequences of disease spread. First, while measles infections can lead to immediate serious illness, they are also understood to partially eliminate previously obtained immunity to other childhood diseases leading to higher incidence of non-measles disease after a measles infection (Gadroen et al., 2018; Mina et al., 2015, 2019; Petrova et al., 2019). This impact on secondary infections makes measles spread an important driver of childhood sickness, with the potential of long-term consequences for human capital development (Atwood, 2022; Driessen et al., 2015; Nandi et al., 2019). Second, measles is extremely infectious with an $R0$ of 16 or higher (Guerra et al., 2017) and it is fully immunizing.[2] The high

[1] See, for example, Acemoglu and Johnson (2007); Ager et al. (2018); Bhalotra and Venkataramani (2015); Bleakley (2007, 2010); Bütikofer and Salvanes (2020); Cutler et al. (2010); Egedesø et al. (2020); Jayachandran et al. (2010); Lazuka (2020); Lucas (2010).

[2] $R0$ refers to the number of people infected by a single contagious individual in a fully

infectiousness implies that in an unvaccinated, fully susceptible population virtually everyone will be infected during the first years of life. Hence there is no socio-economic selection into infections and variation in infection rates at a given point in time are largely driven by the recent location-specific outbreak history. Moreover, the full immunization allows for better epidemiological accounting and a more precise assignment of observed cases across population groups. In sum, measles is understood to be a central driver of severe childhood sickness and is relatively easily accountable, providing a good case for the use of epidemiological modeling to measure the long-term human capital impacts of disease spread.

Our analysis relies on measles case counts between 1950 and 1980 that are available at the state level but without age information. We combine these counts with demographic data on births and population sizes to build an epidemiological dynamic accounting model which distributes observed cases across cohorts in each state. We show that the resulting measles distributions are highly predictive of future outbreaks in line with epidemiological Susceptible, Infectious, Recovered (SIR) models, corroborating the mechanics of our accounting framework. In contrast, the economic reduced-form approach using pre-intervention measles mortality rates has no predictive power of year-to-year outbreaks. We then use the accounting model to calculate the shares in each birth cohort that have ever been infected with measles. Finally, we estimate the impact of measles on long-term outcomes measured in the Census and the American Community Survey (ACS) comparing the epidemiological modeling approach and the reduced-form approach.

Linking outcomes in adulthood to measles infections in early childhood measured by the baseline epidemiological accounting model indicates strongly negative impacts on a broad range of adult outcomes. Cohorts with lower shares of people ever infected with measles have lower high school dropout rates, higher family income, and are less dependent on welfare income and

susceptible population.

food stamps. However, the variation in measles infection shares identifying these effects largely covaries with overall cohort trends and these estimated long-term effects are not robust to the inclusion of cohort fixed effects. The economic reduced-form approach, on the other hand, yields similar estimates as the baseline epidemiological model and is fully robust to controlling flexibly for cohort trends. The reduced form approach also has more predictive power of medium-term cohort mortality.

Focusing on the reduced-form approach, we extend the analysis to health outcomes that are observed in the ACS. We find positive impacts of measles reductions on a broad set of physical health difficulties, including work disability and difficulties with daily physical functioning.

Overall, our analysis suggests that the severity of the disease environment is a central factor determining early life impacts that persist throughout the life course – variation that is captured by the reduced-form approach. Differences in overall infection rates captured by the epidemiological approach that do not account for the severity of the disease burden are less consistently associated with long-term outcomes. Notably, the epidemiological model performs well in predicting year-to-year fluctuations in measles cases but this variation is not the relevant margin for impacts on child development.

Given its extremely high contagiousness and full immunization, measles provide an upper bound for the power of epidemiological modeling to measure the long-term effects of disease spread. Diseases that spread less aggressively or infect people numerous times over their lifetime, such as influenza or COVID-19, are much more difficult to model. Our analysis therefore provides strong support for the use of reduced-form approaches that focus on the severity of the disease environment when measuring the long-term benefits of disease reduction (Acemoglu and Johnson, 2007; Ager et al., 2018; Bhalotra and Venkataramani, 2015; Bleakley, 2007, 2010; Bütikofer and Salvanes, 2020; Cutler et al., 2010; Egedesø et al., 2020; Jayachandran et al., 2010; Lazuka, 2020; Lucas, 2010). We discuss these approaches in more detail in

Section 3.2.

Our results further contribute to a concentrated economic and public health literature that has focused on the human capital consequences of measles infections. Our reduced-form results are strongly in line with Atwood (2022) who studies the long-term effects of the measles vaccine introduction in the United States using a similar reduced-form approach (see Section 3.2 for a more detailed discussion). Positive long-term effects of measles vaccination have also been found for developing countries, such as Ethiopia, India, Vietnam, and Bangladesh (Driessen et al., 2015; Nandi et al., 2019).

Finally, our findings are relevant for policies aimed at increasing vaccination rates in low-income countries and maintaining high coverage in high-income countries (Bärnighausen et al., 2014b; Bloom et al., 2021; Nandi et al., 2020). Our analysis suggests that, from a child development perspective, vaccination efforts are particularly beneficial in contexts with high measles and childhood mortality – conditions that are nowadays mostly found in developing countries. At the same time, measles outbreaks in high-incomes context can be very disruptive reducing the educational inputs of those who fall sick and among healthy students if schools have to shut down to contain outbreaks.[3] Overall, our analysis emphasizes the central role of measles vaccination for child development around the world.

2 Data

Our analysis draws on several data sets. The first set of data concerns the *input parameters* for the epidemiological and the reduced-form models. Our epidemiological model is based on measles counts from weekly U.S. "notifiable diseases" reports obtained from Project Tycho (Van Panhuis et al., 2013)

[3]For news reports on recent school closures in response to measles outbreaks, see, for example: www.nytimes.com/2019/04/15/nyregion/measles-nyc-yeshiva-closing.html and www.sciencealert.com/the-clark-county-measles-outbreak-is-now-forcing-hundreds-of-kids-to-miss-school.

and linked to the Vital Statistics birth counts (Bailey et al., 2016). For vaccination rates in the introduction period we rely on the U.S.-wide numbers provided by Orenstein and Hinman (1999) and Orenstein et al. (2004). The reduced-form approach, instead, is based on measles mortality data from the Vital Statistics (National Center for Health Statistics, 1959–1978). The second set of data informs about our outcomes. Specifically, we rely on the US Census records in 2000 and the American Community Survey (ACS) in the years 2001–2019 (Ruggles et al., 2015) for long-term human capital accumulation of individuals born from 1950–1980.

2.1 Measles counts, birth, and mortality data

Measles counts. The Project Tycho data (Van Panhuis et al., 2013) provide weekly reports of all nationally notifiable diseases for U.S. cities and states from 1888 to 2011. For the case of measles, the data contain weekly state-level counts of measles cases and incidence rates per 100,000 inhabitants between 1928 to 2003, summing up to a total of 18,670,996 reported individual cases. Figure 1 plots weekly incidence rates for the overall United States over time, with the red dashed line indicating 1963, the year the measles vaccine was first introduced. As it can be seen, a large outbreak occurred in 1964, but incidence rates quickly dropped in the following years reaching low levels close to zero in 1968. Only very small annual outbreaks are observed between 1970 and 1980, while there are zero reported measles cases in most weeks after 1980 with only a few isolated small spikes in the following two decades.

Figure 2 plots annual measles incidence rates across all 51 U.S. states. There are three important observations. First, average incidence levels in the period before the vaccine introduction are very different across states, suggesting that there might be large differences in reporting rates. Second, the timing of measles outbreaks (peak years) is different across states. Third, while most states experience a strong reduction in incidence rates around the introduction of the vaccine, there is considerable variation in the year when

the last visible outbreak occurred.

Birth data. To build an epidemiological dynamic accounting model which distributes observed measles cases across cohorts in each state, we also need demographic data on births. Therefore, we rely on US county-level natality data starting in 1915 provided by Bailey et al. (2016) and aggregate these data to the state level.

Mortality data. For the economic reduced-form approach, instead, we need measles mortality data. We therefore use individual-level death counts at the state-level containing information on the cause of death from 1959 to 1979 from the National Vital Statistics System National Center for Health Statistics.[4]

2.2 Long-term human capital and health outcomes

Long-term human capital data. The 2000 Census and the American Community Surveys from 2001–2019 (Ruggles et al., 2015) provide information on the demographics, educational attainment, labor market outcomes, welfare dependency, and long-run health status. Demographic information contains the birthplace as well as the year of birth.[5] This information is crucial in order to align individuals with the level of the treatment (state and year of birth).

Educational information is measured by the respondents' highest year of school or degree completed. We then construct a high school completion dummy indicating, whether an individual has finished the 12th grade or more. As economic and labor market outcomes, we include respondents' employment status in the week prior to the survey as well as log household income (in real 2010 U.S. dollars).

[4]The cause of death is recorded according to the International Classification of Diseases (ICD) Revision 7 in the years 1959–1967 and Revision 8 for the years 1968–1979.

[5]The year of birth has been calculated by subtracting age from the survey year introducing an important source of measurement error (the 2000 Census takes place on April 1, while the ACS is administered throughout the year).

We also construct three binary measures representing welfare dependency, indicating (1) household income below the federal poverty line (2) receipt of welfare income at the individual level, and (3) receipt of food stamps at the family level.

Health outcomes. The Census and the ACS include two types of measures of respondents' health status: a binary variable indicating work disability (only available up to 2007), and a set of five disability questions (reporting cognitive, ambulatory, independent living, self-care, and vision or hearing complications) that we summarize in an equally weighted disability index.

Finally, in order to test our two models against each other for a measure where we have good theoretical predictions, we also study mortality as an outcome. In this analysis, we calculate two versions of mortality rates per 1,000 births by cumulating measles specific and overall deaths for the ages 0 to 5 for each specific birth cohort and diving by the number of births of the respective birth cohort.[6]

2.3 Sample restrictions and descriptive statistics

We restrict our sample to individuals born from 1950 to 1980 in the United States, excluding Alaska, Delaware, District of Columbia, Hawaii, Massachusetts, Nevada, North Dakota, Oregon, Rhode Island, and Wyoming which are excluded due to missing data in birth counts (Alaska and Hawaii), measles mortality data (Delaware, District of Columbia, Nevada, North Dakota, Oregon, Rhode Island, and Wyoming), or measles incidence data (Massachusetts).

With these restriction we yield a sample of 21,473,682 observations. The 2000 Census is a 5% sample of the US population and therefore contributes the most with 4,912,526 observations. The 2001–2004 ACS capture around 394,000 observations each year. Finally, the ACS from 2005–2019 are a 1-

[6]Using this measure as an outcome reduces our number of observations drastically, as (1) our cells are then on the level of state-of-birth x year-of birth and as (2) we can only include birth cohorts 1959–1973 due to data availability.

in-100 national random sample of the population and contribute with an average of around 1 million observations each year to our total sample.

For our analysis, we aggregate the data to cells at the state-of-birth x year-of-birth x age level as treatment in terms of early disease environment occurs at the state- and year-of-birth level. This aggregation results in 25,400 cells. Table 1 reports the descriptive statistics for the entire sample and the sub-samples divided into decades.

3 Approaches and empirical strategy

3.1 Epidemiological dynamic distribution model

In this section we develop a dynamic distribution model based on the SIR framework that distributes all measles counts observed across states and years across cohorts defined by birth year and state. The goal is to calculate the ratio of infections to cohort size which represents the share of a cohort that has ever been infected by measles.

If data were available on all measles infections including age information, calculating cohorts' share of ever infected individuals would be trivial. But, as typically the case for infectious disease data, reporting of measles infections in the analyzed time period was not complete and no information on the age at infection is available. We therefore first need to estimate states' reporting rates to obtain estimates of the total number of infections in a given state in a given year. Second, we need to know how many susceptible individuals, who have not been infected with measles before, remain in each cohort in each year. Our notation follows a discrete model with time units representing years and cohorts aging by one year each period.

3.1.1 Reporting rate estimation

In the pre-vaccination era we can roughly assume that everyone is infected with measles eventually (Finkenstädt and Grenfell, 2000). This implies that cumulative measles cases approximately equal cumulative births (Finkenstädt and Grenfell, 2000). We follow the existing literature and choose to link births to measles cases focussing on the pre-vaccination years 1950–1960. The slope of the regression of cases on births then represents the reporting rate (see Figure A1 for an illustration in the case of Oklahoma). Figure 3 shows the distribution of the estimated reporting rate across states. A significant share of states has reporting rates below 10% and the top state has a rate of 38%. The average reporting rate across states is 13%. Overall, this shows that reporting is fairly low and heterogeneous across states, emphasizing the importance accounting for these differences. Appendix Figure A2 shows the spatial distribution of reporting rates across the United States. Reporting rates seem somewhat lower in the South (with the exception of Texas) than in the North, in particular the North-East. But there is no strong spatial pattern. We can modify the reporting rate (e.g. by taking doubling the rate or cutting it in half) or the shape of the reporting rate (linear or changing over time) without changing the main results in a qualitative way.

3.1.2 Distributing cases across cohorts

Next, we need to distribute the total number of cases (reported cases divided by the reporting rate) across cohorts. We assume that every person can only get infected once and that the risk of infection is independent of an individual's age conditional on prior infection status.[7] Hence, infections in a given state and year are distributed across cohorts based on each cohort's

[7]This implies that cases are equally distributed across all susceptibles in a state regardless of their age. However, more cases will be distributed to younger ages as the share of susceptibles decreases with age. We can modify the assumption of age-independent infection risk, e.g. by assigning higher risks to younger individuals. The main results remain qualitatively unchanged.

share of total susceptibles in the population:

$$I(t,a) = I(t) * \frac{S(t,a)}{\sum S(t,a)} \tag{1}$$

$$\text{with } I(t) = \frac{I_{reported}(t)}{\text{reportingrate}} \ , \tag{2}$$

where t and a refer to year and age, respectively, and the total infections $I(t)$ correspond to the ratio of reported cases divided by the estimated reporting rate.

Unvaccinated cohorts become fully susceptible at age 1 when they lose maternal immunization (Sato et al., 1979)[8] while vaccination is modeled as a reduction in the number of births (Keeling and Rohani, 2011):

$$S(t+1, a=1) = \text{births}(t) \times [1 - \text{vaccination}(t+1)] \text{ if age=1.} \tag{3}$$

After age 1, the stock of susceptibles within a cohort follows the following simple law of motion:

$$S(t+1, a+1) = S(t,a) - I(t,a) \text{ if age>1.} \tag{4}$$

Finally, our dynamic distribution model needs to start in some period, and in this initial period we do not know the number of susceptibles remaining in cohorts that are older than 1, as infections in prior years are not observed. Our initial time period (1935) lies in the pre-vaccination era and we can approximate the share of susceptibles across age groups by average

[8]For a review on passive transmission of antibodies against measles in newborns see Leuridan and Van Damme (2007). Interestingly, immunization duration differs for newborns from mothers who naturally acquired measles or mothers who acquired antibodies vaccine-induced . While in our model we rely on passive immunization from mothers who were naturally infected, women vaccinated with live attenuated measles vaccine generally pass on shorter protection against measles to their children up to the age of 8 months (Lennon and Black, 1986).

susceptibility curves ($AS(a)$) from pre-vaccination era studies, calibrated to match an overall population susceptibility rate of 2.5 percent (Orenstein et al., 1985).[9] Hence, susceptibles across age in the initial period are determined by:

$$S(t = 1, a) = \text{pop}(t = 1, a) * AS(a) \tag{5}$$

3.1.3 Calculating cohorts' share ever infected

Having distributed all infections across cohorts, we then sum up infections over time within each cohort and divide by cohort size. This then gives us the share of individuals ever infected. We run the model at the state level though we use U.S.-wide vaccination rates as there is no state-level data on the vaccine rollout. For a graphical overview of states' share ever infected see Appendix Figure A3.

3.2 Economic "reduced-form" approach

A rich and prominent literature in economics has accumulated evidence that the early life disease environment has strong and persistent impacts on the population's human capital accumulation. Studies from that literature have in common that they typically identify the long-term effects of disease spread by exploiting some type of health intervention or sudden society-wide health improvement and interacting it with a measure of baseline disease exposure (often mortality). These approaches typically are build on a version of the following regression model

$$Y_{s,c} = \alpha + \gamma(Tpre_c \times Pre_s) + \theta_s + \delta_c + \epsilon_{s,c} \ , \tag{6}$$

[9] In the study by Orenstein et al. (1985), the data covers the pre-vaccine era from three different locations and dates. Results are all basically identical, so that there is no reason to suspect that that profile would look much different for any other location or time in the pre-vaccine era.

where s refers to state (or region, or country), c is the birth cohort (or simply time), Pre_s is a proxy of the disease exposure (e.g. mortality) during the pre-intervention, $Tpre_c$ is a dummy for the pre-intervention period, and θ_s and δ_c are state and cohort fixed effects.

Prominent economic studies using such an approach are, among others, Acemoglu and Johnson (2007), Ager et al. (2018), Bhalotra and Venkataramani (2015), Bleakley (2007), Bleakley (2010), Bütikofer and Salvanes (2020), Cutler et al. (2010), Egedesø et al. (2020), Jayachandran et al. (2010), Lazuka (2020), and Lucas (2010).

For instance, Acemoglu and Johnson (2007) interact major international health improvements in the 1940s with pre-1940 mortality rates from 15 leading infectious diseases to study life expectancy on economic performance. Bleakley (2007), more specifically, combines pre-eradication distribution of hookworm disease with the hookworm-eradication campaign to show the beneficial impact on education while Bleakley (2010), Cutler et al. (2010) and Lucas (2010) obtain plausible exogenous variation in child health by relying on malaria eradication in the US, India, or Paraguay and Sri Lanka to study the impact on adult labor market outcomes, consumption, and educational attainment.

In the same vein Ager et al. (2018) combine pre-vaccination variation in smallpox mortality across Swedish counties with the introduction of the smallpox vaccine, while Bütikofer and Salvanes (2020) and Egedesø et al. (2020) interact tuberculosis infection rates and mortality with tuberculosis testing and vaccination in Norway and tuberculosis dispensaries in Danish cities, respectively. Finally, Bhalotra and Venkataramani (2015), Jayachandran et al. (2010), and Lazuka (2020) exploit the introduction of sulfa drugs, a medical innovation that reduced exposure to pneumonia in infancy, interacted with pre-innovation pneumonia mortality to examine the long-term impacts on educational and labor market outcomes of affected cohorts.

In recent contemporaneous work, Atwood (2022) examines the long-run

effects of measles vaccination on earnings and employment interacting the measles vaccine introduction with pre-vaccine levels of reported measles incidence. The long-term effects reported in Atwood (2022) are qualitatively and quantitatively very close to the effects we estimate despite our use of pre-vaccine measles mortality rather than pre-vaccine measles incidence in our baseline economic reduced-form approach. The similarity of our estimates suggests that, in the case of measles, reported incidence rates are a good proxy for disease severity and mortality.

Exceptions in terms of their empirical designs are provided by Driessen et al. (2015) and Nandi et al. (2019), who study measles vaccination impacts on child development in Bangladesh, Ethiopia, India, and Vietnam. Driessen et al. (2015) exploit a slow, staggered roll-out of the measles vaccine across different areas in Bangladesh allowing to study an area acting as a control group while only being treated several years after. Nandi et al. (2019) exploit quasi-experimental variation in vaccine coverage in low-income countries using longitudinal data that includes information on vaccination status. While these designs are not feasible in most developed countries where vaccine coverage is near universal and the initial roll-out has been fast, these studies provide important evidence of positive vaccination impacts on a broad range of developmental outcomes, including child anthropometry, cognition, schooling enrolment, and test scores.

In this paper, we base our reduced-form specification on the most common approach used in the literature exploiting local differences in measles mortality prior to the introduction of the vaccine interacted with the vaccine roll-out. Specifically, we calculate state-level measles mortality for the age groups 0–5 averaged across the years 1959–1963 and interact them with the introduction of the measles vaccine in the United States. States individual pre-vaccine mortality rates are scaled so that the overall average for the United States equals one.

3.3 Empirical strategy

To examine the relationship between the early childhood disease environment and adult long-term outcomes, we estimate two types of linear regressions for age group a in birth cohort c in state of birth s. Specifically, in case of the epidemiological model the equation looks as follows:

$$Y_{sca} = \alpha + \gamma ShareEverInfected_{sc} + \beta_a + \delta_c + \theta_s + \epsilon_{sca} \ , \tag{7}$$

where $ShareEverInfected_{sc}$ is the share of each cohort ever infected with measles derived from the epidemiological distribution model. For the reduced-form approach, we adjust Equation (7) to the following version:

$$Y_{sca} = \alpha + \gamma(Tpre_c \times Pre_s) + \beta_a + \delta_c + \theta_s + \epsilon_{sca} \ , \tag{8}$$

where Pre_s corresponds to states' measles mortality (ages 0–5, in 1959–1963) and $Tpre_c$ is a dummy equal to one for cohorts born in or prior to 1963, the year of the measles vaccine introduction.

In both equations, Y_{sca} stands for the different averaged individual long-term outcomes, such as high school completion, employment status, total household income (in logarithmic terms), poverty status, receiving any welfare transfers or food stamps as well as health outcomes such as a disability status at work and a disability index comprised of impairments in 5 daily-life dimensions. Furthermore, we control for age, birth cohort, and state of birth fixed effects by β_a, δ_c, and θ_s, respectively. We also conduct an alternative specification where we control for cohort polynomials instead of cohort fixed effects. Our coefficient of interest is γ, representing the estimate of the effect of measles exposure in early childhood. Standard errors are clustered at the state of birth level (Bertrand et al., 2004) and observations are weighted by cohort size.

Combining both approaches, we also estimate a Bartik-type regression:

$$Y_{sca} = \alpha + \gamma(Pre_s \times ShareEverInfected_{sc}) + \beta_a + \delta_c + \theta_s + \epsilon_{sca} \ , \qquad (9)$$

where Pre_s is now multiplied with the $ShareEverInfected_{sc}$. All other variables are defined as above.

Figure 4 shows the epidemiological "share-ever-infected" measure, the reduced-form measure, and the combined Bartik measure for five selected states and averaged across the United States. As it can be seen, the variation in the epidemiological measure is driven entirely by the different timing in the reduction of infection shares. All states start with a share close to one and end with a share close to zero, but some states experience reductions earlier than others. The variation in the reduced-form measure, on the other hand, is entirely driven by pre-existing level differences while the time variation is uniform across all states. The Bartik measure combines the two types of variations.

4 Results

4.1 Preditive power of epidemiological measure

Before turning to our main results, we first explore the predictive power of our dynamic distribution model. Our distribution model is based on the SIR framework that is centred around the key assumption that outbreaks are driven by susceptibles who drop out of the susceptible pool once they have been infected. A larger pool of susceptibles in one period therefore increases the risk of an outbreak in the following period. An outbreak in a given period, on the other hand, leads to a smaller pool of susceptibles in the following period. In Table 2 we test these predictions using state-specific annual susceptible counts generated from our dynamic distribution model. The table shows coefficients from a regression of state-level measles counts

on lags and leads of susceptible counts while controlling for state and year fixed effects. In line with the mechanics of the SIR model, lagged susceptible counts strongly predict measles outbreaks (column 1) while these outbreaks are strongly negatively related to susceptible counts in the subsequent period (column 2). Note that our dynamic distribution model generates susceptible based on lagged measles outbreaks, hence the relationship in column (2) is mechanical. However, the information about future outbreaks does not feed into susceptible counts and therefore the result in column (1) shows the predictive power of the model. Column (3) shows the predictive power of the lagged susceptible measure when it is split into deciles. Columns (4–9) test the predictive power within subperiods of our data. Remarkably, our model is predictive of future outbreaks during the pre-vaccine era (columns 4–5), as well as during the period surrounding the introduction of the vaccination (columns 6–7) and during the post-vaccine era when measles were largely eliminated and only idiosyncratic minor outbreaks occurred (columns 8–9). These results validate our model across three periods with very distinct measles environments.

Appendix Table A1 compares the predictive power of the dynamic distribution model to the reduced-form measure. As shown in column (3), the reduced form measure has no predictive power of future outbreaks and is also not correlated with lagged outbreaks. Moreover, since the measure only changes in the year of the measles introduction and the regressions include state fixed effects, the measures' impact cannot be estimated in the pre- and the post-vaccine era (columns 6 and 10).

To sum up, our epidemiological distribution model generates a distribution of measles cases that is highly predictive of measles outbreaks while the reduced-form measure is uncorrelated with year-to-year fluctuations in measles cases.

4.2 Long-term effects on human capital and health outcomes

Figures 5–7 compare the epidemiological approach with the reduced-form approach as well as the combined Bartik approach in terms of the long-term impacts on human capital outcomes. The figures show regression coefficients based on Equations (7)–(9), both for versions with cohort quadratic trends and cohort fixed effects. Corresponding numerical results are reported in Tables 3–5.

The top left panel of Figure 5 plots the estimated effects of early life measles exposure on high school completion. The darker colored bars show the estimates resulting from regressions including quadratic cohort trends and they indicate similar effects across all three approaches. The coefficient for the epidemiological model is -0.539 with a standard error of 0.196 (Table 3, Panel A1, column 1), indicating that a cohort that has been fully exposed to measles early in life (ever infected share equals one) has a 0.54 percentage points lower high school completing rate than a cohort with zero exposure. The point estimate from the reduced form specification (Table 3, Panel B1, column 1), suggests that states with a pre-vaccination era measles mortality around the U.S. average at that time experienced a significant increase in high school completion of about 0.85 percentage points in response to the introduction of the vaccine. The Bartik specification (Table 3, Panel C1, column 1) which combines the two approaches results in a similar point estimate, indicating that decreasing the share of individuals ever infected with measles in a cohort from one to zero, in a state with the average pre-vaccination measles rate, significantly increases high school completion by 1.05 percentage points. When we control for cohort fixed effects instead of quadratic trends, the approaches based on the reduced-form measure result in very similar estimates (Table 3, Panels B2 and C2, column 1). The estimate from the epidemiological model (Table 3, Panel A2, column 1), however, flips sign and looses in significance when cohort fixed effects are included.

The pattern for high school completion is observed similarly for log household income, poverty, welfare income receipt, and food stamp receipt. In the specifications with cohort trends all approaches indicate positive impacts of reducing measles exposure by improving these long-term outcomes. While the reduced form approach is robust to the inclusion of cohort fixed effects, the epidemiological modeling approach is, however, not. An exception is the unemployment rate which is not consistently affected in the specifications with cohort fixed effects.

Figure 6 and Table 4 show the same set of regression results for health measures that are reported in the Census and ACS. Again the reduced-form and the Bartik approach result in very similar point estimates that are robust to the inclusion of cohort fixed effects, indicating negative long-term health consequences of early life measles exposure. For example, the reduced form specification with cohort fixed effects indicates that the measles vaccine introduction reduced work disability, cognitive disability, self-care disability, independent living difficulties, and vision or hearing difficulties by 0.8, 0.4, 0.3, 0.5, and 0.5 percentage points, respectively. Relative to the sample mean of these outcomes, the effect sizes range between 8 and 12.8 percent. The epidemiological modeling approach, on the other hand, is not robust to the inclusion of different cohort effects. Surprisingly, the approach results in significantly negative health impacts (i.e. a higher share of ever infected individuals is associated with less disability and fewer health difficulties).

The negative long-run impacts of early life measles exposure on human capital and health outcomes estimated via the reduced-form approach are in line with the long-term disease impacts known from the economic literature while it is more difficult to reconcile the estimates resulting from the epidemiological modeling approach, especially the reverse health impacts. At the same time, the economic literature relies on similar reduced form approaches, so the comparability of results does not provide an independent assessment. We therefore analyse two additional outcomes that are known

to be directly linked to measles infection rates: Measles mortality and overall mortality in early childhood (ages 0–5). While measles mortality is obviously linked to measles infections, the epidemiological literature has argued that measles infections also increase the risk of childhood mortality from other diseases (Mina et al., 2015).

Figure 7 (estimates reported in Table 5) shows that the reduced-form approach indeed indicates a significantly positive impact of higher disease exposure both on measles mortality and overall childhood mortality. The epidemiological model approach, on the other hand does not result in a significant effect for the cohort trends model while it indicates an implausible mortality reduction in response to higher measles infection shares when cohort fixed effects are controlled for. Notice that the reduced-from measure is based on pre-vaccination measles mortality and therefore more directly linked to mortality changes. But there is no reasonable explanation why a cohort's share of individuals ever infected with measles should be negatively associated with childhood mortality. The mortality results therefore support the notion that the epidemiological modeling approach provides variation in measles infections that are difficult to disentangle from other factors driving cohort outcomes.

4.3 Alternative epidemiological modeling assumptions

Finally, one might wonder to which extent the effect patterns observed for the epidemiological approach depend on the assumptions built into the dynamic distribution model. Appendix Figure A5 plots the estimated impacts of the share ever infected on high school graduation and log household income across 12 models with varying alternative parameter specifications. In particular, we vary the initial share of susceptibles in the population (S_0), we relax the assumption that infection rates among susceptibles are independent of age (assuming higher infection rates below age 6), and we allow for different functional forms when estimating the reporting function (default is the mean

reporting rate, and we show results allowing for a linear time trend, a local polynomial trend (LOESS), and a spline function). As it can be seen in Figure A5, estimates are very similar across the different models and the inclusion of cohort fixed effects has a similar impact on the effect pattern. These results suggest that the dynamic accounting model is not sensitive to the core assumptions.

5 Discussion and conclusion

This paper uses an epidemiological modeling and an economic reduced-form approach to measure the long-term impacts of early-life measles exposure around the introduction of the measles vaccine in the United States. The reduced-form approach reveals large improvements in educational, labor market, and health outcomes in adulthood among cohorts born after the vaccine introduction in states with historically high measles mortality. No consistent effect patterns are detected using the epidemiological modeling approach which exploits differences in cohorts' share of individuals that were ever infected with measles. These results suggest that differences in the severity of measles infections, proxied by measles mortality, matter more for long-term outcomes than the difference in mere infection rates itself.

The long-term impacts of early life measles exposure are in line with the economic literature that has documented, among others, long-term consequences of exposure to hookworm (Bleakley, 2010), malaria (Cutler et al., 2010; Lucas, 2010), and tuberculosis (Bütikofer and Salvanes, 2020; Egedesø et al., 2020). Our finding that the severity of the disease environment matters more than the measles infection rate itself is also in line with the notion that measles infections are particularly harmful for children as they can eliminate previously obtained immunity to other diseases (Gadroen et al., 2018; Mina et al., 2015, 2019; Petrova et al., 2019). This mechanism should matter most in a severe disease environment with high childhood mortality not only for

measles but also other childhood diseases.

Our estimated impacts are also very much in line with the contemporaneous analysis of Atwood (2022). We find that the introduction of the vaccine lead to a 2.7 percent increase in log total family income, a reduction of the poverty rate by 5.7 percent and a non-significant reduction of unemployment of 0.1 percent, while Atwood (2022) reports a family income increase of 1.7 percent, a poverty reduction of 7.3 percent, and an increase of 0.3 percent for employment. A key difference in our approach is that we use pre-vaccine measles mortality as a proxy for disease burden while Atwood (2022) uses pre-vaccine measles cases that are not adjusted for reporting rate differences. The similarity of our reduced-form estimates suggests that, at least in the context of the pre-vaccine United States, reporting rates can be used as a proxy for mortality and disease burden (presumably as more severe cases of measles were more likely to be reported). This insight further emphasizes that the epidemiological distribution model, which fully adjusts for differences in reporting rates across locations, focuses on a very different margin.

Despite the discovery of an effective measles vaccine over half a century ago and its world-wide dissemination, measles outbreaks have been rising again over the past decade. The COVID-19 pandemic is expected to further increase vaccination hesitancy and the risk of future measles outbreaks (Roberts, 2020). Our analysis documents large life-long welfare gains among cohorts that benefited from the introduction of the measles vaccine in the United States emphasizing the importance of a continued vaccination effort (Bärnighausen et al., 2014a; Bloom et al., 2018). However, our comparison of epidemiological and reduced-form approaches indicates that vaccination benefits in terms of long-term human capital impacts are particularly large in environments with severe childhood disease burdens. Our results are therefore particularly supportive of vaccination efforts in low-income countries with high childhood mortality (Bärnighausen et al., 2014b; Nandi et al., 2020).

References

Acemoglu, Daron and Simon Johnson, "Disease and development: The effect of life expectancy on economic growth," *Journal of Political Economy*, 2007, *115* (6), 925–985.

Ager, Philipp, Casper Worm Hansen, and Peter Sandholt Jensen, "Fertility and early-life mortality: Evidence from smallpox vaccination in Sweden," *Journal of the European Economic Association*, 2018, *16* (2), 487–521.

Atwood, Alicia, "The long-term effects of measles vaccination on earnings and Employment," *American Economic Journal: Economic Policy*, 2022, *14* (2), 34–60.

Bailey, Martha, Price Fishback, Michael Haines, Shawn Kantor, Edson Severnini, and Anna Wentz, "US county-level natality and mortality data, 1915–2007," 2016. Inter-university Consortium for Political and Social Research (ICPSR).

Bärnighausen, Till, David E Bloom, Elizabeth T Cafiero-Fonseca, and Jennifer Carroll O'Brien, "Valuing vaccination," *Proceedings of the National Academy of Sciences*, 2014, *111* (34), 12313–12319.

_ , Seth Berkley, Zulfiqar A Bhutta, David M Bishai, Maureen M Black, David E Bloom, Dagna Constenla, Julia Driessen, John Edmunds, David Evans et al., "Reassessing the value of vaccines," *The Lancet Global Health*, 2014, *2* (5), e251–e252.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How much should we trust differences-in-differences estimates?," *The Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Bhalotra, Sonia R and Atheendar Venkataramani, "Shadows of the captain of the men of death: Early life health interventions, human capital investments, and institutions," *Human Capital Investments, and Institutions (August 8, 2015)*, 2015.

Bleakley, Hoyt, "Disease and development: Evidence from hookworm eradication in the American South," *The Quarterly Journal of Economics*, 2007, *122* (1), 73–117.

_ , "Malaria eradication in the Americas: A retrospective analysis of childhood exposure," *American Economic Journal: Applied Economics*, 2010, *2* (2), 1–45.

Bloom, David E, Daniel Cadarette, and Maddalena Ferranna, "The societal value of vaccination in the age of COVID-19," *American Journal of Public Health*, 2021, *111* (6), 1049–1054.

_ , Victoria Y Fan, and JP Sevilla, "The broad socioeconomic benefits of vaccination," *Science Translational Medicine*, 2018, *10* (441), eaaj2345.

Bütikofer, Aline and Kjell G Salvanes, "Disease control and inequality reduction: Evidence from a tuberculosis testing and vaccination campaign," *The Review of Economic Studies*, 2020, *87* (5), 2087–2125.

Cutler, David, Winnie Fung, Michael Kremer, Monica Singhal, and Tom Vogl, "Early-life malaria exposure and adult outcomes: Evidence from malaria eradication in India," *American Economic Journal: Applied Economics*, 2010, *2* (2), 72–94.

Driessen, J, A Razzaque, D Walker, and David Canning, "The effect of childhood measles vaccination on school enrolment in Matlab, Bangladesh," *Applied Economics*, 2015, *47* (55), 6019–6040.

Egedesø, Peter Juul, Casper Worm Hansen, and Peter Sandholt Jensen, "Preventing the white death: Tuberculosis dispensaries," *The Economic Journal*, 2020, *130* (629), 1288–1316.

Finkenstädt, Bärbel F and Bryan T Grenfell, "Time series modelling of childhood diseases: A dynamical systems approach," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2000, *49* (2), 187–205.

Gadroen, Kartini, Caitlin N Dodd, Gwen MC Masclee, Maria AJ De Ridder, Daniel Weibel, Michael J Mina, Bryan T Grenfell, Miriam CJM Sturkenboom, David AMC Van De Vijver, and Rik L De Swart, "Impact and longevity of measles-associated immune suppression: A matched cohort study using data from the THIN general practice database in the UK," *BMJ Open*, 2018, *8* (11), e021465.

Guerra, Fiona M, Shelly Bolotin, Gillian Lim, Jane Heffernan, Shelley L Deeks, Ye Li, and Natasha S Crowcroft, "The basic reproduction number (R0) of measles: A systematic review," *The Lancet Infectious Diseases*, 2017, *17* (12), e420–e428.

Jayachandran, Seema, Adriana Lleras-Muney, and Kimberly V Smith, "Modern medicine and the twentieth century decline in mortality: Evidence on the impact of sulfa drugs," *American Economic Journal: Applied Economics*, 2010, *2* (2), 118–46.

Keeling, Matt J and Pejman Rohani, "Modeling infectious diseases in humans and animals," in "Modeling infectious diseases in humans and animals," Princeton university press, 2011.

Lazuka, Volha, "Infant health and later-life labor market outcomes: Evidence from the introduction of sulfa antibiotics in Sweden," *Journal of Human Resources*, 2020, *55* (2), 660–698.

Lennon, Jeffrey L and Francis L Black, "Maternally derived measles immunity in era of vaccine-protected mothers," *The Journal of Pediatrics*, 1986, *108* (5), 671–676.

Leuridan, E and P Van Damme, "Passive transmission and persistence of naturally acquired or vaccine-induced maternal antibodies against measles in newborns," *Vaccine*, 2007, *25* (34), 6296–6304.

Lucas, Adrienne M, "Malaria eradication and educational attainment: Evidence from Paraguay and Sri Lanka," *American Economic Journal: Applied Economics*, 2010, *2* (2), 46–71.

Mina, Michael J, Tomasz Kula, Yumei Leng, Mamie Li, Rory D De Vries, Mikael Knip, Heli Siljander, Marian Rewers, David F Choy, Mark S Wilson et al., "Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens," *Science*, 2019, *366* (6465), 599–606.

Mina, Michale J., C. Jessica E. Metcalf, Rik L. de Swart, A.D.M.E. Osterhaus, and Bryan T. Grenfell, "Long-term measles-induced immunomodulation increases overall childhood infectious disease mortality," *Science*, 2015, *348* (6235), 694–699.

Murray, Eleanor J, "Epidemiology's time of need: COVID-19 calls for epidemic-related economics," *Journal of Economic Perspectives*, 2020, *34* (4), 105–20.

Nandi, Arindam, Anita Shet, Jere R Behrman, Maureen M Black, David E Bloom, and Ramanan Laxminarayan, "Anthropometric, cognitive, and schooling benefits of measles vaccination: Longitudinal cohort analysis in Ethiopia, India, and Vietnam," *Vaccine*, 2019, *37* (31), 4336–4343.

_ **, Santosh Kumar, Anita Shet, David E Bloom, and Ramanan Laxminarayan**, "Childhood vaccinations and adult schooling attainment: Long-term evidence from India's Universal Immunization Programme," *Social Science & Medicine*, 2020, *250*, 112885.

National Center for Health Statistics, "Vital statistics of the United States. Vol II. Mortality. Part A. Hyattsville, Maryland: US Department of Health and Human Services," Public Health Serivce 1959-1978.

Orenstein, Walter A, Alan R Hinman, and Mark J Papania, "Evolution of measles elimination strategies in the United States," *The Journal of Infectious Diseases*, 2004, *189* (Supplement_1), S17–S22.

_ **and** _ , "The immunization system in the United States—the role of school immunization laws," *Vaccine*, 1999, *17*, S19–S24.

_ **, Roger H Bernier, Timothy J Dondero, Alan R Hinman, James S Marks, Kenneth J Bart, and Barry Sirotkin**, "Field evaluation of vaccine efficacy," *Bulletin of the World Health Organization*, 1985, *63* (6), 1055.

Panhuis, W. G. Van, J. Grefenstette, S. Y. Jung, N.S. Chock, A. Cross, H. Eng, B.Y. Lee, V. Zadorozhny, S. Brown, D. Cummings, and Burke D.S., "Contagious diseases in the United States from 1888 to the present," *The New England Journal of Medicine*, November 2013, *369* (22), 2152–2158.

Petrova, Velislava N, Bevan Sawatsky, Alvin X Han, Brigitta M Laksono, Lisa Walz, Edyth Parker, Kathrin Pieper, Carl A Anderson, Rory D de Vries, Antonio Lanzavecchia et al., "Incomplete

genetic reconstitution of B cell pools contributes to prolonged immunosuppression after measles," *Science Immunology*, 2019, *4* (41), eaay6125.

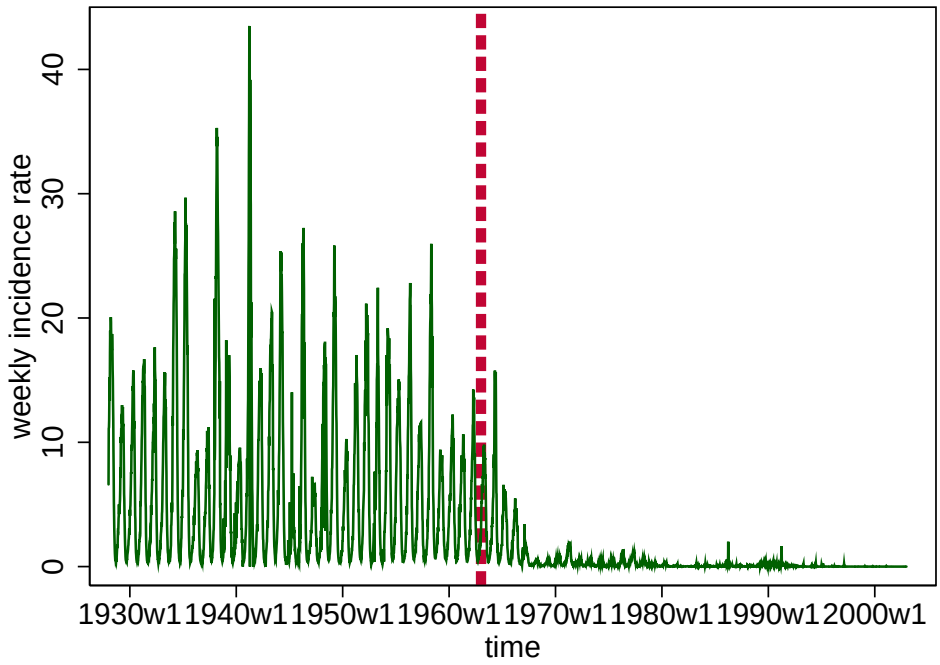
Roberts, Leslie, "Why measles deaths are surging–and coronavirus could make it worse.," *Nature*, 2020, *580* (7804), 446–448.

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek, "Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]. Minneapolis: University of Minnesota," 2015.

Sato, Hiroshi, Paul Albrecht, David W Reynolds, Sergio Stagno, and Francis A Ennis, "Transfer of measles, mumps, and rubella antibodies from mother to infant: Its effect on measles, mumps, and rubella immunization," *American Journal of Diseases of Children*, 1979, *133* (12), 1240–1243.

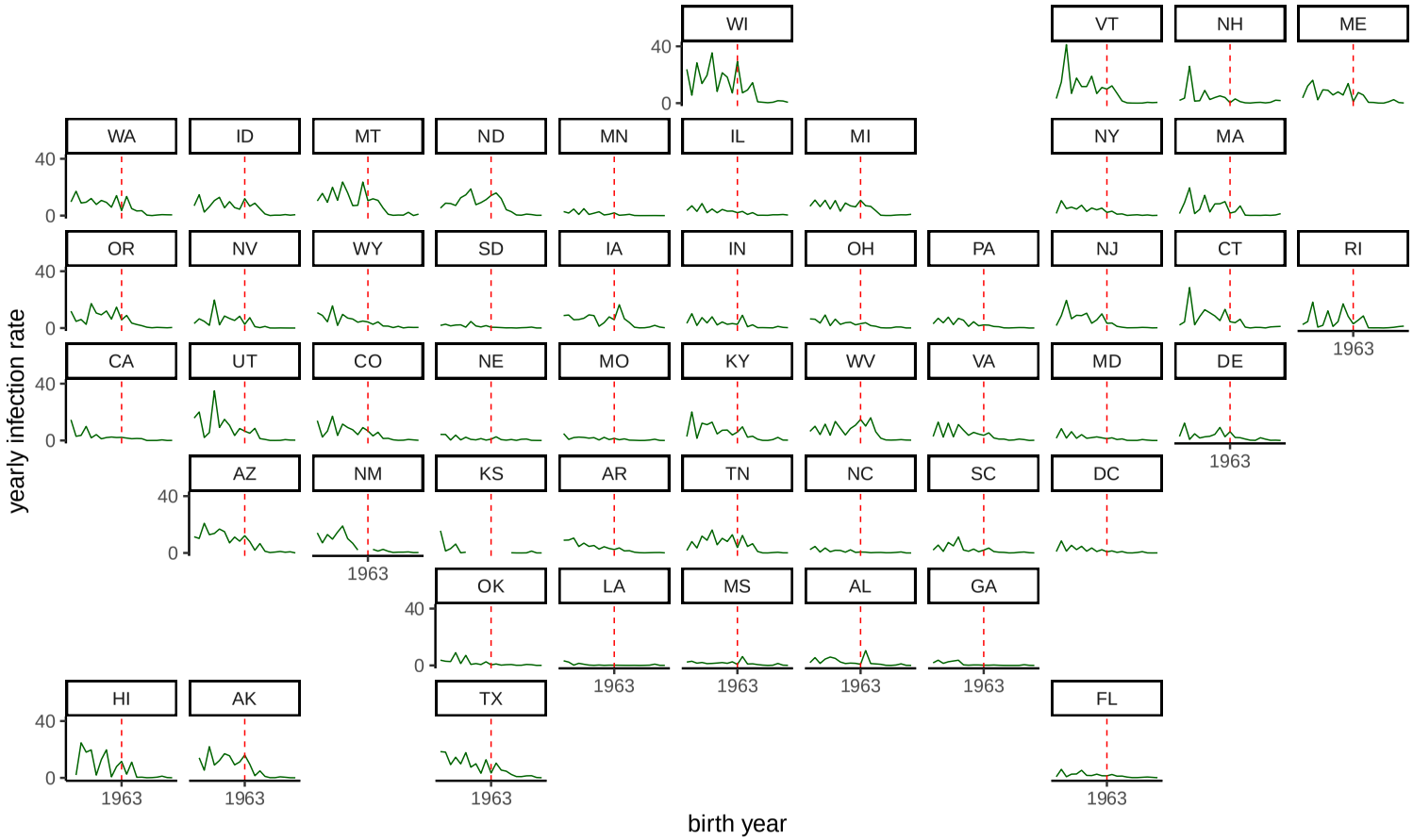
Figures and Tables

Figure 1: Overall US measles incidence rate



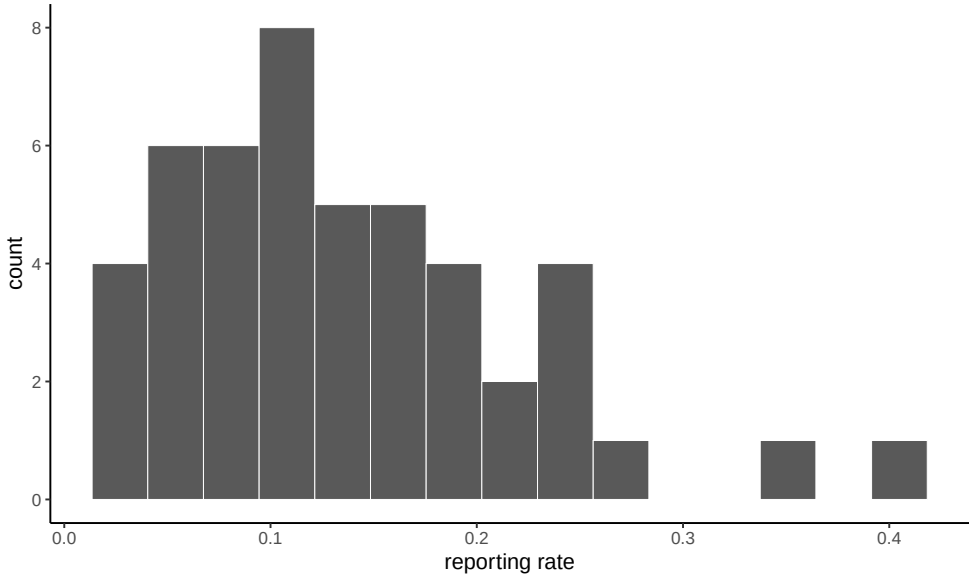
Notes: Overall US measles incidence rate per 100,000 based on weekly reports from 1928 to 2003 in all US states (weighted with interpolated population per state in 1963 to get the overall incidence rate, where population data per state were obtained from the US Census Bureau). The vertical dashed line indicates the year of introduction of the measles vaccine in 1963.

Figure 2: Measles incidence rate by state: 1953–1973



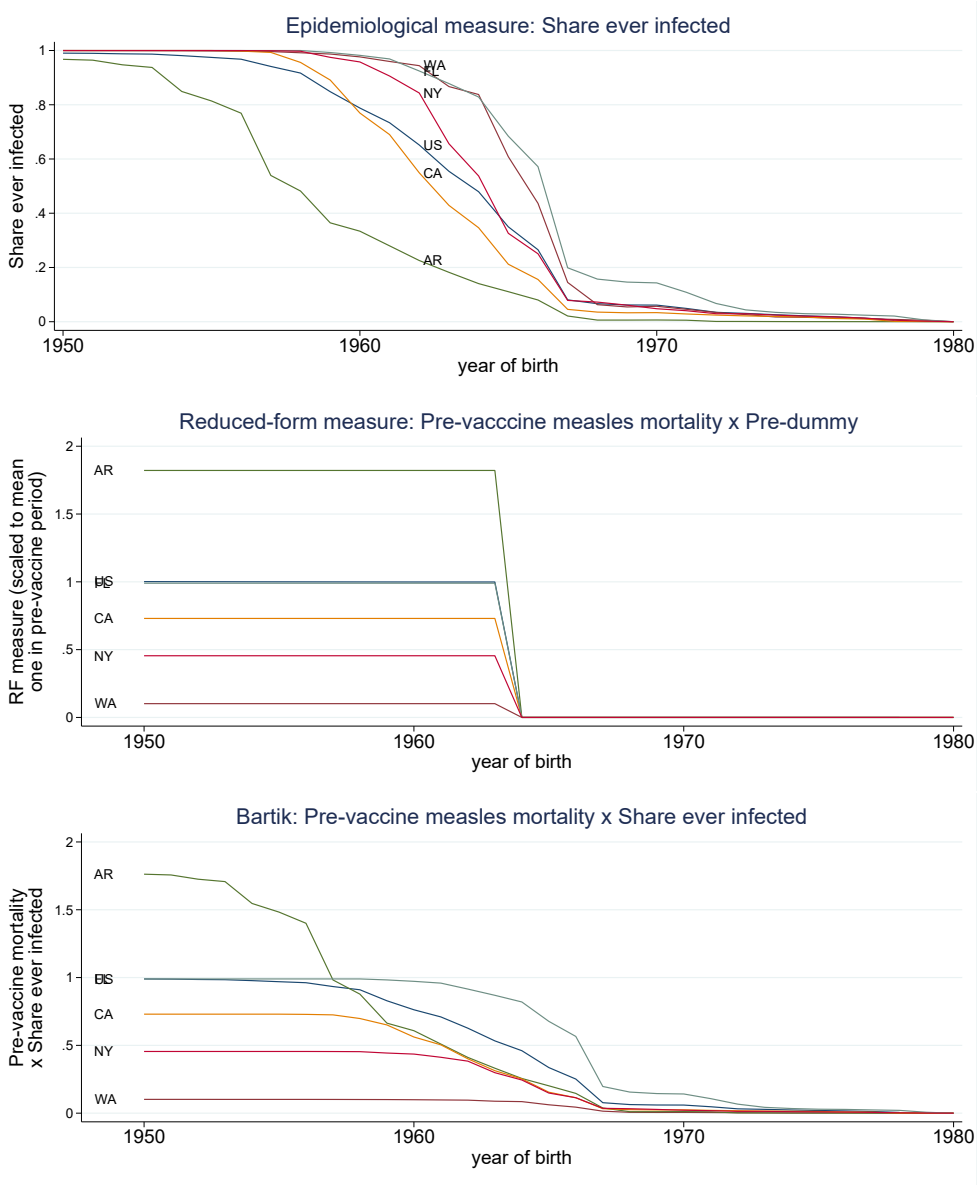
Notes: Incidence rate of measles per 100,000 in the period of 1953 to 1973 for all US states. The vertical dashed line indicates the year of introduction of the measles vaccine in 1963.

Figure 3: Distribution of state-level reporting rates



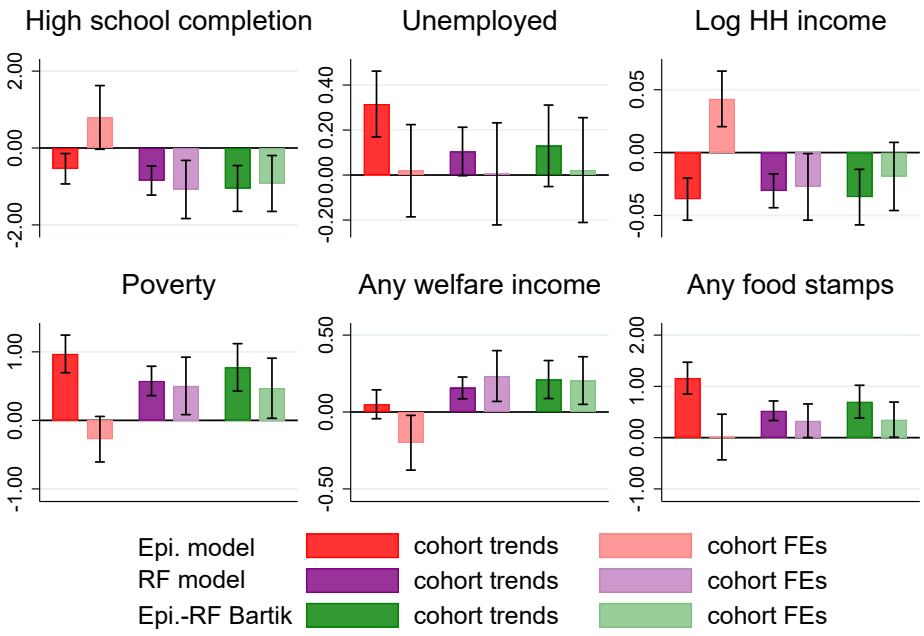
Notes: This figure shows the distribution of reporting rates across U.S. states estimated using the epidemiological approach described in Section 3.1.

Figure 4: Share ever infected and pre-vaccine variation in measles mortality



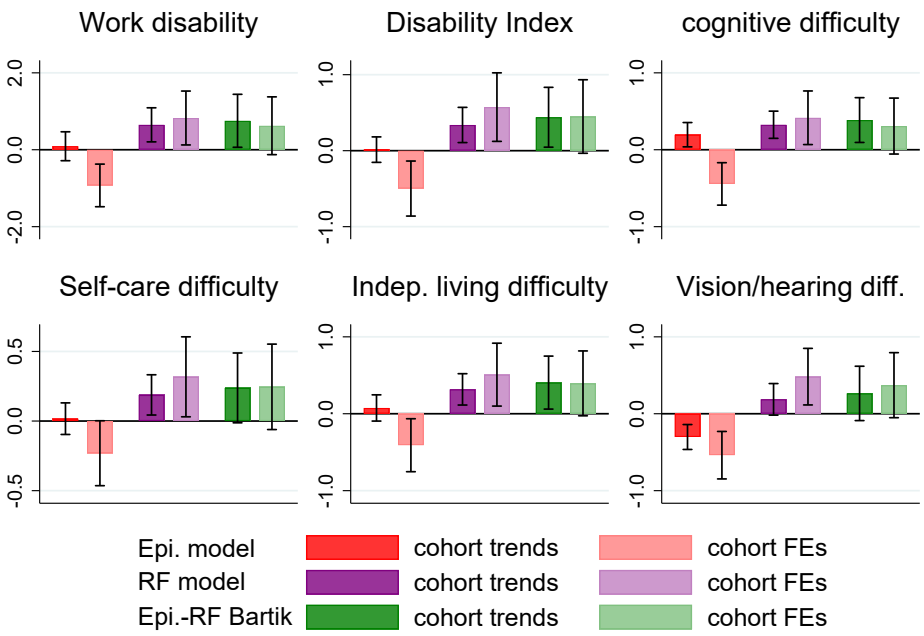
Notes: This figure depicts the three different measures used in the analysis for the US (blue line) and 5 sample states (Arkansas, California, Florida, New York, and Washington). Measures are scaled so that the overall US measure moves from 1 to 0 from 1950 to 1980. Panel A shows the share infected derived from the epidemiological (Epi.) model. Panel B shows the interaction of the states' measles mortality (ages 0—5, in 1959—1963) with a dummy equal to one for cohorts born prior (<=) 1963 (year of measles vaccine introduction), the reduced-form (RF) measure. Panel C shows a combination of both approaches — replacing the dummy in the RF measure with the share infected from the Epi. model — the Epi.-RF Bartik measure.

Figure 5: Effects of early life measles exposure on labor market and welfare outcomes



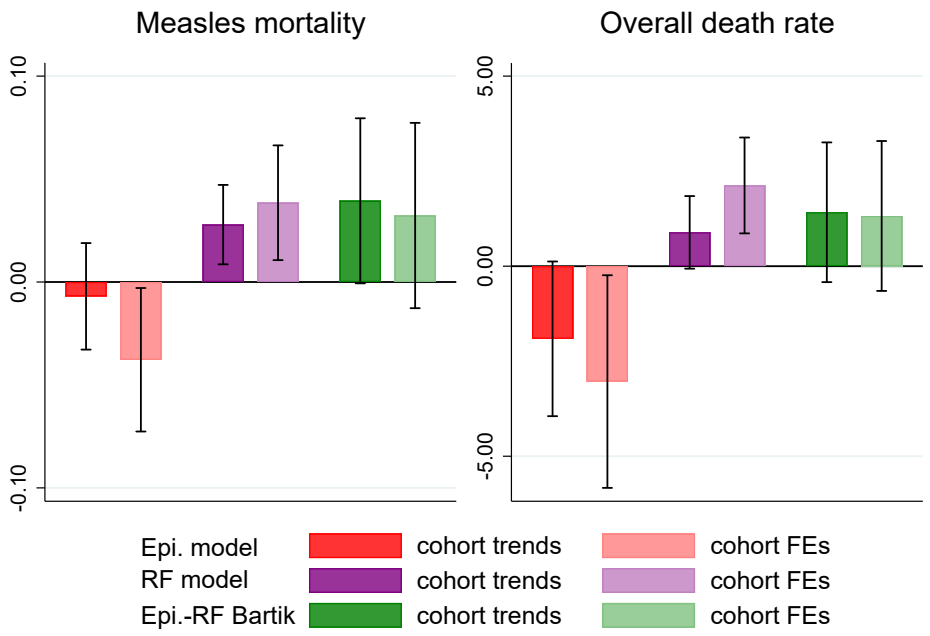
Notes: This figure shows point estimates including 95%-confidence intervals for the dependent variable as indicated in each subfigure from linear regressions in the full sample aggregated to cells at the state-of-birth x year-of-birth x age level and weighted by the cell's population. Explanatory variables are the share infected derived from the epidemiological (Epi.) model shown in red, the interaction of the states' measles mortality (ages 0—5, in 1959—1963) with a dummy equal to one for cohorts born prior (<=) 1963 (year of measles vaccine introduction) for the reduced-form (RF) model shown in violet, and a combination of both approaches for the Epi.-RF Bartik regression shown in green. Regressions either include cohort trends — linear and quadratic terms — as indicated in rich colors or cohort fixed effects as indicated in pale colors. All estimations further control for state and age fixed effects. Standard errors are clustered at the state level.

Figure 6: Effects of early life measles exposure on health outcomes



Notes: This figure shows point estimates including 95%-confidence intervals for the dependent variable as indicated in each subfigure from linear regressions in the full sample aggregated to cells at the state-of-birth x year-of-birth x age level and weighted by the cell's population. Explanatory variables are the share infected derived from the epidemiological (Epi.) model shown in red, the interaction of the states' measles mortality (ages 0—5, in 1959—1963) with a dummy equal to one for cohorts born prior (<=) 1963 (year of measles vaccine introduction) for the reduced-form (RF) model shown in violet, and a combination of both approaches for the Epi.-RF Bartik regression shown in green. Regressions either include cohort trends — linear and quadratic terms — as indicated in rich colors or cohort fixed effects as indicated in pale colors. All estimations further control for state and age fixed effects. Standard errors are clustered at the state level.

Figure 7: Effect of early life measles exposure on medium-run mortality (ages 0–5)



Notes: This figure shows point estimates including 95%-confidence intervals for the dependent variable as indicated in each subfigure from linear regressions in the full sample aggregated to cells at the state-of-birth x year-of-birth x age level and weighted by the cell's population. Explanatory variables are the share infected derived from the epidemiological (Epi.) model shown in red, the interaction of the states' measles mortality (ages 0—5, in 1959—1963) with a dummy equal to one for cohorts born prior (<=) 1963 (year of measles vaccine introduction) for the reduced-form (RF) model shown in violet, and a combination of both approaches for the Epi.-RF Bartik regression shown in green. Regressions either include cohort trends — linear and quadratic terms — as indicated in rich colors or cohort fixed effects as indicated in pale colors. All estimations further control for state and age fixed effects. Standard errors are clustered at the state level.

Table 1: Descriptive Statistics

	(1) 1950–1960 Mean/SD	(2) 1961–1970 Mean/SD	(3) 1971–1980 Mean/SD	(4) Total Mean/SD
Epi. measure (share infected)	0.925	0.328	0.021	0.441
	(0.158)	(0.308)	(0.027)	(0.429)
RF measure	1.133	0.340	0.000	0.512
	(0.593)	(0.613)	(0.000)	(0.690)
Epi.-RF Bartik measure	1.046	0.358	0.021	0.493
	(0.590)	(0.420)	(0.027)	(0.606)
High school completion	92.362	92.613	93.012	92.653
	(4.093)	(3.444)	(3.486)	(3.710)
Unemployed	3.286	4.188	5.242	4.208
	(1.810)	(2.030)	(2.630)	(2.317)
Log HH income	10.986	11.029	10.839	10.952
	(0.171)	(0.158)	(0.318)	(0.240)
Poverty	8.062	8.462	11.205	9.205
	(3.340)	(3.361)	(5.861)	(4.540)
Any welfare income	1.231	1.643	2.020	1.618
	(0.850)	(1.033)	(1.301)	(1.119)
Any food stamps	8.363	10.074	12.518	10.255
	(4.201)	(4.356)	(5.021)	(4.842)
Work disability	10.308	6.633	4.577	7.274
	(3.718)	(2.897)	(2.491)	(3.919)
Disability Index	6.641	4.388	2.814	4.680
	(2.304)	(1.817)	(1.208)	(2.436)
Cognitive difficulty	5.943	5.031	4.139	5.067
	(2.143)	(2.075)	(1.826)	(2.155)
Self-care difficulty	3.525	2.205	1.258	2.368
	(1.678)	(1.379)	(0.991)	(1.676)
Indep. living difficulty	5.802	4.116	2.863	4.310
	(2.401)	(2.024)	(1.593)	(2.380)
Vision/hearing difficulty	6.149	3.803	2.337	4.163
	(3.156)	(2.192)	(1.483)	(2.883)
Number of cohorts (year of birth x state of birth)	451	410	410	1,271

Notes: Columns (1) – (3) show descriptive statistics for the sub-samples born from 1950–1960, 1961–1970, 1971–1980, respectively and column (4) shows the descriptive statistics for the overall population. The overall population includes individuals born in the years 1950–1980 in the US and sampled in the Census 2000 or the American Community Surveys 2001–2019. The data is aggregated to cells at the state-of-birth x year-of-birth x age level. The variable on having a work disability is only available up to the year 2007. Information on receiving food stamps, instead, is only available starting from 2005.

Table 2: Prediction Power of the Dynamic Distribution Model

Dependent variable	1950-1980			1950-1960		1961-1970		1971-1980	
Measles cases per capita	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Susceptibles (t-1)	0.040**			0.287***		0.228***		0.141**	
	(0.016)			(0.080)		(0.021)		(0.066)	
Susceptibles (t+1)		-0.048***							
		(0.011)							
10 quantiles of s_lag=2			0.003		0.008**		-0.000		
			(0.002)		(0.003)		(0.001)		
10 quantiles of s_lag=3			0.002		0.010**		0.003		0.000
			(0.002)		(0.005)		(0.003)		(.)
10 quantiles of s_lag=4			0.003		0.018**		0.004		0.003***
			(0.003)		(0.007)		(0.003)		(0.001)
10 quantiles of s_lag=5			0.002		0.017**		0.006**		0.004***
			(0.003)		(0.007)		(0.003)		(0.001)
10 quantiles of s_lag=6			0.004		0.023***		0.011***		0.005***
			(0.003)		(0.007)		(0.003)		(0.001)
10 quantiles of s_lag=7			0.005*		0.032***		0.015***		0.005***
			(0.003)		(0.006)		(0.003)		(0.002)
10 quantiles of s_lag=8			0.007**		0.039***		0.020***		0.006***
			(0.003)		(0.008)		(0.004)		(0.002)
10 quantiles of s_lag=9			0.005		0.016***		0.022***		0.007***
			(0.004)		(0.005)		(0.004)		(0.002)
10 quantiles of s_lag=10			0.008**		0.031***		0.034***		0.008***
			(0.004)		(0.006)		(0.004)		(0.003)
R2	0.566	0.551	0.568	0.225	0.231	0.655	0.628	0.390	0.331
Observations	1,230	1,230	1,230	410	410	410	410	410	410

Notes: This table tests measles outbreak predictions using lags and leads of state-specific annual susceptible counts generated from the dynamic distribution model. The dependent variable are the measles cases per capita from 1950–1980 (columns 1—3) and in the subperiods as indicated in the top row. Explanatory variables are the susceptibles in the years (t-1) and (t+1) as derived from the epidemiological (Epi.) model, and deciles of the susceptibles in the years (t-1). All regressions control for cohort and state fixed effects. Standard errors are clustered at the state level. ***/**/* indicate statistical significance at the 1%/5%/10%-level.

Table 3: Effects of early life measles exposure on labor market and welfare outcomes

Dependent variable:	(1) High school	(2) Un-employed	(3) ln total family income	(4) Poverty	(5) Any family welfare inc.	(6) Food stamps
Panel A1: Epi. model and cohort trends						
Share infected	-0.539***	0.316***	-0.037***	0.970***	0.050	1.160***
	(0.196)	(0.072)	(0.008)	(0.136)	(0.046)	(0.154)
R^2	0.722	0.396	0.905	0.747	0.325	0.682
Panel A2: Epi. model and cohort fixed effects						
Share infected	0.795*	0.019	0.043***	-0.275	-0.200**	0.012
	(0.410)	(0.101)	(0.011)	(0.164)	(0.088)	(0.220)
R^2	0.730	0.398	0.909	0.751	0.328	0.686
Panel B1: RF model and cohort trends						
Pre measles mort	-0.846***	0.105*	-0.030***	0.576***	0.157***	0.525***
x Pre	(0.187)	(0.053)	(0.007)	(0.106)	(0.035)	(0.094)
R^2	0.729	0.395	0.907	0.748	0.328	0.682
Panel B2: RF model and cohort fixed effects						
Pre measles mort	-1.079***	0.005	-0.027**	0.503**	0.234***	0.329**
x Pre	(0.374)	(0.112)	(0.013)	(0.207)	(0.082)	(0.162)
R^2	0.736	0.398	0.909	0.752	0.331	0.686
Panel C1: Epi.-RF Bartik and cohort trends						
Pre measles mort	-1.051***	0.130	-0.035***	0.774***	0.211***	0.702***
x Share infected	(0.295)	(0.090)	(0.011)	(0.171)	(0.061)	(0.158)
R^2	0.728	0.395	0.906	0.748	0.328	0.682
Panel C2: Epi.-RF Bartik and cohort fixed effects						
Pre measles mort	-0.922**	0.022	-0.019	0.469**	0.204**	0.351**
x Share infected	(0.360)	(0.115)	(0.013)	(0.217)	(0.077)	(0.170)
R^2	0.734	0.398	0.906	0.751	0.330	0.686
Observations	25,420	25,420	25,420	25,420	25,420	24,149
Mean dep. var	93.03	4.359	10.99	8.805	1.621	9.771

Notes: Linear regressions in the full sample aggregated to cells at the state-of-birth x year-of-birth x age level and weighted by the cell's population. Dependent variables are indicated in the top row. Explanatory variables are the share infected derived from the epidemiological (Epi.) model in Panel A, the interaction of the states' measles mortality (ages 0—5, in 1959—1963) with a dummy equal to one for cohorts born prior (<=) 1963 (year of measles vaccine introduction) for the reduced-form (RF) model in Panel B, and a combination of both approaches for the Epi.-RF Bartik regression in Panel C. Regressions in subpanels 1 include cohort trends (linear and quadratic terms) and in subpanels 2 control for cohort fixed effects. All panels further control for state and age fixed effects. Standard errors are clustered at the state level. ***/**/* indicate statistical significance at the 1%/5%/10%-level.

Table 4: Effects of early life measles exposure on health outcomes

Dependent variable:	(1) Work disability	(2) Disability Index	(3) Cognitive difficulty	(4) Self-care difficulty	(5) Indep. living diff.	(6) Vision /hearing diff.
Panel A1: Epi. model and cohort trends						
Share infected	0.091	0.014	0.196**	0.017	0.075	-0.302***
	(0.186)	(0.083)	(0.078)	(0.056)	(0.084)	(0.080)
R^2	0.635	0.843	0.535	0.669	0.667	0.807
Panel A2: Epi. model and cohort fixed effects						
Share infected	-0.928***	-0.499***	-0.444***	-0.231*	-0.409**	-0.539***
	(0.274)	(0.179)	(0.136)	(0.115)	(0.170)	(0.153)
R^2	0.639	0.845	0.539	0.670	0.668	0.809
Panel B1: RF model and cohort trends						
Pre measles mort	0.648***	0.338***	0.325***	0.188**	0.317***	0.187*
x Pre	(0.219)	(0.115)	(0.087)	(0.072)	(0.101)	(0.102)
R^2	0.638	0.846	0.539	0.671	0.669	0.808
Panel B2: RF model and cohort fixed effects						
Pre measles mort	0.825**	0.573**	0.416**	0.318**	0.509**	0.482**
x Pre	(0.347)	(0.223)	(0.172)	(0.142)	(0.202)	(0.182)
R^2	0.641	0.848	0.541	0.672	0.671	0.811
Panel C1: Epi.-RF Bartik and cohort trends						
Pre measles mort	0.753**	0.440**	0.386**	0.238*	0.405**	0.265
x Share infected	(0.341)	(0.195)	(0.144)	(0.124)	(0.170)	(0.175)
R^2	0.637	0.846	0.538	0.670	0.669	0.808
Panel C2: Epi.-RF Bartik and cohort fixed effects						
Pre measles mort	0.624	0.450*	0.308*	0.246	0.395*	0.371*
x Share infected	(0.372)	(0.239)	(0.180)	(0.152)	(0.209)	(0.209)
R^2	0.639	0.846	0.538	0.671	0.670	0.810
Observations	10,168	25,420	25,420	25,420	25,420	25,420
Mean dep. var	7.301	4.635	4.973	2.403	4.338	3.992

Notes: Linear regressions in the full sample aggregated to cells at the state-of-birth x year-of-birth x age level and weighted by the cell's population. Dependent variables are indicated in the top row. Explanatory variables are the share infected derived from the epidemiological (Epi.) model in Panel A, the interaction of the states' measles mortality (ages 0—5, in 1959—1963) with a dummy equal to one for cohorts born prior (<=) 1963 (year of measles vaccine introduction) for the reduced-form (RF) model in Panel B, and a combination of both approaches for the Epi.-RF Bartik regression in Panel C. Regressions in subpanels 1 include cohort trends (linear and quadratic terms) and in subpanels 2 control for cohort fixed effects. All panels further control for state and age fixed effects. Standard errors are clustered at the state level. ***/**/* indicate statistical significance at the 1%/5%/10%-level.

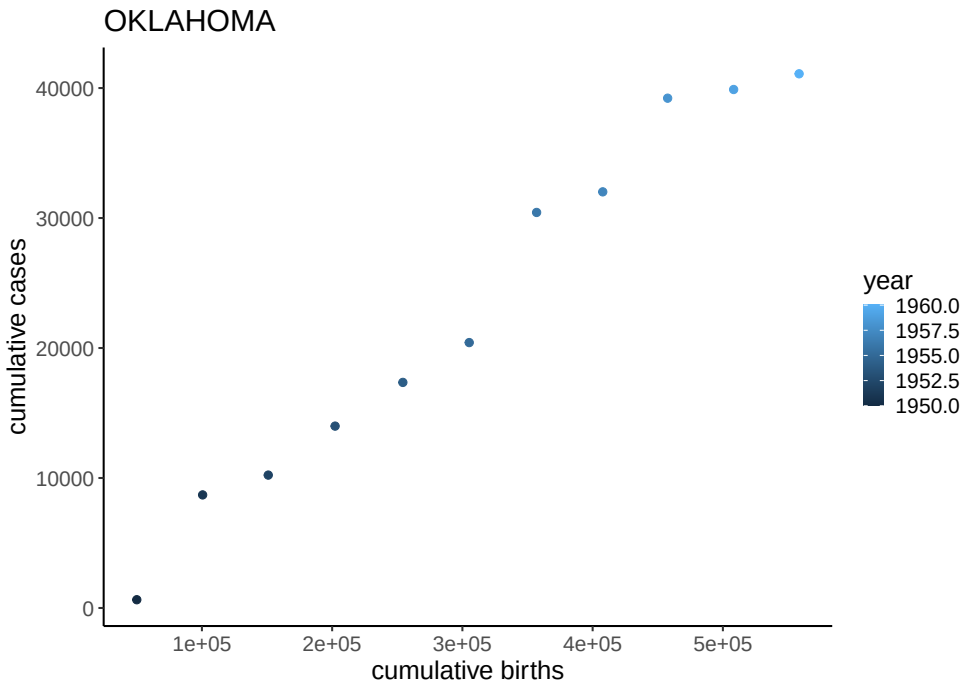
Table 5: Effects of early life measles exposure on medium-term mortality (ages 0–5)

Dependent variable:	(1) Measles mortality	(2) Overall death rate
Panel A1: Epi. model and cohort trends		
Share infected	-0.007	-1.911*
	(0.013)	(1.007)
R^2	0.589	0.889
Panel A2: Epi. model and cohort fixed effects		
Share infected	-0.038**	-3.036**
	(0.017)	(1.384)
R^2	0.626	0.901
Panel B1: RF model and cohort trends		
Pre measles mort	0.028***	0.889*
x Pre	(0.010)	(0.472)
R^2	0.627	0.890
Panel B2: RF model and cohort fixed effects		
Pre measles mort	0.038***	2.124***
x Pre	(0.014)	(0.624)
R^2	0.656	0.906
Panel C1: Epi.-RF Bartik and cohort trends		
Pre measles mort	0.039*	1.418
x Share infected	(0.020)	(0.909)
R^2	0.615	0.889
Panel C2: Epi.-RF Bartik and cohort fixed effects		
Pre measles mort	0.032	1.321
x Share infected	(0.022)	(0.976)
R^2	0.635	0.900
Observations	615	615
Mean dep. var	.03499	26.81

Notes: Linear regressions in the sample including birth cohorts 1959–1973 aggregated to cells at the state-of-birth x year-of-birth level and weighted by the cell's birth counts. Dependent variables are indicated in the top row. Explanatory variables are the share infected derived from the epidemiological (Epi.) model in Panel A, the interaction of the states' measles mortality (ages 0—5, in 1959—1963) with a dummy equal to one for cohorts born prior (<=) 1963 (year of measles vaccine introduction) for the reduced-form (RF) model in Panel B, and a combination of both approaches for the Epi.-RF Bartik regression in Panel C. Regressions in subpanels 1 include cohort trends (linear and quadratic terms) and in subpanels 2 control for cohort fixed effects. All panels further control for state and age fixed effects. Standard errors are clustered at the state level. ***/**/* indicate statistical significance at the 1%/5%/10%-level.

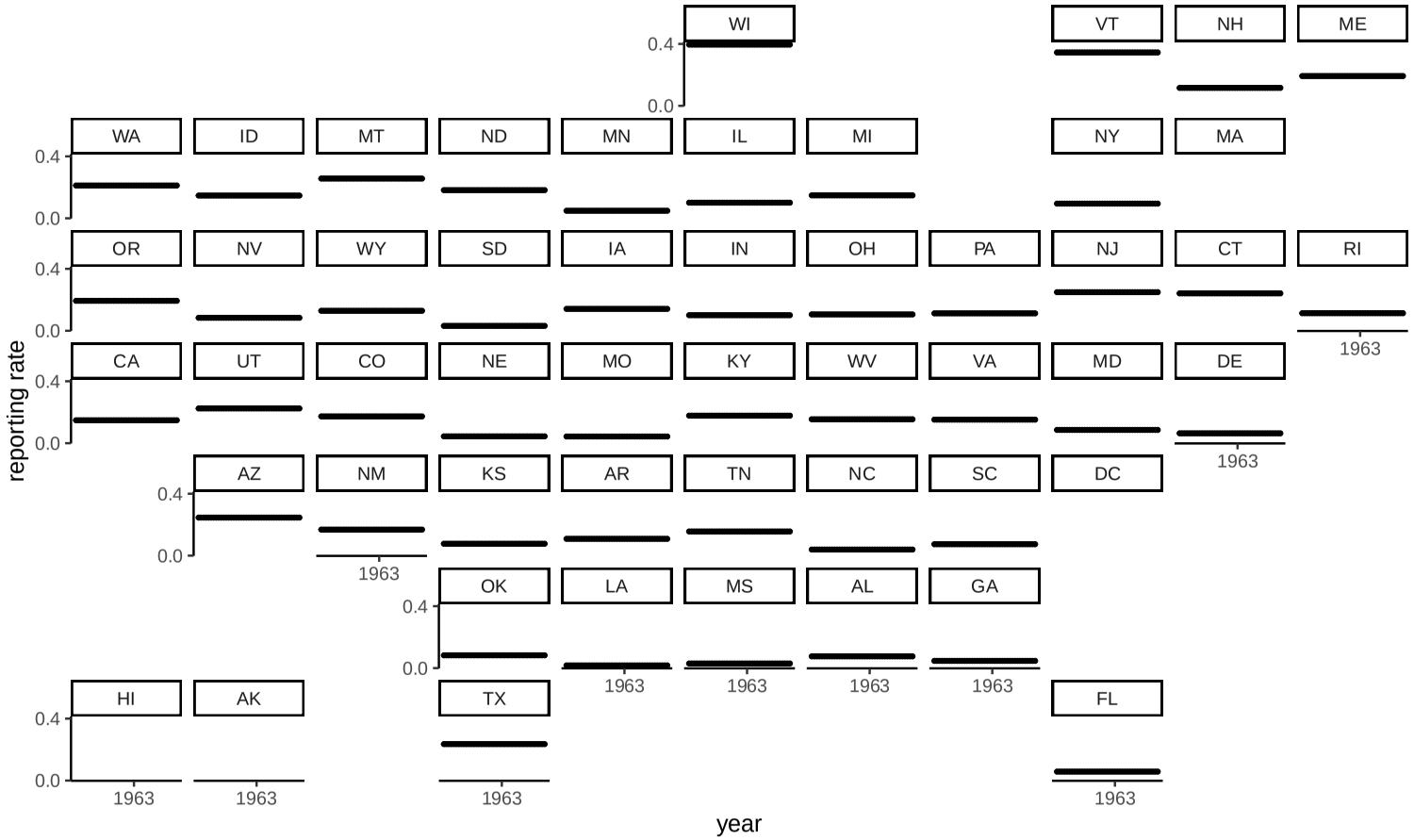
Appendix

Figure A1: Cumulative cases vs cumulative births, Oklahoma 1950–1960



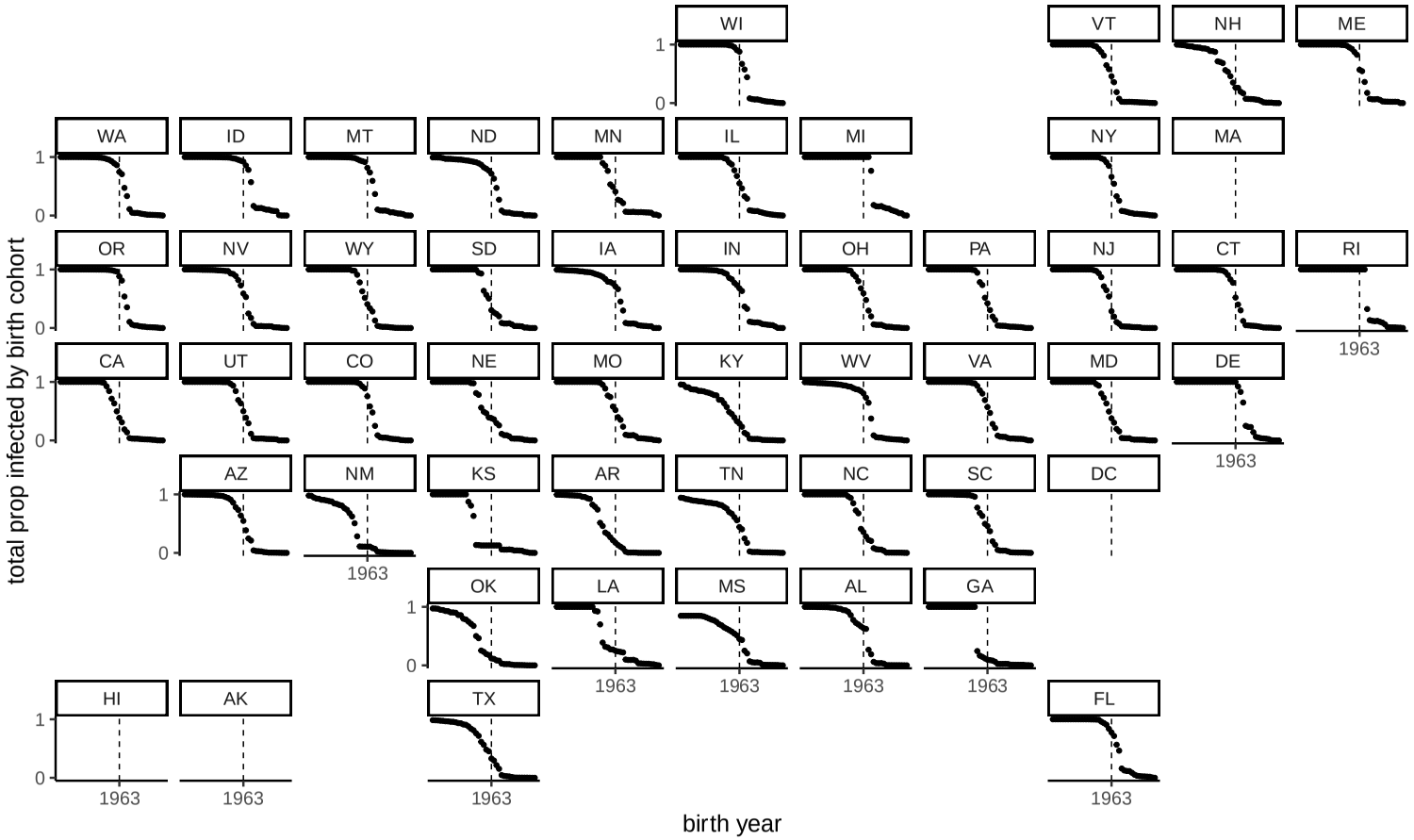
Notes: This figure shows the relationship of cumulative measles cases versus cumulative births in the years 1950–1960 for Oklahoma. The inferred reporting rate for Oklahoma is derived based on the slope of this relationship.

Figure A2: Overview state-level reporting rates



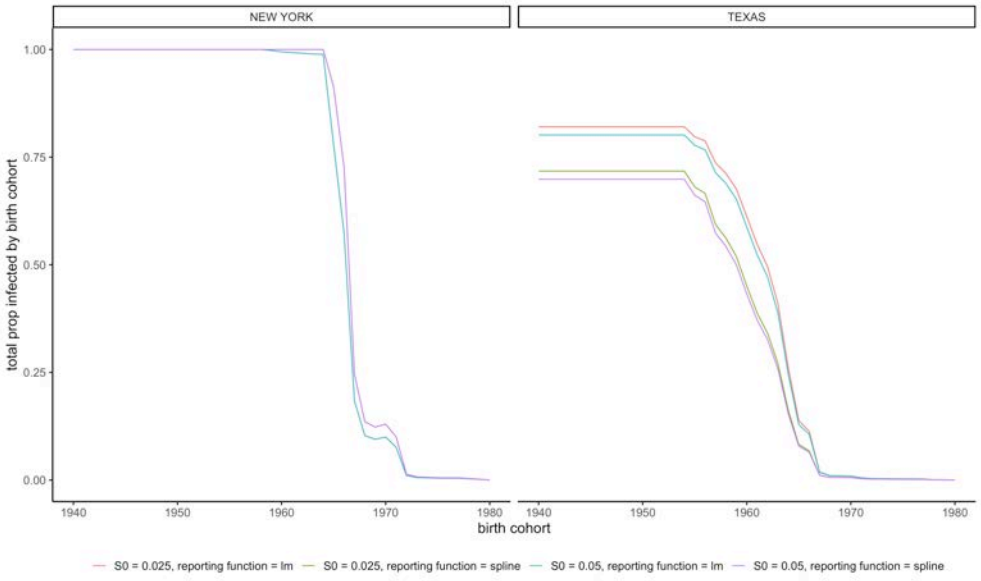
Notes: This figure shows individual reporting rates for all US states using the slope of the relationship of cumulative measles cases versus cumulative births in the years 1950–1960.

Figure A3: Share ever infected across entire U.S.



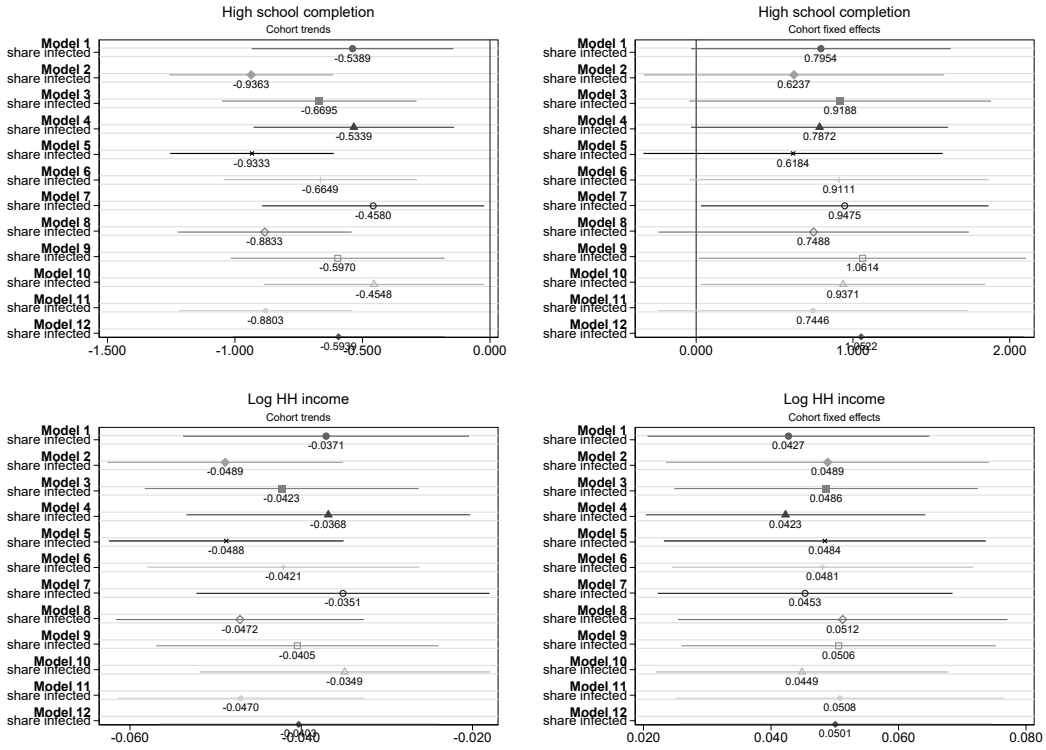
Notes: This figure shows the share ever infected for all US states using the epidemiological approach described in Section 3.1.

Figure A4: Dynamic distribution model - sensitivity



Notes: This figure shows the share ever infected for two sample states, New York and Texas, using the epidemiological approach described in Section 3.1 while varying input parameters such as the initial susceptibility share (0.025 vs. 0.05) and different functional forms to estimate the reporting rate.

Figure A5: Sensitivity of epidemiological model



1 S0 = 0.025; $\text{IR}_{age} = 1; reporting function = linear$
2 S0 = 0.025; $\text{IR}_{age} = 1; reporting function = loess$
3 S0 = 0.025; $\text{IR}_{age} = 1; reporting function = spline$
4 S0 = 0.025; $\text{IR}_{age} = 2; reporting function = linear$
5 S0 = 0.025; $\text{IR}_{age} = 2; reporting function = loess$
6 S0 = 0.025; $\text{IR}_{age} = 2; reporting function = spline$
7 S0 = 0.05; $\text{IR}_{age} = 1; reporting function = linear$
8 S0 = 0.05; $\text{IR}_{age} = 1; reporting function = loess$
9 S0 = 0.05; $\text{IR}_{age} = 1; reporting function = spline$
10 S0 = 0.05; $\text{IR}_{age} = 2; reporting function = linear$
11 S0 = 0.05; $\text{IR}_{age} = 2; reporting function = loess$
12 S0 = 0.05; $\text{IR}_{age} = 2; reporting function = spline$

Notes: Shown are the effects of the share ever infected on high school graduation rates and log household income estimated using regression equation (7). The share ever infected is calculated via the dynamic distribution model based on 12 versions of different alternative parameter estimates. These versions are permutations of changes in the initial susceptibility share (0.025 vs. 0.05), changes in the ratio of infection risk for the population below and above age 6 (infection ratio 1 vs. 2), and different functional forms to estimate the reporting rate.

Table A1: Predictive Power Dynamic Distribution Model vs. Reduced-Form Measure

Dependent variable	1950-1980				1950-1960		1961-1970		1971-1980	
Measles cases per capita	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Susceptibles (t-1)	0.040**				0.287***		0.228***		0.141**	
	(0.016)				(0.080)		(0.021)		(0.066)	
Susceptibles (t+1)		-0.048***								
		(0.011)								
RF measure (t-1)			0.001			0.000		0.001		0.000
			(0.001)			(.)		(0.002)		(.)
RF measure (t+1)				0.001						
				(0.001)						
R2	0.566	0.551	0.561	0.545	0.225	0.152	0.655	0.541	0.390	0.316
Observations	1,230	1,230	1,230	1,230	410	410	410	410	410	410

Notes: This table compares measles outbreak predictions using lags and leads of state-specific annual susceptible counts generated from the dynamic distribution model with lags and leads from the reduced-form measure. The dependent variable are the measles cases per capita from 1950–1980 (columns 1–4) and in the subperiods as indicated in the top row. Explanatory variables are the susceptibles in the years (t-1) and (t+1) as derived from the epidemiological (Epi.) model and the reduced-form (RF) measure in the years (t-1) and (t+1). All regressions control for cohort and state fixed effects. Standard errors are clustered at the state level. ***/**/* indicate statistical significance at the 1%/5%/10%-level.