Estimating Mortality Effects with Incomplete Data

Kevin P. Connolly*

January 6, 2025

Abstract

There is a large and rapidly growing literature estimating treatment effects on older-age mortality using early-life natural experiments. This literature mostly ignores the problem of endogenous sample selection, which occurs when the treatment can change how much information is observed about an individual in the study population. I illustrate the severity of the issue by considering two prominent natural experiments that are able to avoid endogenous sample selection. After replicating the original results, I impose patterns of missing mortality data common in the literature and successively re-estimate the treatment effects. Endogenous sample selection results in bias, artificially increasing the estimated magnitude of the treatment effect in these natural experiments by up to 40% or decreasing it by as much as 99%. At the same time, by artificially shrinking the residual variance, endogenous sample selection artificially shrinks standard errors by as much as 75%. After accounting for endogenous sample selection, the results of many papers in this literature cannot be taken at face value, but rather require a substantial upward standard error correction and a sensitivity analysis. The magnitude of the standard error correction depends on the type of mortality outcome, as well as the type and extent of missing mortality data. The sensitivity analysis consists in showing the sensitivity of the estimated treatment effect to different assumptions about the effect of the treatment on unobserved mortality. I conclude by laying out how researchers can avoid endogenous sample selection arising from missing mortality data.

*University of Chicago (email: kevinconnolly@uchicago.edu). I thank Christopher Avery, Dan Black, David Cutler, Edward Glaeser, Joshua Gottlieb, Sebastian Haneuse, Adriana Lleras-Muney, and Jack Mountjoy for helpful comments and suggestions. I am grateful to J. Clint Carter for help in accessing restricted-use Census data. This paper is based on work supported by the National Institute on Aging through grants T32AG000186 and 5T32AG000243. The views expressed are those of the author and not those of the U.S. Census Bureau or the National Institute on Aging. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 1743 (CBDRB-FY21-P1743-R9271).

1 Introduction

There is a large and rapidly growing literature estimating treatment effects on older-age mortality using early-life natural experiments. All of the mortality datasets used in this literature are incomplete to at least some degree. In many cases individuals are still alive at the time the analysis is being conducted, for example, and in many others, historical mortality data are not obtainable (as is the case with Social Security mortality data, which is incomplete for deaths prior to the mid-1970s). At the same time, there is great variability in the type and degree of incompleteness. The window of fully observed mortality can range from 17 to more than 50 years. Outside this window, information on deaths can be partially missing ("censored") but in most cases is completely missing ("truncated").[1] In all cases mortality information is missing not at random but in a way that depends upon the outcome, age at death, which in turn depends on the treatment. In this way the treatment can change how much information is observed about an individual. I will call this phenomenon *endogenous sample selection*.[2]

I directly test for bias arising from endogenous sample selection using two natural experiments, the Mothers' Pension natural experiment (Aizer et al., 2016) and the U.S. community college openings natural experiment (Connolly, 2024). Both natural experiments have mortality datasets that are comparatively complete: both have a relatively large window of fully observed mortality, and mortality information outside the window is only partially missing (Figure 1). With each of these two datasets I can conduct the analysis as if the mortality dataset were less complete. I drop mortality information from each of the two datasets to match various patterns of missing mortality data in the literature. I then re-estimate the treatment effect using the less complete mortality dataset and assess the impact of endogenous sample selection.

[1] The distinction between censored and truncated mortality data is crucial for estimating valid mortality effects. The Duke-Medicare mortality dataset (Black et al., 2015) is left-truncated at age 65, meaning individuals who died before age 65 do not show up at all in the mortality dataset. By contrast, the Mothers' Pension mortality dataset (Aizer et al., 2016) is left-censored at the point when Social Security mortality data becomes comprehensive (the early 1970s). Individuals who died before the early 1970s still show up in the Mothers' Pension mortality dataset, but their date of death is known only to have occurred prior to the early 1970s.

[2] When I describe a mortality dataset as incomplete, I mean that mortality information is missing in a way that depends on age at death. If one estimates a causal effect on a mortality outcome using an incomplete mortality dataset, the mortality sample is endogenously selected. The only exception is when the incomplete mortality dataset is censored, rather than truncated, and the outcome is survival to an age within the window of fully observed mortality.

I find that endogenous sample selection leads to bias. The headline result of the Mothers' Pension paper is an estimated average treatment effect of one additional year of lifespan (the reported result, in log years of lifespan, is $\beta = 0.0182$, $SE = 0.007$; Table 4, Panel A of Aizer et al. (2016)). I first show that I can exactly replicate this result using the same mortality dataset. I then re-estimate this result as if the authors had used a less complete mortality dataset, the National Archives and Records Administration's Numident dataset ("NARA Numident"). Commonly used to estimate mortality effects, NARA Numident includes individuals who died between 1988 and 2005 with high probability, and those who died before 1988 or after 2005 with low probability.[3] After dropping individuals from the sample who died before 1988 or after 2005, and therefore would be missing from NARA Numident, the treatment effect "estimate" of the Mothers' Pension treatment on log years of lifespan is zero, precisely estimated [$\beta = 0.0002$, $SE = 0.003$; Figure 2]. In this case, the bias from endogenous sample selection artificially reduces the magnitude of the estimated treatment effect by 99%. I apply the same procedure in the case of the community college natural experiment, imposing the NARA Numident pattern of missing mortality data, and find that endogenous sample selection artificially reduces the magnitude of the estimated treatment effect by 83% (Appendix Figure 1).

Not only does endogenous sample selection lead to bias, it also leads to incorrect standard errors. When the sample changes from the full Mothers' Pension dataset to the NARA Numident pattern of missing mortality data, the sample size falls by 57%. The calculated standard error, instead of increasing, *decreases* by more than 50% (Figure 2). Similarly, when the sample changes from the full community college dataset to the NARA Numident pattern of missing mortality data, the sample size falls by 80% but the calculated standard error *decreases* by 38% (Appendix Figure 1).

These calculated standard errors are incorrect because endogenous sample selection is changing the distribution of the regression residuals. NARA Numident consists mostly of individuals who died between 1988 and 2005. When dropping observations from the Mothers' Pension mortality dataset to match the pattern of missing mortality data in NARA Numident, the observations are dropped not at random but selectively, from the tails of the longevity distribution. The resulting treatment effect "estimate" has a very different

[3]Social Security Administration (SSA). 2007. "Numerical Identification Files (NUMIDENT), created, 1936 - 2007, documenting the period 1936 - 2007." [dataset] Washington, DC: Social Security Administration.

distribution of regression residuals, with all of the largest residuals now missing (Figure 3). Changing the distribution of residuals in this way sharply and artificially decreases the residual variance—the numerator in the standard error calculation—yielding a calculated standard error that is artificially smaller than the correct standard error by 75% (Figure 2).

The headline result[4] of the Mothers' Pension paper is itself estimated on an mortality dataset that, while more complete than many mortality datasets in the literature, is still substantially incomplete. Most of the missing mortality data, the authors reasonably assume, can be explained by individuals dying young, before Social Security mortality data became highly comprehensive in the 1970s. I show that if one imputes the missing deaths at young ages (using U.S. cohort mortality curves), doubling the sample size, the standard error *increases* by 25% (Figure 4). A reasonable conjecture is that the true standard error is larger than the reported standard error by 40% (Figure 4). The deaths at young ages generate large residuals, and their absence in the original estimation leads to an artificially small standard error. In light of this conclusion based on imputed data, the original headline result has been re-estimated using actual data for younger-age deaths, collected recently from other mortality data sources. The new sample size is twice as large as the original sample size, and the new standard error is indeed larger than the original standard error.[5]

In addition to this adjustment in the standard error, the original headline result also requires a strong assumption about the treatment effect on the individuals who are missing and assumed to have died young (Figure 5). After accounting for missing mortality data, the original Mothers' Pension mortality dataset tells us less than the paper asserts about the treatment effect on overall lifespans.[6]

There is one special case of incomplete mortality data that entirely avoids the problem of endogenous sample selection. The important consideration is whether deaths outside the window of fully observed mortality are partially missing ("censored") or com-

[4]See the abstract of Aizer et al. (2016), which corresponds to Table 4, Panel A of Aizer et al. (2016).

[5]A. Lleras-Muney, personal communication, October 21, 2024.

[6]The Mothers' Pension paper describes the headline result as "a convenient summary measure of the total effect of the program on longevity," while at the same time acknowledging that "[the specification] does not allow us to easily deal with attrition and places strong restrictions on the shape of the hazard rate" (p. 945). The paper correctly acknowledges that the estimate for survival to age 80 does not have these same strong restrictions and is therefore more credible.

In the community college natural experiment (Connolly, 2024), I give primary emphasis to the results for survival to a specific age, for which the problem of endogenous sample selection is entirely avoided. While I also report results for overall lifespans using an incomplete mortality dataset, these results have secondary emphasis. Such results, I make clear, require additional assumptions that are at best questionable.

pletely missing ("truncated"). Endogenous sample selection can be entirely avoided if all deaths outside the window of fully observed mortality are censored, and one's outcome is survival to a specific age within the window. Examples of this special case include the Mothers' Pension natural experiment (e.g., the treatment effect on survival to age 80; Table 4, Panel D, columns 1-4) and the community college natural experiment (survival to age 68; Table 4).[7] In these cases, one can fully observe the outcome—survival to an age within the mortality window—for every individual in the treatment and control groups, regardless of age at death. For this outcome, the researcher has exactly the same mortality information as she would have if she were estimating the identical regression on a complete mortality dataset.

In several prominent natural experiments, a truncated (rather than censored) mortality sample is used to estimate a treatment effect on survival to a specific age (Bedard and Deschênes, 2006; Clark and Royer, 2013; Black et al., 2015; Hollingsworth et al., 2024). I again use the Mothers' Pension natural experiment as the benchmark to assess the impact of endogenous sample selection. Dropping observations to match the pattern of missing data in the Great Migration natural experiment (Black et al., 2015), I find that endogenous sample selection leads to bias (Figure 6). The magnitude of the estimated treatment effect is artificially increased by 40% or artificially decreased by as much as 93%, depending on whether the outcome is survival to age 70, 75 or 80 (Appendix Figure 6). Endogenous sample selection artificially shrinks the standard error by as much as 20% (Figure 6).

This paper presents evidence from prominent, well-identified natural experiments, but these natural experiments are only two in number, have finite sample sizes and mortality datasets (though more complete than nearly every other paper in the literature) that are still substantially incomplete. To complement the evidence from these two natural experiments, I conduct an analysis of incomplete mortality data using U.S. state mortality curves. For example, I estimate the "effect" of living in Texas in 2019 relative to New Mexico on overall lifespans. After estimating the benchmark effect with the complete state mortality curves, I re-estimate the effect with various patterns of incomplete mortality data found in the literature and then assess the impact of incomplete mortality data.

This analysis has one important drawback—that the state-of-residence mortality effects have no causal interpretation—and three important advantages. The sample size

[7] A third example is Meghir et al. (2018).

can be made arbitrarily large, allowing one to fully distinguish between bias and sampling variability.[8] The mortality curves are complete, enabling one to assess the impact of *any* pattern of incomplete mortality data.[9] There are 50 unique state mortality curves, yielding 1,225 unique pairs of mortality curves—a wide range of (non-causal) mortality effects with which to assess the impact of missing mortality data.

I find that commonly used patterns of incomplete mortality data can result in bias several times larger in magnitude than mortality effects reported in the literature. The bias can lead not only to attenuation errors (as in Figure 2) and amplification errors (as in Appendix Figure 6, panel C), but also to wrong-sign errors. Indeed, it is possible to estimate a statistically significant mortality "effect," of sufficient magnitude to have real-world significance, only for the true effect to have the opposite sign and more than triple the magnitude.

The impact of endogenous sample selection depends a great deal on three factors: the mortality outcome (e.g., overall years of lifespan versus survival to a specified age), the type and extent of missing mortality data, and what one can reasonably assume about the effect of the treatment on mortality in unobserved age ranges. The best approach by far is to avoid the issue of endogenous sample selection entirely, by using censored mortality data and estimating effects on survival to an age within the window of fully observed mortality.

Mortality estimates that depart from this approach require both a standard error correction (as in Figure 4) and a sensitivity analysis (as in Figure 5). The sensitivity analysis should consist in first establishing a range of reasonable assumptions for the treatment effect on mortality in unobserved age ranges, and then showing how varying the assumption within this range affects the estimated overall treatment effect. In Section 2, I apply the standard error correction and sensitivity analysis for the headline result of the Mothers' Pension paper (Aizer et al., 2016).

The problem of endogenous sample selection is econometrically equivalent to the "bad control" problem.[10] Just as one cannot interpret as causal the estimated effect of

[8] One can treat each state mortality curve as the "true" mortality curve, and then sample with replacement to obtain an arbitrarily large sample size.

[9] The patterns of missing mortality data in the Mothers' Pension and community college mortality datasets preclude, for example, a direct test of the effect of the pattern of missing data in the U.K. compulsory schooling natural experiment (see Figure 1).

[10] As Angrist and Pischke (2009) note, the inclusion of a control variable that is itself (potentially) affected by the treatment means that the ostensibly causal estimate "does not have a causal interpretation" (p. 65). There is an econometric equivalence between estimating a mortality treatment effect on an endogenously

education on earnings controlling for occupation, even if one has an ideal randomized experiment with an enormous sample size, so too one cannot interpret as causal the estimated effect of any treatment on one measure of mortality, controlling for a different measure of mortality. By an econometric equivalence (Appendix A), one likewise cannot interpret as causal a mortality estimate on an endogenously selected sample.

Only a few results in the literature entirely avoid the problem of endogenous sample selection arising from missing mortality data. The remaining results cannot be taken at face value, but rather require a first-order increase in the standard error and a sensitivity analysis. Such results include all of the mortality results in the U.S. compulsory schooling natural experiment (Lleras-Muney, 2005), all of the mortality results in the military service natural experiment (Bedard & Deschenes, 2006), the 1947 set of results (but not the 1972 set of results) for U.K. compulsory schooling (Clark & Royer, 2013), the primary mortality results for the Great Migration natural experiment (Black et al., 2015), the headline result of the Mothers' Pension natural experiment (Aizer et al., 2016), the mortality results in the college access natural experiment in Chile (Gonzalez et al., 2024), and the mortality results (excluding the infant mortality results) for the modern hospitals natural experiment (Hollingsworth et al., 2024).

There is a large literature estimating the causal effect of educational attainment on older-age mortality. According to a recent literature review, "An effect of education on mortality exists in some contexts but not in others, and seems to depend on (i) gender; (ii) the labor market returns to education; (iii) the quality of education; and (iv) whether education affects the quality of individuals' peers" (Galama et al., 2018). Nearly all of the estimates in all of the papers in this literature use endogenously selected samples, and therefore require a first-order upward adjustment in the standard error and a sensitivity analysis. The results that avoid the problem of endogenous sample selection arising from missing mortality data include Meghir et al. (2018), part of Clark and Royer (2013), part of Albouy and Lequin (2009), and the primary results in Connolly (2024). For this much smaller body of results, the evidence remains mixed, with null results for the papers that estimate the causal effect of compulsory schooling, and a strong positive effect for college openings.[11]

selected mortality sample, and estimating the same effect on the complete mortality sample while including an indicator variable equal to 1 if an individual is observed in the endogenously selected sample (Appendix A). This indicator variable is the bad control.

[11]Several factors could account for the contrasting estimated effects for compulsory schooling and community college, including the level of educational attainment (high school versus college) and the difference

The rest of this paper is organized as follows. In Section 2, I show how endogenous sample selection leads to bias and incorrect standard errors. Section 3 contains results on bias resulting from incomplete mortality data in the context of U.S. state mortality curves. Section 4 is the conclusion, and in Appendix A I show how the problem of endogenous sample selection is econometrically equivalent to the "bad control" problem.

between compulsory and voluntary attainment. See Connolly (2024).

2 Bias and Incorrect Standard Errors: Evidence from Natural Experiments

Figure 1 displays patterns of missing mortality data in some of the most prominent published papers in the literature. Each mortality dataset has a window of fully observed mortality, and in every case the window covers substantially less than the full range of possible adult lifespans. The window can be the same for all birth cohorts in the mortality dataset, as in the Great Migration natural experiment (Black et al., 2015), or it can vary by birth cohort as in the Mothers' Pension natural experiment (Aizer et al., 2016). Table 1 contains a list of 54 papers in this literature.

In every natural experiment in this literature, a large proportion of the individuals in the study population die outside the window of fully observed mortality. If these individuals do not show up at all in the mortality dataset, then the mortality dataset is said to be *truncated*. Nearly every result in this literature is estimated using a truncated mortality dataset. In a few natural experiments, the individuals are partially observed: they show up in the mortality dataset but their exact dates of death are unobserved. Instead, deaths are only known only to have occurred in a certain interval.[12] This second type of missing data is known as *censoring*. This distinction between truncated and censored mortality data is of great importance for estimating valid mortality effects.

I directly test for bias arising from endogenous sample selection using two natural experiments. In the Mothers' Pension natural experiment (Aizer et al., 2016), the treatment is a cash transfer to single-mother families in the U.S. taking place between 1911 and 1935. The cash transfer typically lasted for three years and constituted 12-25% of family income (Aizer et al., 2016, p. 936). The outcome is the lifespans of male children in single-mother families, and the control group consists of comparable families whose applications for the Mothers' Pension were rejected. The mortality dataset is constructed by linking individuals from administrative records of the Mothers' Pension program to a public version of Social Security Administration death data known as the Death Master File.

In the second natural experiment, the U.S. community college natural experiment (Connolly, 2024), the treatment is access to community college created by a local community

[12]For example, in the Mothers' Pension natural experiment individuals who died before 1975 still show up in the mortality dataset, but their date of death is known only to have occurred at some point in the interval between the time they are observed in childhood and 1975.

college opening in the postwar period (1945 - 1980). Access to the newly opened college varies along two dimensions: the distance from one's childhood town to the exact location of the college, and one's age at the time of the college opening. The treatment group consists of men and women who grew up in a childhood town near a college opening, and who were college-age or younger at the time of the opening. The mortality dataset is constructed using the restricted-access version of the Social Security Administration mortality dataset, which is more complete than either of the two public versions of Social Security mortality data.

The headline result of the Mothers' Pension natural experiment is an estimated average treatment effect of one additional year of lifespan (Aizer et al., 2016, Table 4, Panel A), which I plot in the first column of Figure 2. Using the Mothers' Pension mortality dataset and the authors' replication code, I am able to exactly replicate this result (column 2). I then impose a more restrictive pattern of missing mortality data, the pattern found in the National Archives and Records Administration's Numident dataset ("NARA Numident").[13] That is, I drop individuals from the mortality dataset who died prior to 1988 or after 2005 (Appendix Figure 2). I re-estimate the headline result after imposing the NARA Numident pattern of missing mortality data (column 3 of Figure 2). The difference in the mortality samples causes the magnitude of the estimated effect to shrink by 99%. I apply the same procedure in the case of the community college openings natural experiment (Appendix Figure 1), imposing the NARA Numident pattern of missing mortality data, and find that the difference in mortality samples artificially shrinks the estimated effect by 83%.

When the sample changes from the Mothers' Pension mortality dataset to the NARA Numident pattern of missing mortality data, the sample size shrinks by 57%. Instead of increasing, however, the standard error actually *decreases* by about 50%. In column 4 of Figure 2, I take a simple random subsample of the mortality sample used for the headline result in columns 1 and 2, matching the sample size of 3,419 in the column 3 sample. Re-estimating the treatment effect on this simple random subsample and using the result as a benchmark, the NARA Numident sample restriction leads to a standard error that is smaller than the benchmark standard error by a factor of four.

Similarly, when the sample changes from the community college mortality dataset to the NARA Numident pattern of missing mortality data (Appendix Figure 1), the sample

[13]Social Security Administration (SSA). 2007. "Numerical Identification Files (NUMIDENT), created, 1936 - 2007, documenting the period 1936 - 2007." [dataset] Washington, DC: Social Security Administration.

size shrinks by 80%, while the standard error *decreases* by 38%. Compared with a benchmark with the same sample size (column 4), the NARA Numident sample restriction leads to a standard error that is smaller by a factor of three.

When the sample changes from the original Mothers' Pension mortality sample (columns 1 and 2 of Figure 2) to the NARA Numident pattern of missing mortality data (column 3), it is not only the sample size that changes, but also the distribution of the regression residuals. In Figure 3, I plot the distribution of regression residuals for the original headline result (blue curve) as well as the benchmark result (Figure 2, column 4) that uses a simple random subsample of the original sample (red curve). While the sample size falls by 57%, the distribution of the regression residuals is essentially unchanged. However, when one instead reduces the sample by imposing the NARA Numident pattern of missing mortality data (orange curve), in effect one is selectively dropping observations from the tails of the mortality distribution. The resulting distribution of residuals has a different shape and a much smaller variance. The residual variance is the numerator in the formula for the standard error, and by artificially shrinking the residual variance, the NARA Numident sample restriction artificially shrinks the standard error. Comparing columns 3 and 4 in Figure 2, the endogenous sample selection artificially shrinks the standard error by 75%.

The headline result[14] of the Mothers' Pension paper is itself estimated on a mortality dataset that, while more complete than many mortality datasets in the literature, is still substantially incomplete. A date of death is unobserved for about 50% of the study population. The primary reason for missing mortality data, the authors reasonably argue (pp. 942-43), is younger-age mortality, deaths before the point at which Social Security mortality data became comprehensive in the mid-1970s. In Appendix Figure 3, I plot age at death for the Mothers' Pension mortality dataset (blue line) and age at death for U.S. men born in 1912 (orange line). This figure suggests that the authors are correct, that relatively young deaths account for a large proportion of the missing mortality data. The mortality data is missing not at random but in a way that depends upon the outcome (age at death), raising the possibility of bias and incorrect standard errors from endogenous sample selection.

To assess the potential impact of endogenous sample selection on the standard error, I impute the missing mortality data using U.S. male cohort mortality curves.[15] In

[14]See the abstract of Aizer et al. (2016), which corresponds to Table 4, Panel A of Aizer et al. (2016).

[15]I impute dates of death in the following way. Individuals with missing dates of death are assumed to have died at some point between the age of 18 and 1975. For a given individual born in 1900, for example,

Appendix Figure 4, I plot the original mortality curve for the study population (orange line), as well as the mortality curve after imputing missing dates of death (blue curve). In Figure 4, column 2, I again replicate the headline result of the Mothers' Pension natural experiment. In column 3, I re-estimate the effect after imputing the missing dates of death. Compared with the original result, the sample size increases by 53%. The standard error, rather than decreasing, *increases* by about 30%. In column 4, I match the original sample size (N = 7,860) by taking a simple random subsample of the (column 3) sample with imputed dates of death. A comparison of columns 2 and 4 suggests that endogenous sample selection is artificially shrinking the standard error of the Mothers' Pension headline result by about 40%.

To assess the potential for bias from endogenous sample selection, I impute missing dates of death under different assumptions about the distribution of lifespans among individuals whose date of death is missing. In Appendix Figure 5, I plot the original mortality curve for the Mothers' Pension control group (blue curve), as well as three alternative control group mortality curves. I create the first alternative mortality curve by imputing missing dates of death according to the 1900 U.S. male cohort mortality curve, then do the same using the 1912 and 1925 U.S. male cohort mortality curves.[16] I then re-estimate the headline result, imputing missing deaths in the treatment and control groups according to one of these three U.S. cohort mortality curves (Figure 5). I find that the treatment effect of the Mothers' Pension on overall lifespans is sensitive to the assumptions that one can make about the missing mortality data. Specifically, the point estimates range from having the same sign and more than triple the magnitude (column 3), to having the opposite sign (column 5). To interpret the original result, therefore, one must first apply a first-order standard error correction, and then assess the sensitivity of the result to different assumptions about the distribution of the treatment effect on unobserved younger-age mortality.[17]

In several prominent natural experiments, a truncated (rather than censored) mor-

the date of death is determined using the U.S. cohort mortality curve for men born in 1900. The date of death is randomly drawn from a probability distribution corresponding to the part of the cohort mortality curve between age 18 and age 75.

[16] I have chosen the 1900, 1912, and 1925 mortality curves because the Mothers' Pension study population is a subset of men in the U.S. born between 1900 and 1925.

[17] The Mothers' Pension paper describes the headline result as "a convenient summary measure of the total effect of the program on longevity," while at the same time acknowledging that "[the specification] does not allow us to easily deal with attrition and places strong restrictions on the shape of the hazard rate" (p. 945). The paper correctly acknowledges that the estimates for survival to age 80 do not have these same strong restrictions and are therefore more credible.

tality sample is used to estimate a treatment effect on survival to a specific age (e.g., Bedard and Deschênes, 2006; Clark and Royer, 2013; Black et al., 2015; Hollingsworth et al., 2024). In Figure 6, column 1, I plot the Mothers' Pension estimated treatment effect for the outcome "survival to at least age 80," an effect that is estimated on a censored mortality sample. I am able to exactly replicate this result using the original Mothers' Pension mortality sample (column 2). In column 3, I impose a truncated pattern of missing mortality data analogous to several papers in the literature,[18] dropping all individuals from the sample who died prior to age 70. The sample size falls by 67%, and the point estimate shrinks by 92%. In column 4, I match the sample size in column 3 by taking a simple random subsample of the sample in columns 1 and 2. In this case, comparing the standard errors in columns 3 and 4, endogenous sample selection artificially shrinks the standard error by about 20%.

I make the same comparison for the outcomes of survival to age 80, 75 and 70 in Appendix Figure 6. In each case, I impose the same sample restriction—that individuals are truncated (completely missing) from the mortality sample if they died more than 10 years prior to the survival age of interest. When the outcome is survival to age 80, the point estimate is severely attenuated, falling by 92% (Panel A). When the outcome is survival to age 75, the point estimate is attenuated to a substantially smaller degree, falling by 52%. In Panel C, when the outcome is survival to age 70, the point estimate is not attenuated at all, but rather amplified by 39%. These results illustrate the difficulty (if not the impossibility) of predicting the direction and magnitude of the bias resulting from endogenous sample selection.

[18] Black et al. (2015); Johnson & Taylor (2019); Aaronson et al. (2021).

3 Bias from Incomplete Mortality Data: Evidence from U.S. States

This paper presents evidence from prominent, well-identified natural experiments on the extent to which incomplete mortality data can cause bias and incorrect standard errors. However, these natural experiments are only two in number, have finite sample sizes and mortality datasets (though more complete than nearly every other paper in the literature) that are still substantially incomplete. To complement the evidence from these two natural experiments, I conduct an analysis of missing mortality data using mortality curves for U.S. states.

The CDC collects and publishes annual mortality data for the resident population of each U.S. state (Arias et al., 2022). Consider the mortality curves in Figure 7. The Texas curve, for example, consists of a probability of death in 2019 for male Texas residents at each specific year of age, from age 18 to age 99. Known as a period mortality curve, this type of mortality curve is used to calculate life expectancy.[19] With an ideal mortality dataset,[20] one could estimate a complete mortality curve for any subpopulation of individuals, including the treatment and control groups of any natural experiment. In practice, mortality datasets for natural experiments are incomplete and have limited sample sizes. The shapes of the underlying mortality curves, though unobserved, determine the extent to which incomplete mortality data causes bias.

One starting point is to simply assume *arguendo* that mortality curves in natural experiments are similar to U.S. state mortality curves, at least in terms of the extent to which incomplete mortality data causes bias. This assumption could fail to hold because the true mortality curves in a given natural experiment are better-behaved, in the sense of causing less bias than U.S. state mortality curves. It could also fail to hold because the true mortality curves in a given natural experiment are worse-behaved, causing a greater degree of bias. This analysis of U.S. state mortality curves provides a benchmark for thinking about the bias that can arise from incomplete mortality data.

[19] A cohort mortality curve, by contrast, gives a probability of death at each age for a cohort of individuals born in a specific year. I use cohort mortality curves in Section 2 to impute missing mortality data in the Mothers' Pension natural experiment.

[20] By an ideal mortality dataset, I mean a mortality dataset that is complete and has a sufficiently large sample size.

Using the CDC state mortality curves for the year 2019, I estimate state mortality effects, such as the (non-causal) effect of living in Texas relative to New Mexico on overall lifespans. After estimating the benchmark effect with the complete state mortality curves, I re-estimate the effect with various patterns of incomplete mortality data found in the literature and then assess the impact of incomplete mortality data. If the mortality effect based on a given pattern of incomplete mortality data is never biased to any substantial degree, such results could be taken as (suggestive) evidence that the bias caused by that pattern of incomplete mortality data is small in practice.[21]

Figure 7 is a plot of 2019 male mortality curves for four U.S. states.[22] These mortality curves are estimated using the full 2019 mortality data for each state, a far larger sample size than is typically available in a natural experiment. The four mortality curves all have a similar general shape. At the same time, each mortality curve has important deviations from the general shape, deviations that could lead to biased mortality estimates when using incomplete mortality data.

In Figure 8 I plot the survival rates from age 18 to age 80—a key outcome in the Mothers' Pension natural experiment—for the four states in Figure 7. The survival rates in the left panel are based on incomplete mortality data. Specifically, the mortality dataset is left-truncated at age 70, meaning individuals who died before age 70 do not show up at all in the mortality dataset. The survival rate to age 80 for male Texas residents in 2019 is 64.9%. Compared with Texas, the survival rates for the other three states are greater by 2.2-2.5 percentage points.

The right panel of Figure 8 is identical to the left panel, except that the survival rates are now based on a complete mortality dataset. The results look very different. The survival rate to age 80 in Texas now exceeds that of New Mexico by 1.7 percentage points. The apparent mortality effect (left panel of Figure 8) of living in New Mexico relative to Texas is a wrong-sign error. The magnitude of the bias is 4.2 percentage points. For Idaho relative to Texas, the apparent mortality effect is an attenuation error, meaning the sign of

[21]The results from Section 2 suggest that there is substantial bias from incomplete mortality data. Once the necessary standard error correction is made, however, sample size limitations prevent one from fully distinguishing between bias and sampling variability.

[22]I have chosen these four specific states for the purposes of illustrating (in Figure 8) the three different types of bias that can arise from incomplete mortality data. I later show results for every state. I have chosen the year 2019 because it is the most recent year before the Covid-19 pandemic. The general conclusions about bias from state mortality curves are unchanged if one uses mortality curves for women instead of men, or a calendar year other than 2019.

the mortality effect is correct but the magnitude is biased (by 2.0 percentage points or 44%) towards zero. For Alaska relative to Texas, the apparent mortality effect is an amplification error, meaning the sign is correct but the magnitude is biased (by 0.4 percentage points, or 22%) away from zero. For all six possible pairs of these four states, the apparent mortality effect is biased to a substantial degree.

In Figure 9, I plot state-of-residence mortality effects for all 1,225 possible pairs of U.S. states. The outcome is the log odds of survival from age 18 to age 80, one of the key outcomes in the Mothers' Pension natural experiment. Each point is a unique state pair, with the x-coordinate equal to the mortality effect that one would obtain with a complete mortality dataset and an infinitely large sample. The y-coordinate is the same mortality effect, except the mortality dataset is left-truncated at age 70, meaning anyone who died before age 70 is completely missing from the mortality dataset. The bias from missing mortality data is the vertical distance from the 45-degree line. The light-gray box contains all the state-of-residence effects for which the true mortality effect (i.e., x-coordinate) falls within the 95% confidence interval found in the Mothers' Pension natural experiment. The dark-gray box contains all the mortality effects for which both the true mortality effect (x-coordinate) and apparent mortality effect (y-coordinate) fall within this confidence interval.

In nearly every instance the incomplete data in Figure 9 causes bias to at least some degree. When the true mortality effect (x-coordinate) in Figure 9 is greater than 0.3 in absolute value, the type of bias is always attenuation bias. On the other hand, when the true effect is between zero and 0.2 in absolute value, there is a substantial probability of both amplification errors and wrong sign errors. Likewise, when the apparent mortality effect (y-coordinate) exceeds 0.4 in absolute value, the type of bias becomes exclusively attenuation bias. There is a substantial probability of amplification bias and wrong-sign errors when the magnitude of the apparent effect is less than 0.2 in absolute value. In the literature of early-life natural experiments on older-age mortality, the reported magnitudes of reduced-form mortality effects typically have an absolute value that is less than 0.2.

The standard error of the Mothers' Pension natural experiment is large relative to most papers in the literature. If the sample size of the Mothers' Pension natural experiment were larger, such that the standard error fell by half, both the light-gray box and the dark-gray box would shrink by half. At the same time, the locations of the points in the Figure 7 scatterplot (being infinite-sample results) would not change. Accordingly, as the sample

size increases there is a *decreasing* likelihood that the confidence interval for the apparent mortality effect contains the true mortality effect.

Figure 10 is identical to Figure 9 except that the y-coordinate now is based on a mortality dataset that is left-censored (rather than left-truncated) at age 70. Individuals who died before age 70 now show up in the mortality dataset, but their dates of death are not known exactly. All one can observe is that such individuals died at some point between age 18 and age 69. The mortality information lost from left-censoring at age 70 is important for some outcomes, such as overall lifespans. In this special case, however, when the outcome is survival to age 80, the mortality information lost from left-censoring at age 70 is completely irrelevant. For all 1,225 state pairs, the true mortality effect (x-coordinate) is exactly equal to the apparent mortality effect (y-coordinate). For this special case the bias from incomplete mortality data is necessarily zero.

In Figure 11, the outcome is now the logarithm of overall lifespans, the outcome used for the headline result of the Mothers' Pension natural experiment. The apparent mortality effect (x-coordinate) is based on a mortality dataset that is left-truncated at age 63, approximating the Mothers' Pension pattern of incomplete mortality data.[23] The true mortality effect (y-coordinate) is again based on complete U.S. state mortality data. The apparent mortality effect is almost always biased to a substantial degree. When the apparent mortality effect (x-coordinate) is greater than about 0.3 in absolute value, the type of bias is exclusively attenuation bias. When less than 0.3 in absolute value, there is a substantial probability of amplification bias, and when less than 0.2, a substantial probability of a wrong-sign error. Reduced-form mortality effects on overall lifespans reported in the literature seldom exceed 0.2 in absolute value, and usually fall below 0.1 in absolute value.

In Figure 11, as the magnitude of the apparent mortality effect (x-coordinate) approaches zero, it becomes largely uninformative about the value of the true mortality effect (y-coordinate). For example, suppose one learns that the apparent mortality effect (x-coordinate) is positive and between 0.0 and 0.005 in magnitude. In that interval of apparent mortality effects, the true mortality effect(y-coordinate) ranges from -0.45 to +0.5.

[23]The age of truncation varies by birth cohort, but for the median birth cohort (1912) the age of truncation is 63. The Mothers' Pension dataset is truncated when the outcome is overall lifespans, but censored when the outcome is survival to age 80. The reason for this distinction is that the individuals whose date of death is missing and assumed to have died young are dropped from the mortality dataset when estimating the effect on overall lifespans (N = 7,860), but they are included and assumed to have died young when the outcome is survival to age 80 (N = 16,069; Table 4, Aizer et al., 2016).

I repeat the analysis in Figure 11 for patterns of incomplete mortality data from two other prominent natural experiments. In Appendix Figure 8, I use the pattern of incomplete mortality data found in the 1947 subset of results of the U.K. compulsory schooling natural experiment (Clark & Royer, 2013). The evidence suggests that incomplete mortality data could cause bias that is an order of magnitude larger than the reported standard error. In Appendix Figure 9, I use the pattern of missing mortality data found in the primary results of the Great Migration natural experiment (Black et al., 2015). Again, the results suggest that the bias can be larger than the reported standard error by an order of magnitude.

4 Conclusion

In this paper I assess the extent to which incomplete mortality data can cause bias and incorrect standard errors when estimating treatment effects on older-age mortality. Using two prominent natural experiments, I find evidence of substantial bias arising from common patterns of incomplete mortality data. I also find that incomplete mortality data can artificially shrink the standard error. A first-order upward adjustment of the standard error is required when interpreting causal estimates based on incomplete mortality datasets, therefore, as well as an analysis of the sensitivity of the causal estimate to reasonable assumptions about unobserved mortality. After applying the standard-error correction and sensitivity analysis, mortality estimates based on incomplete mortality data become substantially less informative. Both problems arising from incomplete mortality data can be avoided in one special case—when the mortality dataset is censored, rather than truncated, and the outcome is survival to a specific age within the window of fully observed mortality. When possible, therefore, researchers should estimate mortality effects using censored mortality datasets.

References

Aaronson, D., Mazumder, B., Sanders, S. G., & Taylor, E. J. (2021). Estimating the Effect of School Quality on Mortality in the Presence of Migration: Evidence from the Jim Crow South. *Journal of Labor Economics*, *39*(2), 527–558.

Aizer, A., Eli, S., Ferrie, J., & Lleras-Muney, A. (2016). The Long-Run Impact of Cash Transfers to Poor Families. *American Economic Review*, *106*(4), 935–971.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Arias, E., & Xu, J. (2022). United States Life Tables, 2019. *National Vital Statistics Reports*, *70*(19).

Atherwood, S. (2022). Does a Prolonged Hardship Reduce Life Span? Examining the Longevity of Young Men who Lived Through the 1930s Great Plains Drought. *Population and Environment*, *43*(4), 530–552.

Bedard, K., & Deschênes, O. (2006). The Long-Term Impact of Military Service on Health: Evidence from World War II and Korean War Veterans. *American Economic Review*, *96*(1), 176–194.

Black, D. A., Hsu, Y.-C., & Taylor, L. J. (2015). The Effect of Early-Life Education on Later-Life Mortality. *Journal of Health Economics*, *44*, 1–9.

Black, D. A., Sanders, S. G., Taylor, E. J., & Taylor, L. J. (2015). The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South. *American Economic Review*, *105*(2), 477–503.

Breen, C. F. (2024). The Longevity Benefits of Homeownership: Evidence from Early 20th-Century US Male Birth Cohorts.

Buckles, K., Hagemann, A., Malamud, O., Morrill, M., Wozniak, A., et al. (2016). The Effect of College Education on Mortality. *Journal of Health Economics*, *50*(100), 99–114.

Clark, D., & Royer, H. (2013). The Effect of Education on Adult Mortality and Health: Evidence from Britain. *American Economic Review*, *103*(6), 2087–2120.

Connolly, K. (2024). The Effects of U.S. Community Colleges on Educational Attainment and Longevity [*Unpublished Manuscript*].

Coppock, A., Gerber, A. S., Green, D. P., & Kern, H. L. (2017). Combining Double Sampling and Bounds to Address Nonignorable Missing Outcomes in Randomized Experiments. *Political Analysis*, *25*(2), 188–206.

Cowan, B. W., & Tefft, N. (2024). College Access and Adult Health. *American Journal of Health Economics*.

Davies, N. M., Dickson, M., Davey Smith, G., Van Den Berg, G. J., & Windmeijer, F. (2018). The Causal Effects of Education on Health Outcomes in the UK Biobank. *Nature Human Behaviour*, *2*(2), 117–125.

Davies, N. M., Dickson, M., Davey Smith, G., Windmeijer, F., & Van Den Berg, G. J. (2023). The Causal Effects of Education on Adult Health, Mortality and Income: Evidence from Mendelian Randomization and the Raising of the School Leaving Age. *International Journal of Epidemiology*, *52*(6), 1878–1886.

Fletcher, J., & Noghanibehambari, H. (2023). Toxified to the Bone: Early-Life and Childhood Exposure to Lead and Men's Old-Age Mortality. Working Paper 31957. *National Bureau of Economic Research*.

Fletcher, J., & Noghanibehambari, H. (2024a). The Effects of Education on Mortality: Evidence Using College Expansions. *Health Economics*, *33*(3), 541–575.

Fletcher, J., & Noghanibehambari, H. (2024b). The Siren Song of Cicadas: Early-Life Pesticide Exposure and Later-Life Male Mortality. *Journal of Environmental Economics and Management*, *123*, 102903.

Frisvold, D., & Golberstein, E. (2011). School Quality and the Education–Health Relationship: Evidence from Blacks in Segregated Schools. *Journal of Health Economics*, *30*(6), 1232–1245.

Frisvold, D., & Golberstein, E. (2013). The Effect of School Quality on Black-White Health Differences: Evidence from Segregated Southern Schools. *Demography*, *50*(6), 1989–2012.

Galama, T. J., Lleras-Muney, A., & Van Kippersluis, H. (2018). The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence. *Oxford Research Encyclopedia of Economics and Finance*.

Glied, S., & Lleras-Muney, A. (2008). Technological Innovation and Inequality in Health. *Demography*, *45*(3), 741–761.

González, F., Martínez, L. R., Muñoz, P., & Prem, M. (2023). Higher Education and Mortality: Legacies of an Authoritarian College Contraction. *Journal of the European Economic Association*.

Halpern-Manners, A., Helgertz, J., Warren, J. R., & Roberts, E. (2020). The Effects of Education on Mortality: Evidence from Linked US Census and Administrative Mortality Data. *Demography*, *57*(4), 1513–1541.

Helgertz, J., & Warren, J. R. (2023). Early Life Exposure to Cigarette Smoking and Adult and Old-Age Male Mortality. *Demographic Research*, *49*, 651–692.

Hollingsworth, A., Karbownik, K., Thomasson, M. A., & Wray, A. (2024). The Gift of a Lifetime: The Hospital, Modern Medicine, and Mortality. *American Economic Review*, *114*(7), 2201–2238.

Huber, M. (2012). Identification of Average Treatment Effects in Social Experiments under Alternative Forms of Attrition. *Journal of Educational and Behavioral Statistics*, *37*(3), 443–474.

Jacks, D. S., Pendakur, K., Shigeoka, H., & Wray, A. (2024). Later-life Mortality and the Repeal of Federal Prohibition [*Forthcoming*]. *Journal of Public Economics*.

Karger, E. (2021). The Long-Run Effect of Public Libraries on Children: Evidence from the Early 1900s. *SocArXiv. Doi*, *10*.

Kocks, G. (2023). School Desegregation and Long-Run Health.

Lacroix, G., Laliberté-Auger, F., Michaud, P.-C., & Parent, D. (2021). The Effect of College Education on Health and Mortality: Evidence from Canada. *Health Economics*, *30*, 105–118.

Lillard, D. R., & Molloy, E. (2010). Live and Learn or Learn and Live: Does Education Lead to Longer Lives. *Available at SSRN 1880767*.

Lleras-Muney, A. (2005). The Relationship Between Education and Adult Mortality in the United States. *The Review of Economic Studies*, *72*(250), 189.

Malamud, O., Mitrut, A., & Pop-Eleches, C. (2023). The Effect of Education on Mortality and Health: Evidence from a Schooling Expansion in Romania. *Journal of Human Resources*, *58*(2), 561–592.

Manski, C. F. (1989). Anatomy of the Selection Problem. *Journal of Human Resources*, 343–360.

Mazumder, B. (2008). Does Education Improve Health? A Reexamination of the Evidence from Compulsory Schooling Laws. *Economic Perspectives*, *32*(2).

Meghir, C., Palme, M., & Simeonova, E. (2018). Education and Mortality: Evidence from a Social Experiment. *American Economic Journal: Applied Economics*, *10*(2), 234–256.

Moorthy, A., Figinski, T. F., & Lloro, A. (2022). Revisiting the Effect of Education on Later Life Health. *FEDS Working Paper*.

Noghanibehambari, H. (2022). In Utero Exposure to Natural Disasters and Later-Life Mortality: Evidence from Earthquakes in the Early Twentieth Century. *Social Science & Medicine*, *307*, 115189.

Noghanibehambari, H., & Engelman, M. (2022). Social Insurance Programs and Later-Life Mortality: Evidence from New Deal Relief Spending. *Journal of Health Economics*, *86*, 102690.

Noghanibehambari, H., & Fletcher, J. (2023a). Childhood Exposure to Birth Registration Laws and Old-Age Mortality. *Health Economics*, *32*(3), 735–743.

Noghanibehambari, H., & Fletcher, J. (2023b). Early-Life Income Shocks and Old-Age Mortality: Evidence from World War I Veterans' Bonus. *Available at SSRN 4499916*.

Noghanibehambari, H., & Fletcher, J. (2023c). In Utero and Childhood Exposure to Alcohol and Old Age Mortality: Evidence from the Temperance Movement in the US. *Economics & Human Biology*, *50*, 101276.

Noghanibehambari, H., & Fletcher, J. (2023d). Long-Term Health Benefits of Occupational Licensing: Evidence from Midwifery Laws. *Journal of Health Economics*, *92*, 102807.

Noghanibehambari, H., & Fletcher, J. (2023e). The Early Bird Catches the Worm: The Effect of Birth Order on Old-Age Mortality. *Population and Development Review*, *49*(3), 531–560.

Noghanibehambari, H., & Fletcher, J. (2024a). Dust to Feed, Dust to Gray: The Effect of in Utero Exposure to the Dust Bowl on Old-Age Longevity. *Demography*, *61*(1), 87–113.

Noghanibehambari, H., & Fletcher, J. (2024b). Unequal Before Death: The Effect of Paternal Education on Children's Old-Age Mortality in the United States. *Population Studies*, 1–27.

Noghanibehambari, H., Fletcher, J., Schmitz, L., Duque, V., & Gawai, V. (2024). Early-Life Economic Conditions and Old-Age Male Mortality: Evidence from Historical County-Level Bank Deposit Data. *Journal of Population Economics*, *37*(1), 1–33.

Noghanibehambari, H., & Noghani, F. (2023). Long-Run Intergenerational Health Benefits of Women Empowerment: Evidence from Suffrage Movements in the US. *Health Economics*, *32*(11), 2583–2631.

Salm, M. (2011). The Effect of Pensions on Longevity: Evidence from Union Army Veterans. *The Economic Journal*, *121*(552), 595–619.

Sanchez, Y. (2010). The Longevity Gains of Education. *Manuscript, Univ. Chicago, Dept. Econ.*

Sansani, S. (2011). The Effects of School Quality on Long-Term Health. *Economics of Education Review*, *30*(6), 1320–1333.

Savelyev, P. A., Ward, B. C., Krueger, R. F., & McGue, M. (2022). Health Endowments, Schooling Allocation in the Family, and Longevity: Evidence from US Twins. *Journal of Health Economics*, *81*, 102554.

Seltzer, N. (2021). The Intergenerational Consequences of Climate Change on Later Life Mortality: Evidence from Individual-Level US Death Records.

Taylor, S. (2020). Growing Up in a Steel Town: Early-Life Pollution Exposure and Later-Life Mortality. *Social Science History*, *44*(1), 117–142.

Vu, H., Noghanibehambari, H., Fletcher, J., & Green, T. (2023). Prenatal Exposure to Racial Violence and Later Life Mortality among Males: Evidence from Lynching. *Available at SSRN 4613372*.

Appendix A: Endogenous Sample Selection is Econometrically Equivalent to the Bad Control Problem

Consider an ideal randomized experiment for some early-life treatment, with the outcome $Y = \textit{Lifespan in Years}$. One can estimate a treatment effect β_1 using the following regression equation:

$$Y_i = \beta_1 Treat_i + \epsilon_i$$

One can estimate separate effects for demographic groups:

$$Y_i = \beta_1 Treat_i \times Female_i + \underbrace{\beta_2 Treat_i}_{\substack{\text{treat. effect} \\ \text{for men}}} + \beta_3 Female_i + \epsilon_i$$

One can likewise estimate separate effects for groups defined by date of death ("DOD"):

$$Y_i = \beta_1 Treat_i \times 1(DOD \notin [1988, 2005])_i + \underbrace{\beta_2 Treat_i}_{\substack{\text{"treat. effect"} \\ \text{for DOD}\in \\ [1988, 2005]}} + \underbrace{\beta_3 1(DOD \notin [1988, 2005])_i}_{\text{Bad Control}} + \epsilon_i$$

The variable $1(DOD \notin [1988, 2005])_i$ is a "bad control": a variable that is itself an outcome in the experiment. The estimand β_2 therefore "does not have a causal interpretation" (Angrist & Pischke, 2009, p. 65). At the same time, estimating β_2 in the above equation is econometrically equivalent to estimating α_1 (below) if the sample is restricted to individuals with a date of death between 1988 and 2005:

$$Y_i = \alpha_1 Treat_i + \epsilon_i$$

Under this sample restriction, therefore, α_1 does not have a causal interpretation.

Patterns of Missing Mortality Data in Published Economics Literature

Figure 1

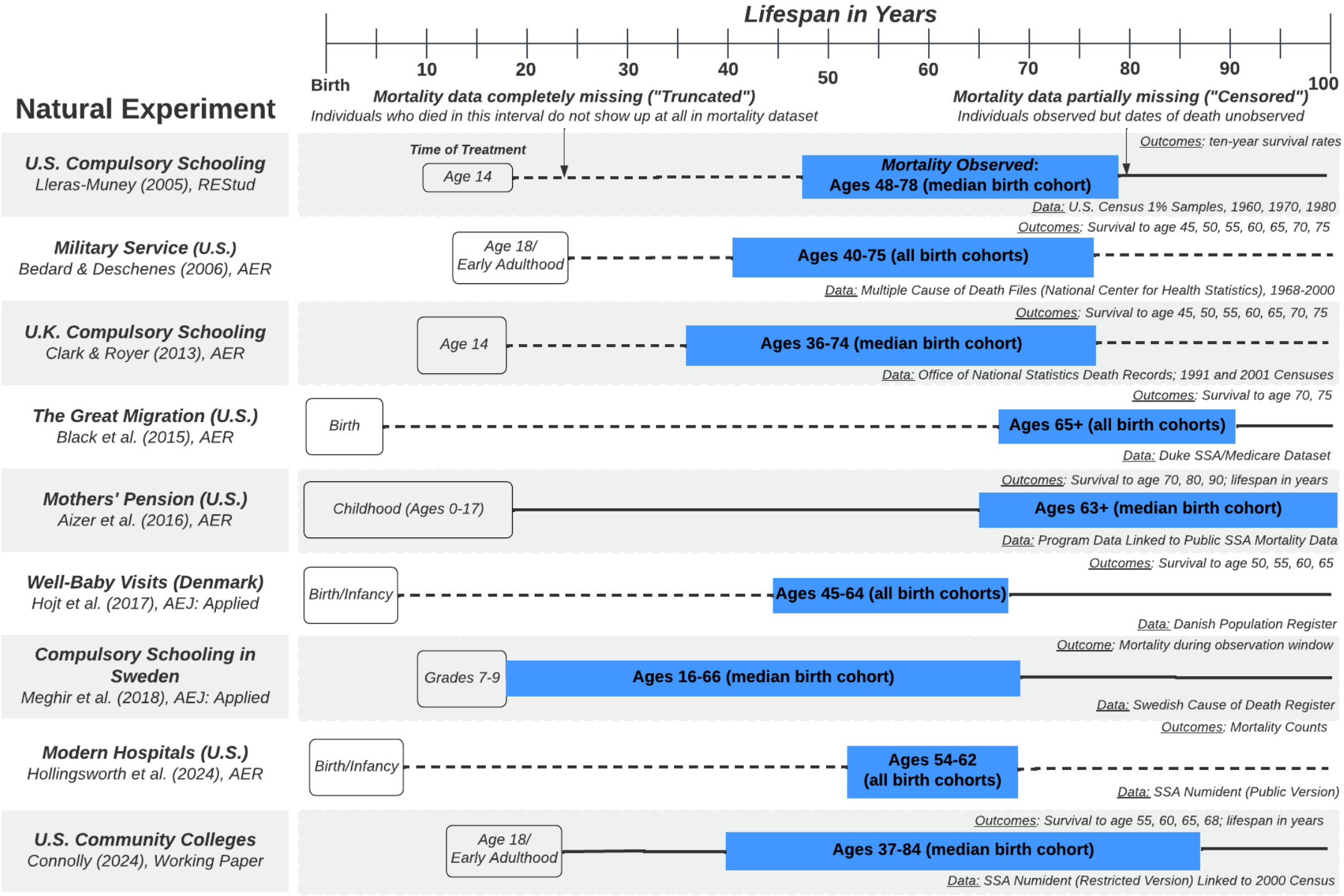
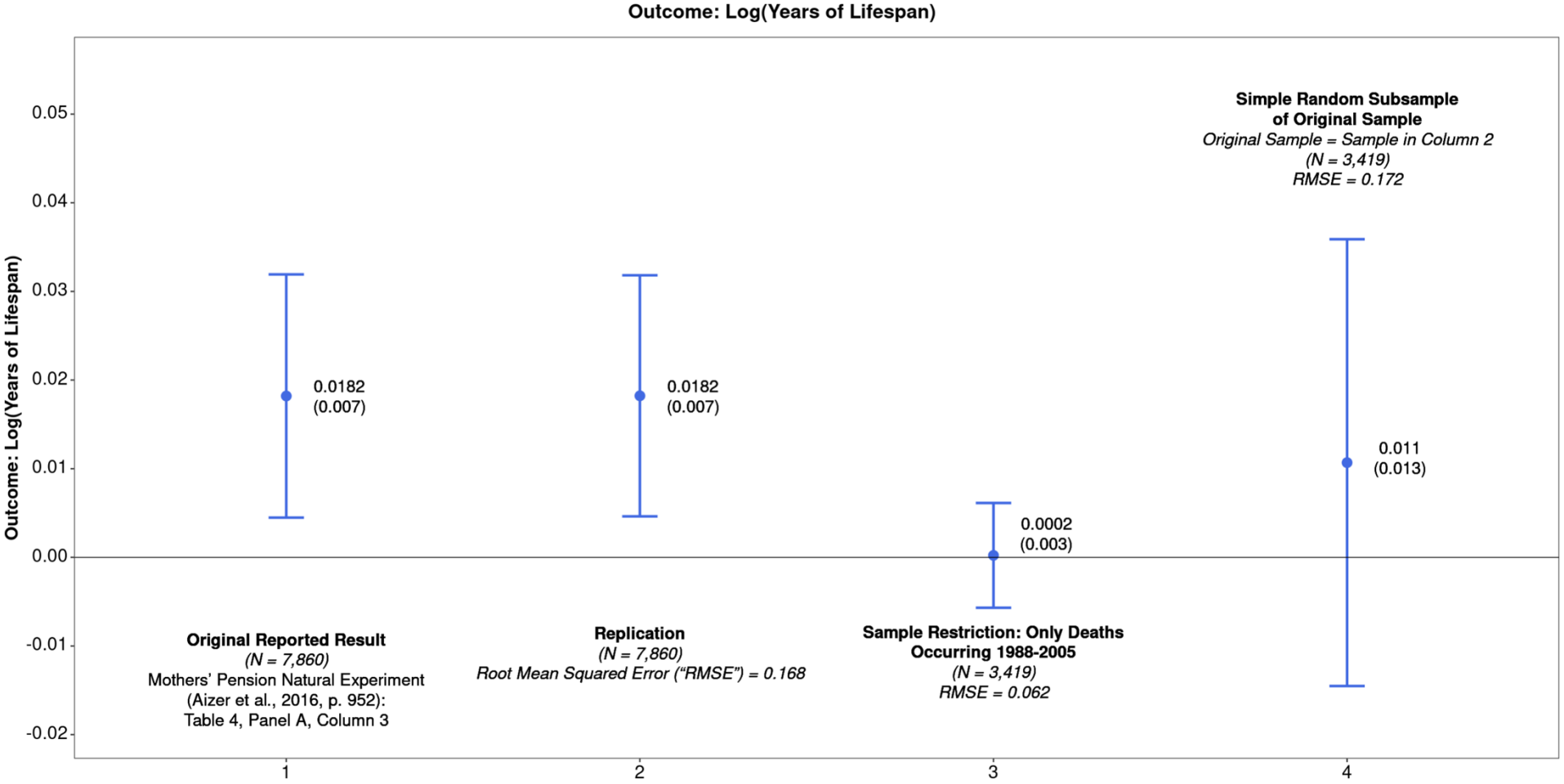


Figure 2

Endogenous Sample Selection Can Cause Problems for Estimating Mortality Effects

The estimated treatment effect of the Mothers' Pension on longevity changes when the sample is restricted to individuals who died between 1988 and 2005, suggesting that the latter estimate is biased with respect to the estimand of interest. The sample restriction also artifically shrinks the standard error.



Notes: The only difference between the estimating equations in columns 2 and 3 is the sample, which in column 3 includes only individuals whose year of death falls between 1988 and 2005, as in Fletcher & Noghanharambari (2024) and papers that use the NARA Numident dataset (e.g., Hollingsworth et al., 2024). Aizer et al. (2016). 'The Long-Run Impact of Cash Transfers to Poor Familes.' American Economic Review, 106(4), 935-971.

Figure 3

Endogenous Sample Selection Can Change the Distribution of Residuals, Leading to Incorrect Standard Errors

Plot of Residuals from Mortality Regressions

Outcome = Log(Lifespan in Years)

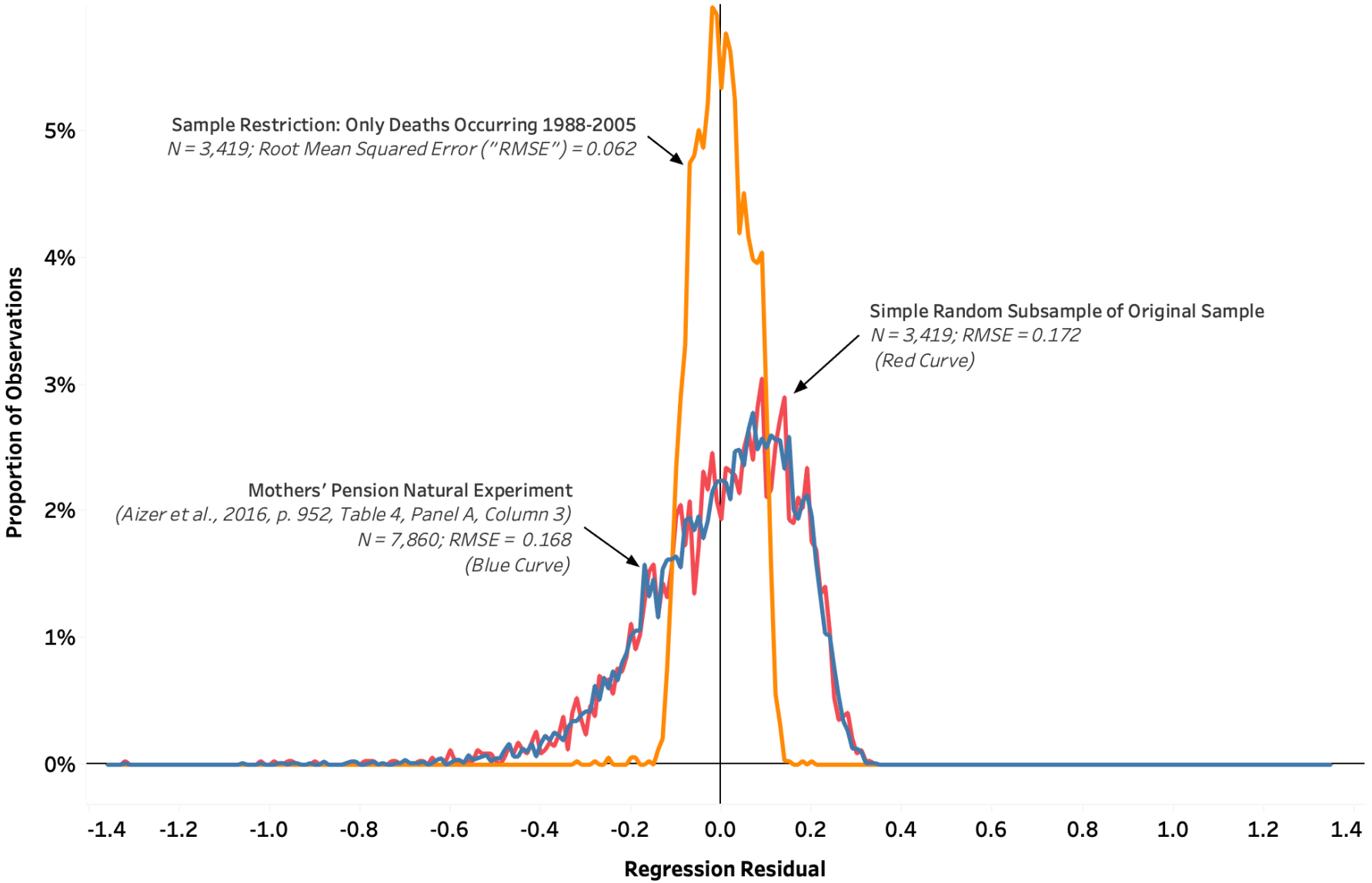
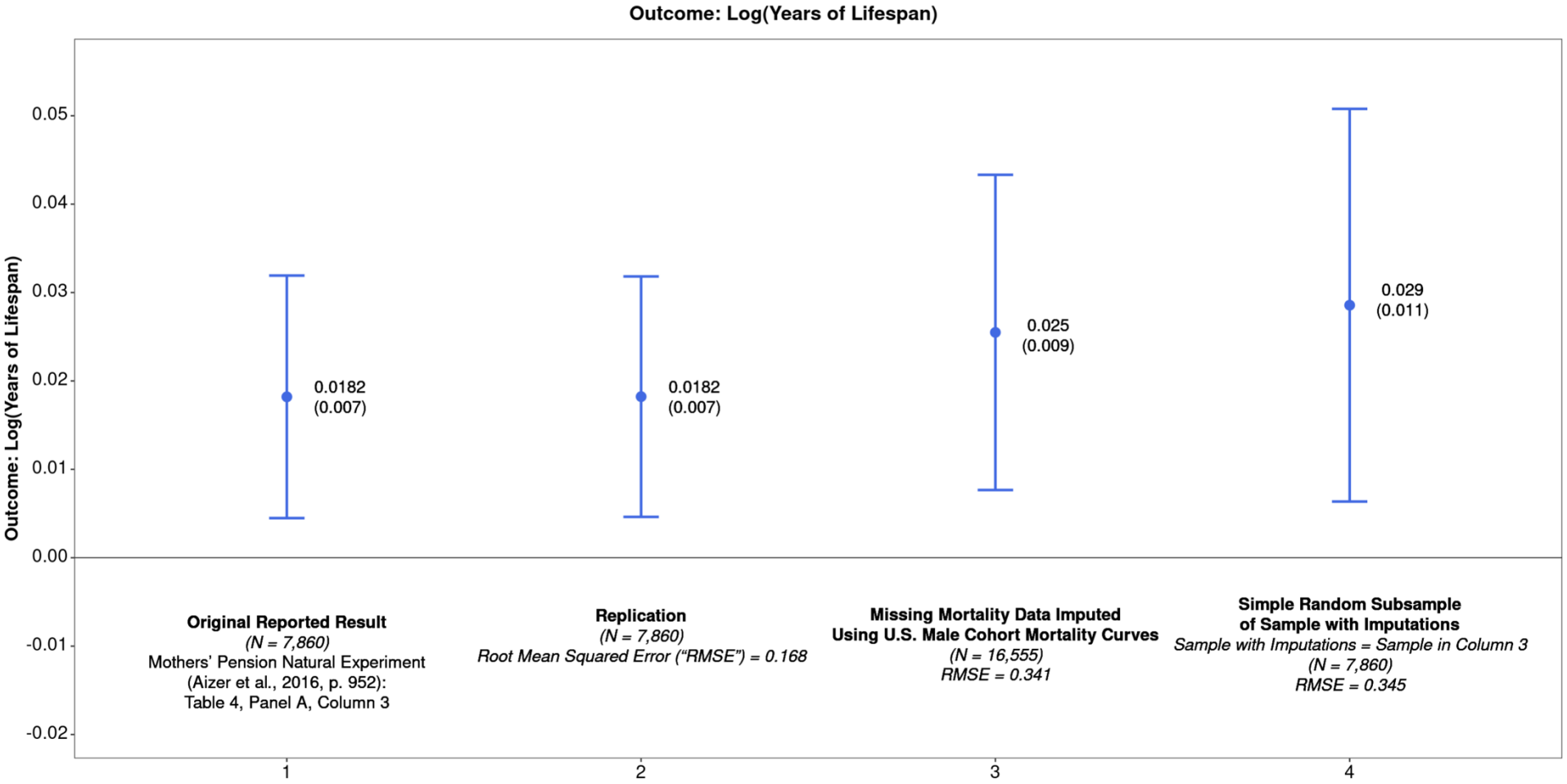


Figure 4

Endogenous Sample Selection Can Cause Problems for Estimating Mortality Effects

*If one imputes the missing left-tail mortality data, the sample size doubles but the standard error **increases** by 30%. That is, the missing left-tail mortality data artificially shrink the standard error.*

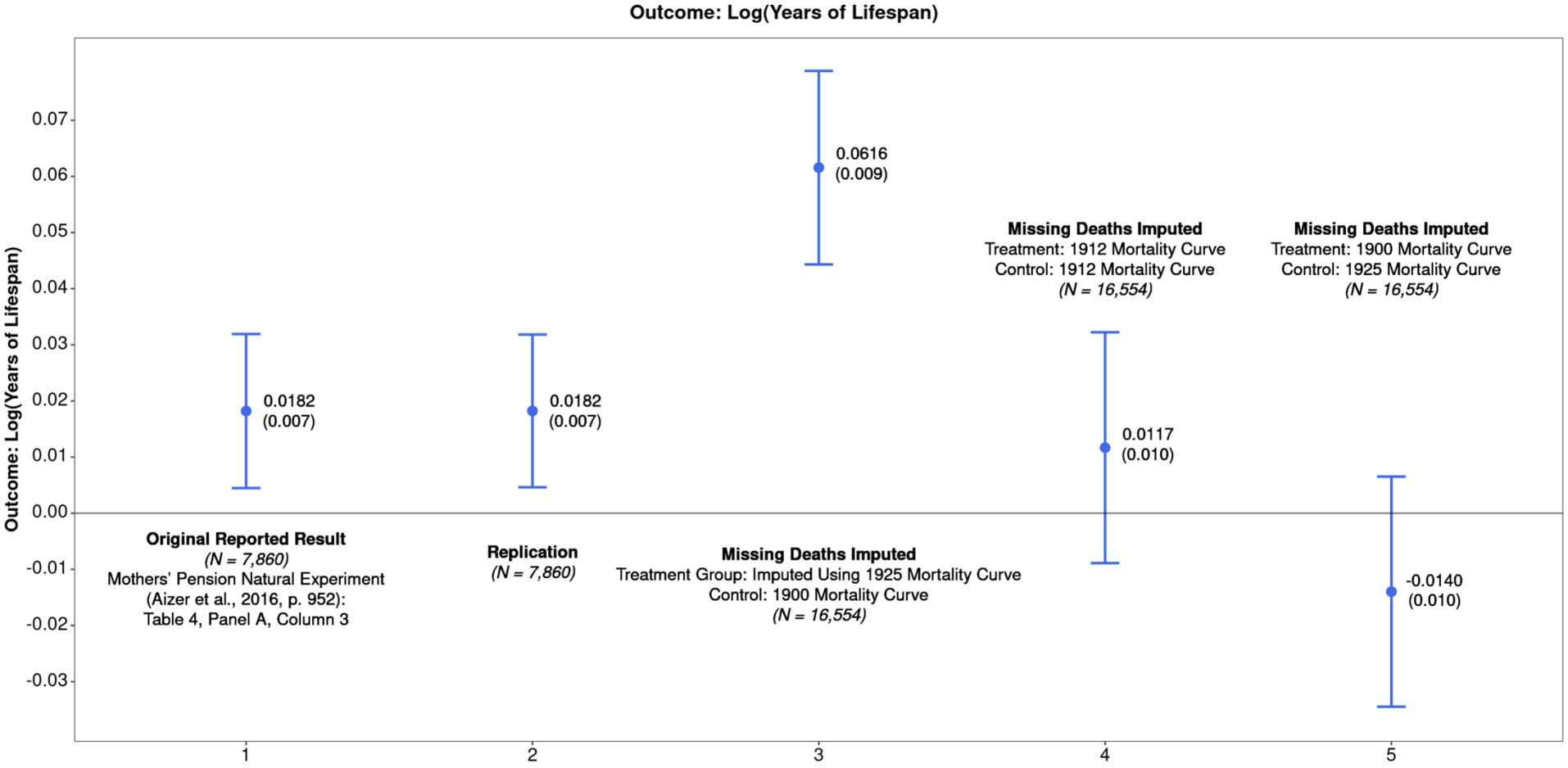


Notes: For the sample in column 3, individuals with missing dates of death are randomly assigned a year of death according to U.S. male cohort mortality curves, conditional on year of birth, surviving to age 18, and dying before the year 1975 (the year 1975 is when Social Security mortality data becomes highly comprehensive in the publicly available Death Master File).
Aizer et al. (2016). 'The Long-Run Impact of Cash Transfers to Poor Familes.' American Economic Review, 106(4), 935-971.

Figure 5

Treatment Effect of Mothers' Pension on Overall Lifespans is Sensitive to Assumptions About Missing Left-Tail Mortality Data

In estimations 3-5, below, I impute missing mortality data for individuals who are assumed to have died young (i.e., prior to 1975). The estimated effect of the Mothers Pension program on overall lifespans is sensitive to what one assumes about the distribution of missing left-tail mortality data.

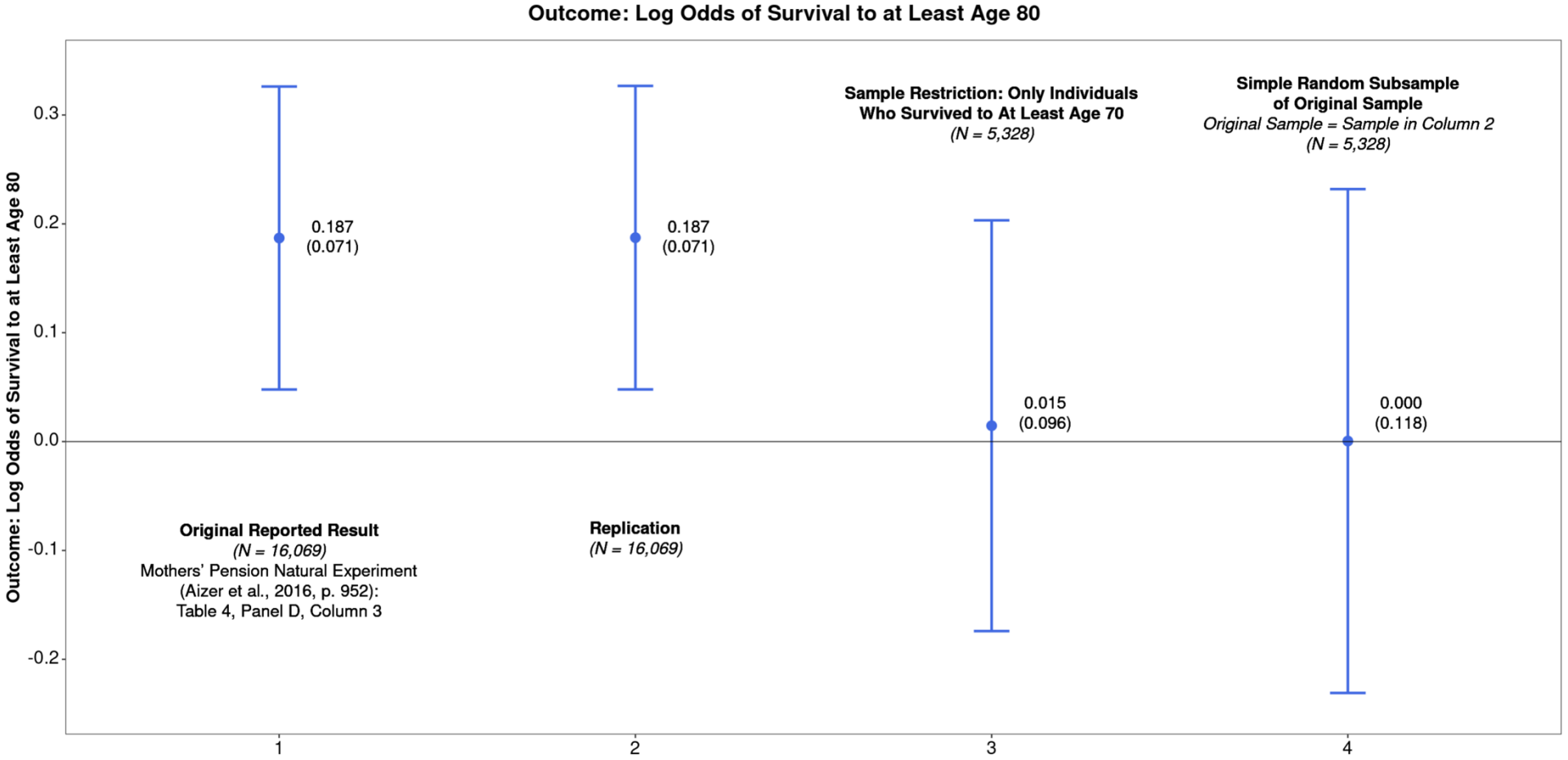


Notes: This figure plots point estimates and 95% confidence intervals. For the imputations of missing left-tail mortality data, age at death is randomly drawn from the U.S. male cohort mortality curve for the indicated birth cohort (e.g., 1912 mortality curve). Consider an individual born in 1910 whose date of death is missing. Using the 1912 U.S. male cohort mortality curve for imputation, the date of death is randomly drawn from the part of the 1912 mortality curve between ages 18 and 65. As the authors note, deaths occurring after 1975 were likely to have been observed in the Mothers' Pension mortality data.
Aizer et al. (2016). "The Long-Run Impact of Cash Transfers to Poor Familes." American Economic Review, 106(4), 935-971.

Figure 6

Endogenous Sample Selection Can Cause Problems for Estimating Mortality Effects

The estimated treatment effect of the Mothers' Pension on survival to (at least) age 80 changes when the sample is restricted to individuals who survived to at least age 70, suggesting that the latter estimate is biased with respect to the estimand of interest. The sample restriction also artifically shrinks the standard error by about 20%.



Notes: The only difference between the estimating equations in columns 2 and 3 is the sample, which in column 3 includes only individuals who suvived to at least age 70. This pattern of missing mortality data is analogous to Black et al. (2015) and Aaronson et al. (2021), which estimate effects of early-life natural experiments on survival to age 75 conditional on survival to age 65.
Aizer et al. (2016). 'The Long-Run Impact of Cash Transfers to Poor Familes.' American Economic Review, 106(4), 935-971.

Figure 7

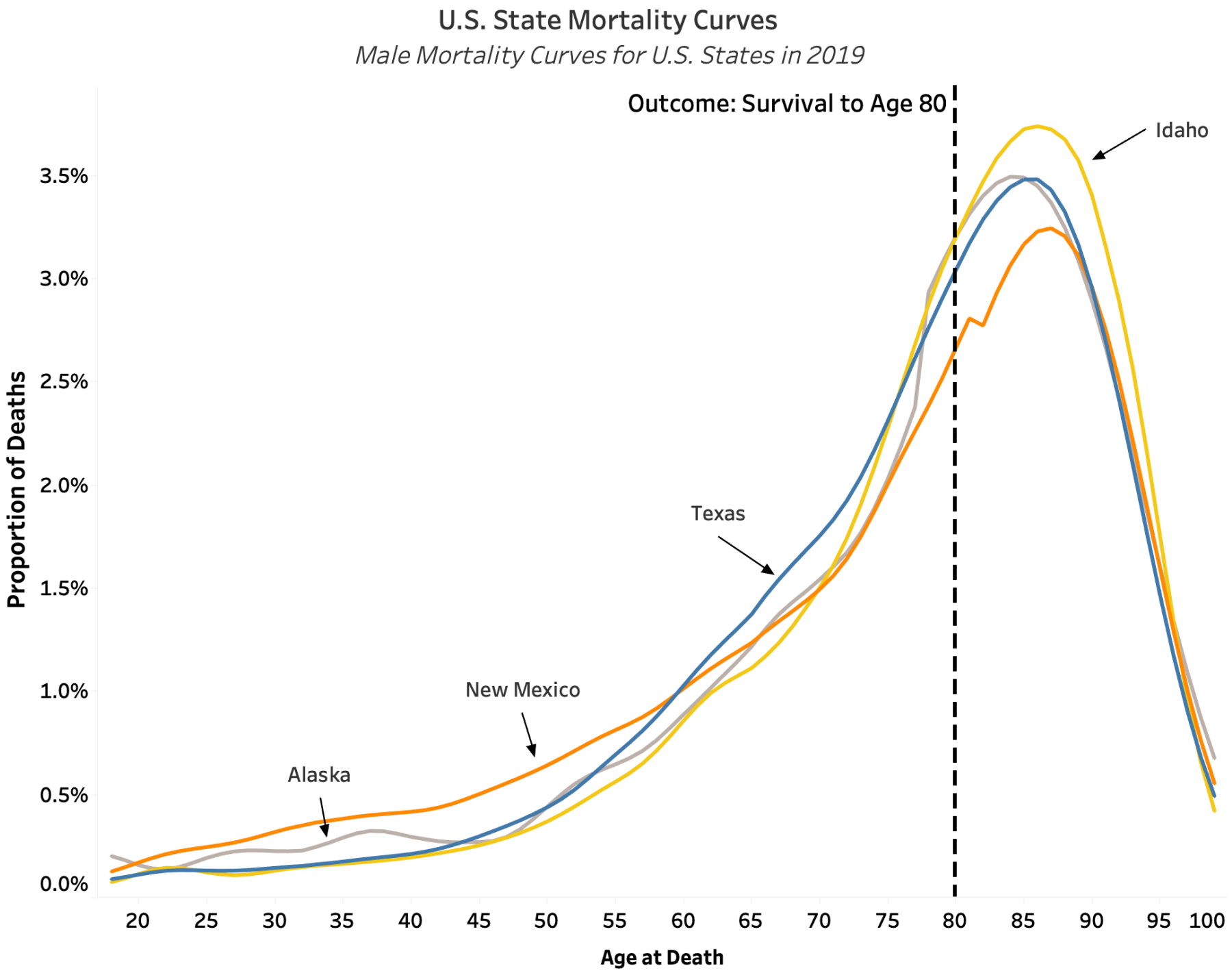

U.S. State Mortality Curves
Male Mortality Curves for U.S. States in 2019
Outcome: Survival to Age 80
Idaho
Texas
New Mexico
Alaska
Proportion of Deaths
3.5%
3.0%
2.5%
2.0%
1.5%
1.0%
0.5%
0.0%
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
Age at Death

Figure 8

Incomplete Mortality Data Leads to Biased Estimates of Survival Rates

If one uses a mortality dataset in which deaths before age 70 are completely missing (left panel), the rates of survival from age 18 to age 80 are similar for the states of Idaho, Alska and New Mexico. With complete mortality data from age 18 through 100 (right panel), there is a sharp divergence for these states in the observed survival rates from age 18 to age 80. If one were estimating effects relative to Texas, then bias from missing mortality data would lead to a wrong-sign error (New Mexico), an attenuation error (Idaho), and an amplification error (Alaska).

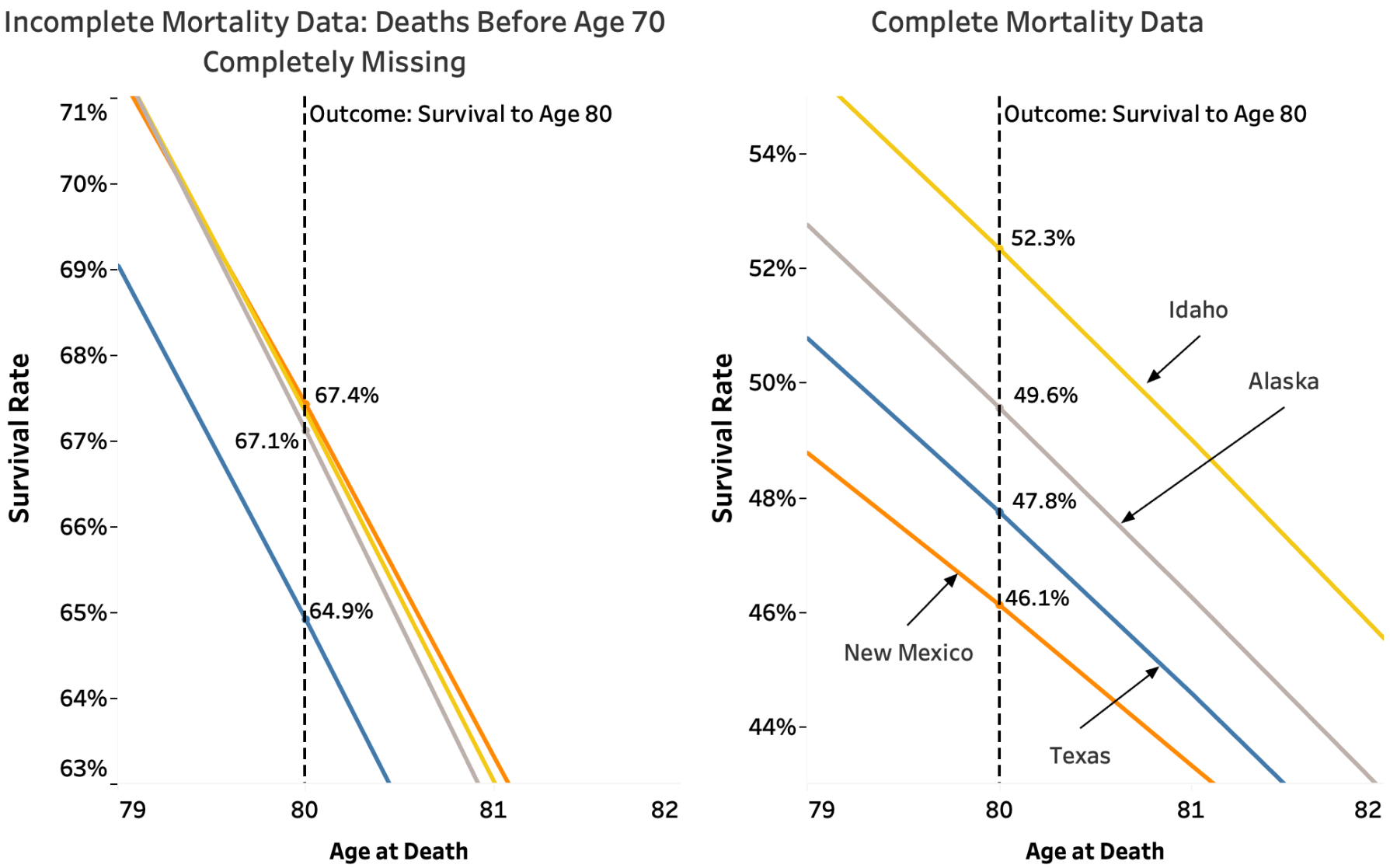
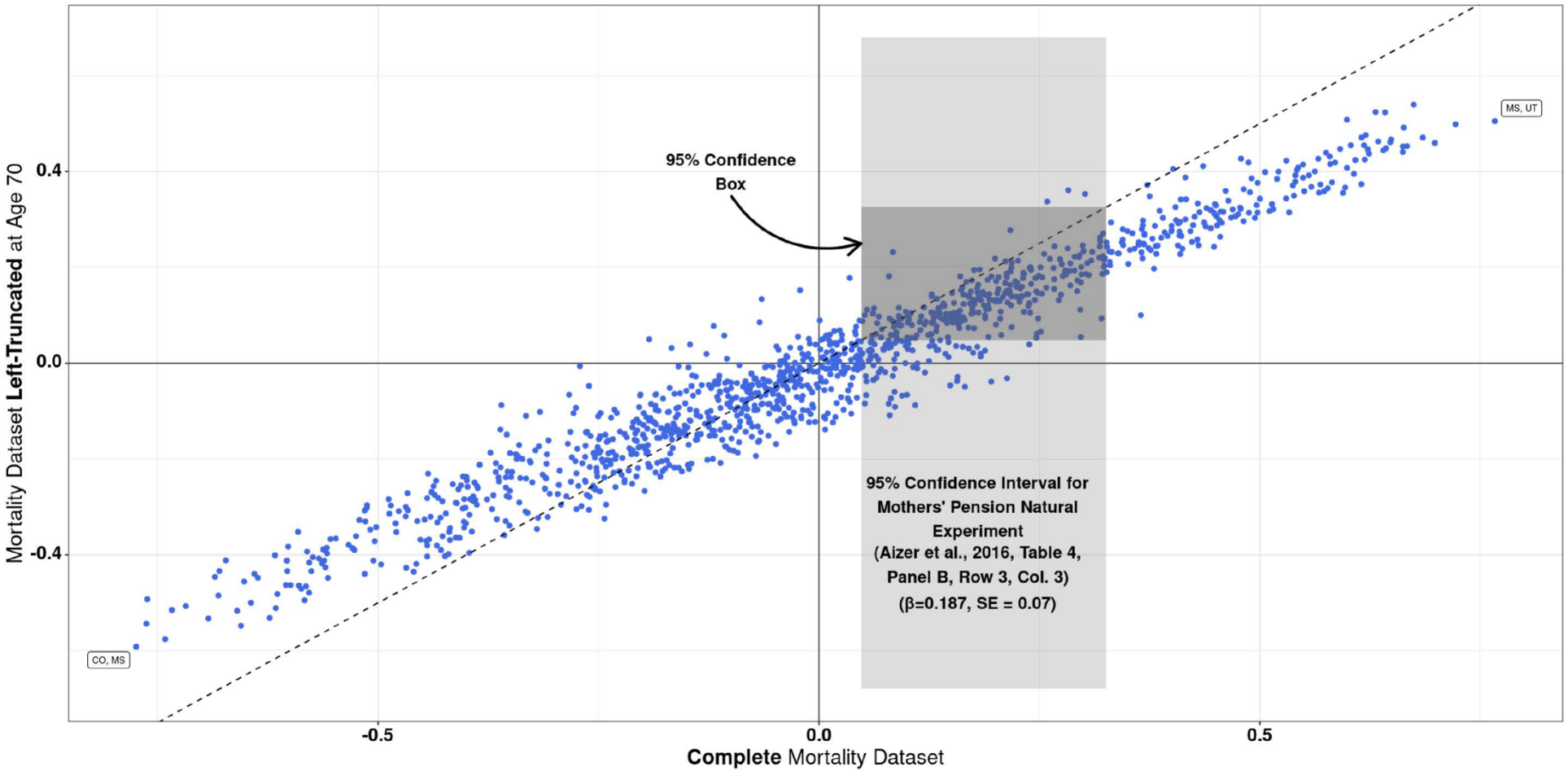


Figure 9

Truncated Mortality Datasets Lead to Bias

I estimate state-of-residence effects for U.S. men in 2019. For each pair of states (X, Y), I estimate the effect of living in state Y relative to state X on survival from age 18 to 80 using a complete mortality dataset (x-axis), then estimate the same effect using a mortality dataset that is left-truncated at age 70 (y-axis). There are no confidence intervals because, taking each state mortality curve as the "true" curve, I sample with replacement to obtain the result for an infinite sample.

Outcome for Both Axes: Log Odds of Survival from Age 18 to Age 80



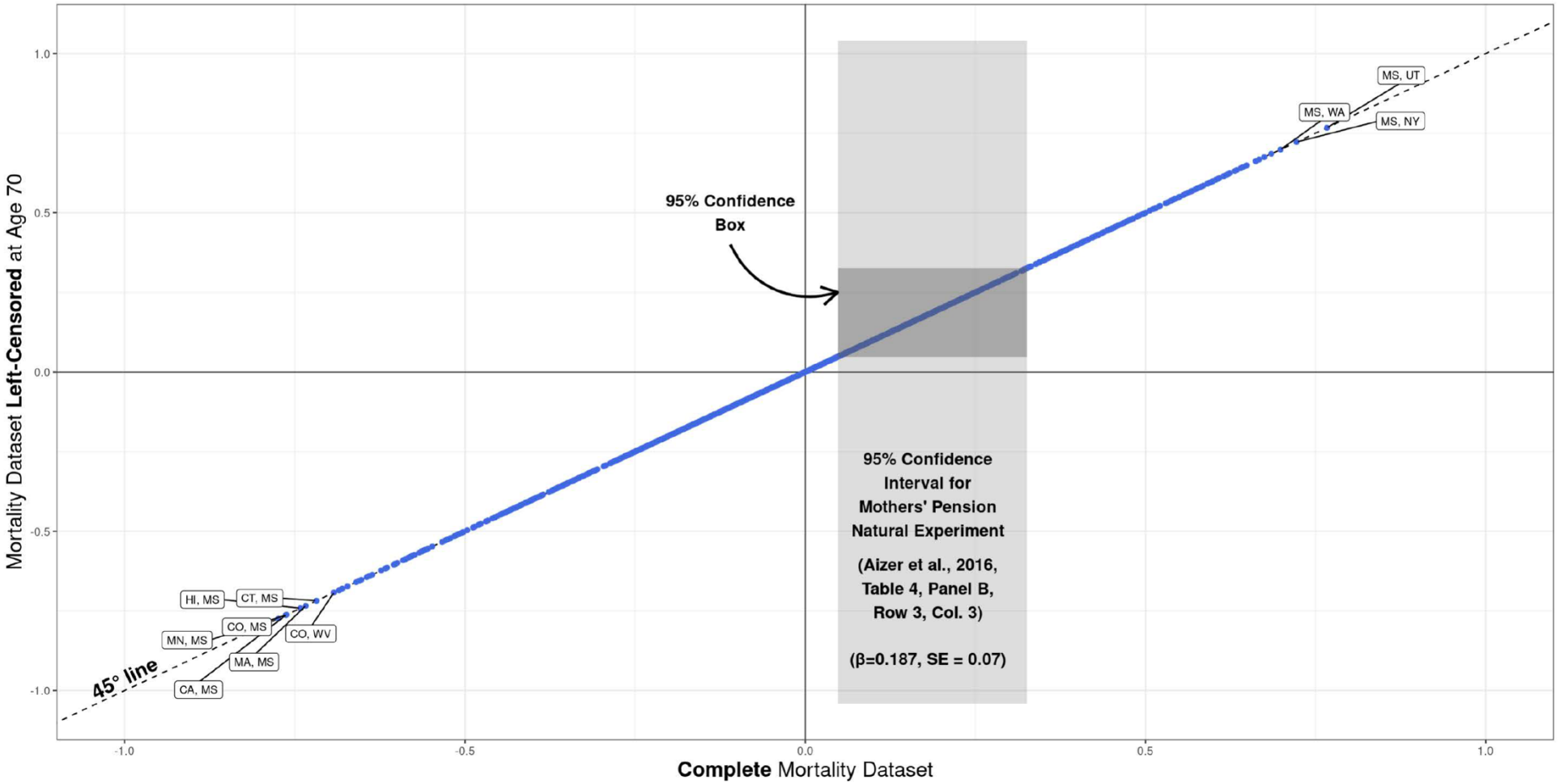
The truncation pattern—left-truncation ten years prior to the survival age of interest—has been used in several papers in the literature (see Figure 6).
Data Source: Arias E, Xu JQ, Tejada-Vera B, Bastian B. U.S. State Life Tables, 2019. National Vital Statistics Reports; vol. 70 no. 18. Hyattsville, MD: National Center for Health Statistics. 2022.

Figure 10

Bias Can Be Entirely Avoided with Censored Mortality Datasets

I estimate state-of-residence effects for U.S. men in 2019. For each pair of states (X, Y), I estimate the effect of living in state Y relative to state X on survival from age 18 to 80 using a complete mortality dataset (x-axis), then estimate the same effect on survival from age 18 to 80 using a mortality dataset that is left-censored at age 70 (y-axis). There are no confidence intervals because, taking each state mortality curve as the "true" curve, I sample with replacement to obtain the result for an infinite sample.

Outcome for Both Axes: Log Odds of Survival from Age 18 to Age 80



Data Source: Arias E, Xu JQ, Tejada-Vera B, Bastian B. U.S. State Life Tables, 2019. National Vital Statistics Reports; vol. 70 no. 18. Hyattsville, MD: National Center for Health Statistics. 2022.

Figure 11

Truncated Mortality Data Leads to Bias

I estimate state-of-residence effects for U.S. men in 2019. For each pair of states (X, Y), I estimate the effect of living in state Y relative to state X on overall lifespans, using a mortality dataset that is left-truncated at age 63 (x-axis), then estimate the same effect using a complete mortality dataset (y-axis). There are no confidence intervals because, taking each state mortality curve as the "true" curve, I sample with replacement to obtain the result for an infinite sample.



The truncation pattern is chosen to approximate that of the Mothers' Pension natural experiment, which is left truncated in 1975, the year when the median birth cohort (born 1912) was age 63.
Data Source: Arias E, Xu JQ, Tejada-Vera B, Bastian B. U.S. State Life Tables, 2019. National Vital Statistics Reports; vol. 70 no. 18. Hyattsville, MD: National Center for Health Statistics. 2022.

Table 1

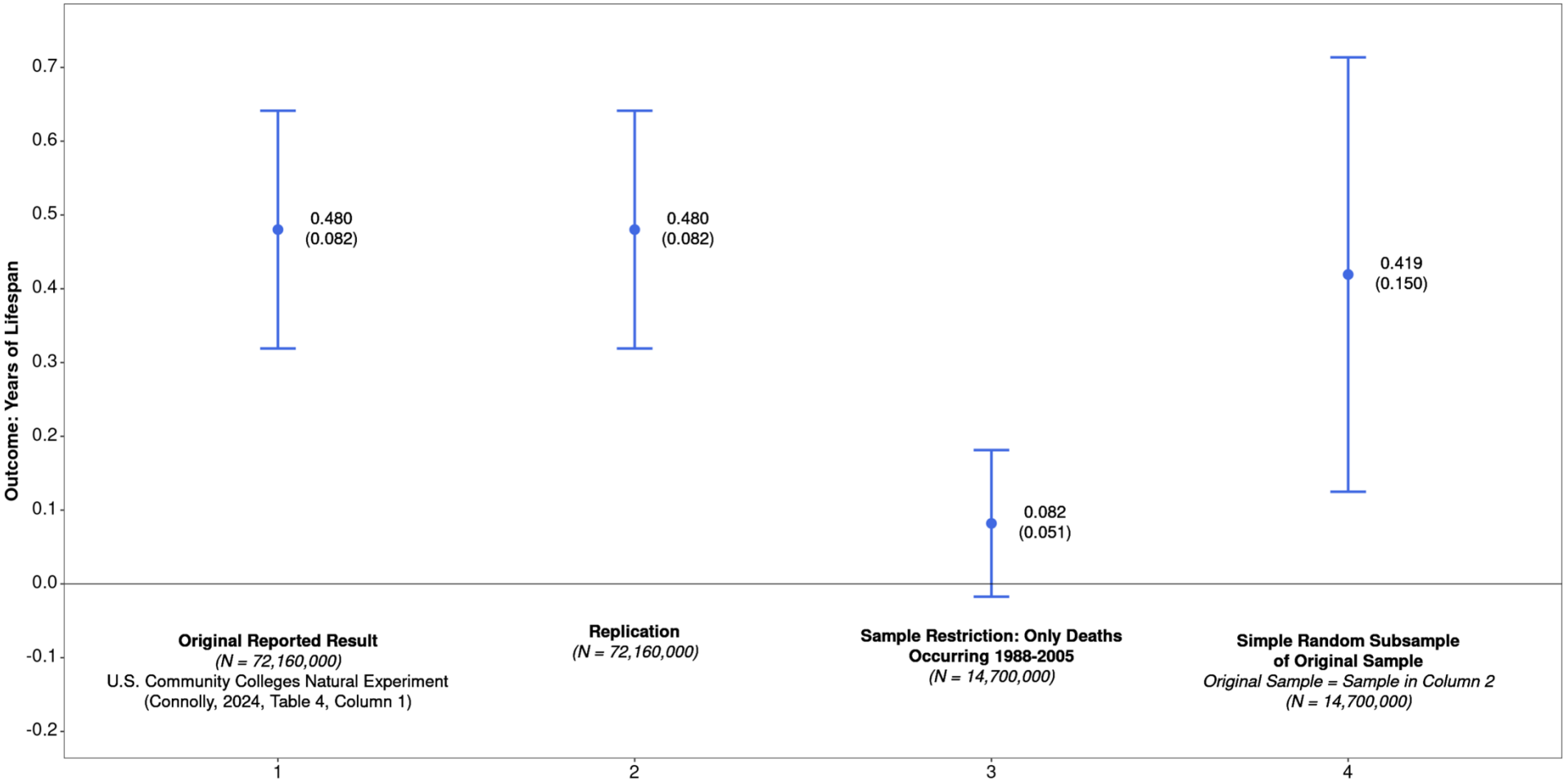
Papers With Key Results Affected by Problem of Endogenous Sample Selection

Journal	Number of Papers at Journal	Cumulative Number of Papers	List of Papers
American Economic Review	5	5	Bedard & Deschenes (2006); Clark & Royer (2013), Black et al. (2015), Aizer et al. (2016), Hollingsworth et al. (2024)
Review of Economic Studies	1	6	Lleras-Muney (2005)
American Economic Journal: Applied Economics	1	7	Hjort et al. (2017)
Journal of Labor Economics	1	8	Aaronson et al. (2021)
Journal of Public Economics	1	9	Jacks et al. (forthcoming)
Journal of the European Economic Association	1	10	Gonzalez et al. (2023)
The Economic Journal	1	11	Salm (2011)
Journal of Human Resources	2	13	Van Kippersluis et al. (2011), Malamud et al. (2023)
Journal of Health Economics	7	20	Albouy and Lequien (2009); Frisvold & Golberstein (2011); Black, Hsu et al. (2015); Buckles et al (2016); Noghanharambari & Engleman (2022); Noghanharambari & Fletcher (2023); Savelyev et al. (2022);
Health Economics	4	24	Noghanharambari & Fletcher (2023); Noghanharambari & Noghani (2023); Fletcher & Noghanharambari (2024), Lacroix et al. (2021)
American Journal of Health Economics	1	25	Cowan & Tefft (2024)
Economics of Education Review	1	26	Sansani (2011)
Demography	5	31	Frisvold & Golberstein (2013); Glied & Lleras-Muney (2008); Halpern-Manners et al. (2020); Noghanharambari & Fletcher (2024); Breen (2024);
Economic Perspectives	1	32	Mazumder (2008)
Quantitative Economics	1	33	Johnson & Taylor (2019)
Social Science & Medicine	1	34	Fletcher (2015)
Nature Human Behavior	1	35	Davies et al. (2018)
International Journal of Epidemiology	1	36	Davies et al. (2023)
Social Science History	1	37	Taylor (2020)
Journal of Population Economics	1	38	Noghanharambari et al. (2024)
Population and Development Review	1	39	Noghanharambari & Fletcher (2023)
Economics & Human Biology	1	40	Noghanharambari & Fletcher (2023)
Population Studies	1	41	Noghanharambari & Fletcher (2024)
Journal of Environmental Economics	1	42	Fletcher & Noghanharambari (2024)
Demographic Research	1	43	Helgertz & Warren (2023)
Population and Environment	1	44	Atherwood (2022)
NBER Working Papers	2	46	Masuda & Shigeoka (2023), Fletcher & Noghanharambari (2024)
Federal Reserve Working Paper	1	47	Moorthy et al. (2022)
SSRN Working Papers	3	50	Noghanharambari & Fletcher (2023); Saavedra (2013); Vu et al. (2023)
Unpublished Dissertations	1	51	Sanchez (2010);
Unpublished Working Papers	3	54	Lillard & Malloy (2010); Kocks (2023); Seltzer (2021)

Appendix Figure 1

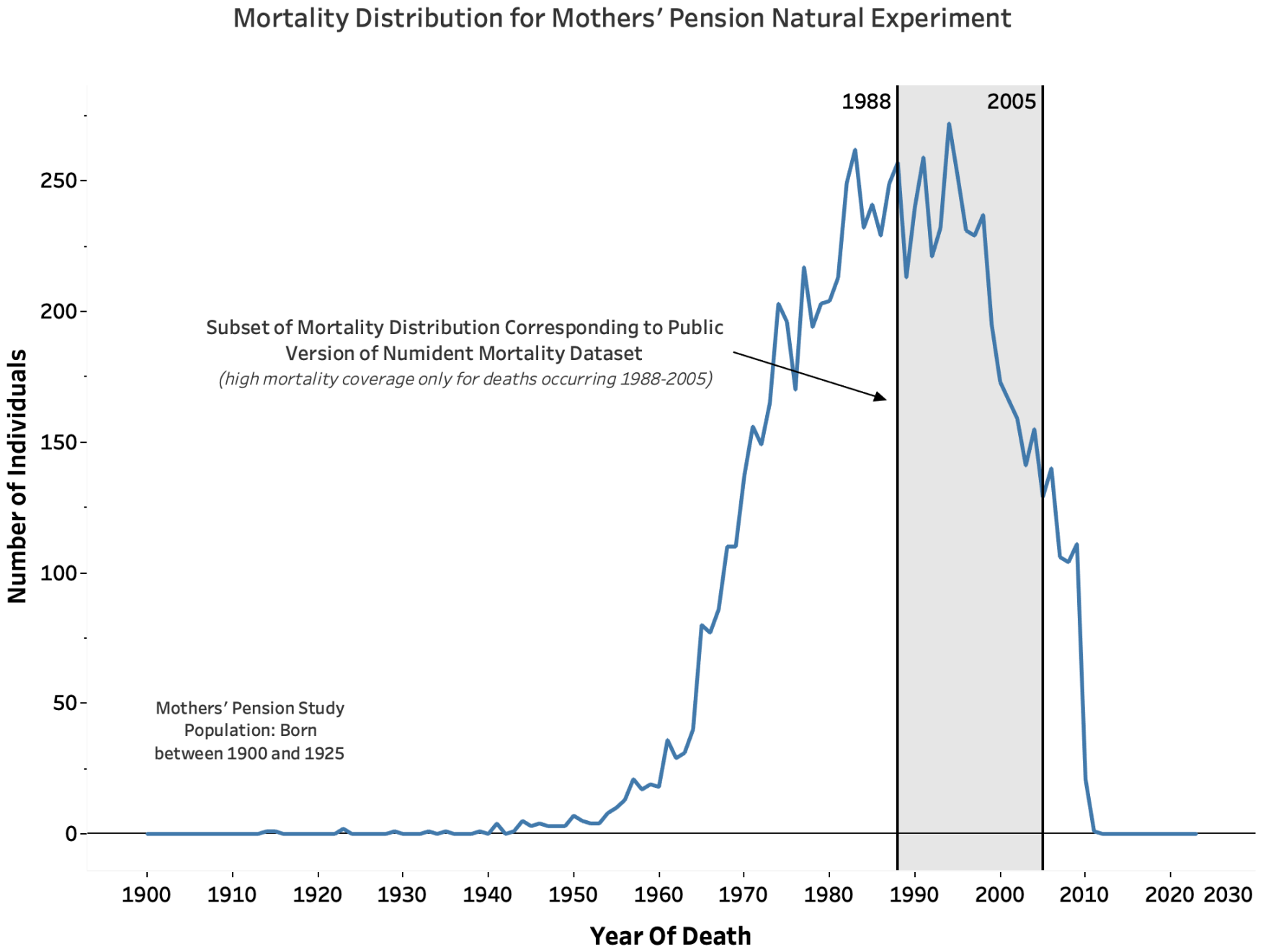
Endogenous Sample Selection Can Cause Problems for Estimating Mortality Effects

The estimated treatment effect of the community college openings on longevity changes when the sample is restricted to individuals who died between 1988 and 2005, suggesting that the latter estimate is biased. The sample restriction also artifically shrinks the standard error.



Notes: This figure plots point estimates and 95% confidence intervals. The only difference between the estimating equations in columns 2 and 3 is the sample, which in column 3 includes only individuals whose year of death falls between 1988 and 2005, as in Fletcher & Noghanharambari (2024) and papers that use the NARA Numident dataset (e.g., Hollingsworth et al., 2024).
Connolly (2024). "The Effects of U.S. Community Colleges on Educational Attainment and Longevity." Working Paper.

Appendix Figure 2

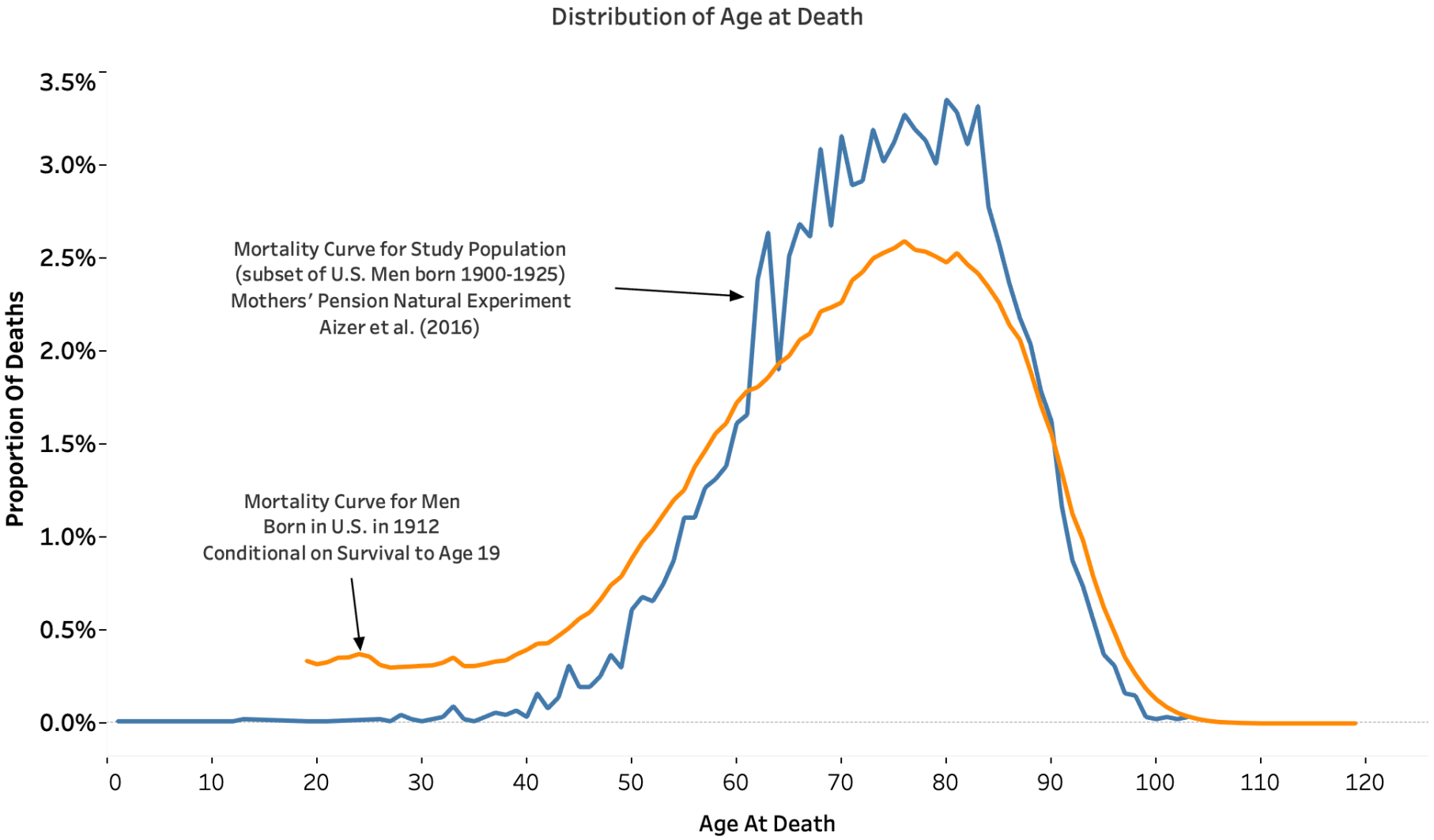


Aizer et al. (2016). "The Long-Run Impact of Cash Transfers to Poor Familes." *American Economic Review*, 106(4), 935-971.

Appendix Figure 3

Missing Mortality Data in Mothers' Pension Natural Experiment

There is substantial missing mortality data in the Mothers' Pension natural experiment (Aizer et al., 2016), and the authors argue (p. 942) that the primary reason is younger-age mortality, deaths before 1975 that were never recorded in Social Security mortality data. A comparison of the Mothers' Pension mortality dataset with the 1912 U.S. male cohort mortality curve suggests that this assumption is correct.

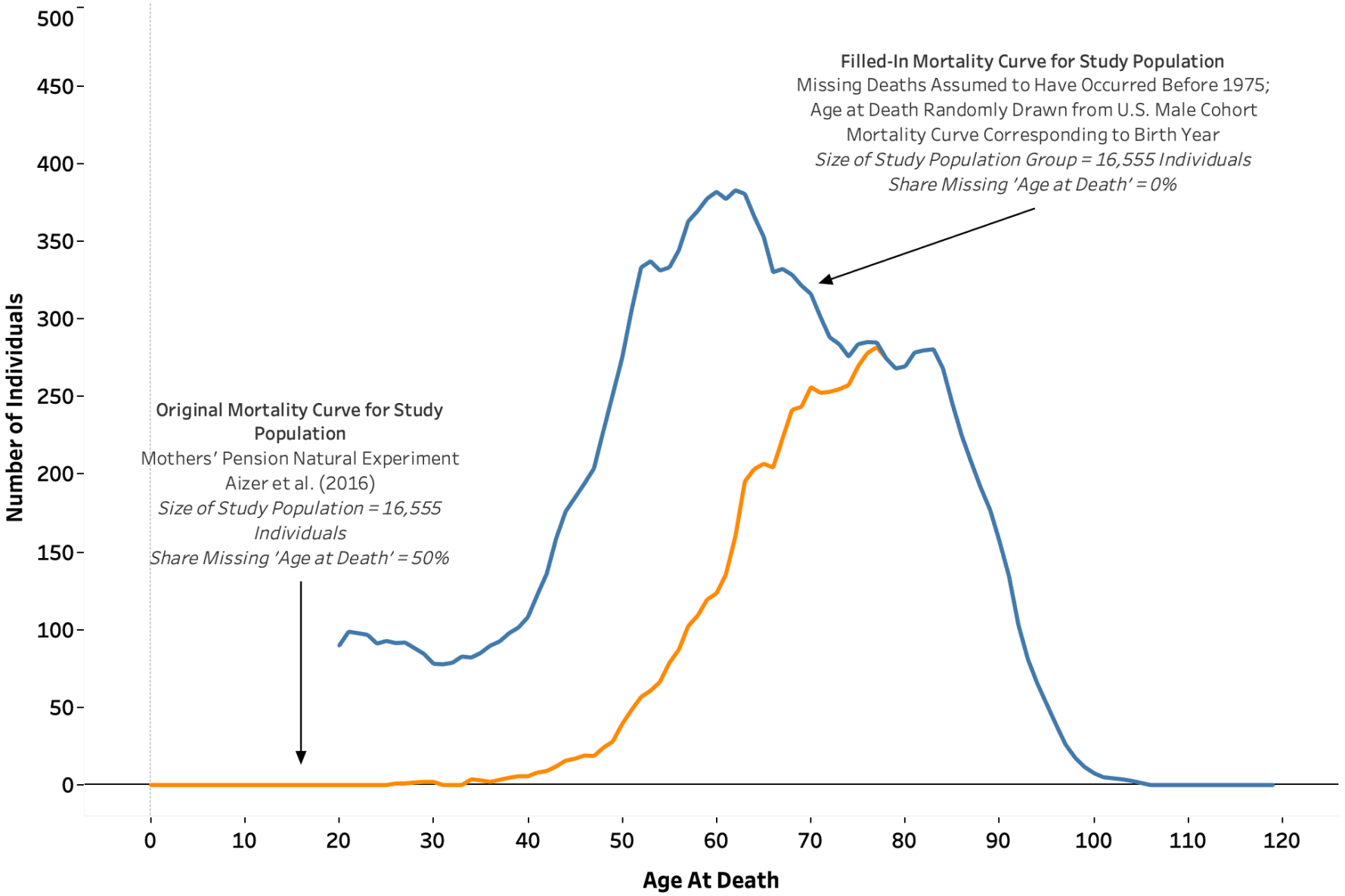


Cohort Life Tables, Social Security Administration, 1900-2095. <https://www.ssa.gov/oact/HistEst/CohLifeTables/2020/CohLifeTables2020.html>
Aizer et al. (2016). "The Long-Run Impact of Cash Transfers to Poor Familes." *American Economic Review*, 106(4), 935-971.

Appendix Figure 4

Imputing Missing Left-Tail Mortality Data Using U.S. Cohort Mortality Curves

Mortality Curves are Displayed as Three-Year Rolling Averages

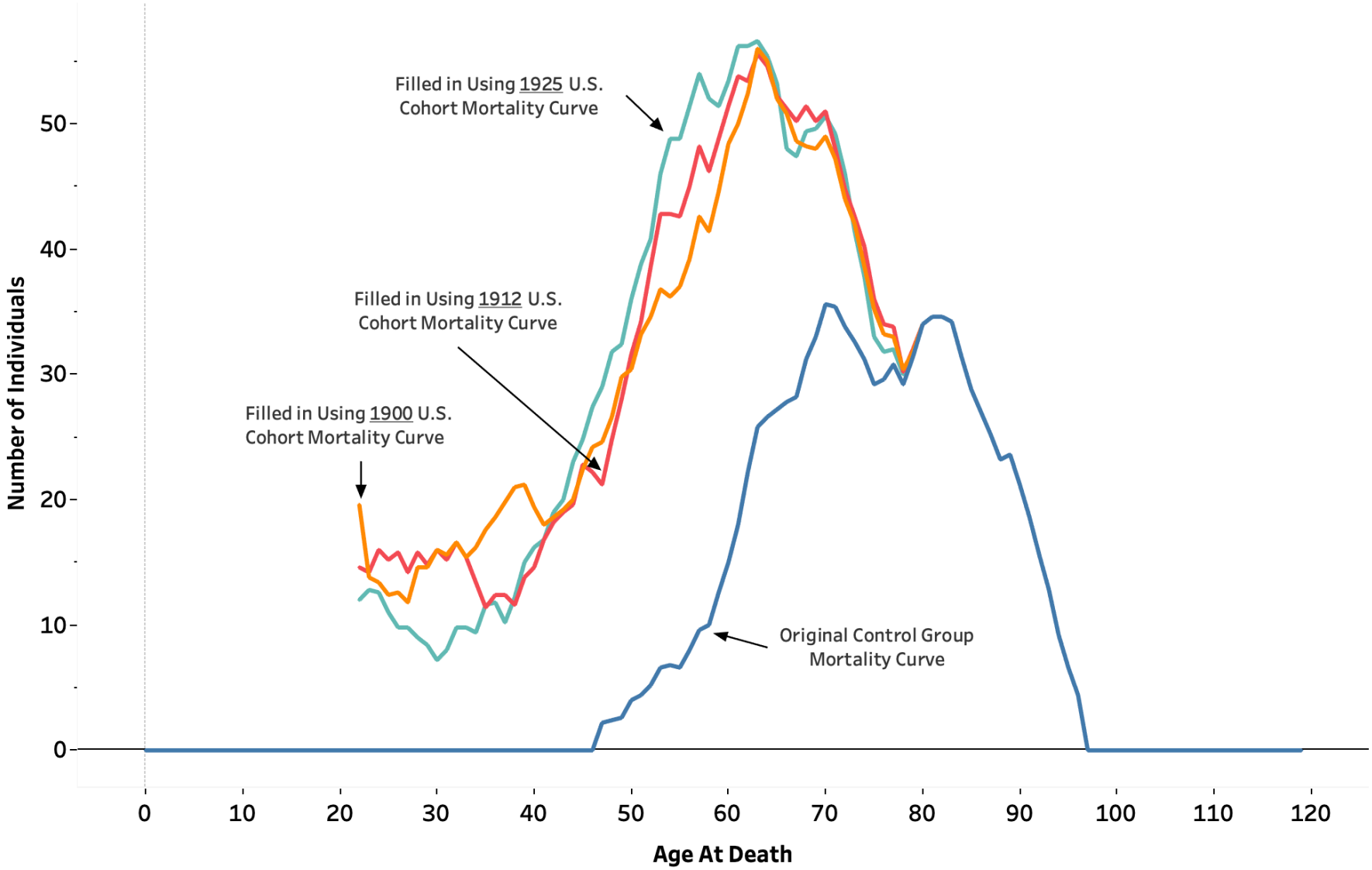


Aizer et al. (2016). "The Long-Run Impact of Cash Transfers to Poor Familes." *American Economic Review*, 106(4), 935-971.

Appendix Figure 5

Imputing Missing Left-Tail Mortality Data Using U.S. Cohort Mortality Curves

Mortality Curves are Displayed as Five-Year Rolling Averages

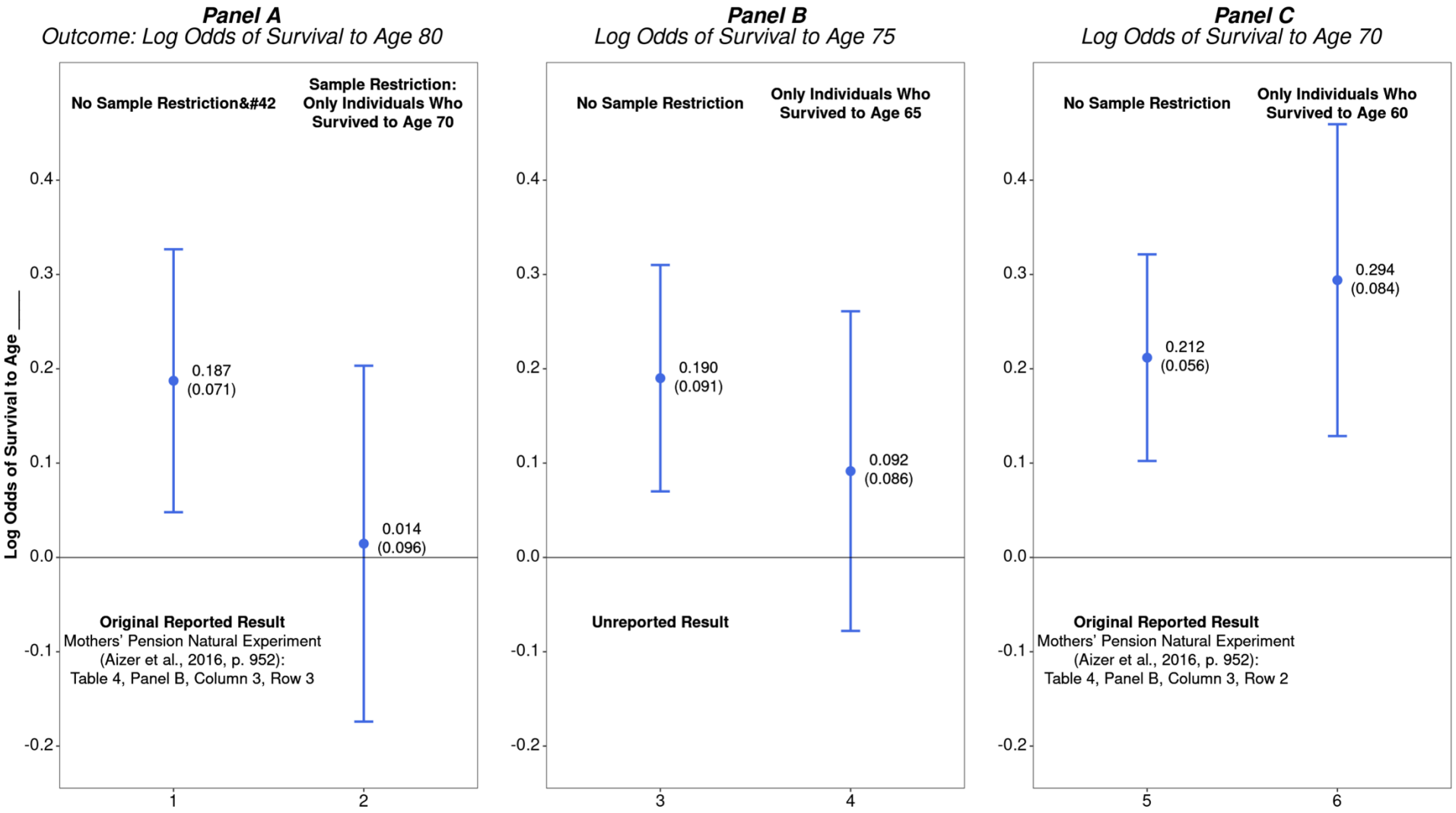


Missing deaths are assumed to have occurred before 1975, the point at which Social Security mortality data becomes comprehensive. Age at death randomly drawn from U.S. male cohort mortality curve of the indicated year.

Aizer et al. (2016). "The Long-Run Impact of Cash Transfers to Poor Familes." *American Economic Review*, 106(4), 935-971.

Appendix Figure 6

Endogenous Sample Selection Does Not Necessarily Attenuate Point Estimates

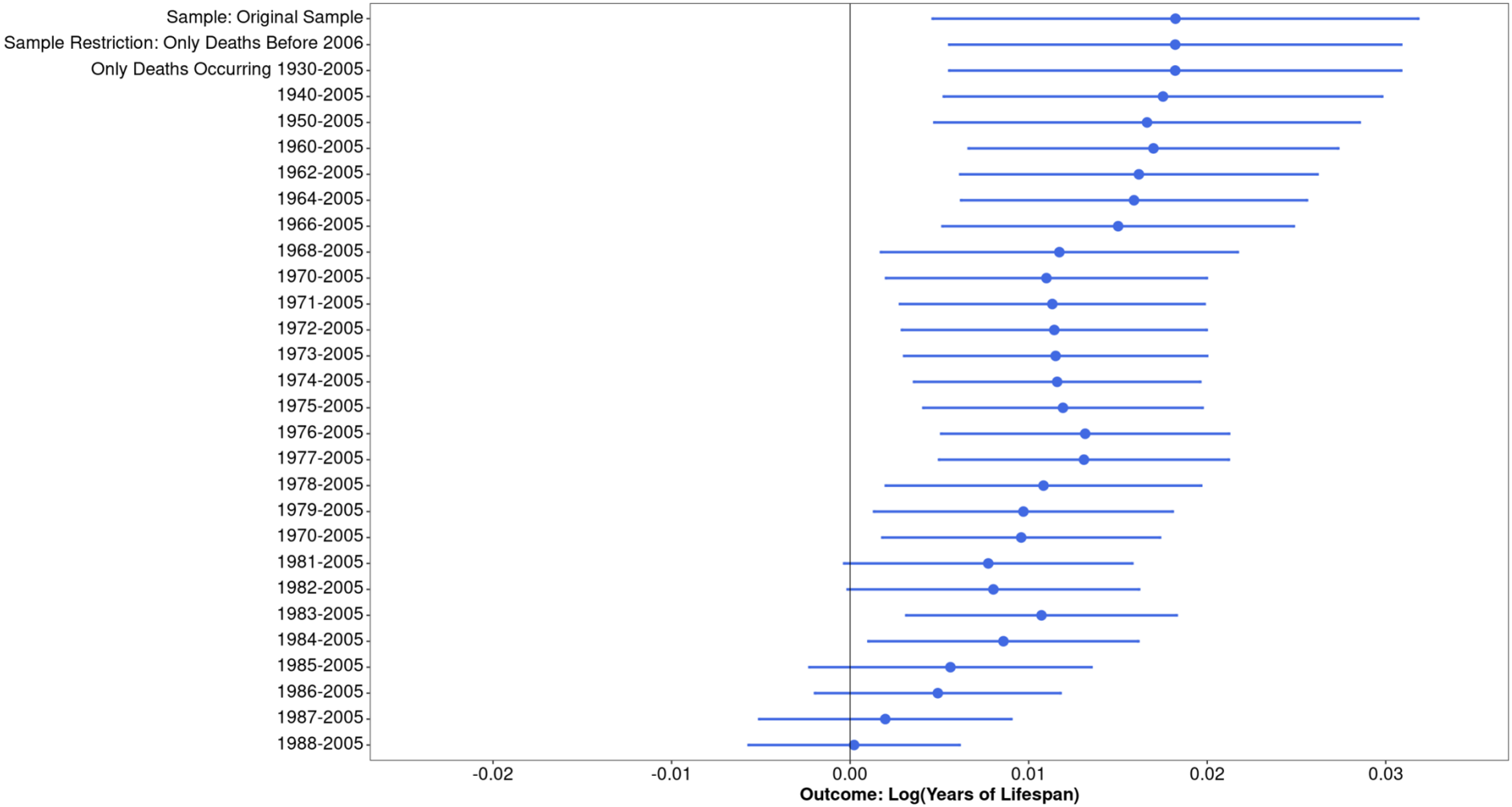


Aizer et al. (2016). "The Long-Run Impact of Cash Transfers to Poor Familes." American Economic Review, 106(4), 935-971.

Appendix Figure 7

Endogenous Sample Selection Can Cause Problems for Estimating Mortality Effects

How various patterns of missing mortality data change the estimated causal effect of the Mothers' Pension program on longevity.

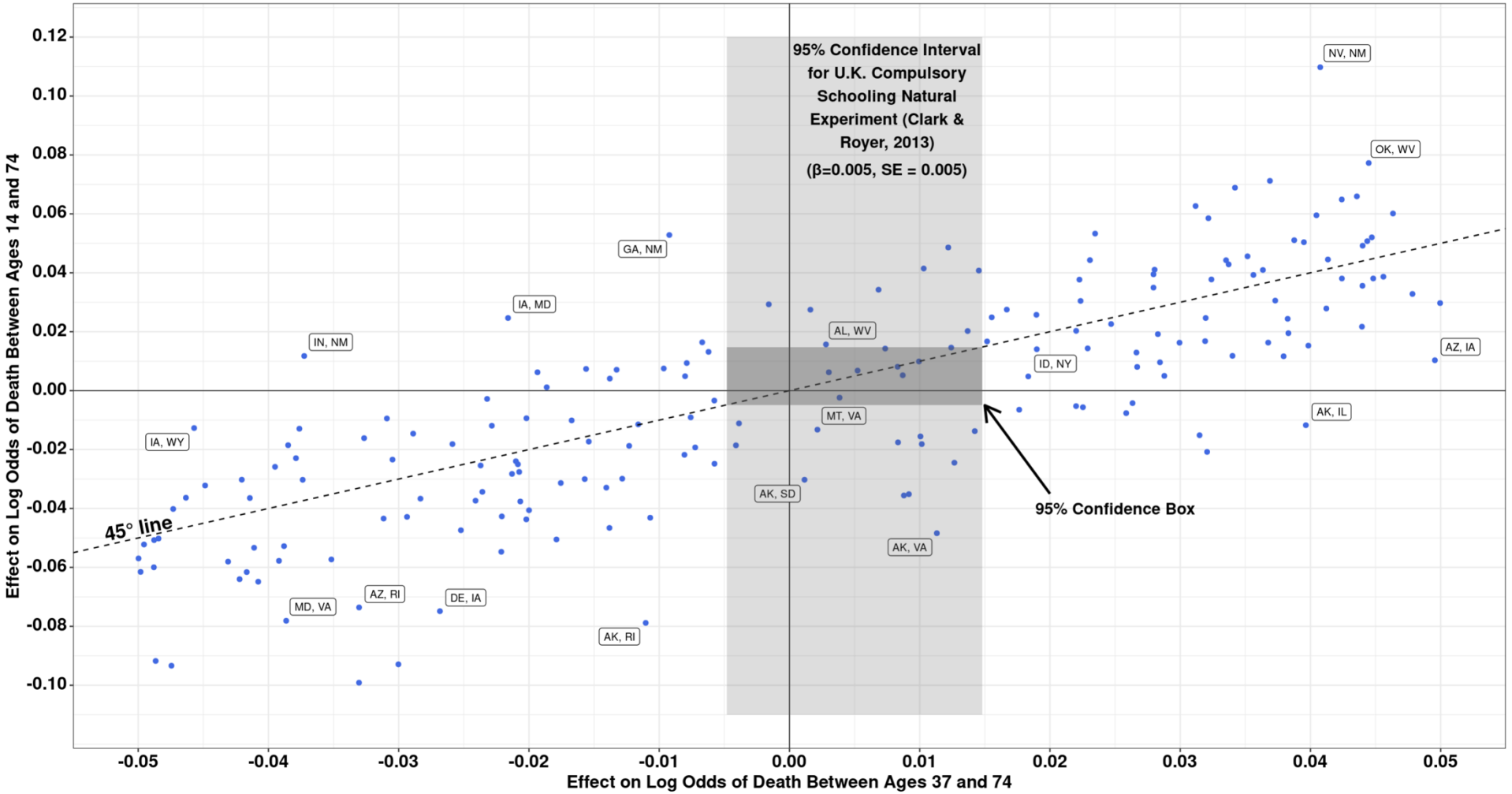


Notes: Original Result is Table 4, Panel A, Column 3 of Aizer et al. (2016). in A pattern of mortality data with high coverage for only the years 1988-2005 is used in several papers, including Fletcher & Noghanharambari (2024) and other papers that use the NARA Numident dataset (e.g., Hollingsworth et al., 2024).
Aizer et al. (2016). 'The Long-Run Impact of Cash Transfers to Poor Familes.' American Economic Review, 106(4), 935-971.

Appendix Figure 8

Mortality From Age 37 to 74 is a Biased Proxy Outcome for Mortality From Age 14 to 74

I estimate state-of-residence effects for U.S. men in 2019. For each pair of states (X, Y), I estimate the effect of living in state Y relative to state X on mortality from age 37 to 74 (x-axis), then estimate the same effect on survival from age 14 to 74 (y-axis). There are no confidence intervals because, taking each state mortality curve as the "true" curve, I sample with replacement to obtain the result for an infinite sample.

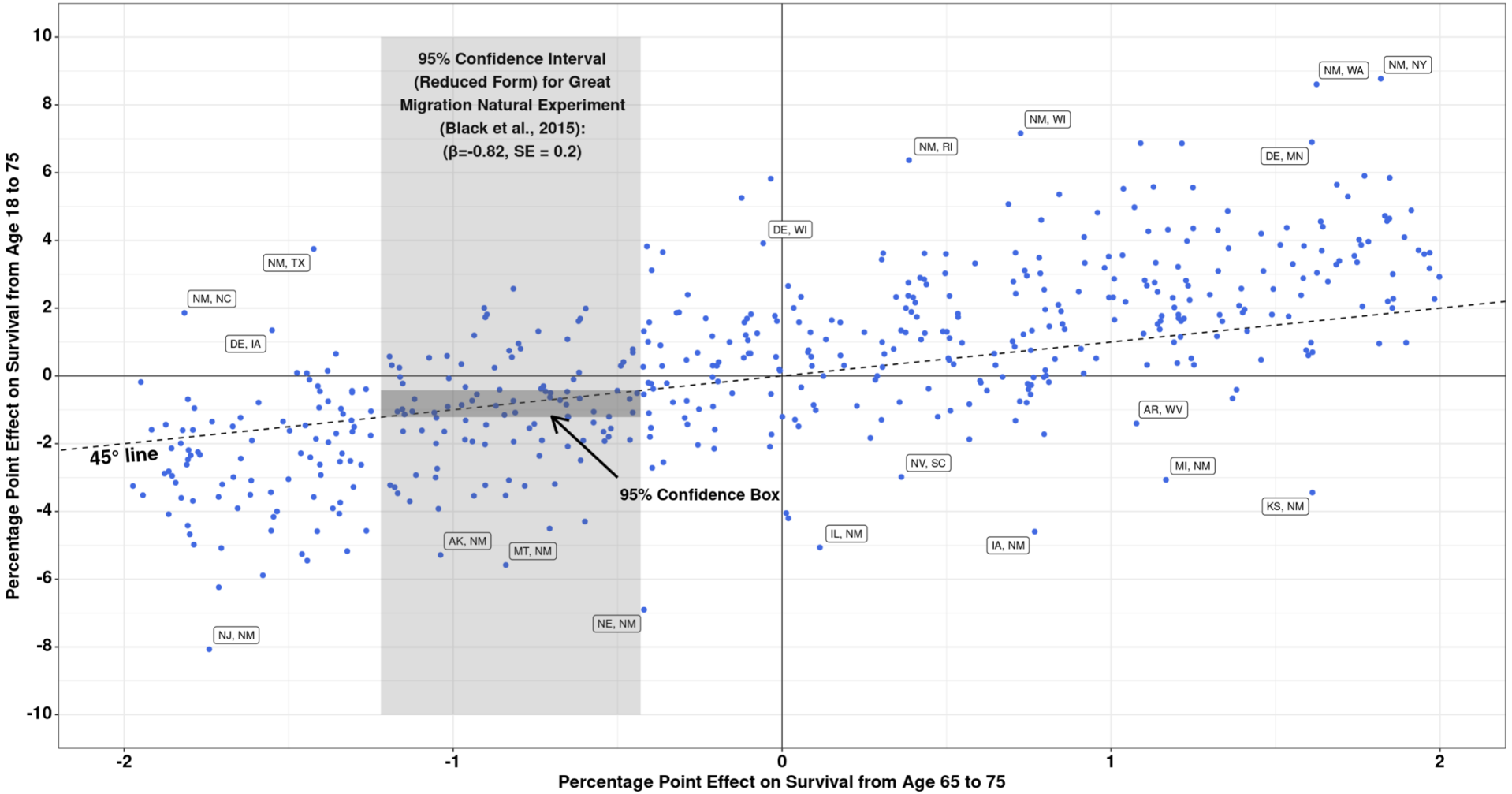


Data Source: Arias E, Xu JQ, Tejada-Vera B, Bastian B. U.S. State Life Tables, 2019. National Vital Statistics Reports; vol. 70 no. 18. Hyattsville, MD: National Center for Health Statistics. 2022.

Appendix Figure 9

Survival From Age 65 to 75 is a Biased Proxy Outcome for Survival from Age 18 to 75

I estimate state-of-residence effects for U.S. men in 2019. For each pair of states (X, Y), I estimate the effect of living in state Y relative to state X on survival from age 65 to 75 (x-axis), then estimate the same effect on survival from age 18 to 75 (y-axis). There are no confidence intervals because, taking each state mortality curve as the "true" curve, I sample with replacement to obtain the result for an infinite sample.



Data Source: Arias E, Xu JQ, Tejada-Vera B, Bastian B. U.S. State Life Tables, 2019. National Vital Statistics Reports; vol. 70 no. 18. Hyattsville, MD: National Center for Health Statistics. 2022.