NBER WORKING PAPER SERIES

# REVISING INFANT MORTALITY RATES FOR THE EARLY 20TH CENTURY UNITED STATES

Katherine Eriksson
Gregory T. Niemesh
Melissa Thomasson

Working Paper 23263
http://www.nber.org/papers/w23263

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2017

We are particularly grateful to William Even and Analisa Packham for helpful suggestions. We would also like to thank Jeremy Atack, William J. Collins, Dora Costa, Gordon Hanson, Adriana Lleras-Muney, Seth Sanders, Marianne Wanamaker, and Sven Wilson for comments when portions of this work were included in the paper "Death In the Promised Land: The Great Migration and Black Infant Mortality". Brian Lee and Man-Ting Chang provided excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Revising Infant Mortality Rates for the Early 20th Century United States
Katherine Eriksson, Gregory T. Niemesh, and Melissa Thomasson
NBER Working Paper No. 23263
March 2017
JEL No. I19,J11,N32

**ABSTRACT**

Accurate vital statistics are required to understand the evolution of racial disparities in infant health and the causes of rapid secular decline in infant mortality during the early twentieth century. Unfortunately, infant mortality rates prior to 1950 suffer from an upward bias stemming from a severe underregistration of births. At one extreme, African-American births in Southern states went unregistered at the rate of 15 to 25 percent. In this paper, we construct improved estimates of births and infant mortality in the United States for the 1915-1940 period using recently released complete count decennial census microdata combined with the counts of infant deaths from published sources. We check the veracity of our estimates with a major birth registration study completed in conjunction with the 1940 Decennial Census, and that the largest adjustments occur in states with less complete birth registration systems. An additional advantage of our census-based estimation method is the extension back of the birth and infant mortality series for years prior to published estimates of registered births, enabling previously impossible comparisons and estimations. Finally, we show that underregistration can bias effect estimates even in a panel setting with specifications that include location fixed effects and place-specific linear time trends.

Katherine Eriksson
Department of Economics
University of California, Davis
One Shields Avenue
Davis, CA 95616
and NBER
kaeriksson@ucdavis.edu

Gregory T. Niemesh
Miami University of Ohio
Economics Department
Farmer School of Business
800 E. High Street
Oxford, OH 45056
niemesgt@miamioh.edu

Melissa Thomasson
Miami University/FSB
Department of Economics MSC 1035
800 E. High Street, Rm. #2054
Oxford, OH  45056
and NBER
mthomasson@miamioh.edu

## I. Introduction

Vital statistics form the foundation of our understanding of health trends for the nation, and have come to be regarded as indispensable when targeting effective public health programs and evaluating interventions. As early as the late-19th century, public health officials recognized the importance of statistics coming from the vital registration system as an important resource in the fight against infectious disease (Cassedy, 1965). For modern researchers in economics, demography, and public health, vital statistics from the early 20th-century provide a rich data source to understand trends in mortality and longevity, socioeconomic correlates with health, and estimate causal impacts of health interventions.[1] Unfortunately, estimates of live births, infant mortality rates, and maternal mortality rates for years prior to 1950 suffer from an upward bias stemming from a severe underregistration of births. Not only are rates incorrect, but the measurement error varies over races and locations in ways that are potentially correlated with variables of interest.

In this paper, we construct improved estimates of live births, infant mortality, and maternal mortality for the United States using newly released decennial census microdata. We focus on infant mortality to demonstrate the importance of using the new estimates. To obtain these estimates, we revise the number of births, while leaving the published counts of infant deaths unchanged. Thus, any differences between published and revised rates arises from a different estimate of live births. In addition to improving upon published estimates, our method enables us to extend the existing series backwards in time. While current state-level infant mortality rates begin only after a state enters the Birth Registration Area (BRA), we are able to construct a series based on when a state entered the Death Registration Area (DRA), which generally occurred prior to a state's entrance into the BRA.[2] As a result, our series allows for previously impossible comparisons of fertility and infant mortality across groups and analyses of earlier interventions.

Infant mortality rates (IMR) are computed by dividing registered deaths of infants by the number of registered live births occurring during a calendar year. Bias can enter the calculation through an incorrect estimate of infant deaths (the numerator) or an incorrect estimate of births (the denominator). Contemporary evidence suggests that severe underregistration of births biased IMR estimates at least until 1940, with the bias varying by region and race (Grove, 1943). Bias in the numerator from unregistered deaths was believed to be a minor issue. Thus, IMR estimates using

[1] Examples of recent papers that use births estimates that suffer a bias from underregistration include: Collins and Thomasson (2004); Cutler and Miller (2005); Thomasson and Treber (2008); Jayachandran, Lleras-Muney and Smith (2010); Clay, Troesken and Haines (2013); Hansen (2014); Moehling and Thomasson (2014); Bhalotra and Venkataramani (2015); Eriksson and Niemesh (2016)

[2] States entered the Death Registration Area as early as 1880, while the Birth Registration Area did not begin until 1915. Table A3 lists the entry dates for each state into the BRA and the DRA

registered events will vary inversely with the completeness of birth registration.[3]

The main source of national evidence of birth registration completeness during the period comes from an infant card test conducted concurrently with the 1940 decennial census, as summarized by Grove (1943). Enumerators were instructed to fill out a special infant card for any child born during the four months prior to the census date. The Census Bureau then attempted to match each card – and registered infant death – to a birth certificate filed in the state registrar's office. The completeness of registrations was estimated as the proportion of infant cards and deaths for which a birth certificate could be found. For the nation as a whole, 92.5 percent of births were found to be registered, but the total hid large differences between races (94.0 percent for whites and 82.0 percent for blacks), cities and rural areas (96.9 percent in cities above 10,000 in population and 88.0 percent in small cities and rural areas combined), and whether the birth occurred in an institution (98.5 percent in hospitals and 86.1 percent outside of hospitals). The test suggested that underregistration in some states was quite severe in total, and particularly poor for blacks. For example, only 77.6 percent of births were registered in South Carolina versus 99.4 percent in Connecticut. In general, the South had the highest level of underregistration.[4]

We produce annual, 2-year, and 5-year adjusted estimates of births, infant mortality rates, and maternal mortality rates by state and race, as well as at the national level, to account for the well-documented severe underregistration of births.[5] The adjustment procedure estimates births by adding together the enumeration of live children in the census, the number of infant deaths, and the number of non-infant deaths. We start with the enumeration of children in the decennial census for each *state of birth by year of birth by race* cell, using newly released complete census microdata for the 1920, 1930, and 1940 decennial census from IPUMS (Ruggles et al., 2015). Infant deaths are allocated to the state and year of occurrence. Finally, deaths of children after infancy, but prior to the subsequent decennial census enumeration, are allocated to year and state of birth. The two major sources of bias in our estimates come from migration of children who

[3]Researchers at the time understood the biases present in infant mortality rates. For example, a former presidient of the Population Association America, P.K. Whelpton, wrote in the *Journal of the American Statistical Association* in 1934, "If birth registration is equally deficient in various states, only absolute values for birth rates and infant mortality are affected. However, if there are large differences in completeness between states, the comparative standing of states in these respects will vary correspondingly when they are ranked on an adjusted instead of an unadjusted basis." (Whelpton, 1934).

[4]A subsequent test conducted in 1950 showed major improvements over the decade with 97.8 percent registered for the nation as a whole, although some states lagged behind (Shapiro and Schachter, 1952). The likely explanation for this rapid improvement is that registration completeness is highly correlated with the percent of births delivered in a hospital. Completeness eventually reached close to 100 percent by the mid to late 1960s as hospital deliveries approached 100 percent of all births after the integration of hospitals in the South (U.S. Bureau of the Census, 1973)

[5]City-level infant mortality estimates are also biased from birth underregistration. However, we are loathe to use our procedure to adjust rates at the city-level. The census records only state of birth, not city. Any allocation of census enumerations back to a city of birth would be plagued by bias from cross-city migration. Because state of birth is known, this migration bias does not affect our state based revised rates.

subsequently die and underenumeration of children in the decennial census. Bias from migration stems from children dying outside their state of birth; however, we show that a rough allocation of these non-infant deaths back to their state of birth does not meaningfully alter the estimates. Underenumeration, when the census undercounts the number of children alive at the census date, poses a more challenging problem. To partially resolve this issue, we use the number of children counted in a later census rather than the first census in which a child should be enumerated (e.g. 11 year olds in the 1940 census rather than 1 year olds in the 1930 census for the 1929 birth year). In the end, we are not able to fully account for underenumeration, but the revision of births becomes a race between underregistration of births and underenumeration of children in the census. The revision procedure provides more accurate birth estimates for states where underregistration was more severe than underenumeration (i.e. the South). In the few instances where underenumeration was a larger problem, we restrict the adjusted rates to be no higher than published rates.

In general, estimates are biased to the extent that underregistration is correlated with the intervention or group attribute on which a comparison is made. Any cross-sectional or time-series comparison is biased when registration completeness varies across groups or over time. Importantly, we show that underregistration potentially biases results even in a panel setting with specifications that include location fixed effects and place-specific linear time trends. These added controls do not fully explain the difference between the revised rates and the published rates, suggesting that a meaningful amount of measurement error remains.

We end by briefly discussing a number of important implications of the census-based adjustment method. First, at the national level the revised estimates suggest a lower black IMR relative to those in published sources, with larger differences prior to 1925: 11.1 percentage points in 1915 versus 18.9 in the published data. The lower initial level in 1915 also implies slower *progress* in black infant health. IMR declined by 11.2 percentage points between 1915 and 1940 in the published data, but only by 4.9 percentage points in the revised estimates. Because underregistration of births was not as severe for whites, revised estimates of white IMR do not deviate from published estimates as much. The largest difference occurs in 1915 and is only 1.3 percentage points.

The large variation across states in the quality of birth registration data leads to significant revisions of cross-sectional comparisons, as evidenced by the change in rankings of states based on infant mortality. The effect of rank changes extends to regional differences and subsequent convergence. The South initially had a mortality advantage over the North for black infants, but rates converged as the urban penalty gradually declined over the course of the early 20th century.

Three main implications for regional convergence emerge from using the revised estimates. First, the southern mortality advantage widens as the adjustment method primarily lowers black IMR in the South. Second, starting from a lower initial IMR in the South implies a faster *convergence rate* between the regions. Finally, the level shift downward in southern IMR delays the North overtaking South until the late 1930s, if at all before 1940.

The paper is organized as follows. First, we describe the development of the birth registration area and evidence of a severe underregistration of births that varies with race and geography. Next, we describe the procedure used to revise birth and infant mortality estimates to account for underregistration of births. We provide a detailed discussion of the potential biases that enter the revised estimates. Finally, the paper discusses some important implications of the revised series for national trends and regional convergence in the black-white IMR gap, and the power of socio-economic status in explaining the gap by revisiting Collins and Thomasson (2004).

## II. Development of the Birth Registration Area and Evidence of Completeness

The first registration law for vital events was adopted by the Massachusetts legislature in 1842, with six other states enacting similar legislation by 1851. These early systems, however, operated in only a few localities and suffered from lax enforcement (Lunde, 1980). Despite the known flaws in the system, public health professionals realized the importance of vital statistics reporting in their efforts to combat and eradicate infectious disease in the latter half of the 19th century. The federalism of the time slowed the growth of the registration system, as it imposed a piece-meal state-by-state approach that eventually created nationally representative statistics.[6] The Death Registration Area (DRA) began in 1880 with two states, the District of Columbia, and several large cities. In 1900, the Census Bureau established a national DRA that initially included 10 states, mainly from the Northeast and Midwest. The DRA and was completed in 1933 with the entrance of Texas.

It took longer to establish the Birth Registration Area (BRA). Public health officials viewed mortality data as being more helpful for preventive medicine than birth data, and registrars believed enforcement of birth registration to be more difficult than for deaths (Cassedy, 1965). However, once started in 1915 with 10 states and the District of Columbia, the BRA was completed relatively quickly over a period of 18 years. Again, states in the Northeast, Middle Atlantic, and Midwest

[6]Vital registration systems were and remain the responsibility of the several states. The Federal Government's role is limited to the promotion of state registration systems and to work with the states to produce national level statistics. An act of Congress in 1902, put the national system on a firm footing by making the Census Bureau a permanent agency and providing the authority to collect information on births and deaths.

joined first, with most of the remainder of the country entering in the 1920s. Southern states lagged the others, and the BRA was not completed until 1933 with the entrance of Texas. A list of entrance dates for each state can be found in Appendix Table A3.

States seeking entrance to the BRA had to overcome two hurdles. First, the state legislature needed to enact and enforce registration laws in a manner deemed sufficient by the Census Bureau. The more difficult second hurdle was to show evidence that registrations were at least 90 percent complete (Lunde, 1980; Moriyama, 1990). All tests of registration completeness proceeded by first obtaining a list of children born during a fixed period of time, and then determining whether birth certificates had been filed for those children. Various methods of obtaining the list of names were used by the Census Bureau over the course of the early 20th century. At the advent of the BRA, the test was conducted under the direction of the Census Bureau, and consisted of comparing birth registrations to collected lists of births from postmasters, newspapers, death registers, and church records. Contemporaries acknowledged early on that the tests used to enter the BRA were woefully inadequate (Whelpton, 1934). Cressy Wilbur, Chief Statistician for Vital Statistics of the United States for 1906-1914, believed that the use of lists of births collected by postmasters to be a highly biased sample for a test (Wilbur, 1916). Deacon (1937) relates the story of how after finding a 100 percent registration rate from names provided by a postmaster, he came to find that the postmaster received the list directly from the local registrar. Later evidence showed the sources used to create the list of children - death registrations, hospital births, and newspaper announcements - were likely a highly selected sample of births; children born to urban, educated, and wealthier parents were more likely to appear in these sources, and also more likely than the population to register a birth (Moriyama, 1990). The selected sample caused the tests to overestimate the completeness of the registration system. Nevertheless, entrance to the BRA was granted after a positive test result.

In the mid-1920s, the Census Bureau switched to a testing procedure based on postal infant test cards, which were sent out in mass mailings to every known household. Residents were asked to list the occurrence of any deaths or births that occurred during the prior 12 months, with returned cards checked against birth registers. While believed to be an improvement over collected lists, the postal test card method suffered from its own biases. Errors entered the lists from memory lapses inherent in any recall method. More importantly, households with unregistered events were less likely to return the cards, as were households with low education and incomes (Moriyama, 1990). A 1934 test in Georgia and Maryland used the postal card test method and compared the results from a canvas of enumerators. The relevant findings include: 1.) Registrations were more complete for households with higher incomes, with more education, lived in cities, delivered their birth in a hospital, and were white, 2.) The postal card method led to overstatements of completeness as mail carriers were more likely to deliver the cards to households receiving other mail - meaning households with higher income and education levels, 3.) Finally, higher income and higher educated households were more likely to return the cards (Hedrich, Collinson and Rhoads, 1939). Postal test cards, generally thought of as an improved method of testing for entrance into the BRA, grossly overstated the completeness of birth registrations. Officials at the Census Bureau recognized by the 1930s the need for a nationwide test built on proper sampling procedures.

In addition to biased samples, public health officials worried about the subsequent quality of registrations after the entrance test (Wilbur, 1916). The early policy called for periodic retests using the collected lists methodology to ensure the 90 percent cutoff continued to be met (Davis, 1925). However, retests were infrequent - once in sixteen years in the case of Michigan - and poor results rarely led to a state exiting the BRA (Deacon, 1937). Only two states were ever expelled despite evidence that a number of states were well under the 90 percent cutoff: Rhode Island in 1919 (re-entering in 1921), and South Carolina in 1925 (re-entering in 1928) (Wilcox, 1933). By the mid-1930s, the Census Bureau's policy was that retests were for the sole purpose of helping to improve the registration systems of underperforming states, not to threaten removal from the BRA (Lenhart, 1943).

### *A. 1940 test of birth registration completeness*

The opportunity arose with the 1940 Decennial Census to develop a nationwide test that would greatly improve knowledge about the accuracy of the birth registration system (Grove, 1943). Officials believed that census enumerators could provide a more representative list of children born during a sample period than previous methods. Enumerators were instructed to fill out a special infant card for any child born during the four months prior to the census date.[7] The Census Bureau then matched each infant card and recorded death of an infant to birth certificates filed in state registrar offices. The completeness of registrations was then estimated as the proportion of infant cards and registered deaths for which a birth certificate had been filed.

For the nation as a whole, 92.5 percent of births were found to be registered, but the total hid large differences between races (94.0 percent for whites and 82.0 percent for blacks), cities and rural areas (96.9 percent in cities above 10,000 in population and 88.0 percent in small cities and rural

[7]The recall period for the infant test cards was limited to 4 months to reduce bias from memory lapses.

areas combined), and whether the birth occurred in an institution (98.5 percent in hospitals and 86.1 percent outside of hospitals) (Grove, 1943; Moriyama, 1946). The geographic variation in birth registration completeness for all races combined can be seen in Figure 1.[8] Registrations were more complete in northern states compared to the southern states. Much of the regional difference was suspected to arise from regional differences in urbanization and rates of hospital births (Moriyama, 1946). The racial disparity can also be explained by a largely rural black population in South with very low rates of hospital delivery. Although diminished, the racial difference remains when comparing within cities and within category of place of delivery.

The test suggests that births - especially black births - were systematically under-registered at higher rates in southern states relative to northern states, inducing an upward bias in reported infant mortality for the South. Regional comparisons and empirical strategies relying on cross-sectional variation may provide results contaminated by this systematic bias. We take this as our prime motivation for revising mortality estimates from 1910-1940.

### *B. Improvement in birth registration completeness*

Continued urbanization and increases in the proportion of births delivered in hospitals eventually reduced the number of births that went unregistered. Additionally, the value to the individual of holding a birth certificate rose as proof of age was increasingly required for receipt of government benefits, school attendance, and other privileges such as drivers licenses. Subsequent tests for registration completeness were conducted at a national scale in conjunction with the 1950 census and in the late 1960s using household surveys such as the Current Population Survey and the Health Information Survey (Shapiro and Schachter, 1952; U.S. Bureau of the Census, 1973). The results of the tests between 1940 and 1950 suggest large improvements in birth registration at the national level: from 92.5 percent in 1940 to 97.8 percent in 1950. The national average, however, belied large regional differences for minorities.[9] Completeness for southern non-whites only increased to 92 percent by 1950. For states in the Mountain census region with large Native American populations, the non-white completeness rate lagged at 78 percent.[10] By at least 1968, after the integration of hospitals in the South, the proportion of births delivered in a hospital converged to almost 99 percent nationwide for all races combined, and the birth registration system covered close to the

[8] Appendix Table A1 reports results from the test by region and whether delivery occurred in a hospital.

[9] Appendix Figure A1 plots the proportion of all births registered from the 1950 test against that from the 1940 test. We can clearly see that all states increased the quality of their published birth data. However, some improved more than others. One plausible reason is that low rates of out-of-hospital births persisted for non-whites in the South and West.

[10] The Mountain census region includes the following states: Arizona, New Mexico, Nevada, Utah, Colorado, Wyoming, Idaho, and Montana.

entire universe of all births: 99.4 percent for whites and 98.0 percent for nonwhites.

## III. Adjusting Infant Mortality Rates

In this section we outline the method and data sources used to revise infant mortality rates and birth estimates to account for the underregistration of births. We then graphically present the adjusted rates for different subcategories and discuss differences with the published vital statistics. The results of the exercise consist of a set of tables of infant mortality rates by subcategory for 1-,2-, and 5-year averages for use by researchers.[11] In the end, we provide two additional estimates of infant mortality in addition to those in the published vital statistics: one using the census-based adjustment method, and a second series where births are scaled by the extent of underregistration in the 1940 test in Grove (1943).

Published infant mortality rates are constructed from registered deaths before the age of 1 and registered births using the following formula:

$$\text{(Published)} \qquad \text{IMR}_{s,r,t}^{PUB} = \frac{\text{Pub Deaths}_{s,r,t}}{\text{Pub Births}_{s,r,t}}$$

where s denotes state of occurrence, r denotes race, and t denotes calendar year. IMR is often reported as deaths per thousand live births, but we choose to report in percentage points for simplicity. We know from contemporary evidence that ($\text{Pub Births}_{s,r,t}$) is biased downward in a way that leads to an upward bias in infant mortality rates for blacks and southern states.

To revise these rates, we rely on newly available complete count census microdata for 1920, 1930, and 1940 as the main source of information on the number of children who remained alive, published age-specific deaths for each state and race to account for non-infant deaths, and, finally, deaths of infants from published sources. In all estimates, the numerator of the IMR calculation, infant deaths, is held constant and comes from the published counts of registered deaths. Thus, any differences from the published mortality rates arises from an alternate estimate of live births. Our method provides a distinct improvement for understanding infant mortality during the early 20th century United States.

Our revisions can be expressed as:

$$\text{(Adjusted)} \qquad \text{IMR}_{s,r,t}^{ADJ} = \frac{\text{Pub Deaths}_{s,r,t}}{\text{Adj Births}_{s,r,t}}$$

[11] A full set of machine readable tables can be found on line at http://www.gregoryniemesh.net/data.

so that any difference with the published rates are entirely driven by differences in birth estimates. Our adjustment uses the complete count census datasets from IPUMS to estimate the number of live children by race, birth state, and birth year (Ruggles et al., 2015). To this, we then add the number of infant and non-infant deaths during intervening years between the birth year and the census year, both of which come from published tables. The data appendix contains a lengthy discussion of the data sources used and additional detail on the construction of estimates.

With this method, underenumeration of young children, especially infants, provides a potential bias in the census counts of live children, with the extent of underenumeration more severe for black infants (Brunsman, 1953). We use census counts for a given birth year from a *later* census to partially account for this underenumeration. For example, the number of live children born in 1929 (children less than one) is under counted in the 1930 census, but can be replaced with the count of ten year olds from the 1940 census. We implement the strategy by race, state of birth, and birth year using the counts of older children (10, 11, 12,... year olds) from a later census whenever their numbers exceed that in the earlier census. However, in Northern states where underenumeration presented a greater problem than underregistration, our revisions may yield birth estimates that are lower than the published estimates. Thus, we use the published registered births for any cell in which our census based method results in a smaller number of births. As such, revised infant mortality rates will never be higher than the published rates. Figure 4 plots the bias in the published rates (Published - Adjusted) against the extent of underregistration in the 1940 test form Grove (1943). States with higher levels of underregistration do in fact see larger reductions in IMR using our census-based method, just as we would expect. Over time, the size of the adjustment falls and the relationship between extent underregistration and the bias in published rates weakends. We interpret this set of facts as evidence of gradual improvement in the birth registration system over time.

Additionally, the complete census indexes provide a method to address a number of potential biases from the migration of young children out of their state of birth. Importantly, our estimate of live children includes those born in state $s$ regardless of the state of residence at the time of the census. The potential migration of children outside of their state of birth does not bias our estimates of births downward as long as they remain alive to the next census, or to the subsequent census when using counts of live children aged 10 or 11 years old. In the absence of a complete death index for the entire country, we do not have complete information on children that died outside of their state of birth. Bias enters in the case where states experience differential net migration rates, or differential mortality rates. We split the discussion of bias from migrant deaths into two parts: infant deaths and non-infant deaths. Infant deaths are allocated to the state of occurrence regardless of the child's birth state as we have no information on state of birth for deaths in this age group. As the out-of-state migration rate for infants is small (less than 1 percent in 1940), and most infant deaths occur in the first 30 days of life, and the likelihood of migration with a sick infant is relatively small, we believe the potential bias from this source is limited.

Both the cumulative likelihood of migration and the hazard rate increase with age, implying an increased potential for non-infant children to die outside of their state of birth. Thus, deaths of children aged 1 and above outside of their state of birth present a larger pathway for bias to enter the estimates. Working in the opposite direction, however, is the fact that mortality rates decrease rapidly after the first year of life, as do cross-state differentials in age-specific mortality. In practice, bias from migrant deaths is small. Figure 2 plots adjusted infant mortality when allocating non-infant deaths to state of birth versus adjusted infant mortality when allocating non-infant deaths to state of occurrence.[12] The methods have a tight almost, one-for-one, relationship. Differences do arise, however, from the high amounts of out-migration from southern states with large black populations during the "Great Migration". Nevertheless, these differences are small. As such, we choose to allocate non-infant deaths to states of birth in the revised estimates, but emphasize the limited importance of migration in this context.

Two potential biases remain as concerns: additional underenumeration in the census that using later censuses does not account for, and the underregistration of deaths. It is well known that the decennial censuses of the early 20th century under counted young children (Greville, 1947; Brunsman, 1953). Our estimates of the number of births are biased downward to the extent underenumeration occurred. Whether our adjustments are an improvement over the published aggregate mortality rates depends on the amount of underregistration of births relative to the amount of underenumeration in the census.[13] Underregistration was severe in the Southern states, and this is exactly where we see a large change in mortality rates from our adjustment, on the

[12] Rates with non-infant deaths allocated to state of birth is our preferred revised rate and corresponds to ($ADJ^4$) in the appendix. The procedure allocates the number of age-specific reported non-infant deaths in each *state of occurrence* to *states of birth* using the age-birth-state breakdown in the complete count censuses. For example, if 10 percent of black eight year olds living in Illinois in the 1940 census were born in Mississippi, then 10 percent of black non-infant deaths in Illinois are apportioned to black births in Mississippi for the 1932 birth year. Rates with non-infant deaths recorded in the state of occurrence corresponds to ($ADJ^2$) in the appendix. Appendix figure A2 plots the relationship for non-infant deaths.

[13] Walter F. Wilcox, former president of both the American Statistical Association and American Economic Association, wrote in 1933,"A Federal census is more likely to understate that to overstate the population because omissions are more common that false returns or double entries, but a census is probably nearer the truth than American registration of births...I should not consider that the registration of births was fairly complete unless the births exceeded the population under1 year of age by at least 10 per cent," (Wilcox, 1933). Our method makes a similar comparison of enumerated children to reported births.

order of a 20 percent increase in the number of births in the case for South Carolina. On the other hand, adjusted births slightly undershoot registered births in Northern states, because registration was nearly complete in those areas. These results confirm our expectations. The adjustment provides improved estimates of infant mortality where underregistration was more severe than underenumeration. Note that we impose a restriction that revised IMR can never be higher than published IMR, to account for the few instances in Northern states where registration was more complete than enumeration.

Bias from non-registered deaths presents a more difficult issue. When a parent decides against registering a death, no record of the event exists, and thus no direct means to assess the size of death underregistration is available to the researcher (Greville, 1947). To our knowledge, no contemporary evidence exists for the special case of the extent of underregistration for infant deaths. Contemporaries clearly believed the issue was less severe than for birth registration (Wilbur, 1916; Whelpton, 1934). Supporting this view, incentives were in place for death registration that were absent for birth registration. A cemetery burial, with the family or in churchyard, required a burial permit, which was only issued after a death had been registered and a certificate created. In the absence of a direct assessment of the potential bias from death underregistration, our revised rates provide a lower bound on infant mortality in the presence of death underregistration, whereas published rates provide an upper bound.

As an additional robustness check, and to help illuminate the sources of potential bias across estimation methods, we present a second adjusted series in which registered births in every year are scaled by the extent of underregistration from the 1940 test reported in Grove (1943). The adjusted IMR by scaling births can be expressed as:[14]

$$\text{(Scaled)} \qquad \text{IMR}^{SCALED}_{s,r,t} = \frac{\text{Pub Deaths}_{s,r,t}}{\text{Adj}_{SCALED}\ \text{Births}_{s,r,t}}$$

Biases in scaled rates stem from changes over time in the extent of completeness of birth registration.[15] The 1940 estimate of underregistration provides an increasingly uncertain or inaccurate method of adjustment the more distant the year of birth cell is from 1940. The processes that lead to registration evolve gradually over time (e.g. states placing importance on birth registration, and

[14]This scaled rate corresponds to adjustment 5 in the appendix.

[15]For births during the six months prior to the April census date in 1940, the estimates of registration completeness contained in Grove (1943) provide an accurate measure of the bias in infant mortality calculations. As such, we are confident in their use to make adjustments at the state level for 1939 and 1940. See the discussion of Adjustment 3 in the appendix for more information.

the proportion of births in hospitals or attended by a physician).[16] Underregistration, then, likely followed a downward trend, introducing some bias into the scaled IMR estimates.

How should a researcher choose between the revised and published estimates? Comparing the potential sources of bias and how they vary across time and place is helpful to distinguish the proper estimate. Biases in ($IMR^{ADJ}$) enter from underenumeration of young children, whereas bias from time-varying registration completeness affects ($IMR^{SCALED}$). In the end, we suggest using both IMR estimates, as well as the original published rates, to check any results for robustness. The bias present in any one of the three suggested estimates behave differently in the cross-section and over time, and thus showing that an estimate of a casual effect is robust to the choice of series provides convincing evidence of a true effect. Alternatively, the various rates can be used to provide a range of values for trends or group differentials.

Finally, we want to emphasize that a major contribution of our work is to produce infant mortality rates for states prior to entering the Birth Registration Area. Most states entered the Death Registration Area before meeting the requirements to enter the Birth Registration Area. We use the reported infant death counts in the mortality statistics volumes and our own estimates of births to construct infant mortality estimates for states prior to their entrance to the BRA.[17] The additional data allows researchers to extend analysis further into the past.

## IV. When does underregistration bias matter?

In general, IMR differentials and treatment effect estimates are biased to the extent that underregistration is correlated with the intervention or group attribute. Answering this question is simplified if we consider three scenarios. First, sometimes the researcher would like to know the true IMR for a given place and time without making any comparisons. In this simple scenario, *any* underregistration of births will bias the estimate of IMR. Secondly, researchers frequently make comparisons across locations, groups, or time. IMR differences arising from a cross-sectional comparison partially reduce the bias as long the extent of underregistration remains constant across the groups being compared. However, underregistration appears to vary in important ways across groups and locations (e.g. higher bias in the IMR for blacks and in southern states). Later, we provide two applications of cross-sectional comparisons where this bias can dramatically change

[16] The proportion of births registered clearly varies over time within a state. A simple way to argue the point is to notice the large differences in registration rates by whether the birth occurred in a hospital, and that the proportion of hospital births increased rapidly over time. The 1940 test showed that 98.5 percent of all hospital births were registered versus 86.1 percent of births outside of hospitals. Moriyama (1946) estimates that only 36.9 percent of births were hospital deliveries in 1935, but increased to 55.8 percent in 1940 and to 75.6 percent in 1944.

[17] Appendix Table A2 lists the years and states for which new estimates are available.

results. The first shows the impact on the pace and timing of regional convergence in the North-South difference in black IMR from Eriksson and Niemesh (2016). We close by revisiting Collins and Thomasson (2004) to conduct Oaxaca-Blinder decomposition of the national black-white IMR gap using socio-economic status measures as explanatory variables.

In the third scenario, researchers use panel data with observations for each location taken over multiple points in time. In this final scenario of a panel setting, location fixed effects and location-specific trends potentially account for any mis-measurement of IMR from differential completeness of the birth registration system. To explore this possibility, we estimate a series of regressions to determine the ability of state fixed effects and state-specific linear time trends to explain the gap between the published and adjusted infant mortality estimates.[18] We use three measures for the gap that correspond to three specifications for IMR commonly used in the literature: the difference $\left(IMR^{PUB} - IMR^{ADJ}\right)$, the ratio $\left(\frac{IMR^{PUB}}{IMR^{ADJ}}\right)$, and the natural log of the ratio $\left(\ln \frac{IMR^{PUB}}{IMR^{ADJ}}\right)$. Additionally, we split the sample into black, white, and total. No matter how the gap is specified, or on which sample the regression is run, between 16 and 26 percent of the variation in the gap remains after including state fixed effects. After state-specific linear time trends are included, the remaining variation in the gap ranges from 12 to 16 percent for whites and the combined sample, and 7 to 8 percent for the black sample. The scope for bias from this source is large. The standard deviation of the residuals from specifications that include linear trends range between 3 and 6 percent of the level of IMR depending on the sample and how the gap is measured. Using a panel setting to difference out unobservables or allowing for differential trends in unobservables does not fully remove the potential bias from causal estimates in the presence of birth underregistration.

The remainder of the section briefly discusses a number of implications that arise from using revised infant mortality rates in place of the published estimates. We begin by graphically showing national trends in IMR by race and the black-white gap. The most important changes from using revised estimates are on cross-sectional comparisons, such as the pace and timing of regional convergence in the North-South differential. We close by revisiting Collins and Thomasson (2004) to conduct a Oaxaca-Blinder decomposition of the national black-white IMR gap using state-level socio-economic status measures as explanatory variables.

[18]Table A4 reports results from this exercise.

### A. *Implications for national-level IMR*

Figure 5 plots three IMR series: published, restricted sample adjusted, and full sample adjusted. The restricted sample adjusted series limits the sample to state-year observations that are also in the published series (i.e in the BRA). The full sample adjusted series lifts that restriction and includes state-year observations for which our method fills a hole in the the published series (i.e. the state is part of the DRA, but not the BRA). Differences between the published and restricted sample adjusted series arise solely from differences in birth estimates, not changes in the composition of states. Differences between the published and full sample adjusted series arise from both changes in the composition of states and birth estimates.

Holding the sample of states constant between series, Panel A of Figure 5 suggests that adjustments to black rates lead to a level shift in IMR, but not to any meaningful change in the trend. Prior to 1925, this meant primarily the Northeastern and Midwestern states, where blacks experienced elevated rates of mortality compared to the southern states that were not yet included. However, adding the low IMR southern states, as in the full sample adjusted series, reduces IMR substantially in early years: by 42 percent in 1915. As more southern states enter the BRA, the “Full Adj” and “Restricted Adj” series converge and become identical when the entrance of Texas completes the BRA in 1933. The evidence suggests that black health was not as poor as contemporaries thought, but also implies that black health progress proceeded at a slower rate: a fall of 11.2 percentage points from 1915 to 1940 as compared to 4.9 percentage points in the published data.

Because black births were much more likely than white births to go unregistered, adjustments clearly reduce black infant mortality rates relative to those of whites at the national level, as seen in Figure 6. The figures make clear that adjustments lead to a shift in the level of both the absolute and relative black-white gap in IMR, but not to a revision in the trend. Thus, we find that the gap started from a smaller initial level, but fell at roughly the similar rate in terms of percentage points. Our understanding of national trends in the IMR gap does not seem to be much changed.

### B. *Implications for cross-state comparisons of IMR*

The large variation across states in the quality of birth registration data, however, leads to significant revisions of cross-sectional comparisons. Figure 7 illustrates the number and magnitude of rank changes between the published and revised rates, capturing the impact on cross-sectional comparisons. The left y-axis ranks states by published IMR and the right y-axis ranks states by revised IMR with the values for a state connected by a line. A downward slope in the line implies an improvement in rank. From Panel A of Figure 7, we can clearly see a large number of rank changes, many of a large magnitude. In general, the southern states for which the revision lowered IMR experience improvements in rank at the expense of states in the northeast and Midwest.

The effects of rank changes extend to regional differences and any subsequent convergence. In 1915, the South initially had a mortality advantage over the North for black infants, as shown in Figure 8.[19] Much of the gap is explained by the existence of a black urban-rural penalty combined with the fact that Northern blacks lived in cities, but were primarily rural in the South.[20] Northern infant mortality rates converged with the South as the urban penalty gradually declined over the course of the early 20th century. In the published data, the North overtook the South by the early 1930s in terms of black infant health. Three main implications emerge from using the revised estimates. First, the southern mortality advantage widens as the adjustment method primarily lowers black IMR in the South. Second, starting from a lower initial IMR in the South implies a faster *convergence rate* between the regions. Finally, the level shift downward in southern IMR delays the North overtaking South until the late 1930s, if at all before 1940.

To illustrate the importance of our adjustments to cross-place comparisons, we reprint IMR comparisons from Eriksson and Niemesh (2016). In that paper, we estimate the effect on the subsequent birth outcomes of infants to southern-born black parents after moving North during the first half of the "Great Migration". Here, we are concerned solely with the observed differences in black IMR across regions as an indicator of the health environments from which blacks left and in which they settled. Table 1 reports regional comparisons with published estimates and revised estimates. The change in inference induced by the bias from underregistration of births is clear. In the published data, black infant mortality was initially 33 percent higher (4.4 percentage points) in the North, with the southern mortality advantage declining to only 9 percent (1.1 percentage points) by the late 1920s and disappearing completely in the 1930s. The revised data widens the initial gap so that IMR in the North is 55 percent higher than in the South, and increases the southern mortality advantage in all decades (rows labeled "Diff" in Table 1). Additionally, we find that infant mortality rates were identical in 1940 rather than the North overtaking the South as in the published data. Finally, the last row shows the bias in the regional comparison, calculated as

[19]We do not observe a similar regional convergence for whites as the urban penalty for infant whites had disappeared by 1920.

[20]According to the 1940 decennial census, 89 percent of blacks lived in urban areas in the North census region, whereas 34 percent were urban dwellers in the South census region. Data underlying these calculations come from the full count 1940 census microdata from IPUMS.

the regional difference in the published data minus the regional difference in the revised data. The magnitude of the negative bias in each period is large: 38 percent, 154 percent, and 108 percent of the published regional IMR difference. Clearly, accounting for underregistration bias with our revised rates dramatically changes the interpretation of the differential health risks faced by black infants across the two regions.

### *C. Replication of Collins and Thomasson (2004)*

Finally, we use the revised state-level infant mortality rates to revisit Collins and Thomasson (2004), which decomposes explanatory factors of the racial gap in infant mortality for the period from 1920 to 1970. Their main findings include the fact that measures of income, urbanization, women's education, and physicians per capita (broadly interpreted as SES) explain a large portion of the black-white IMR gap prior to 1945, but a vanishingly small portion afterwards. We show that once the underregistration of births is accounted for in the revised IMR estimates, the interpretation of the decomposition dramatically changes.[21]

Collins and Thomasson (2004) run an Oaxaca-Blinder decomposition of the black-white IMR gap from 1920-1970. Using observations taken every five years at the state and race level, the natural log of IMR is first regressed on physicians per capita and race-specific measures for income, women's education, and urban status, and a set of year fixed effects. The $\beta$'s are averaged over race for the decomposition. Table 2 juxtaposes the results of the decomposition of published and revised infant mortality rates. Using published IMR estimates, the explained gap makes up between 75 to 96 percent of raw difference prior to 1945, with SES (income and education) providing the majority of explanatory power. Three major differences in the findings emerge when an identical decomposition procedure is conducted on revised rates.

1) A smaller raw black-white IMR gap emerges, not surprisingly, because the adjustment procedure lowers IMR relatively more for blacks than for whites.

2) The percent "explained" by controls is significantly reduced, up to 40 percent in some years, because of a change in estimated $\beta$'s. By reducing infant mortality for southern blacks - the low-income region for blacks - the strong correlation between income and IMR found in the original data is weakened.

3) The contribution of racial income differences to the IMR gap is reduced by close to a factor

[21] The authors were well aware of the underregistration of births and provide a discussion of how potential bias might enter their estimates. However, at the time no direct way of accounting for the bias was available.

of 10. Education, on the other hand, is only slightly reduced and remains the most important explanatory factor. Physicians per capita increases in importance.

In summary, the use of corrected IMR rates can change conclusions in meaningful ways in empirical exercises originally conducted with published vital statistics rates that include bias from the underregistration of births.

## V. Conclusion

Accurate vital statistics play an important role to target, execute, and evaluate public health interventions. Biases from underregistration of births hamper our understanding of public health crises, trends, and the evolution of racial health disparities. In this paper, we document large regional and racial disparities in the completeness of birth registration in the first half of the 20th century, which eventually disappeared in the 1960s after the racial integration of hospitals and almost all births were delivered in institutions.

To account for the severe underregistration of births, we construct adjusted infant mortality rates using a method based on the census enumeration of live children to obtain improved estimates of the number of births. The method is essentially a horse race between the extent of underregistration of births in the vital statistics and the extent of underenumeration of children in the census. In states and for races for which underregistration of births is more severe than underenumeration, our method provides a more accurate estimate of infant mortality than the published values.

## REFERENCES

**Bhalotra, Sonia, and Atheendar Venkataramani.** 2015. “Shadows of the Captain of the Men of Death: Health Innovation, Human Capital Investment, and Institutions.” Working Paper.

**Brunsman, Howard G.** 1953. *Infant Enumeration Study 1950. Procedureal Studies of the 1950 Censuses*, U.S. Bureau of the Census.

**Cassedy, James H.** 1965. “The Registration Area and American Vital Statistics.” *Bulletin of the History of Medicine*, 39: 221–223.

**Clay, Karen, Werner Troesken, and Michael R Haines.** 2013. “Lead and Mortality.” *Review of Economics and Statistics*, 96(3): 458–470.

**Collins, William J, and Melissa A Thomasson.** 2004. "The Declining Contribution of Socioeconomic Disparities to the Racial Gap in Infant Mortality Rates, 1920-1970." *Southern Economic Journal*, 70(4): 746–776.

**Cutler, David, and Grant Miller.** 2005. "The role of public health improvements in health advances: The twentieth-century United States." *Demography*, 42(1): 1–22.

**Davis, William H.** 1925. "Necessity for Completing the Registration Area by 1930." *American Journal of Public Health*, 399–404.

**Deacon, W J V.** 1937. "Tests and Promotion of Registration of Births and Deaths." *American Journal of Public Health*, 27(5): 492–498.

**Eriksson, Katherine, and Gregory T Niemesh.** 2016. "Death in the Promised Land: the Great Migration and Black Infant Mortality."

**Greville, Thomas.** 1947. *United States Life Tables and Actuarial Tables, 1939-1941.* U.S. National Office of Vital Statistics.

**Grove, Robert.** 1943. "Studies in the Completeness of Birth Registration, Part 1. Completeness of Birth Registration in the United States, December 1, 1939 to March 31, 1940." *Vital Statistics - Special Reports Series, National Office of Vital Statistics, U.S. Public Health Service*, 17(18): 223–296.

**Hansen, Casper Worm.** 2014. "Cause of death and development in the US." *Journal of Development Economics*, 109: 143–153.

**Hedrich, A.W., John Collinson, and F.D. Rhoads.** 1939. "Comparison of Birth Tests by Several Methods in Georgia and Maryland." *Vital Statistics - Special Reports Series, National Office of Vital Statistics, U.S. Public Health Service*, 7(60): 681–699.

**Jayachandran, Seema, Adriana Lleras-Muney, and Kimberly V Smith.** 2010. "Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs." *American Economic Journal: Applied Economics*, 2(2): 118–146.

**Lenhart, Robert F.** 1943. "Completeness of Birth Registration in the United States in 1940." *American Journal of Public Health*, 33(6): 685–690.

**Lunde, Anders S.** 1980. "The Organization of the Civil Registration System of the United States." International Institute for Vital Registration and Statistics, Technical Report No. 8, Bethesda, MD.

**Moehling, Carolyn, and Melissa Thomasson.** 2014. "Saving Babies: The Impact of Public Health Education Programs on Infant Mortality." *Demography*, 51(2): 367–386.

**Moriyama, Iwao.** 1946. "Estimated Completeness of Birth Registration: United States, 1935 to 1944." *Vital Statistics - Special Reports Series, National Office of Vital Statistics, U.S. Public Health Service*, 23(10): 223–227.

**Moriyama, Iwao.** 1990. "Measurement of Birth and Death Registration Completeness." International Institute for Vital Registration and Statistics, Technical Report No. 43, Bethesda, MD.

**Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek.** 2015. "Integrated Public Use Microdata Series." *Version 6.0 [Machine-readable database].*

**Shapiro, Sam, and Joseph Schachter.** 1952. "Birth Registration Completeness United States, 1950." *Public Health Reports (1896-1970)*, 67(6): 513–524.

**Thomasson, Melissa, and Jaret Treber.** 2008. "From Home to Hospital: The Evolution of Childbirth in the United States, 1928-1940." *Explorations in Economic History*, 45(1): 76–99.

**U.S. Bureau of the Census.** 1973. *Test of Birth Registration Completeness 1964 to 1968. Evaluation and Research Program PHC(E)-2*, U.S. Government Printing Office.

**Whelpton, P.K.** 1934. "The Completeness of Birth Registration in the United States." *Journal of the American Statistical Association*, XXIX(186): 125–136.

**Wilbur, Cressy.** 1916. *The Federal Registration Service of the United States: Its Development, Problems, and Defects.* U.S. Bureau of the Census.

**Wilcox, Walter F.** 1933. *Introduction to the Vital Statistics of the United States, 1900 to 1930.* Bureau of the Census, U.S. Department of Commerce.

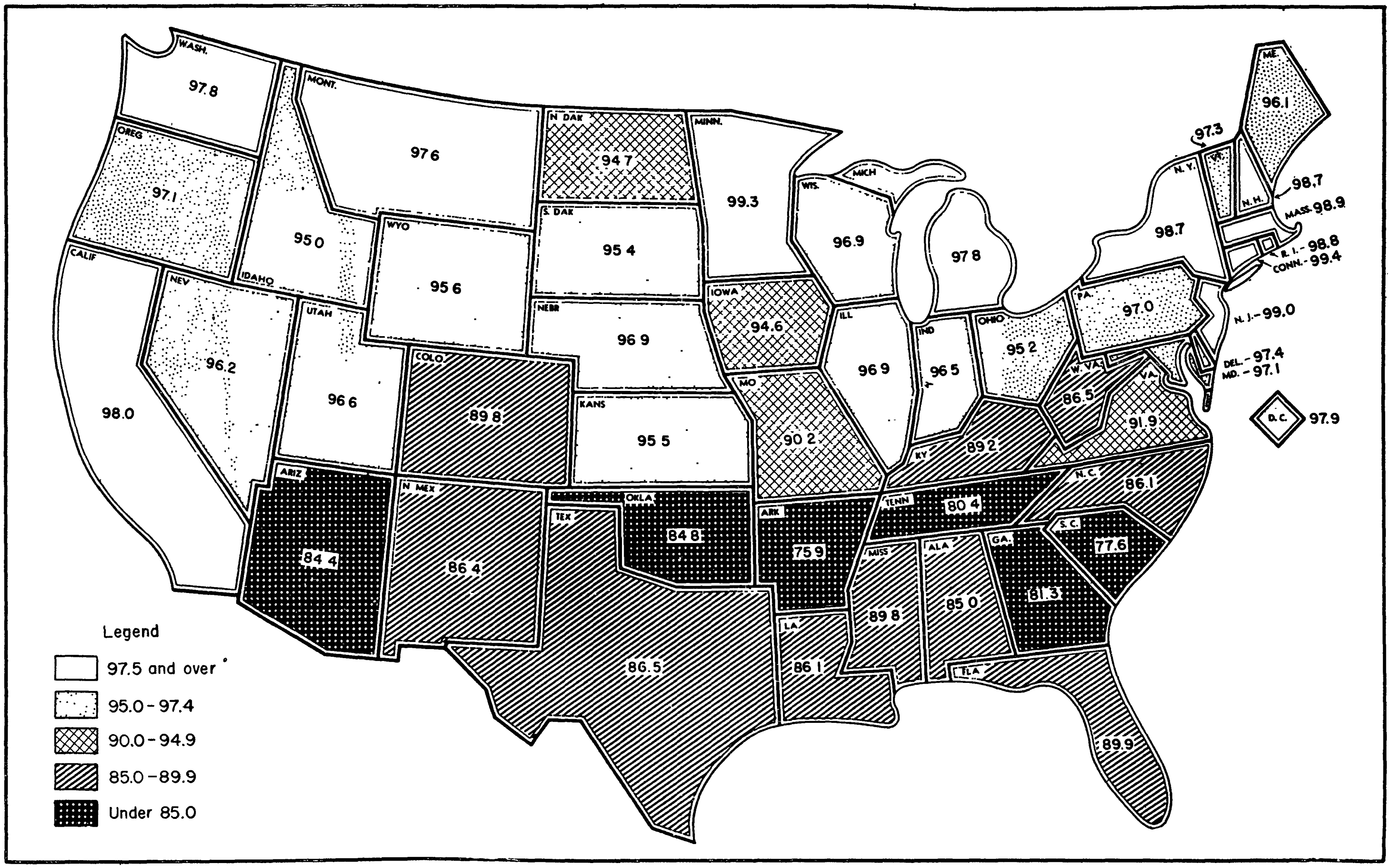
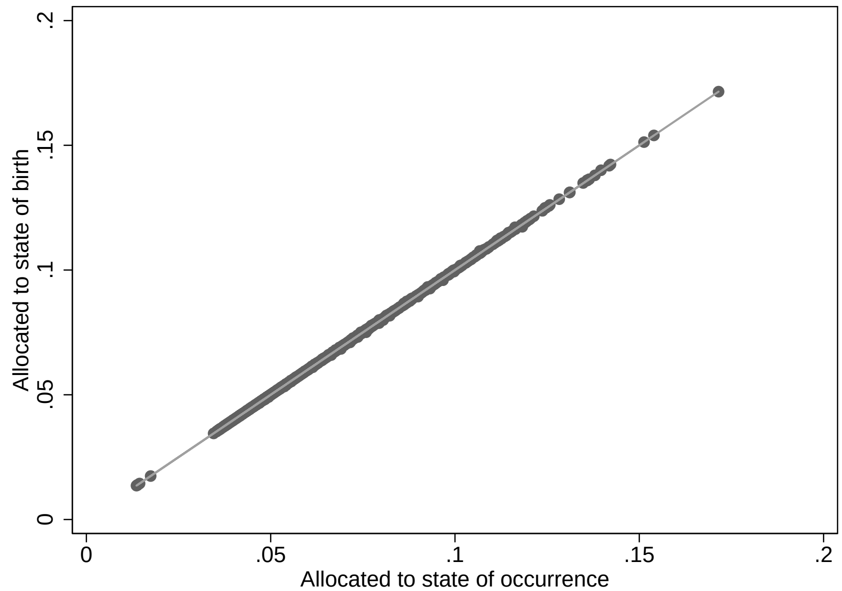
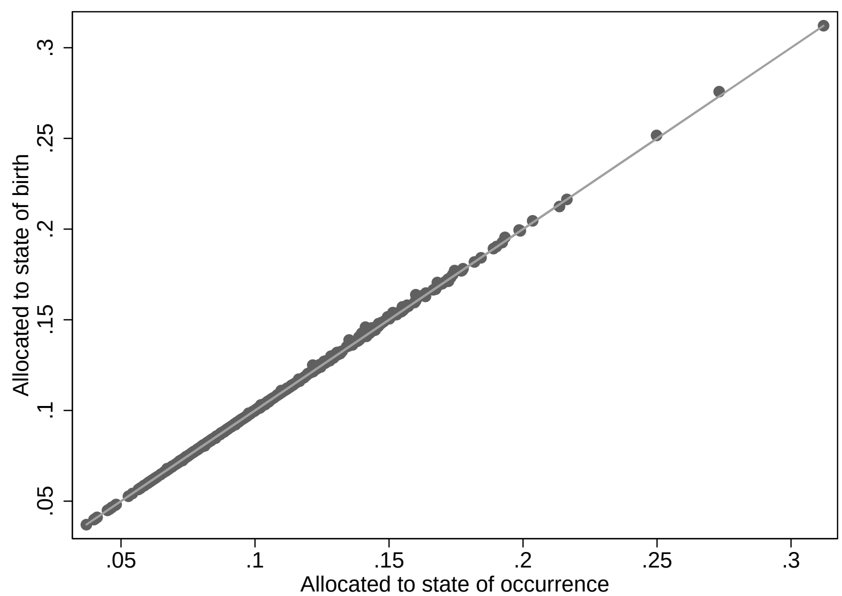



Figure 1. : Percent completeness of birth registration by state, December 1, 1939 to March 31, 1940

*Source:* Data from underlying test of birth registration completeness described in Robert Grove (1943)



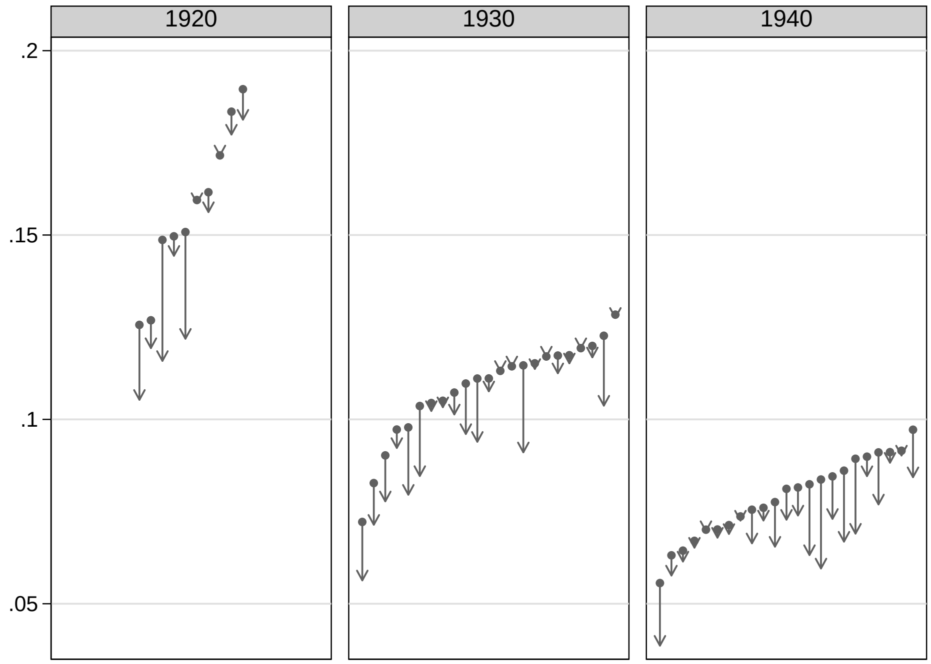

(a) White



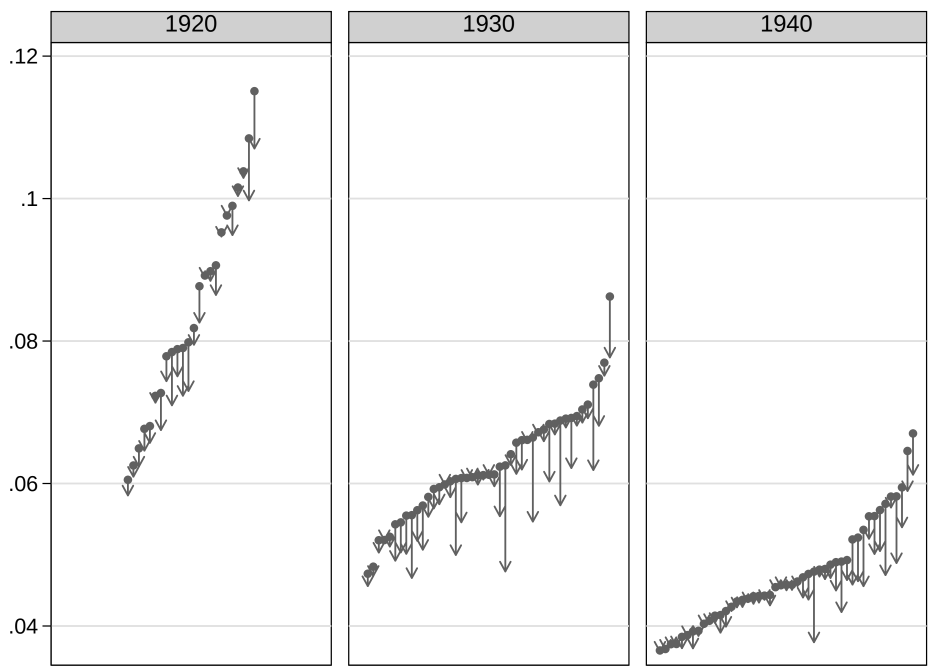

(b) Black

Figure 2. : Relation between infant mortality allocating non-infant deaths to state of birth versus state of occurrence

*Note:* Deaths of non-infants reported by age, race, and state of occurrence in the published tables are allocated to states of birth using the proportion of residents in each state from each state of birth in the complete count census microdata for 1920, 1930, and 1940 by race and age. Observations are limited to states with at least 1,000 births for the figure of black rates.
*Sources*: See data appendix for a discussion of author's calculations and sources used.



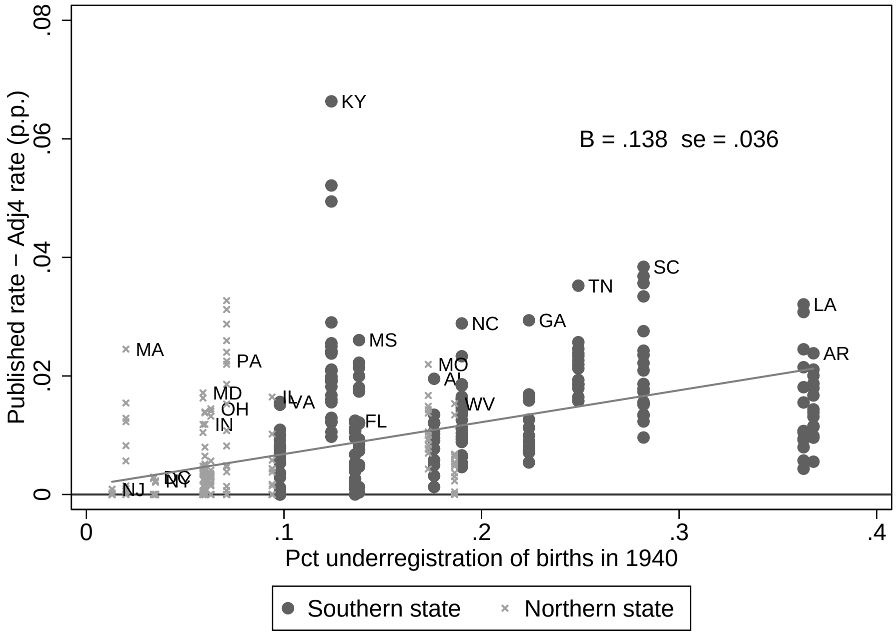

(a) Black



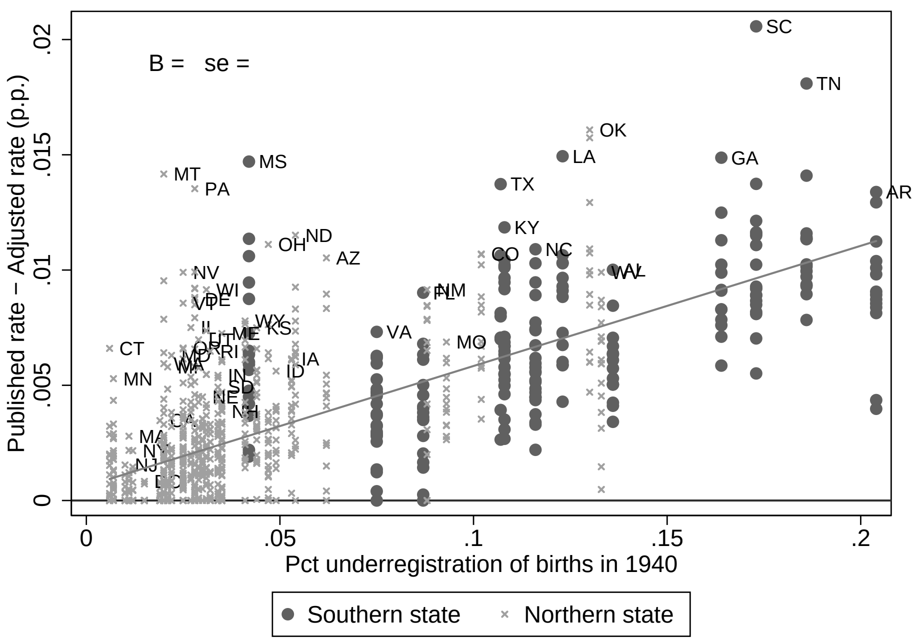

(b) White

Figure 3. : Change in five-year average IMR by race and time period from published to adjusted rates for states in the BRA

*Note:* Dots denote published IMR and the placement of the arrow denotes adjusted IMR. IMR is the average over the five years ending in the census year. States enter the sample when they enter the Birth Registration Area.



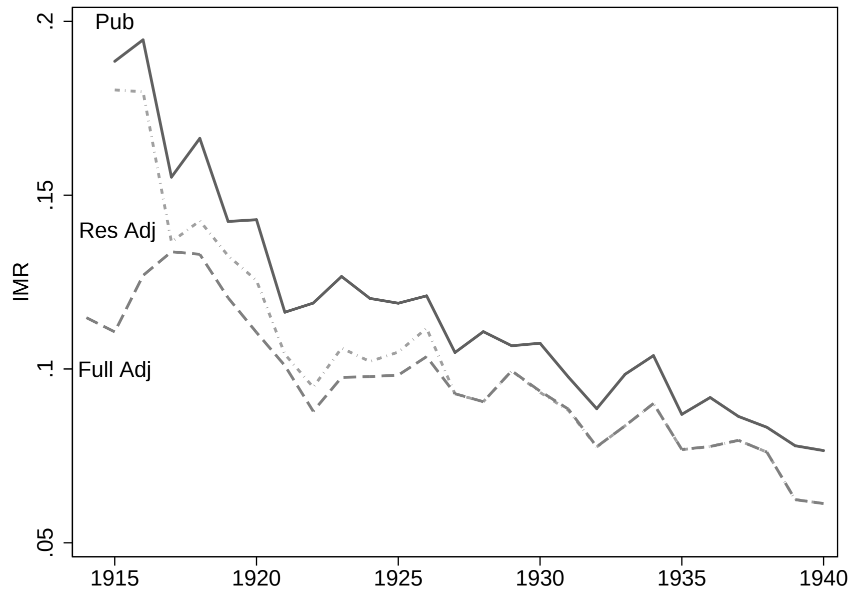

(a) Black



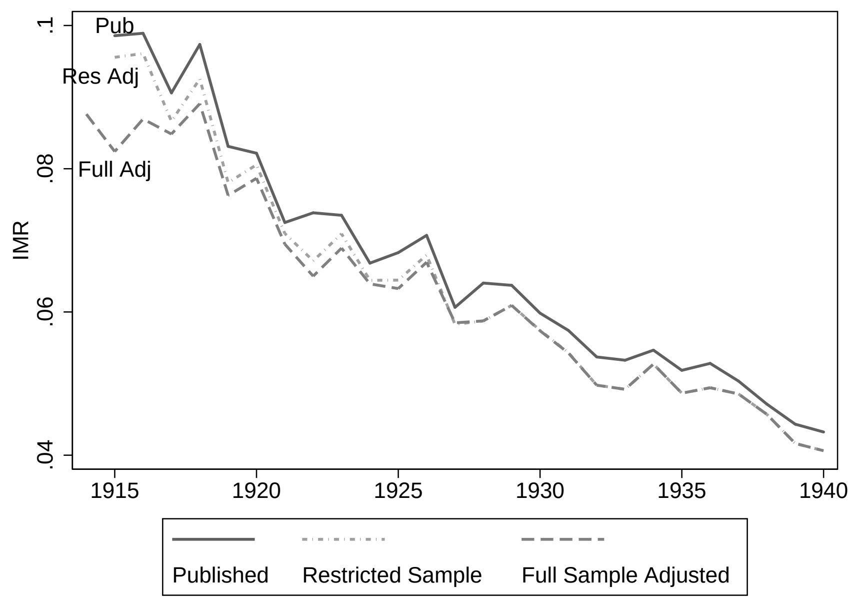

(b) White

Figure 4. : Bias in published IMR relative to percent underregistration - single year IMR

*Note:* Each single-year difference between adjusted rates and the published infant mortality rate is plotted against the percent of underregistration from the 1940 test for each race-state cell. Underregistration is measured only once for each state, and thus all observations for a state are on the same vertical line. The slope coefficient and standard error from the regression line are estimated with controls for year of birth.
*Sources:* See data appendix for a discussion of author's calculations and sources used.




(a) Black

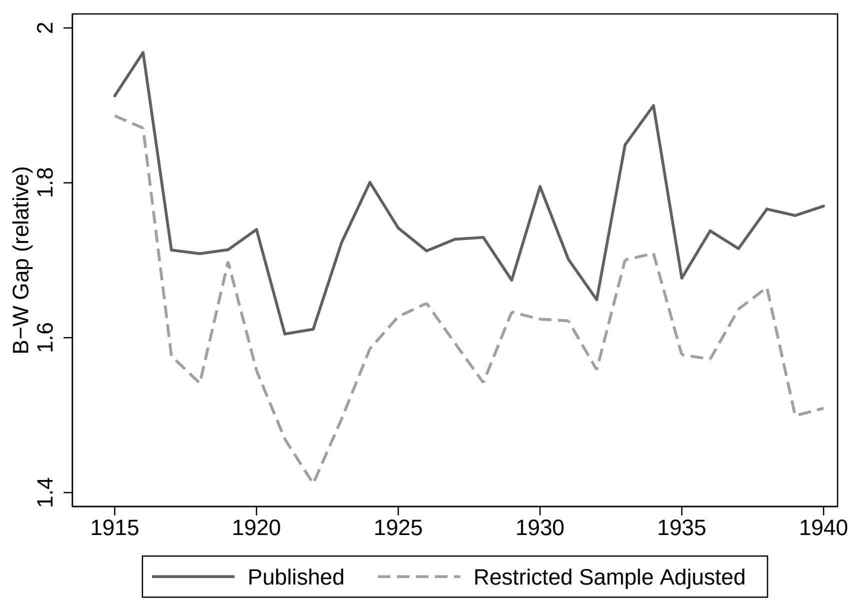



(b) White

Figure 5. : Published vs. adjusted national level rates

*Note:* Published rates include states in the Birth Registration Area. The Restricted Sample Adjusted series consists of the same set of states used in the calculation of the Published series. The Full Sample Adjusted series includes all states for which new rates exist.
*Source*: See data appendix for a discussion of author's calculations and sources used.

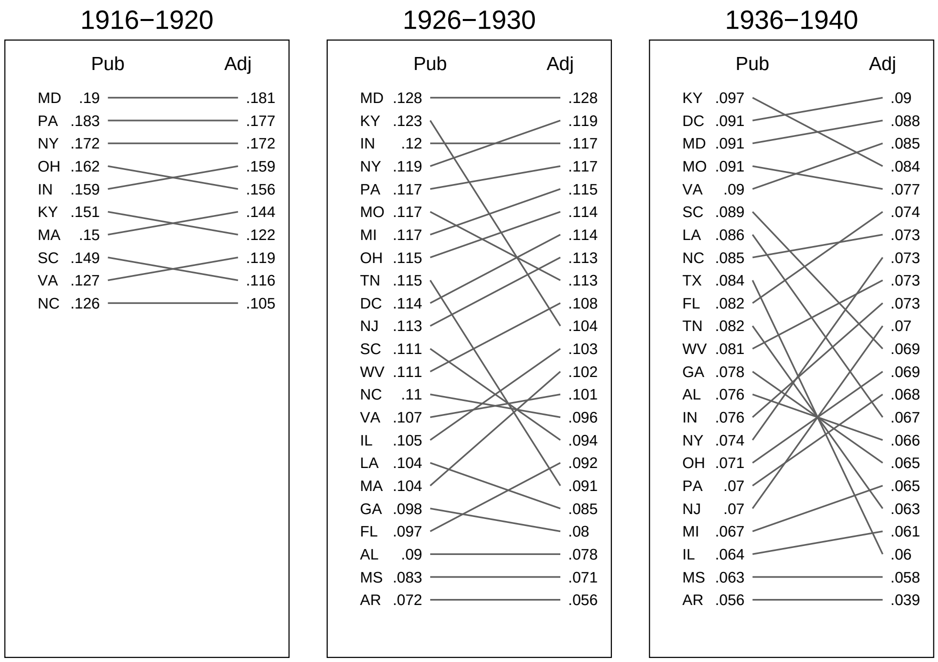

(a) Absolute difference (p.p.)



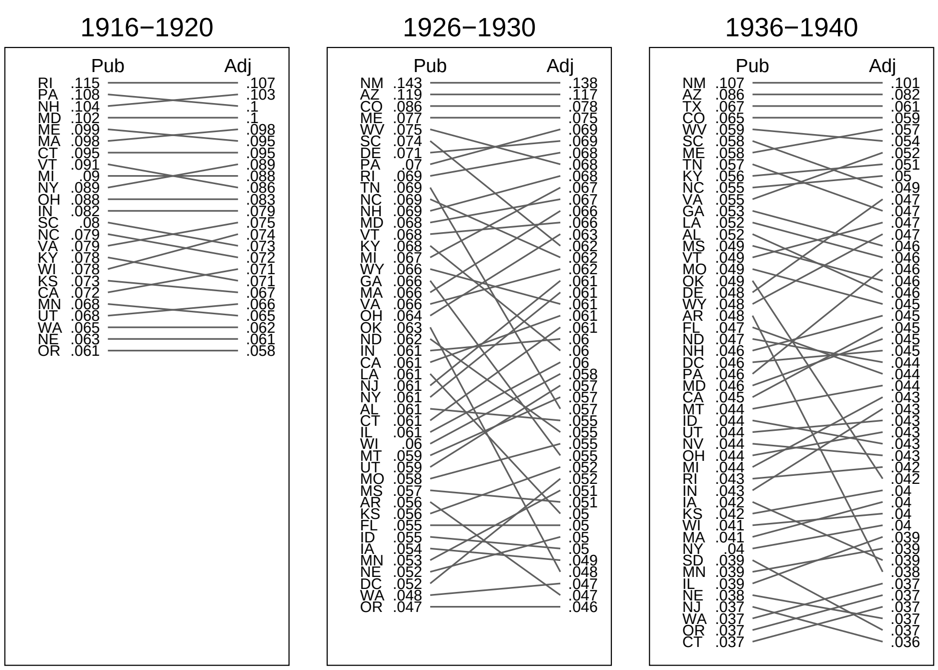

(b) Relative difference

Figure 6. : Black-White gap in IMR - published vs. adjusted rates

*Note:* The same set of states are used in the adjusted rates as in the published rates. States enter the sample as they enter the Birth Registration Area.
*Sources:* See data appendix for a discussion of author's calculations and sources used.




(a) Black




(b) White

Figure 7. : Change in five-year average IMR ranking from published to adjusted rates

*Note:* For each state in the Birth Registration Area for a given time period and with at least 5,000 births for each race over the five years, the chart ranks each state by published IMR on the left and adjusted IMR on the right.
*Sources:* See appendix for a discussion of author's calculations and sources used.

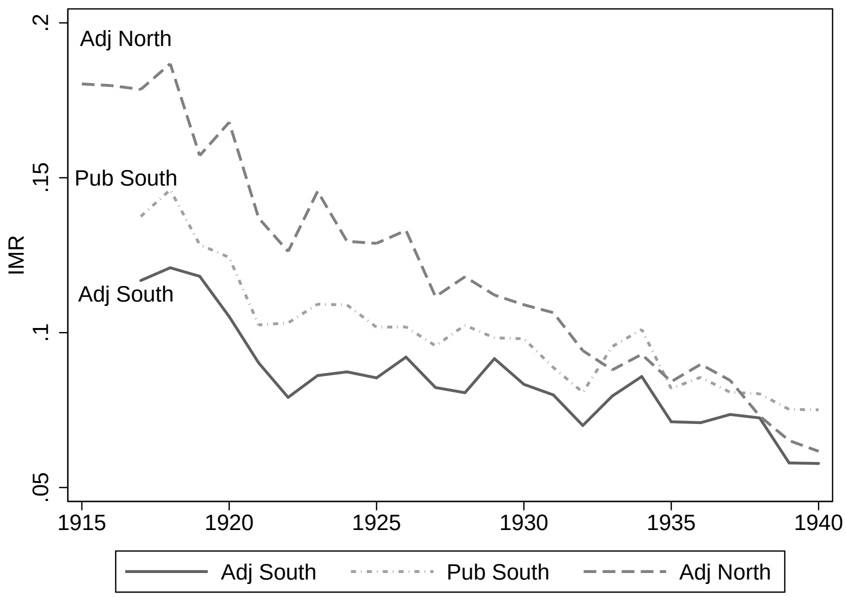



Figure 8. : Regional convergence of black IMR between southern and northern states

*Note:* States are included in calculations as they enter the Birth Registration Area. Published rates for the North are not shown as they are almost identical to adjusted rates.

Table 1—: Regional comparison of black IMR (p.p.)

| | 1916-1920 | 1926-1930 | 1936-1940 |
|---|---|---|---|
| *Published* | | | |
| North | 17.7 | 12.3 | 7.9 |
| South | 13.3 | 11.2 | 9.1 |
| Diff | 4.4 | 1.1 | -1.2 |
| | | | |
| *Adjusted* | | | |
| North | 17.2 | 12.1 | 7.6 |
| South | 11.1 | 9.2 | 7.6 |
| Diff | 6.1 | 2.8 | 0.0 |
| | | | |
| Bias | -1.7 | -1.7 | -1.3 |

*Note:* The difference is northern minus southern IMR. Bias is calculated as the published rate minus adjusted rate. States included in the sample are Illinois, Ohio, and Pennsylvania for the North, and North Carolina, South Carolina, and Tennessee for the South. The sample is limited to these states because of data limitations discussed in Eriksson and Niemesh (2016).

Table 2—: Black-White decomposition results from Collins and Thomasson (2004) and using revised rates

| | 1920 | 1925 | 1930 | 1935 | 1940 | 1945 | 1950 | 1955 | 1960 | 1965 | 1970 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Panel A: Collins and Thomasson (2004)[Published rates]* | | | | | | | | | | | |
| **Total Log Gap** | **0.455** | **0.462** | **0.493** | **0.447** | **0.521** | **0.454** | **0.496** | **0.583** | **0.626** | **0.630** | **0.559** |
| Gap “Explained” by: | | | | | | | | | | | |
| Income | 0.110 | 0.106 | 0.134 | 0.131 | 0.132 | 0.116 | 0.111 | 0.102 | 0.097 | 0.091 | 0.082 |
| Education | 0.337 | 0.293 | 0.343 | 0.329 | 0.307 | 0.265 | 0.223 | 0.166 | 0.121 | 0.093 | 0.069 |
| Urban | -0.106 | -0.079 | -0.082 | -0.069 | -0.056 | -0.037 | -0.014 | 0.000 | 0.011 | 0.019 | 0.029 |
| Physicians | 0.034 | 0.027 | 0.034 | 0.038 | 0.042 | 0.037 | 0.029 | 0.022 | 0.016 | 0.012 | 0.007 |
| **Total “Explained” Gap** | **0.375** | **0.347** | **0.429** | **0.429** | **0.425** | **0.381** | **0.349** | **0.290** | **0.245** | **0.215** | **0.187** |
| **Total “Unexplained” Gap** | **0.079** | **0.115** | **0.064** | **0.017** | **0.097** | **0.073** | **0.147** | **0.293** | **0.381** | **0.414** | **0.372** |
| **Percent “Explained”** | **0.83** | **0.75** | **0.87** | **0.96** | **0.81** | **0.84** | **0.70** | **0.50** | **0.39** | **0.34** | **0.33** |
| *Panel B: Revised rates* | | | | | | | | | | | |
| **Total Log Gap** | **0.388** | **0.426** | **0.434** | **0.408** | **0.378** | **0.454** | **0.496** | **0.583** | **0.626** | **0.630** | **0.559** |
| Gap “Explained” by: | | | | | | | | | | | |
| Income | 0.014 | 0.013 | 0.015 | 0.014 | 0.015 | 0.013 | 0.012 | 0.011 | 0.011 | 0.010 | 0.009 |
| Education | 0.293 | 0.250 | 0.287 | 0.269 | 0.259 | 0.214 | 0.18 | 0.134 | 0.098 | 0.075 | 0.055 |
| Urban | -0.125 | -0.087 | -0.082 | -0.066 | -0.056 | -0.035 | -0.014 | 0.000 | 0.010 | 0.018 | 0.027 |
| Physicians | 0.071 | 0.051 | 0.055 | 0.059 | 0.067 | 0.057 | 0.045 | 0.034 | 0.024 | 0.018 | 0.010 |
| **Total “Explained” Gap** | **0.253** | **0.226** | **0.275** | **0.277** | **0.284** | **0.249** | **0.224** | **0.179** | **0.143** | **0.121** | **0.102** |
| **Total “Unexplained” Gap** | **0.136** | **0.199** | **0.159** | **0.131** | **0.094** | **0.205** | **0.272** | **0.404** | **0.483** | **0.508** | **0.457** |
| **Percent “Explained”** | **0.65** | **0.53** | **0.63** | **0.68** | **0.75** | **0.55** | **0.45** | **0.31** | **0.23** | **0.19** | **0.18** |

*Note:* The “Total IMR Gap” is the difference between the average (weighted by population) log white and nonwhite infant mortality rates in each year. Each component of the ”Gap Explained By” section is the product of the difference in the variables average value for whites and nonwhites (in that year) and the relevant coefficients from a race specific regression of infant mortality for the full 1920-1970 sample period. See Table 2 in Collins and Thomasson (2004).
*Source:* See text for construction of revised IMR for1920-1940, and Collins and Thomasson (2004) for a description of the variables used in the decomposition and source of IMR for 1945-1970.

## Appendix A: Data sources and construction of estimates

### A1. Infant Deaths

In all estimates, we use the counts of registered deaths under 1 year of age by race and state reported in the published volumes of the Vital Statistics of the United States. In the calculations shown below, let infant deaths for each state, race, and birth year be denoted as $D_{s,r,t}$.

Sources:

> *Mortality Statistics, 19XX*, Table 4. Bureau of the Census: U.S. Department of Commerce. 1910-1936.
>
> *Vital Statistics of the United States, 19XX: Part I*, Table 16. Bureau of the Census: U.S. Department of Commerce. 1937-1940.

Three states vary how deaths of Hispanics are categorized over the period. Initially including Hispanics in the white category, Arizona (1930-1934), California (1931-1934), and Colorado (1931-1934) moved Hispanics to the other races category. Then in 1935, all states placed Hispanics back in the white category for the purpose of vital statistics reporting. To keep the series consistent over time, we place Hispanics in the white category in all years. The deaths listed as "Other" are assumed to be Hispanics in these three states for the years listed above.

### A2. Registered Births

Birth counts contained in the published volumes originate from the births registered with each state vital statistics office. Births of Hispanics in Arizona, California, and Colorado suffer from the same issues as the registered deaths in published reports. We choose to categorize Hispanic births as "white" to keep the birth series consistent. In the calculations shown below, let published births for each state, race, and birth year be denoted as $B_{s,r,t}$.

Sources:

> *Birth, Stillbirth, and Infant Mortality Statistics for the Birth Registration Area of the United States, 19XX*, Table 2. Bureau of the Census: U.S. Department of Commerce. 1910-1936.
>
> *Vital Statistics of the United States, 19XX: Part I*, Table 2. Bureau of the Census: U.S. Department of Commerce. 1937-1940.

### *A3. Enumeration of Children at Census Date*

We use the complete count census microdata for 1920, 1930 and 1940 provided by IPUMS to count the number of children of each age ($< 20$), race, and state of birth. Age is measured in years. Cells are allocated to years of birth by $(CensusDate) - age - 1$. Because infants and young children are more likely to go unenumerated, in some instances we choose to use the second census after the birth of the child as providing better information on the true size of the state of birth, race, year of birth cell. For example, black children born in 1929 in South Carolina face underenumeration on the order of 28 percent (Grove, 1943). Instead of using black 0 year olds from the 1930 census, we choose to use black 10 year olds from the 1940 census. In practice, we compare the number of children enumerated in each cell in the first and second census after the child's birth, and use whichever count is larger. In the calculations shown below, let census enumerations of live children for each state, race, and birth year be denoted as $CE_{s,r,t}$.

Sources:

Steven Ruggles, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek.*Integrated Public Use Microdata Series: Version 6.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2015.

### *A4. Non-infant Deaths*

The number of deaths of children aged 1 and older that occur for a given birth cohort are added back into the adjusted birth counts. For a given state and race, we count the number of deaths occurring in state s of 1 year olds in 1937, 2 year olds in 1938, and 3 year olds in 1939, which are then added to the adjusted birth estimate for the 1936 birth cohort. Using the state of occurrence may bias the estimate for states experiencing net in- or out-migration of children, which then face a lower (or higher) risk of death. Ideally, we could fully adjust the counts of non-infant deaths for migration with a complete death index covering the the entire United States. However, this death index does not exist. Instead, we leverage the 1920-1940 complete count census microdata from IPUMS to construct the proportion of all children of age "x" in each state of residence in the census year from each potential state of birth. The non-infant deaths accruing to each birth cohort in a state of residence are then apportioned to states of birth using the calculated proportions.

The published vital statistic tables report deaths for single years of age below 5, but group ages 5 to 9. We would like to have death counts for each age individually. Digitized state death indexes contain the complete listing of the underlying death certificates, which can be used to disaggregate vital statistics in the published tables. FamlySearch.org kindly provided death indexes for Illinois, North Carolina, Ohio, South Carolina, and Tennessee. Estimates of the proportion of deaths in the full 5-9 age group fall in each single age are calculated. We estimate the proportions for each decade (1915-1920, 1921-1930, 1931-1940), race (white, black), and region (North, South) cell. Ohio and Illinois provide death counts for the North, while the remaining states provide death counts for the South.

Additionally, the Census Bureau combined the 1-4 age groups in the 1930-1934 reporting years. We choose to use a state-based estimation strategy for these young ages rather than the process used for older children, as deaths in young children are more likely and this process does not impose the same trend on states within a region. The age breakdown of deaths in the 1-4 age group in state s in the two years on either side of the missing data are averaged. The average proportions are then scaled to sum to 1 and then applied to the published death counts in the combined 1-4 age group.

In the calculations below, the state of occurrence based non-infant death counts are denoted as $NID^{occurrence}_{s,r,t}$ and non-infant deaths apportioned to state of birth are denoted as $NID^{sob}_{s,r,t}$.

Sources:

> *Mortality Statistics, 19XX*, Table 4. Bureau of the Census: U.S. Department of Commerce. 1910-1936.
>
> *Vital Statistics of the United States, 19XX: Part I*, Table 16. Bureau of the Census: U.S. Department of Commerce. 1937-1940.

### *A5. Completeness of Birth Registration*

The Census Bureau conducted a test of the birth registration system in every state of the union covering the the six months prior to the 1940 census date. While conducting the regular duties of the decennial census, enumerators were asked to fill out special cards for infants born during the prior four months. The infant cards were then checked against the state birth registration files. The extent of underregistration of births was determined as the percent of infant cards that were successfully matched to a birth certificate. We use the estimates reported by state and race. The registration test is discussed in more detail in the main text. The percent of completeness is denoted as $pct\_complete$.

Source:

> Grove, Robert. *Studies in the Completeness of Birth Registration: Part 1 - Completeness of Birth Registration in the United States, December 1, 1939 to*

*March 31, 1940.* Vital Statistics - Special Reports Series, 17(18), April 1943. Table 1.

*A6. Construction of infant mortality rates*

We construct various estimates of the infant mortality rate (in percentage points) based on different methods to estimate births. Rates and births are calculated for each state, race, and year of birth cell.

1) The published rate: $\text{IMR}^{pub}_{s,r,t} = \frac{D_{s,r,t}}{B_{s,r,t}}$

2) Adjusted rate 1 uses census counts, infant deaths, and non-infant deaths in state of occurrence as the estimate of births: $\text{IMR}^{adj1}_{s,r,t} = \frac{D_{s,r,t}}{CE_{s,r,t}+NID^{occurrence}_{s,r,t}+D_{s,r,t}}$

3) Adjusted rate 2 replaces non-infant deaths in state of occurrence by allocating non-infant deaths to state of birth: $\text{IMR}^{adj2}_{s,r,t} = \frac{D_{s,r,t}}{CE_{s,r,t}+NID^{sob}_{s,r,t}+D_{s,r,t}}$

   Adjustment 2 directly accounts for the potential of migration to bias results from the deaths of non-infant children. Whereas Adjustment 1 allocates the deaths of children aged 1 and above to the state of occurrence, Adjustment 2 allocates these deaths to the state of birth using the proportion that each state of birth accounts for in each age-race-state-of-residence cell in the census. The movement from Adjusted rate 1 to Adjusted rate 2 does not seem to make a meaningful difference. Appendix Figure A2 plots non-infant deaths allocated to the state of occurrence against non-infant deaths allocated to state of birth. The two series are almost indistinguishable from one another for whites. However, differences emerge in southern states with large black populations that left the south in large numbers during the Great Migration. Figure 2 in the main text plots the relationship for estimated infant mortality rates and the differences are even smaller between the two series.

4) Adjusted rate 3 replaces births in 1939 and 1940 by scaling up published births by the percent complete from the 1940 test: $\text{IMR}^{adj3}_{s,r,t} = \begin{cases} \frac{D_{s,r,t}}{CE_{s,r,t}+NID^{sob}_{s,r,t}+D_{s,r,t}} & t < 1939 \\ \frac{D_{s,r,t}}{B_{s,r,t}\frac{1}{pct\_complete}} & t \in 1939, 1940 \end{cases}$

   This adjustment revises the 1939 data for the severe underenumeration of infants, and allows us to construct a revised estimate for 1940. Adjustment methods 1 and 2 accounted for underenumeration in the census of 0-1 year olds by using a subsequent census count for the same birth years (10-11 year olds). In the absence of complete microdata for the 1950 decennial census, we are unable to make a similar adjustment for 1939. Instead, we scale up registered births in each cell by the extent of registration incompleteness reported in Grove (1943) for the 1939 and 1940 birth years.

Moving from Adjustment 2 to Adjustment 3 reduces infant mortality rates in 1939 and 1940 in amounts inversely proportional to amount of registration completeness.Appendix Figure A3 makes clear the large impact on infant mortality rates in 1939 and 1940 for two illustrative cases. In southern states where registration of births was generally more of a problem than underenumeration, we see that our adjustment method predicts lower infant mortality than the published volumes for both blacks and whites, except for in 1939 and 1940. Scaling published births for these two years seems to keep the relative difference between Adjusted rate 2 and the published mortality rate consistent for for pre-1949 and 1939-1940. For northern states, registration wins the race against enumeration: birth registrations are more accurate than census enumerations. As such, we do not observe large differences between Adjusted rate 2 and published rates in northern states. Again, we see in the North that scaling births by the extent of underregistration keeps a consistent difference between the revised series and published series.

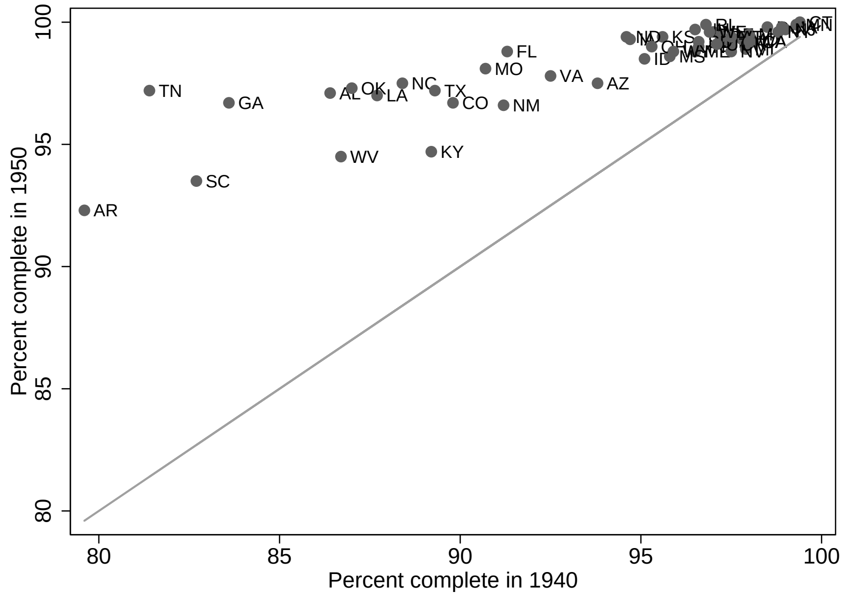
5) Adjusted rate 4 replaces any cell where the adjusted births from method 3 is less than the number of published births, and uses the published births instead:

$$\mathrm{IMR}^{adj4}_{s,r,t} = \begin{cases} \frac{D_{s,r,t}}{CE_{s,r,t}+NID^{sob}_{s,r,t}+D_{s,r,t}} & t < 1939 \\ \frac{D_{s,r,t}}{B_{s,r,t}} & CE_{s,r,t} + NID^{sob}_{s,r,t} + D_{s,r,t} < B_{s,r,t} \\ \frac{D_{s,r,t}}{B_{s,r,t}\frac{1}{pct\_complete}} & t \in 1939, 1940 \end{cases}$$

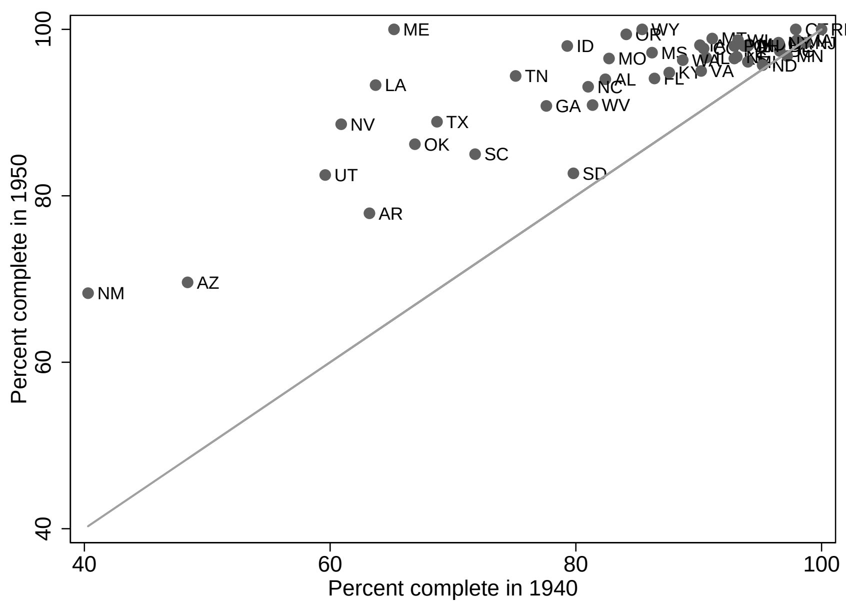
The original purpose for undertaking this study was to construct accurate infant mortality rates in a cross-section of states to make regional comparisons. The fact that method 3 allows for an *increase* in estimated IMR relative to the published rates is concerning. In those instances, registered births provide a better estimate of the truth than does our census based method, and adjusting does not lead to improvements in accuracy. As such, Adjustment 4 takes the results from Adjustment 3 and replaces the census based births with registered births in any state-year-race cell where registered births are greater.

We believe that Adjustment 4 provides the best improvement in accuracy to make cross-sectional comparisons. Figure 4 plots the bias in the published rates (Published - Adj 4) against the extent of underregistration of births from the 1940 test. States with higher levels of underregistration do in fact see larger reductions in IMR using our census-based method, just as we would expect.

6) Adjusted rate 5 replaces births in *all years* with the published births scaled by the percent complete from the 1940 test: $\text{IMR}^{adj5}_{s,r,t} = \frac{D_{s,r,t}}{B_{s,r,t}\frac{1}{pct\_complete}}$



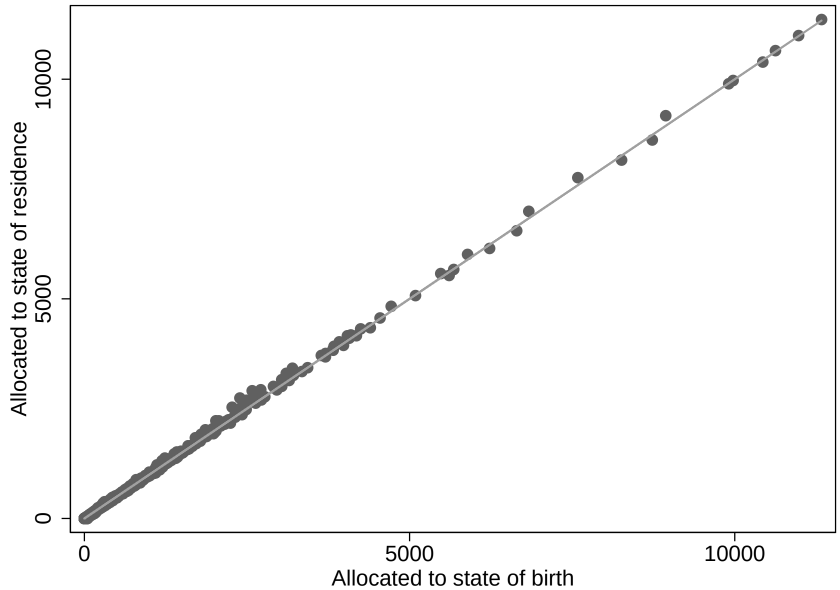

(a) White



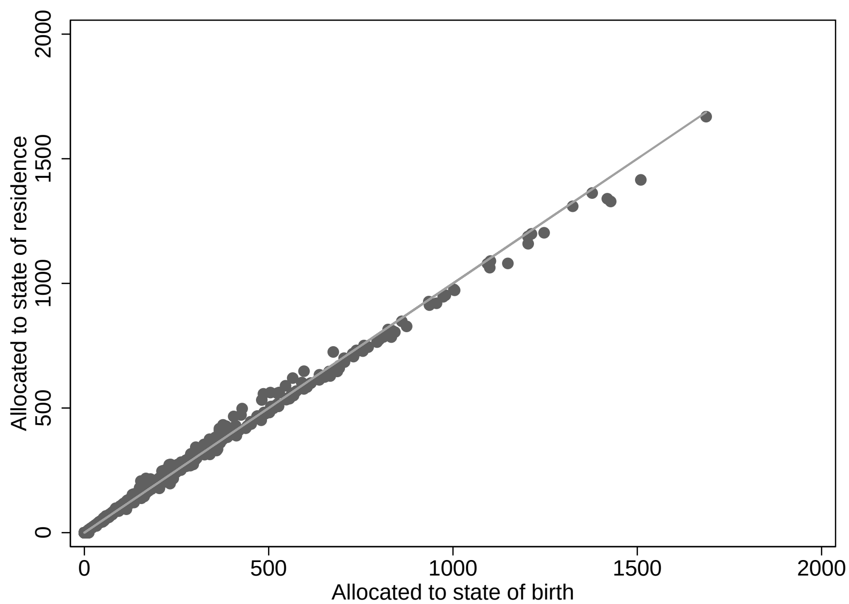

(b) Nonwhite

Figure A1. : Relation between birth registration in the 1950 and 1940 tests

*Note:* The solid line plots the one-to-one relationship for 1940. A point above this line represents an improvement in the proportion of births registered.
*Source:* Data from Shapiro and Schachter (1952).



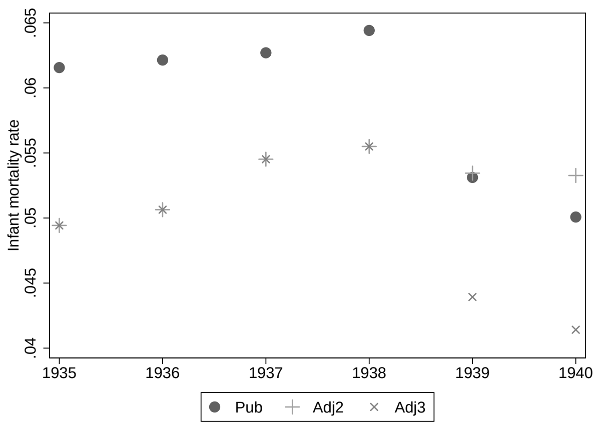

(a) White



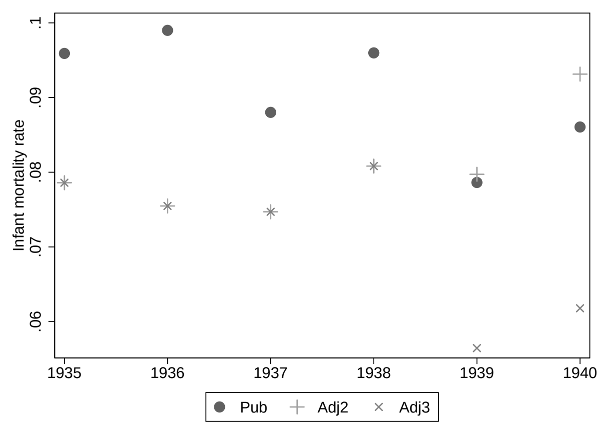

(b) Black

Figure A2. : Relation between non-infant deaths allocated to state of occurrence and allocated to state of birth

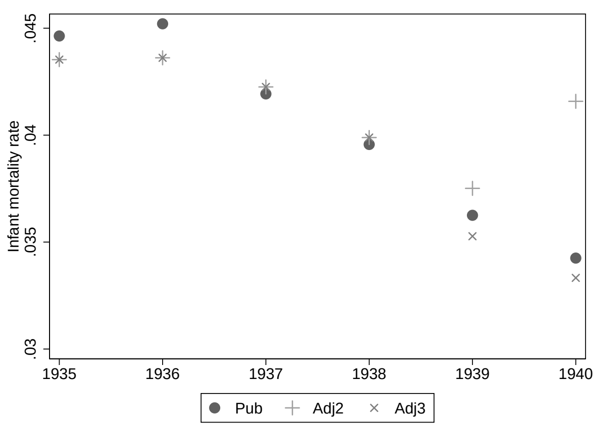
*Note:* Deaths of non-infants reported by age, race, and state of occurrence in the published tables are allocated to states of birth using the proportion of residents in each state from each state of birth in the complete count census microdata for 1920, 1930, and 1940 by race and age. Observations are limited to states with at least 1,000 births for the figure of black deaths.
*Sources*: Table 4 of *Morality Statistics of the United States* for years 1910-1936 and the complete count decennial census microdata for 1920-1940 provided by IPUMS (Ruggles et al., 2015).



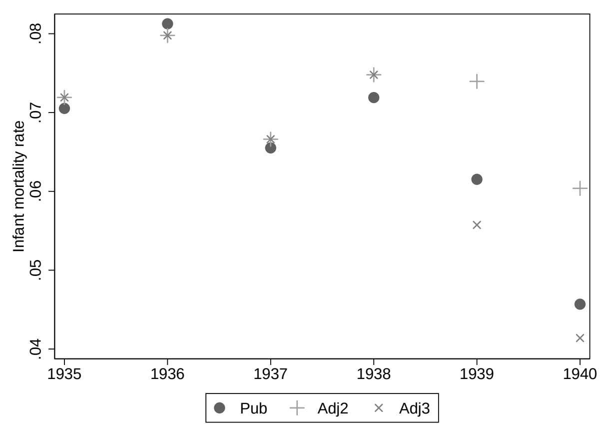

(a) South Carolina - White




(b) South Carolina - Black




(c) Illinois - White




(d) Illinois - Black

Figure A3. : Comparing published rates to Adjusted Rate 2 and Adjusted Rate 3 for two example states

*Note:* Adjusted rate 3 takes values from adjusted rate 2 and replaces births for 1939 and 1940 by scaling up the published births by the extent of of underregistration recorded in the 1940 test.
*Sources*: See data appendix for a discussion of author's calculations and sources used.

Table A1—: Regional and racial differences in the extent of registration in the 1940 test

| Area | Total | | | In institutions | | | Not in institutions | | |
|---|---|---|---|---|---|---|---|---|---|
| | All races | White | Nonwhite | All races | White | Nonwhite | All races | White | Nonwhite |
| United States | 92.5 | 94.0 | 82.0 | 98.5 | 98.6 | 96.3 | 86.1 | 88.2 | 77.2 |
| Geographic Divisions: | | | | | | | | | |
| New England | | | | 99.5 | 99.5 | 99.1 | 95.7 | 95.8 | 90.8 |
| Middle Atlantic | | | | 99.2 | 99.3 | 97.8 | 94.5 | 94.9 | 88.4 |
| East North Central | | | | 98.7 | 98.8 | 96.6 | 93.6 | 93.8 | 89.1 |
| West North Central | | | | 98.2 | 98.4 | 93.7 | 91.1 | 91.5 | 77.0 |
| South Atlantic | | | | 96.7 | 96.8 | 95.8 | 82.4 | 84.4 | 78.0 |
| East South Central | | | | 98.2 | 98.3 | 97.4 | 83.0 | 83.8 | 81.0 |
| West South Central | | | | 96.4 | 96.6 | 94.3 | 78.5 | 81.7 | 68.1 |
| Mountain | | | | 97.9 | 98.0 | 95.0 | 83.2 | 87.9 | 37.1 |
| Pacific | | | | 99.1 | 99.2 | 97.1 | 91.4 | 91.4 | 91.7 |

*Note:* Reprinted from (Grove, 1943; Moriyama, 1946).

Table A2—: List of states and years for which adjusted rates fill gaps in published vital statistics

| State | Years | State | Years |
|---|---|---|---|
| AL | 1925 - 1926 | NC | 1910 - 1916 |
| AR | - | ND | - |
| AZ | - | NE | - |
| CA | 1910 - 1918 | NH | 1910 - 1914 |
| CO | 1910 - 1927 | NJ | 1910 - 1920 |
| CT | 1910 - 1914 | NM | - |
| DE | 1919 - 1920 | NV | - |
| FL | 1919 - 1923 | NY | 1910 - 1914 |
| GA | 1919 - 1923 | OH | 1910 - 1916 |
| IA | 1923 | OK | - |
| ID | 1922 - 1925 | OR | 1918 |
| IL | 1918 - 1921 | PA | 1910 - 1914 |
| IN | 1910 - 1916 | RI | 1910 - 1914 |
| KS | 1914 - 1916 | SC | 1916 - 1918, 1925 - 1927 |
| KY | 1911 - 1916 | SD | 1930 - 1931 |
| LA | 1918 - 1926 | TN | 1917 - 1926 |
| MA | 1910 - 1915 | TX | - |
| MD | 1910 - 1915 | UT | 1912 - 1916 |
| ME | 1910 - 1914 | VA | 1913 - 1916 |
| MI | 1910 - 1914 | VT | 1910 - 1914 |
| MN | 1910 - 1914 | WA | 1910 - 1916 |
| MO | 1912 - 1926 | WI | 1910 - 1916 |
| MS | 1919 - 1920 | WV | - |
| MT | 1910 - 1921 | WY | - |

*Note:* The table lists the years and states for which our process of estimating births allows us to construct infant mortality rates that are not reported in the published volumes in the *VSUS* for the Birth Registration Area. States tended to enter the Death Registration Area prior to entering the Birth Registration Area, allowing us to use the published counts of infant deaths and our own estimates of births. Beginning in 1925, South Carolina was removed from the Birth Registration Area for low levels of registration completeness during the previous years. Rhode Island first entered the BRA in 1915, was removed in 1919, and re-entered in 1921. The Census Bureau allowed Rhode Island to re-enter in 1928 after showing a 90 percent registration rate using test cards.

Table A3—: Entry Date to the Birth Registration Area

| | | | | | |
|---|---|---|---|---|---|
| 1915 (original) | Minnesota (1910) | 1917 | N. Carolina (1916) | 1927 | Missouri (1911) |
| | Michigan (1900) | | Virginia (1913) | | Arkansas (1927) |
| | District of Columbia (1880) | | Ohio (1909) | | Louisiana (1918) |
| | Pennsylvania (1906) | 1919 | Oregon (1918) | | Tennessee (1917) |
| | New York (1890) | | California (1906) | | Alabama (1925) |
| | Rhode Island (1890) | 1920 | Nebraska (1920) | 1928 | Colorado (1906) |
| | Connecticut (1890) | 1921 | Mississippi (1919) | | Oklahoma (1928) |
| | Massachusetts (1880) | | Delaware (1919) | | Georgia (1928) |
| | New Hampshire (1890) | | New Jersey (1880) | | S. Carolina (1916) |
| | Vermont (1890) | 1922 | Montana (1910) | 1929 | Nevada (1929) |
| | Maine (1900) | | Wyoming (1922) | | New Mexico (1929) |
| 1916 | Maryland (1906) | | Illinois (1918) | 1932 | S. Dakota (1930) |
| 1917 | Washington (1908) | 1924 | N. Dakota (1924) | 1933 | Texas (1933) |
| | Utah (1910) | | Iowa (1923) | | |
| | Kansas (1914) | | Florida (1919) | | |
| | Wisconsin (1908) | 1925 | W. Virginia (1925) | | |
| | Indiana (1900) | 1926 | Idaho (1922) | | |
| | Kentucky (1911) | | Arizona (1926) | | |

*Source: Birth, Stillbirth, and Infant Mortality Statistics of the United States, 1933.* Page 2.

Table A4—: Explanatory power of state fixed effects and linear trends for the gap between published and revised IMR

| | $\left(IMR^{PUB} - IMR^{ADJ}\right)$ | $\frac{IMR^{PUB}}{IMR^{ADJ}}$ | $\ln\left(\frac{IMR^{PUB}}{IMR^{ADJ}}\right)$ |
|---|---|---|---|
| Black | | | |
| Year fixed effects | 0.038 | 0.059 | 0.073 |
| State fixed effects | 0.736 | 0.816 | 0.834 |
| State-specific linear trends | 0.926 | 0.927 | 0.917 |
| Standard deviation of residuals | 6% | 5% | 4% |
| White | | | |
| Year fixed effects | 0.119 | 0.111 | 0.114 |
| State fixed effects | 0.777 | 0.809 | 0.814 |
| State-specific linear trends | 0.839 | 0.867 | 0.869 |
| Standard deviation of residuals | 5% | 6% | 2% |
| Total | | | |
| Year fixed effects | 0.087 | 0.096 | 0.099 |
| State fixed effects | 0.795 | 0.808 | 0.816 |
| State-specific linear trends | 0.851 | 0.863 | 0.867 |
| Standard deviation of residuals | 3% | 3% | 3% |

*Note:* Each entry reports the R-squared from a separate specification of the gap between published and adjusted infant mortality rates from an unbalanced panel of states in years in which both published and adjusted rates exist. We run regressions on the Black sample, White sample, and the combined Total sample. The first row of for each sample is from a specification that includes only year fixed effects. The second row adds state fixed effects, and the third row adds state-specific linear trends. The final row reports the standard deviation of residuals from the specification that includes state-specific linear trends as a percent. In the first column, the standard deviation is relative to the mean adjusted IMR in the first column.