No consistent effects of prenatal or neonatal exposure to Spanish flu on late-life mortality in 24 developed countries

Alan A. Cohen

John Tillinghast

Johns Hopkins Bloomberg School of Public Health

Abstract

The fetal origins hypothesis suggests that events during development can affect health and mortality later in life. Malnutrition and disease are thought to be important such events but are hard to tease apart using population-level data because they often co-occur. However, the 1918-1919 Spanish flu pandemic provides an excellent opportunity to test the effects of disease in the absence of major malnutrition relative to surrounding years. We used publicly available data from the Human Mortality Database for 24 countries to test the hypotheses that prenatal or neonatal exposure to severe influenza would either increase or decrease mortality rates at ages 45-80. We also used fertility statistics and half-year cohorts to test for finer-scale effects in France, Italy, and Switzerland. Across the 24 countries, after controlling for age, period, and sex effects, residual mortality rates did not differ systematically for the cohorts born in 1918 or 1919 relative to cohorts born in surrounding years. We calculate at most a 20-day reduction in life expectancy for those born in flu cohorts; likely values are much smaller. Similarly, there was no notable difference in 1918-1919 half-year cohort mortality in France or Italy, though it is possible there was a difference in Switzerland. Estimates of influenza incidence during the pandemic suggest that exposure was high enough for this to be a robust negative result. Potentially, early exposure to influenza specifically or disease generally has little effect on late-life health and mortality; alternatively, there could be countervailing effects of exposure that cancel each other out in population-aggregate measures. Lastly, we detected substantial heterogeneity in late-life mortality rates in different cohorts within countries, though not across countries or cohorts generally, suggesting that local conditions early in life may be important for late-life mortality in ways yet to be elucidated.

Introduction

It has long been hypothesized that events early in life could affect the developmental process and thereby health at later ages. Such effects are well-known in animal models, and have been observed in humans for a number of specific diseases [1-3]. Finch and Crimmins [4] showed an association between early and late mortality in cohorts from Sweden, and suggest that this could imply an effect on lifespan of high inflammation resulting from infections early in life. However, links between early and late life mortality could also result from famine or other factors not directly related to early inflammation; data on the effects of famine are mixed [5-7]. Generally speaking, disease and malnutrition tend to co-occur, making it hard to tease apart these factors with population data.

One potential way to distinguish effects of disease from malnutrition is to use the Spanish influenza pandemic of 1918-1919. Although following close on the heels of World War I and the potential food shortage accompanying war, the pandemic affected neutral countries as well and did not occur at a time of generalized food shortages more severe than what had occurred in

the previous years of the war. Thus, any significant differences in late-life mortality of 1918 and 1919 cohorts relative to surrounding cohorts is likely attributable to the pandemic, especially if such effects can be generalized across countries. Mortality from the pandemic is estimated at 50-100 million worldwide, the largest well-recorded demographic event on record [8]; while the damage was somewhat less in developed countries, no country was immune and effects were severe even in Denmark, the country with lowest excess mortality [9].

The simple prediction, following Crimmins and Finch [10], is that exposure to disease *in utero* or early in life should increase late-life mortality (i.e., reduce adult life expectancy). There are a number of mechanisms that could cause such an effect in addition to an increase in inflammation. In particular, pre- or peri-natal exposure to disease could have effects via changes in health of the mother, who might allocate scarce resources to maintaining her own health rather than ensuring proper development of the fetus. Developmental processes are sensitive to perturbation and can have diverse, long-lasting downstream effects. For example, a slight problem in neural development might result in marginally lower intelligence, which could in turn result in lower social status and all its concomitant health problems. Additionally, developing organisms often look to their environment for cues as to how to develop best in an unstable world ["phenotypic plasticity," 11]; early exposure to disease could prompt the immune system to develop in expectation of high disease burden throughout life, potentially resulting in high constitutive inflammation levels. There may be other pathways as well by which early exposure to disease could affect late-life mortality.

If early exposure to influenza increases mortality via inflammation pathways, we can make several predictions. To start with, the effect should be strongest in those hit by the flu post-natally, when there is less reliance on the maternal immune system [12]. In contrast, other pathways should be strongest for pre-natal exposure, since disturbance earlier in development is likely to have more severe long-term consequences in most cases. Because the timing of the flu pandemic was precise and roughly synchronized worldwide, with most mortality occurring in fall 1918 followed by a smaller wave in early 1919 [8], we should expect 1918 cohort mortality (i.e., mostly those already born before the flu hit) to be higher than 1919 cohort mortality (mostly those still *in utero* during exposure) if inflammation is the primary pathway of action; conversely, if more general developmental disturbance through under-allocation of maternal resources is the primary pathway, the 1919 cohort should have higher late-life mortality. Also, we would expect a slightly delayed cohort effect in Australia, which managed to delay onset of the flu to early 1919 by a maritime quarantine [8].

Potentially confounding these predictions, there are also pathways by which early exposure to disease could actually lower late-life mortality. First, high exposure to disease could cull the weakest infants from a cohort; the survivors should be the most robust individuals and therefore predisposed to live longer. Such selection effects have been documented in wild animals such as Leach's Storm-Petrel (*Oceanodroma leucorhoa*) [13]. Second, early exposure to disease could actually improve the long-term health of infants. Hunter-gatherers are likely exposed to much higher disease burdens than modern humans and are known to have constitutively high inflammation levels [14]. If this is representative of the environment in which our immune systems evolved, it may well be that proper development of the immune system requires exposure to pathogens for some "priming" effect. Along these lines, there are hypotheses that modern epidemics of allergies and asthma result from early under-exposure of the immune system [12, 15]; it might be reasonable to expect similar such effects in other aspects of the immune system late in life, including auto-immune disease and regulation of inflammation.

All of the above mechanisms for effects of early exposure to disease on late-life mortality could operate simultaneously. To the extent that countervailing processes have similar degrees of effect, our ability to detect differences with population-level data will be minimal and detailed physiological data will be necessary to determine pathways of action. Nonetheless, it is possible that some of these effects are much stronger than others, and if there are population-level effects, it is unlikely that a better natural experiment will be found than the 1918-1919 flu pandemic. Analyses of economic and health outcomes were poorer for cohorts born in the U.S. during the flu, but mortality has not been examined [16, 17].

Here, we take publicly available data from the online Human Mortality Database to test for differences in late-life mortality between Spanish flu cohorts and surrounding cohorts. We restrict analyses to ages 45 to about 80, depending on country, and to cohorts born between 1911 and 1923. To ensure that all ages tested are included for all cohorts, we limit countries analyzed to those with mortality data dating back to 1956 (= 1911 + 45), a set of 24 countries including most of Europe (but not the Soviet bloc), the U.S., Canada, Australia, Japan, and New Zealand. We use residual mortality rates after accounting for age, period, and sex effects; age and period splines are used to account for non-linear trends. Life expectancy is also analyzed. Finer-scale cohort analyses are performed for France, Italy, and Switzerland. We show that in most analyses there is no significant difference between 1918 or 1919 cohort mortality and surrounding cohort mortality in either direction.

Methods

Data

We used publicly available data for 24 countries taken from the Human Mortality Database (www.mortality.org; Table 1). We used cohort data on deaths and exposures to calculate mortality rates for each cohort at each age (deaths/exposures). We used all countries in the database as of January 2008 that met our criteria for age and year ranges except Iceland and Luxembourg, for which population size was too small to make reliable inferences. There were several trade-offs to consider in selecting which cohorts and ages to use, as follows. First, the later the last cohort used in our comparison, the younger the oldest age we could use. We used cohorts born through 1923. Most countries had data available through 2003-2005, so we could use a maximum age of 79-82 (depending on country). Second, the younger the youngest age used and the earlier the first cohort used, the fewer countries could be included in the analysis based on how far back their mortality statistics date. Additionally, greater range of ages and cohorts increases the "sample size" of mortality rates available for analysis and thus improves statistical power. Ideally, we would have liked to use mortality starting at age 30 for cohorts born in 1910 or later; of the 33 available countries, 16 had old enough data regardless of first age and cohort used, and two (Slovenia and Taiwan) clearly did not have good coverage at the youngest ages. However, 15 countries had mortality data starting between 1947-1959, which would limit the youngest age and/or earliest cohort used. We ran the analyses with different combinations for all countries and found consistent results; here, we present data using cohorts from 1911 (to include some pre-World War I data) starting at age 45. This results in exclusion of Poland and the former Soviet republics; we were not confident in the quality of the Soviet republic data, but there were no clear patterns in these excluded countries that contradict our broader findings. Thus, we used a cohort range of 1911-1923, an age range of 45-~81 (depending on last year of data available for each country), and a set of 24 countries (Table 1). Roughly 22,000 mortality data points were incorporated into the analyses.

The Human Mortality Database provides only annual data, but the Spanish flu pandemic was worst from the fall of 1918 through spring of 1919, suggesting that one-year cohorts might not be sufficiently specific to pick up cohort effects, especially if age of the infant or stage of pregnancy during influenza exposure are important factors in determining late-life effects. In order to address this, we used monthly fertility data for France (1914-1919), Italy (1911-1923), and Switzerland (1910-1923) combined with Lexis data from the Human Mortality Database for these countries to test for effects at half-year cohort time scales (see *Analysis*) [18-20]. The monthly fertility data for Italy were pooled for the years 1911-1913 and 1921-1923; we ran analyses with and without these years included, but found no substantive differences and present the only the analysis with all years included.

Analyses

24-country multilevel models

Mortality rates (deaths per exposure per year) were log-transformed so that the increase with age was roughly linear. Henceforth we refer to log-mortality rates simply as "mortality rates." We used regression (general linear models) to calculate residual log-mortality after accounting for age, period, and sex. Year of death (period) was calculated as year of birth plus age at death plus one, which is the average time of death assuming constant fertility rates within each year of birth and constant death rates within each one-year age class. We centered age and period and then calculated polynomial B-spline bases with knots every 10 years (R v. 2.6.0, "splines" package, "bs" function). Regression models calculated mortality as a function of age splines, period splines, sex, and all two-way interactions among them. This model should represent a significant improvement over typical age-period-cohort models [e.g. 21, 22] because (a) it allows for interaction terms, which often turn out to be highly significant; (b) there are far fewer spline terms and interactions than data points, so it is easily identified and retains substantial power; (c) it accounts for continuity (i.e., non-independence) of adjacent age-classes and periods; and (d) it does not make any of the assumptions of parametric models. Depending on country, variance explained by this model ranged from 99.51% to 99.95%. Residuals from this model ("residual mortality") were used in subsequent analyses, and represent mortality after accounting for age, period, and sex effects. Residual mortality rates were calculated separately for each country using R v. 2.6.0 [23].

We tested for heterogeneity in residual mortality rates across 24 countries and 13 birth cohorts (1911-1923) using multilevel random effects models. The model used was a "crossed" model; i.e., each mortality rate was nested within a cohort and within a country, but neither cohorts were nested within countries nor countries within cohorts. The specified model was:

$$m_{r(ijk)} = \beta_0 + \beta_{1(k)} + \beta_{2(j)} + \varepsilon_{(ijk)} \quad (1)$$

where m_r is residual morality, i is individual mortality data point (for a given age, cohort, sex, and country), j is country, k is cohort, β_0 is overall mean residual mortality (expected to be approximately 0), β_1s are normally distributed cohort-specific deviations from β_0, β_2s are normally distributed country-specific deviations, and ε is a normally distributed error term. Variances of β_1, β_2, and ε were also estimated and assumed to be gamma-distributed. In order to calculate deviations of specific combinations of country and cohort from the overall average, we used a country-within-cohort model, similar to equation (1) except with the β_2s nested within the β_1s:

$$m_{r(ijk)} = \beta_0 + \beta_{1(k)} + \beta_{2(jk)} + \varepsilon_{(ijk)} \quad (2)$$

Shrunk estimated Bayes values were calculated for each country and cohort (from the crossed model, equation (1)) and for cohort-country combination (from equation (2)). Multilevel models were fit using Bayesian Gibbs sampling and Markov Chain-Monte Carlo simulations (4000 iterations, first 1000 discarded to allow for convergence) in WinBUGS v. 1.4.3 [24].

Life expectancy

In order to examine male-female differences and also to get a summary metric in relevant units ("years" as opposed to "residual log-mortality"), we repeated the analyses using cohort life expectancy from ages 45-79, calculated for each country-cohort combination from the mortality rates used above via standard methods. This quantity shows how many years the average person surviving to age 45 would expect to live through age 79. Because period trends in life expectancy result in roughly linear increases across the 1911-1923 cohorts, we calculated residual life expectancy for each cohort within each country – i.e., given the increase in life expectancy over time in a given country, how much longer or shorter did each cohort live than expected? We then averaged these values across countries to calculate mean residual life expectancy for each cohort. Analyses were run separately for males and females.

France, Italy, and Switzerland: finer-scale cohorts

We took a somewhat different approach to analyzing the finer-scale data from France, Italy, and Switzerland. Because we used Lexis data rather than cohort data, we were able to calculate two mortality rates for each age-cohort-sex-country combination: one for people who died in the earlier possible year, and one for people who died in the later possible year. For example, a woman who was born in 1918 and died at age 60 could have died in 1978 or 1979, depending on when in 1918 she was born and exact age at death. The Lexis data allows us to separate those who died in 1978 from those who died in 1979. The probability of dying in 1978 versus 1979 depends on when in 1918 the woman was born: the closer to the beginning of 1918, the higher the probability she would die in 1978 (given death at age 60). This relationship can be used to calculate distributions of birthdays for those who died in the earlier versus later possible years. Assuming equal birth rates throughout a year and equal death rates throughout a one-year age class, those dying in the earlier possible year have a mean birthday May 1, with 75% born in the first half of the year. Those dying in the later possible year have a mean birthday September 1, with 75% born in the second half. Thus any large differences in mortality rates between half-year cohorts should be apparent (though not precisely estimable) in differences between these "Lexis cohorts" – those that died in the earlier versus later possible year, for a given one-year age class and one-year birth cohort. If age is a and year of birth is c, the two Lexis cohorts are thus those who died in year $a + c$ and those who died in year $a + c + 1$. We will refer to these cohorts as "spring births" and "fall births" respectively, though of course there are individuals in each cohort born at all times of year; the seasons are probabilistic averages. The method is developed at greater length in Online Appendix A.

Death counts but not death rates or population exposures are available in Lexis format from the Human Mortality Database, so calculation of Lexis cohort death rates is only possible if we divide full-year cohort exposure data in half, assuming equal sizes of the Lexis cohorts (i.e., constant birth rates throughout the year). However, there was a baby boom at the end of 1919 in many countries in our database, apparently a consequence of the flu [25]; the resulting difference in the cohort sizes confounds our ability to calculate accurate Lexis cohort mortality rates for that year, one of the two of most interest for the flu pandemic. We circumvented this by gathering

monthly fertility rates for France, Italy, and Switzerland, allowing us to calculate much more precisely the percent of full year exposures that should be assigned to each of the two Lexis cohorts within a given year. For example, those born in January are born on average around January 15, or $1/24^{th}$ of the way through the year. Thus, only 1/24 of their lives are spent exposed to risk in year $a + c + 1$, and $23/24^{ths}$ are spent in year $a + c$. Similarly, for February the proportions are $3/24^{ths}$ and $21/24^{ths}$, etc. The baseline sizes of the spring and fall cohorts can each be calculated as the sum across months of the product of births in that month and the proportion of life lived in year $a + c$ or year $a + c + 1$, as appropriate for that month and cohort. Full-year exposures can then be allocated to the Lexis cohorts not by dividing in half, but by multiplying by the percentage of births that were allocated to each respective Lexis cohort. This method does not account for immigration or emigration, and assumes that differences in death rates between the Lexis cohorts do not affect cohort sizes; nonetheless, it should yield a reasonable approximation of the exposures for each Lexis cohort, allowing calculation of Lexis-cohort-specific death rates (deaths/exposures).

Because the years for which monthly fertility data were available differed between France, Italy, and Switzerland, we ran the countries separately rather than including country as a level in a multilevel model. For each, we generated functions predicting mortality as a function of age and period using splines as outlined above. However, fertility data by month were not sex-specific, so we used total death rates and exposures and do not include a sex term in the model. We then used these functions to generate predicted mortality for each age and year of death; residuals were calculated as the difference between observed and predicted values. We then modeled these residuals as a function of spring versus fall birth (i.e. Lexis cohort) and calculated final residual values. Analyses up to this point were conducted in R; we then imported the residuals into WinBUGS, modeling the heterogeneity of residual mortality across cohorts as above but separately for each country (i.e., equation (1) with no β_2 terms).

Sensitivity Analyses

We ran multiple sensitivity analyses. In most cases, we found no substantial difference from the primary analysis. We present a selection of these results in the Online Appendix B. The one analysis that showed an important difference was excluding three of the Commonwealth countries (England and Wales, New Zealand, and Canada) that had apparently aberrant data for 1919 and 1920 (see Results). We present the multilevel and life expectancy analyses with and without those countries. Additional sensitivity analyses include: (a) varying initial parameters in the multilevel models; (b) using different age ranges (45-70 and 60-80) to check for more age-specific effects; (c) pooling the countries before running the age-period-sex model so that more country variation was preserved in the multilevel model; and (d) using cohort-within-country (equation (3)) and country-within-cohort (equation (2)) models instead of the crossed model (equation(1)).

$$m_{r(ikj)} = \beta_0 + \beta_{1(j)} + \beta_{2(kj)} + \varepsilon_{(ikj)} \quad (3)$$

Lastly, we worried that flu cohorts, who would have been young adults during World War II, might have suffered a specific mortality pattern then that could bias our results; for example, if World War II mortality were normally distributed across cohorts with a peak for the 1919 cohort, there could have been culling of the most healthy individuals (i.e., those fit for battle) in a way that would affect our estimates. We addressed this concern by running the multilevel model including only females and by using subsets of countries with low (<0.5%) or very low (<0.1%) mortality during the war (as a percent of 1939 population, as specified on Wikipedia

http://en.wikipedia.org/wiki/World_War_II_casualties). Again, there were no substantive differences and results are presented in Online Appendix B.

Results

Twenty-four-country mortality analysis

There was little evidence for differences in mortality rates among cohorts (Figure 1a; β_1s from equation (1)), though it is possible that mortality was slightly higher in the 1918 and 1920 cohorts but lower in the 1922 cohort. However, there appears to be substantial heterogeneity across countries within cohorts: out of 312 countries within cohorts, 42 had 95% credible intervals that did not include zero (22 above, 20 below, Figure 2; β_2s from equation (2)). False discovery rates are not generally a concern with Bayesian sampling methods.

Figure 2 also shows that the three lowest mortality rates are for the 1919 cohorts of England and Wales, New Zealand, and Canada; the same three countries have some of the highest mortality rates in the following year, 1920. We believe these results are likely a data artifact. There is no reason to suspect that 1920 mortality should be particularly high for these three countries, which, being members of the Commonwealth, would share more in terms of record-keeping than in terms of demography. If there were misallocation of births from 1920 to 1919 (either in the original records or based on smoothing done in the Human Mortality Database), 1919 and 1920 mortality rates would appear symmetrically low and high respectively, as we observe. We cannot definitively determine if this is an artifact; nonetheless, we felt it prudent to run the analysis excluding these countries. When we do, the 1919 residual mortality rates increase and the 1920 rates decrease, as expected (Figure 1b). Neither 1918 nor 1919 rates differ markedly from zero, and neither are as high as 1922 rates are low. Nonetheless, there does appear to be a weak trend for high mortality in the 1918-1919 cohorts, a pattern not seen when all 24 countries are included.

Lastly, we tested whether the 1918 and 1919 cohort-specific residual mortality rates within each country (subset of β_2s from the model in equation (3)) correlated with severity of the pandemic in each country as measured by excess flu mortality [cite Murray et al.]. The correlation was not significant in either year (1918: r=-0.29, p=0.23; 1919: r=0.27, p=0.26), so there does not appear to be an association with severity of the pandemic by country and late-life mortality rates.

Life expectancy

Life expectancy analyses confirmed the results of the mortality analysis (Figure 3). There were marginal negative effects of being born in 1918 or 1919 only when England and Wales, New Zealand, and Canada were excluded from the analyses (Figure 3b). Males and females showed similar patterns (Pearson's r=0.83). The largest effect observed was for females born in 1919 excluding the three Commonwealth countries, about 19 days less life expectancy. If we extrapolate that roughly half of babies born in 1919 were *in utero* for substantial periods when the pandemic was raging, that roughly one-third of these were actually exposed to the flu, and that all of the 1919 effect is to the pandemic, we can calculate that life expectancy is lowered 114 days by exposure to the flu *in utero*. However, this is a maximum estimate: it is not clear that the low life expectancy of the 1919 cohort is any different than might be expected based on random yearly variation (as opposed to the flu) nor that we should exclude the Commonwealth countries, and even if the effect is real the estimate for males is only about half as large.

France, Italy, and Switzerland with half-year Lexis cohorts

Using the Lexis cohorts described above and the appropriate adjustments for fertility throughout birth year, we were able to test for differences in mortality across half-year cohorts, potentially detecting effects of the influenza pandemic on late-life mortality that are not apparent at a one-year time scale. As in the above analysis, some cohorts in some countries had mortality that differed significantly from the overall average for that country (Figure 4). Across the three countries, there is no consistent trend for high or low mortality in the three cohorts of primary interest (second half of 1918, both halves of 1919). In France (Fig. 4a) all of these three cohorts have 95% credible intervals spanning zero, and in Italy (Fig. 4b) the first half of 1919 appears to have low mortality but is well within the range of variation seen across other cohorts. In Switzerland (Fig. 4c), the second half of 1918 and first half of 1919 have high mortality and the second half of 1919 has low mortality, with the latter two cohorts having the largest mean differences from the overall average of any cohorts in any of the three countries. However, the magnitude of the deviance is similar for both, and we cannot exclude the possibility that the apparent effect is due to some estimation problem of the proper cohort sizes for that year. A similar issue could account for the symmetrically high and low mortality rates of the two 1921 cohorts in Italy.

Discussion

We failed to detect any consistent effect of the 1918-1919 influenza pandemic on late-life mortality for cohorts exposed to the pandemic early in life. Because so much of the population was exposed to the flu that year – roughly 25% in the US and 45% in Norway, for example [25, 26] – any major effect on late-life mortality should have been apparent from the population-level data we used. Moreover, the flu struck young adults, neonates, and the elderly most heavily [27], implying that incidence was higher still among the fetuses (via pregnant women) and neonates that were of primary interest in this study. We made varying predictions that would have been supported by positive or negative effects, differential effects on neonates and those *in utero* during the pandemic, or some effect on all cohorts born before the pandemic, with strongest effects in the children youngest in 1918-1919. None of these patterns were apparent in our data.

There was potential evidence of an effect of the pandemic on late-life mortality when we excluded three Commonwealth countries with apparently spurious data – England and Wales, New Zealand, and Canada – from the analysis. (More detail on why we distrust the data from these countries is available in Online Appendix C.) The trend was still not significant at α=0.05 and the effect was small (less than 20 days lost life expectancy for females born in 1919), but we cannot exclude the possibility of a weak effect. There was also high mortality in Switzerland for those born in the second half of 1918 and first half of 1919, but low mortality for those born in the second half of 1919. This would be consistent with negative effects on the long-term health of neonates but a selection effect on those early *in utero* during the pandemic. However, the symmetry of the 1919 effects suggests they could also be attributable to the same error in estimating the relative sizes of those cohorts that we believe biased the estimates in the three Commonwealth countries. Similar effects were not apparent in France or Italy, so it would be premature to draw any strong conclusions from the Swiss data. Moreover, the Swiss data does not confirm the Commonwealth-excluded analysis because the effects for the two halves of 1919 cancel each other out: none of the three countries we examined with Lexis cohorts bolsters the case for a weak overall 1919 effect.

There are a number of reasons we might have failed to detect major effects of the flu pandemic on late-life mortality. Most obviously, inflammation early in life may not systematically affect late-life mortality. Our results are certainly not sufficient to conclude this, but they at least raise the possibility that lifelong inflammation patterns are not set by developmental events. Second, there might be an effect of early inflammation on late-life mortality that is too weak to detect. A several-week bout of acute illness may not be sufficient to induce detectable changes in lifelong inflammation levels even if those levels are affected by peri-natal levels. If this is the case, it is unlikely that any specific childhood disease would yield major inflammation sequellae that would last through adulthood, but it is possible that environmental factors that affect general propensity to disease could have such an effect. Lastly, it is possible that there are both positive and negative effects on health that cancel each other out in the net population statistics. As noted above, negative effects could arise due to allocation of resources that negatively affect development or due to high inflammation levels; positive effects could arise from selection or from priming of the immune system. At a disease-specific level, it could be that any increases in mortality due to heart disease are balanced by, say, decreases in cancer rates that might result from better functioning of the immune system against cancers. While it is not possible to say which of these explanations accounts for the lack of association we document here, it is clear that simplistic hypotheses about early exposure to disease having large effects on late-life mortality are not supported by this analysis.

Our results stand in contrast to several studies showing long-term effects of the flu. Azambuja [28] and Mamelund [29] showed potential negative health consequences for those who were adults or young adults when the pandemic hit. However, the lack of clear age specificity of such proposed effects makes it hard to fully distinguish period and cohort effects, and the trends must remain suggestive rather than conclusive of the flu as a causal mechanism. In a careful analysis incorporating both quarter of birth and state, Almond [16] showed consistently poorer outcomes for the flu cohorts in the 1960, 1970, and 1980 censuses on economic measures including educational attainment, annual income, neighbors' income, and disability status. A similar analysis of health outcomes such as self-reported health, history of stroke or diabetes, and functional impairment also found some worse outcomes for flu cohorts, but in many cases the effects were not consistent among flu cohorts born in different quarters, even for relatively similar measures such as Trouble Lifting or Trouble Walking At All [17]. Because worse economic and health outcomes should lead to higher mortality, it is somewhat surprising that we did not detect such effects in mortality.

There are several potential explanations for the discrepancy between our results and these other studies. First, many of the economic outcomes could be determined by factors earlier in life: educational attainment, for example, is almost certainly not caused by health later in life, since nearly all individuals in these cohorts would have reached their maximum level of education by their mid-twenties. While it is conceivable that early health status could affect both educational attainment and late-life health, there is not a requisite causal link. Second, if the causal pathway is Poor Early Health $\rightarrow$ Low Educational Attainment $\rightarrow$ Low Economic Status $\rightarrow$ Poor Health $\rightarrow$ High Mortality or something similar, we should expect the effects to be diluted with each successive causal link (arrow), with the link to mortality being hardest to detect. Third, mortality is a composite of many processes, and it is quite possible that countervailing effects might average themselves out here (e.g. lower cancer rates but higher heart disease) in ways that are not predicted by the economic measures. Fourth, it is possible there was a real effect in the U.S. but no effect in all the countries aggregated together. In the U.S., mortality was

significantly higher in the 1918 cohort (t=6.67, p<0.0001) and lower in the 1919 cohort (t=-3.65, p=0.0005) compared to all other cohorts. There was no net effect pooling the two cohorts (t=1.08, p=0.28), but the 1918 effect was larger, potentially indicating a pattern in the U.S. that we missed by using one-year cohorts.

Although it was not our goal to test more generally for cohort effects, our data show clear evidence of fine-scale differences among cohorts. Within France, Italy, and Switzerland, many of the cohorts had mortality significantly above or below average. Forty-two of 312 country-cohort combinations also had mortality above or below the overall average. While the differences among cohorts in the crossed model were not striking, the heterogeneity was significant. Lastly, the strong correlation between cohort-specific male and female life expectancy suggests real trends rather than random deviations. It is possible that the cohort effects are an artifact of imperfect control for age and period, but given that we detect them in multiple analyses and that our age-period-sex model explained more than 99% of the variation, we view this as unlikely. Real cohort effects are not likely attributable to events later in life, since there are few events that would discretely affect one age-class but not another adjacent to it. The events that do have such discrete effects are almost always legal in nature and thus not generalizable across countries. The cohort effects we detect are thus likely attributable to environmental factors very early in life. It is paradoxical that we detected no specific effect of the flu, one of the most salient early-life environmental factors on record, but detected more general evidence for such effects operating at a very fine scale. It is not clear what sorts of variation are causing these cohort effects.

Finally, we believe that the Lexis cohort approach we developed here may be useful in other situations for distinguishing finer-scale cohorts than are initially apparent in the data. For example, several studies have shown that fall babies have higher life expectancy than spring babies [the "Doblhammer effect," 30, 31], but because linked date-of-birth and date-of-death data are hard to acquire, the effect has only been shown in several countries. A Lexis cohort approach should be able to substantially increase the number of countries and time periods for which inferences can be made, allowing analysis of historical trends in the strength and distribution of this effect. We are currently pursuing these analyses.

Acknowledgments

We would like to thank Karen Bandeen-Roche, members of the Johns Hopkins Center on Aging and Health, and [XX] anonymous reviewers for comments on the methods and manuscript. AAC and JT were supported by post-doctoral fellowships from the National Institute on Aging's Behavioral and Social Research program.

Literature Cited

1 Barker DJP. The fetal and infant origins of adult disease. *British Medical Journal* 1990;**301**:1111.

2 Barker DJP. The fetal origins of coronary artery disease. *European Heart Journal* 1997;**18**:883-4.

3 Ozanne SE. Diabetes in Old Male Offspring of Rat Dams Fed a Reduced Protein Diet. *International Journal of Experimental Diabetes Research* 2001;**2**:139-43.

4 Finch CE, Crimmins EM. Inflammatory Exposure and Historical Changes in Human Life-Spans. *Science* 2004;**305**:1736 - 9.

5 Kannisto V, Christensen K, Vaupel JW. No increased mortality in later life for cohorts bom during famine. *American Journal of Epidemiology* 1997;**145**:987-94
6 Ravelli A, van der Meulen J, Michels R*, et al.* Glucose tolerance in adults after prenatal exposure to famine. *Lancet* 1998;**351**:173-7.
7 Stanner SA, Bulmer K, Andrès C*, et al.* Does malnutrition in utero determine diabetes and coronary heart disease in adulthood? Results from the Leningrad siege study, a cross sectional study. *British Medical Journal* 1997;**315**:1342–8.
8 Johnson NPAS, Mueller J. Updating the accounts: global mortality of the 1918–1920 "Spanish" influenza pandemic. *Bulletin of Historical Medicine* 2002;**76**:105–15.
9 Murray CJL, Lopez AD, Chin B*, et al.* Estimation of potential global pandemic infl uenza mortality on the basis of vital registry data from the 1918–20 pandemic: a quantitative analysis. *Lancet* 2006;**368**:2211-8.
10 Crimmins EM, Finch CE. Infection, inflammation, height, and longevity. *Proceedings of the National Academy of Sciences (USA)* 2006;**103**:498-503.
11 Stearns SC. The evolutionary significance of phenotypic plasticity: phenotypic sources of variation among organisms can be described by developmental switches and reaction norms. *BioScience* 1989;**39**:436-45.
12 Holt PG. Postnatal maturation of immune competence during infancy and childhood. *Pediatric Allergy and Immunology* 1995;**6**:59 - 70.
13 Mauck RA, Huntington CE, Grubb TC. Age-specific reproductive success: evidence for the selection hypothesis. *Evolution* 2004;**58**:880-5.
14 Gurven M, Kaplan H, Winking J*, et al.* Aging and Inflammation in Two Epidemiological Worlds. *The Journals of Gerontology;* 2008;**63A**:196-9.
15 Folkerts G, Walzl G, Openshaw PJM. Do common childhood infections 'teach' the immune system not to be allergic? *Immunology Today* 2000;**31**:118-20.
16 Almond D. Is the 1918 influenza pandemic over? Long-term effects of *in utero* influenza exposure in the post-1940 U.S. population. *Journal of Political Economy* 2006;**114**:672-712.
17 Almond D, Mazumder B. The 1918 influenza pandemic and subsequent health outcomes: An analysis of SIPP data. *American Economics Review* 2005;**95**:258-62.
18 Calot G. *Two centuries of Swiss demographic history*. Neuchâtel and Paris: Swiss Federal Statistical Office and Observatoire démographique européen 1998.
19 Huber M. *La Population de la France Pendant la Guerre*. New Haven, CT: Yale University Press 1931.
20 Mortara G. *La Salute Pubblica in Italia Durante e Dopo la Guerra*. New Haven, CT: Yale University Press 1925.
21 Caselli G, Capocaccia R. Age, period, cohort and early mortality: An analysis of adult mortality in Italy. *Population Studies* 1989;**43**:133-53.
22 Horiuchi S. The long-term impact of war on mortality: old-age mortality of the First World War survivors in the Federal Republic of Germany. *Population Bulletin of the United Nations* 1983;**15**:80-92.
23 Team RDC. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing 2007.
24 Spiegelhalter D, Thomas A, Best N*, et al.* WinBUGS. http://www.mrc-bsu.cam.ac.uk/bugs 2007.
25 Mamelund S-E. Can the Spanish influenza pandemic of 1918 explain the baby boom of 1920 in neutral Norway? *Population* 2004;**59**:229-60.

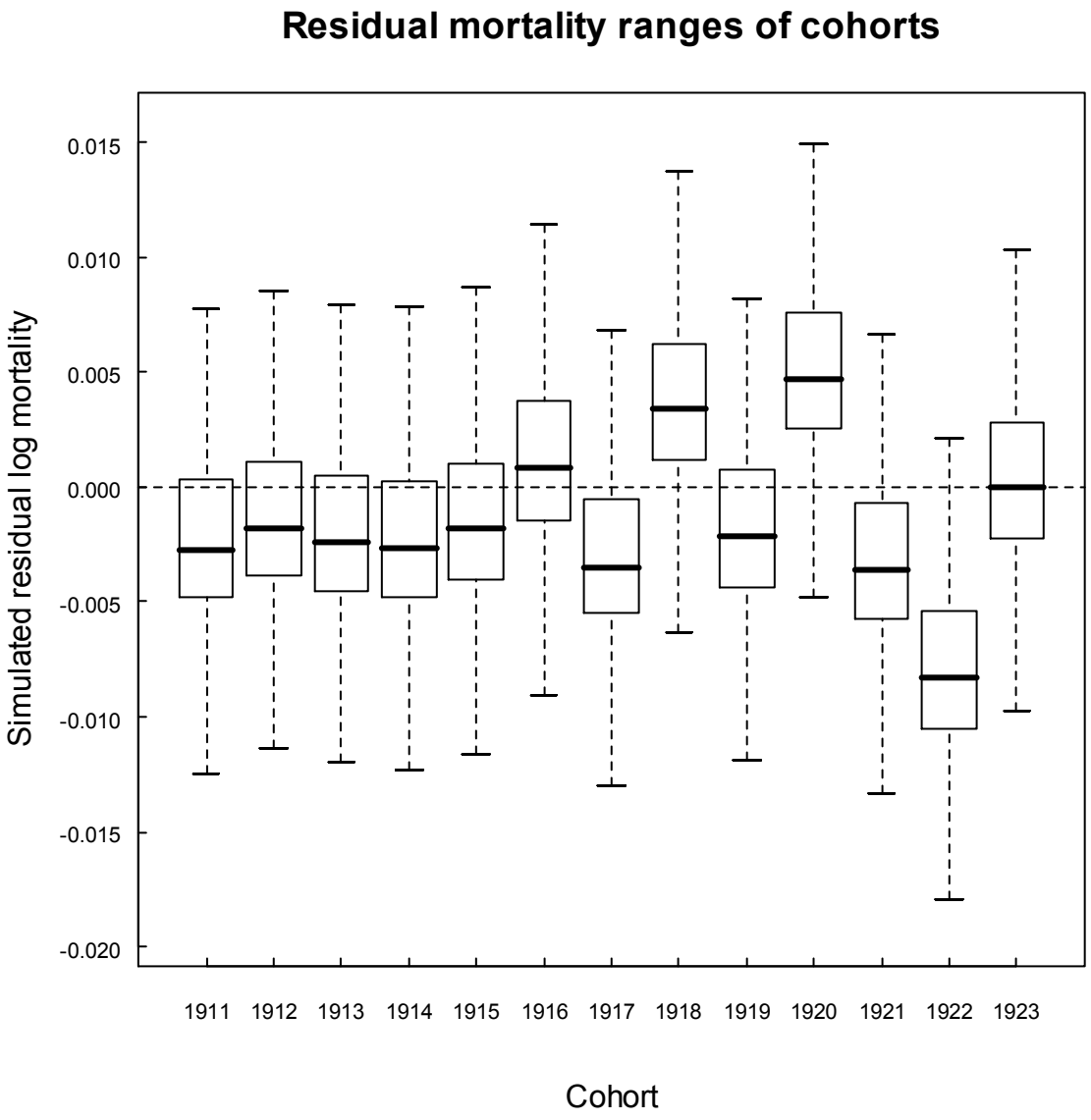
26 Crosby AW. *America's forgotten pandemic: the influenza of 1918*. New York: Cambridge University Press 1990.
27 Noymer A, Garenne M. The 1918 influenza epidemic's effects on sex differentials in mortality in the United States. *Population and Develiopment Review* 2000;**26**:565-81.
28 Azambuja MIR. Spanish flu and early 20th-century expansion of a coronary heart disease-prone subpopulation. *Texas Heart Institute Journal* 2004;**31**:14-21.
29 Mamelund S-E. Effects of the Spanish influenza pandemic of 1918-19 on later life mortality of Norwegian cohorts born about 1900. Oslo: Department of Economics, University of Oslo 2003.
30 Doblhammer G. *The late life legacy of very early life*. Berlin: Springer 2004.
31 Doblhammer G, Vaupel JW. Lifespan depends on month of birth. *Proceedings of the National Academy of Sciences (USA)* 2001;**98**:2934-9

Table 1: Countries used for this analysis and their data ranges

	Country	Min age	Max age	Cohort range	Earliest period data
1	Australia	45	80	1911-1923	1921
2	Austria	45	81	1911-1923	1947
3	Bulgaria	45	81	1911-1923	1947
4	Canada	45	80	1911-1923	1921
5	Switzerland	45	82	1911-1923	1876
6	Czech Republic	45	80	1911-1923	1950
7	Denmark	45	82	1911-1923	1835
8	England and Wales	45	79	1911-1923	1841
9	Spain	45	81	1911-1923	1908
10	Finland	45	82	1911-1923	1878
11	France	45	81	1911-1923	1899
12	West Germany	45	80	1911-1923	1956
13	East Germany	45	80	1911-1923	1956
14	Hungary	45	81	1911-1923	1950
15	Italy	45	80	1911-1923	1872
16	Japan	45	82	1911-1923	1947
17	The Netherlands	45	82	1911-1923	1850
18	Norway	45	82	1911-1923	1876
19	New Zealand	45	79	1911-1923	1846
20	Portugal	45	81	1911-1923	1940
21	Slovakia	45	82	1911-1923	1950
22	Sweden	45	82	1911-1923	1751
23	Belgium	45	81	1911-1923	1841
24	USA	45	80	1911-1923	1933

Numbers are used in subsequent figures to identify countries.

Figure 1a


Residual mortality ranges of cohorts
Simulated residual log mortality
0.015
0.010
0.005
0.000
-0.005
-0.010
-0.015
-0.020
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
Cohort

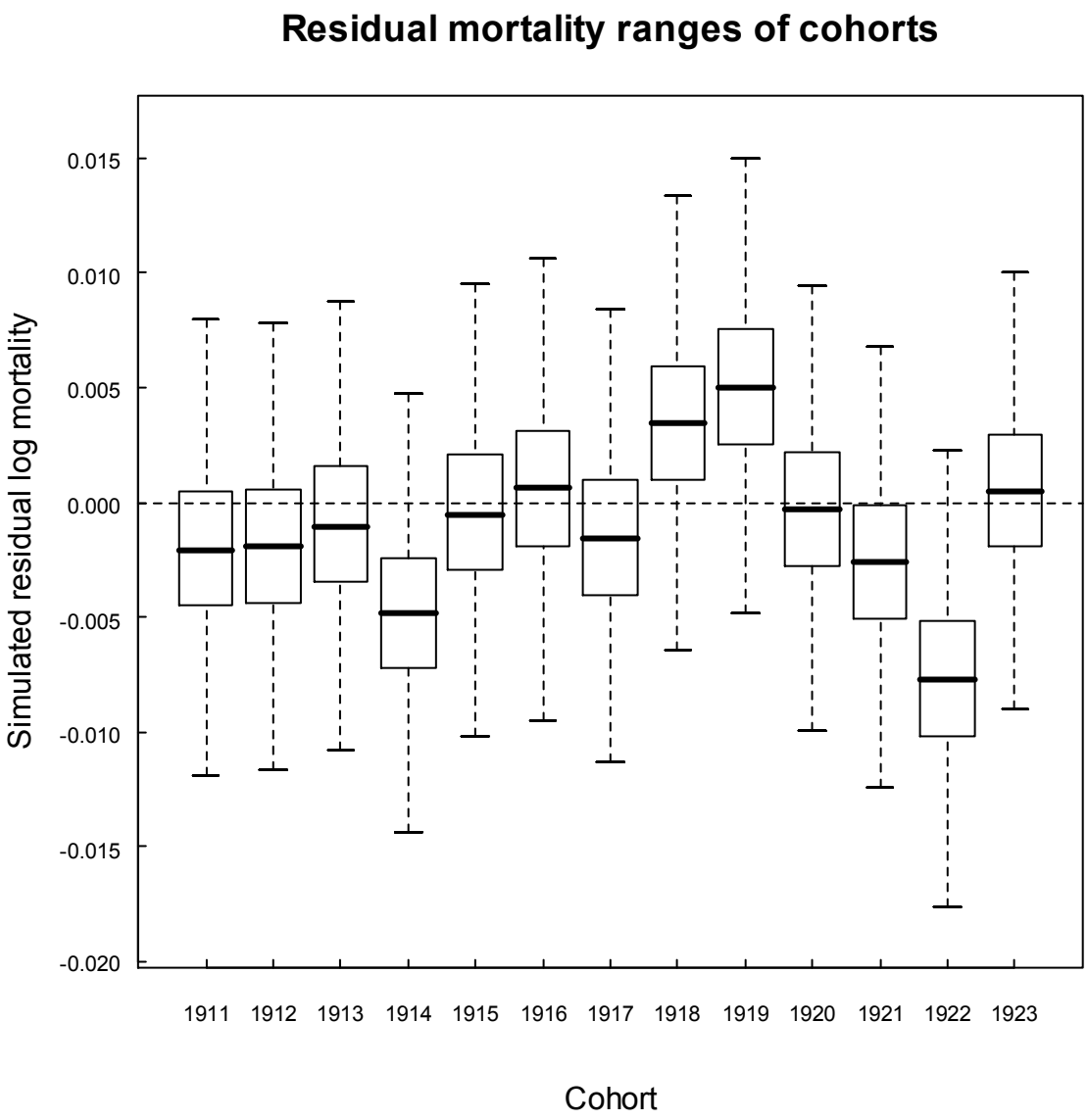


Figure 1b

Figure 1: Estimated Bayesian deviations of residual mortality rates with 50% credible intervals (boxes) and 1.5 times interquartile range (whiskers) for the cohorts 1911-1923 a) for all 24 countries, and b) for 21 countries, excluding England and Wales, New Zealand, and Canada.

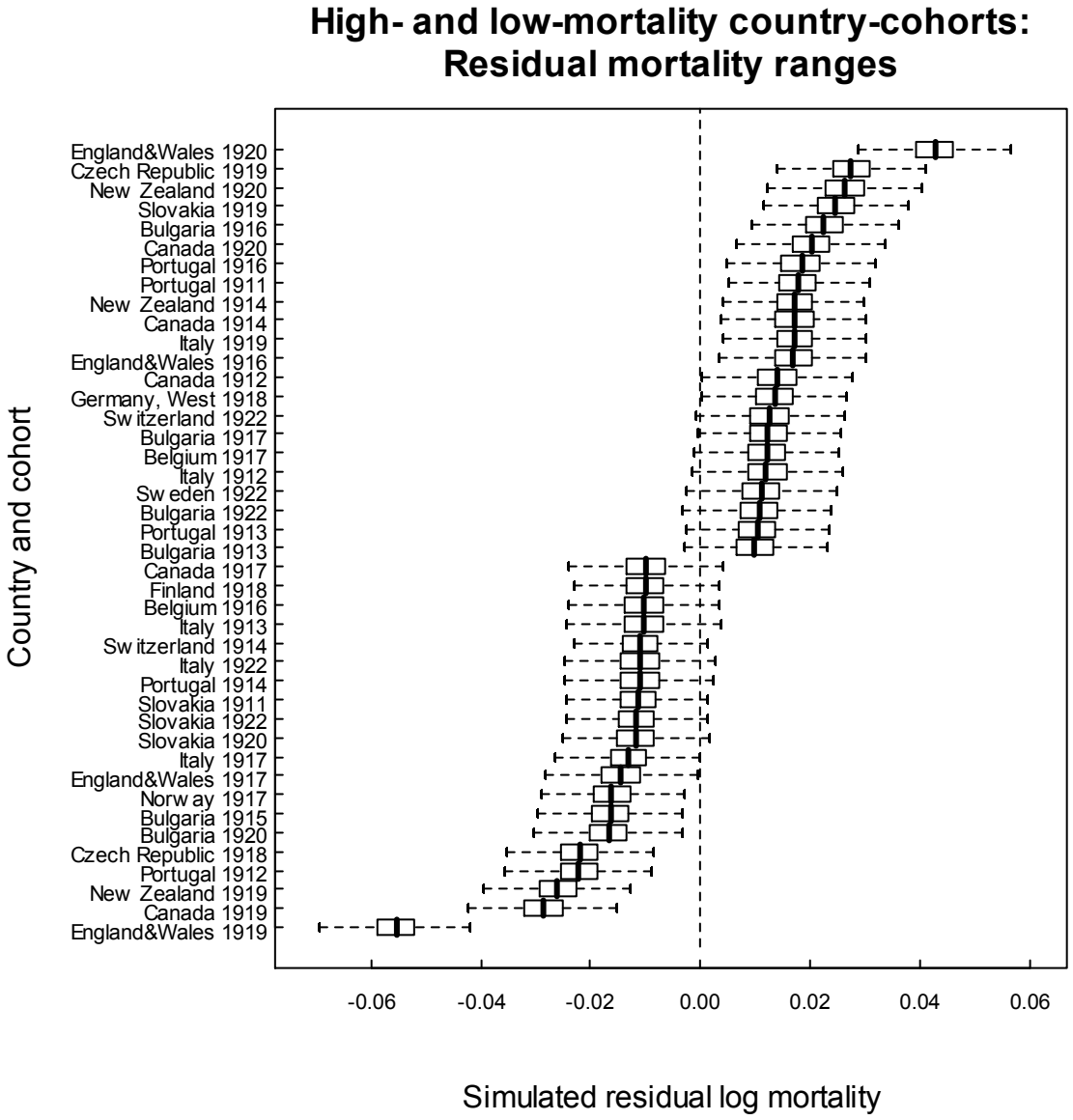


Figure 2: Estimated Bayesian deviations of residual mortality rates with 50% credible intervals (boxes) and 1.5 times interquartile range (whiskers) for the 42 country-cohort combinations whose 95% credible intervals do not span zero. The 270 country-cohorts that do not differ from zero are not shown here, but would fall in the center of the figure between Bulgaria 1913 and Canada 1917. Estimates are from the country-within-cohort model (equation (2) in text), and show substantial heterogeneity.

Figure 3a

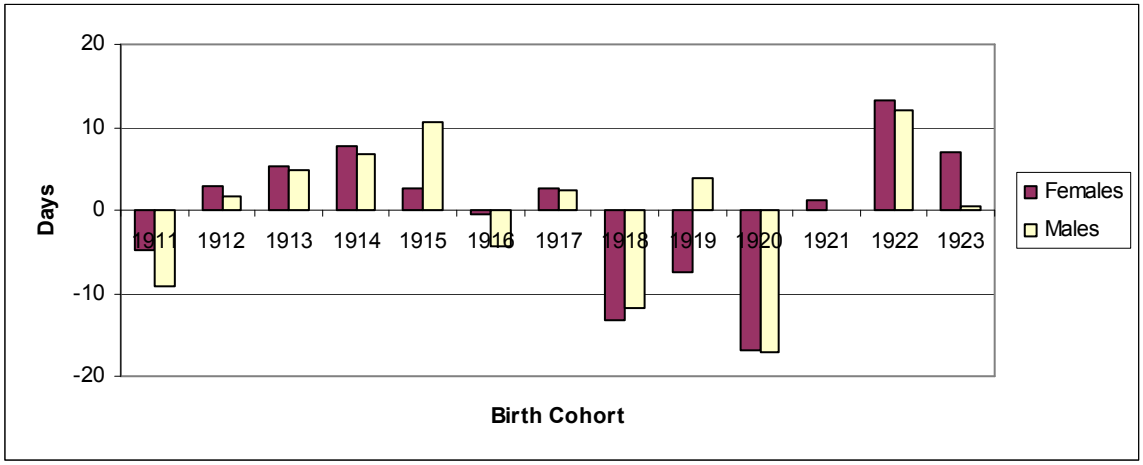


Figure 3b

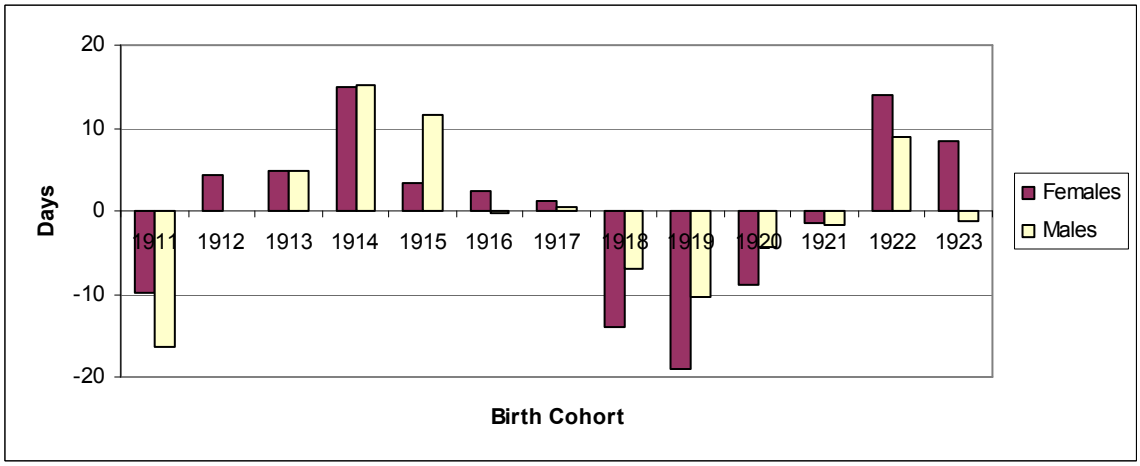


Figure 3: Mean residual life expectancy (by cohort, averaged across countries) for a) all 24 countries and b) 21 countries excluding England and Wales, New Zealand, and Canada.

Figure 4a France 1914-1919

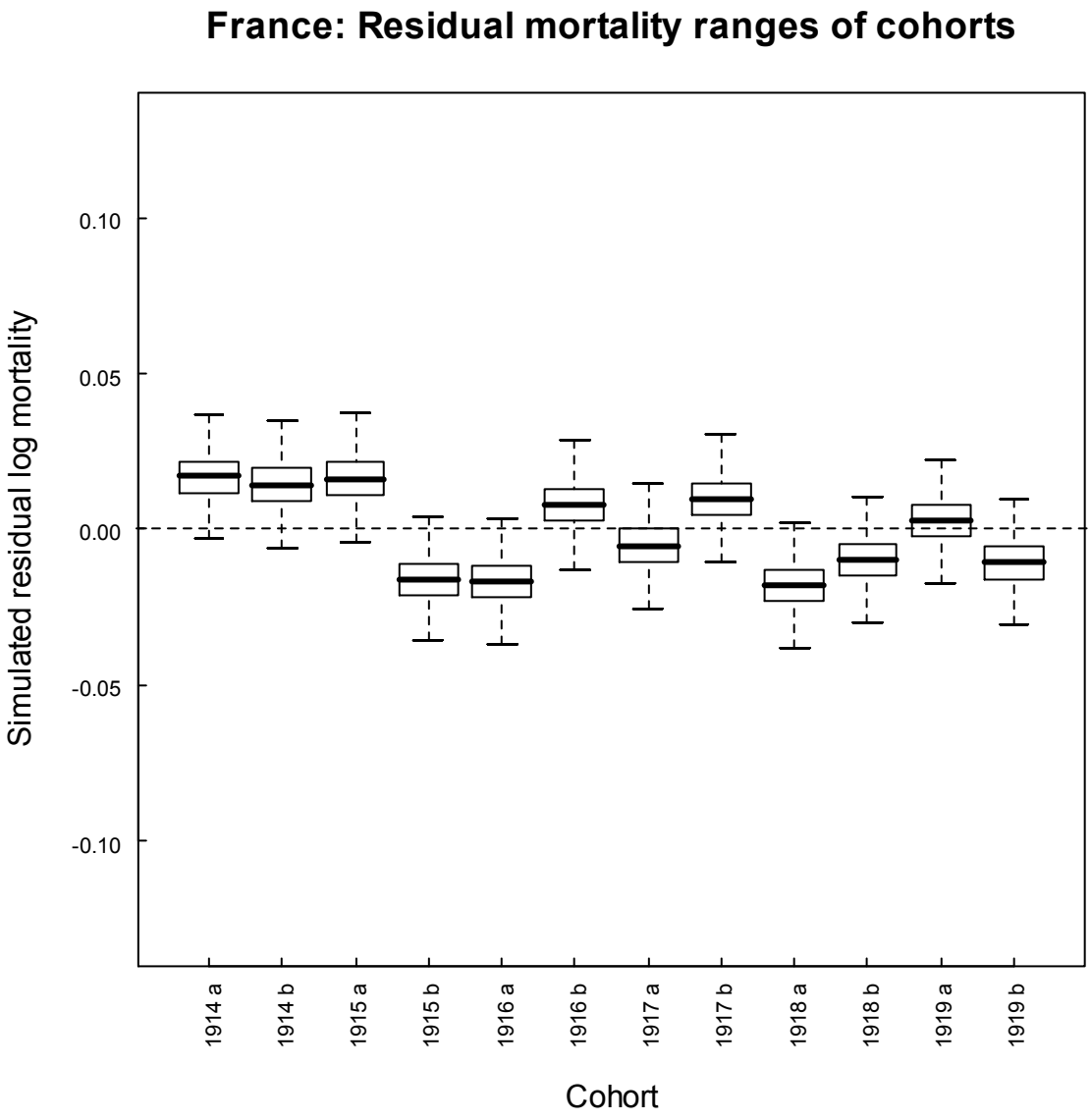


Figure 4b Italy 1911-1923

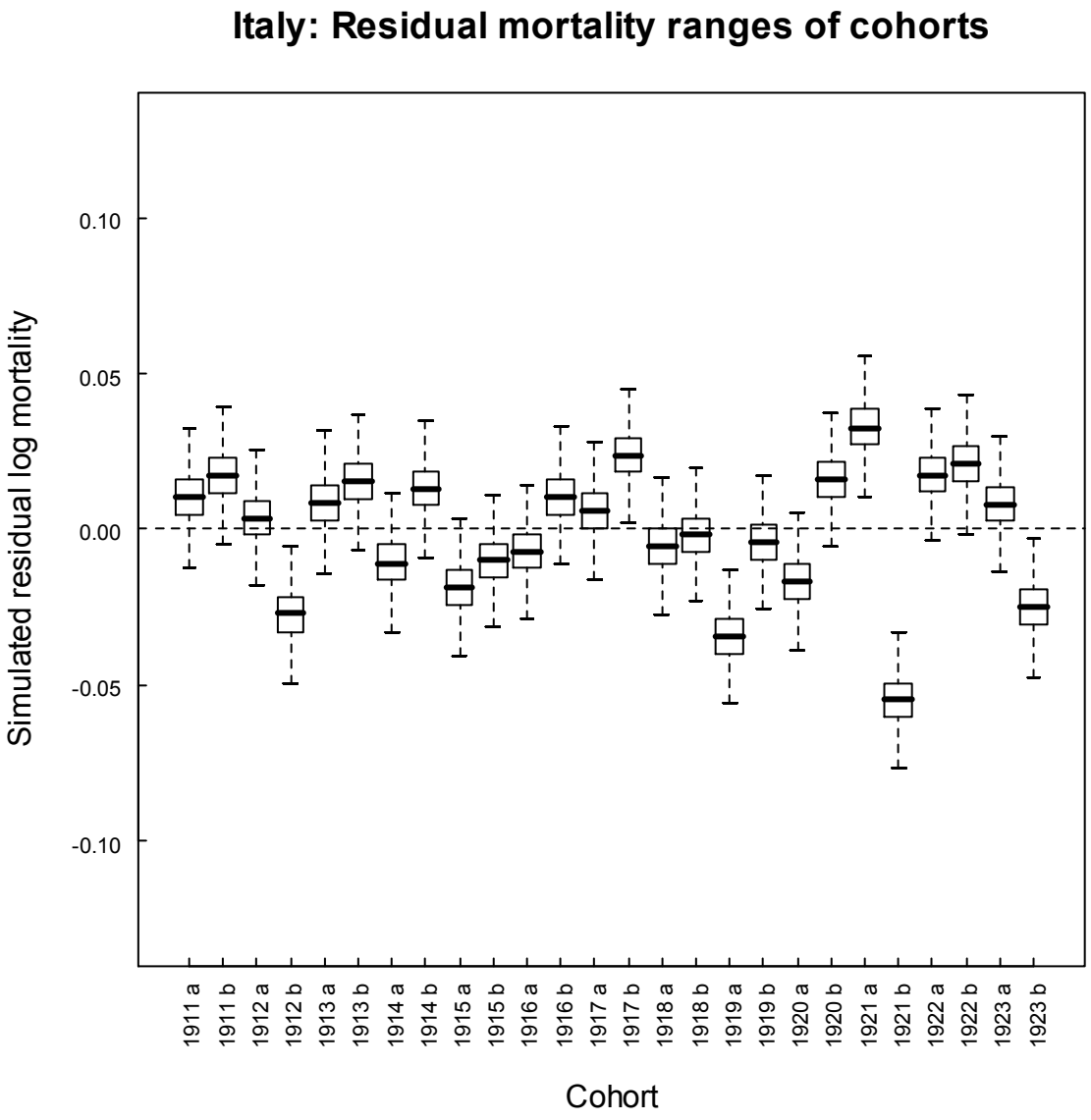


Figure 4c Switzerland 1910-1923

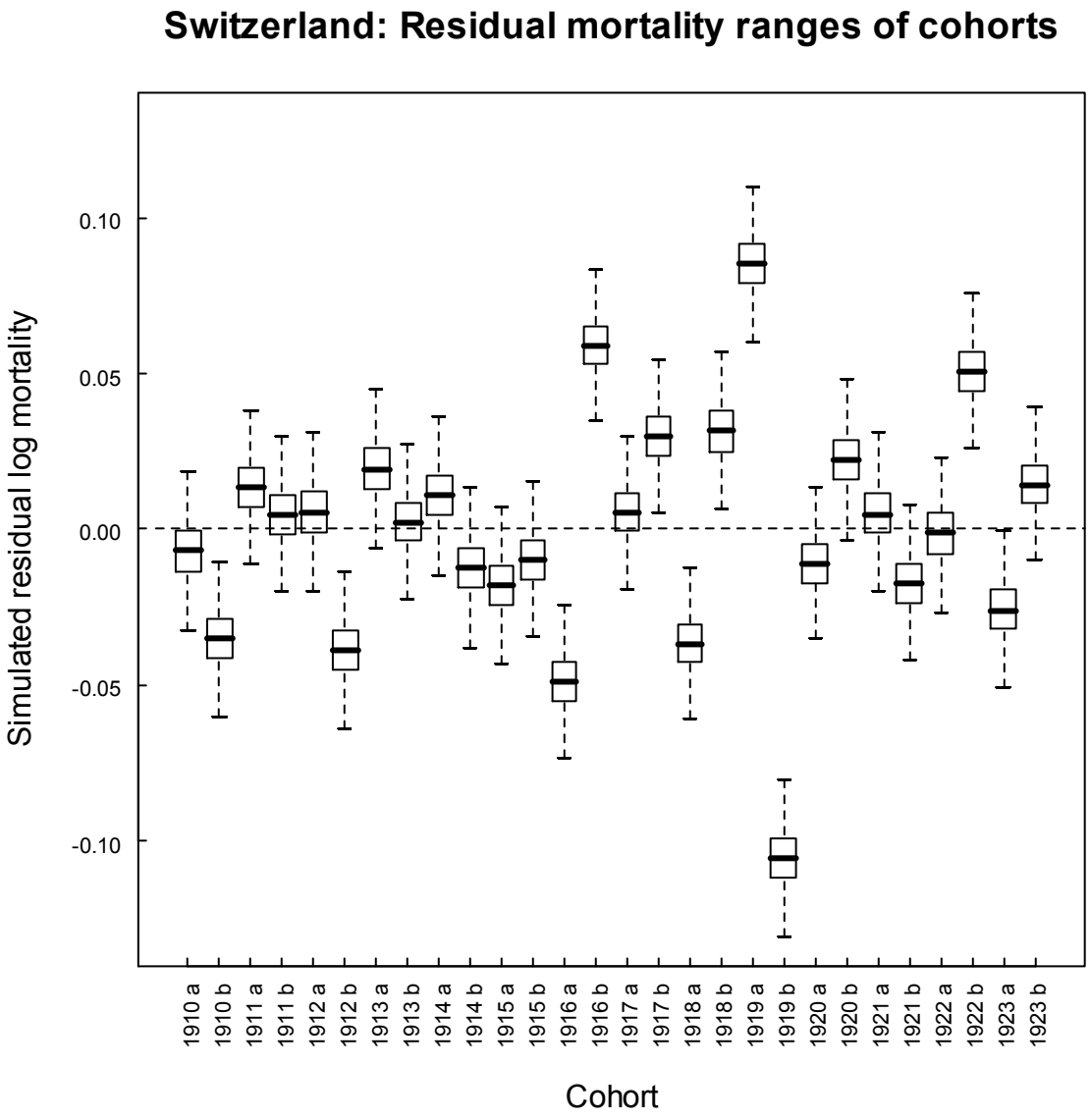


Figure 4: Estimated Bayesian deviations of residual mortality rates with 50% credible intervals (boxes) and 1.5 times interquartile range (whiskers) for a) France 1914-1919; b) Italy 1911-1923; and c) Switzerland 1910-1923.

Appendix A:
Lexis Cohort Methodology and Details

Traditional demographic analyses often rely on making a clear choice between using period and cohort analyses. If we have data on death year and age at death, we have period data; if we have data on birth year and age we have cohort data (Fig. A1). The difference arises because a person born in year *t* (say, 1918) who dies at age *a* (say, 60) could die in either year *a*+*t* or *a*+*t*+1 (1978 or 1979), depending on when the birthday fell in the birth year and how close the person made it to age *a*+1. For example, someone dying at age 60.95 is still recorded as dying at age 60, but probably died in 1979. When we wish to assess the effect of being born in a given year, we use cohort data; when we wish to assess the effect of events during a year on deaths across age-classes, we use period data.

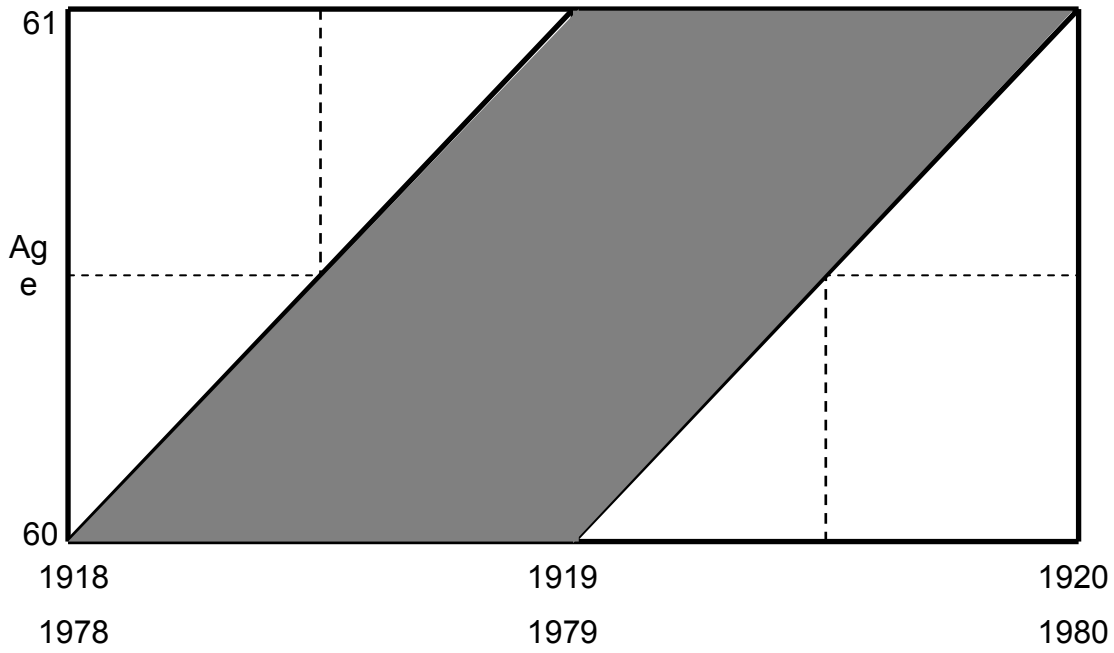


Fig. A1: Lexis diagram of what is known with cohort and age at death ("Cohort data" from the Human Mortality Database).

The problem that can arise here is one of scale of the divisions among cohorts, periods, or age classes. We might wish to detect a pattern with a clear threshold point that may be partway through a year, or with cyclical effects operating more rapidly than can be detected by dividing the data up by year. So we'd like half-year cohorts (Fig. A2). If we have only period and cohort data by year, there is generally no way to get information on a finer scale. However, sometimes we have Lexis data available, in which case both birth year and death year are present. This provides some additional information that we should be able to utilize for detection of finer scale patterns (Fig. A3); the topic of this appendix is how to maximize the information we can gain by utilizing Lexis data.

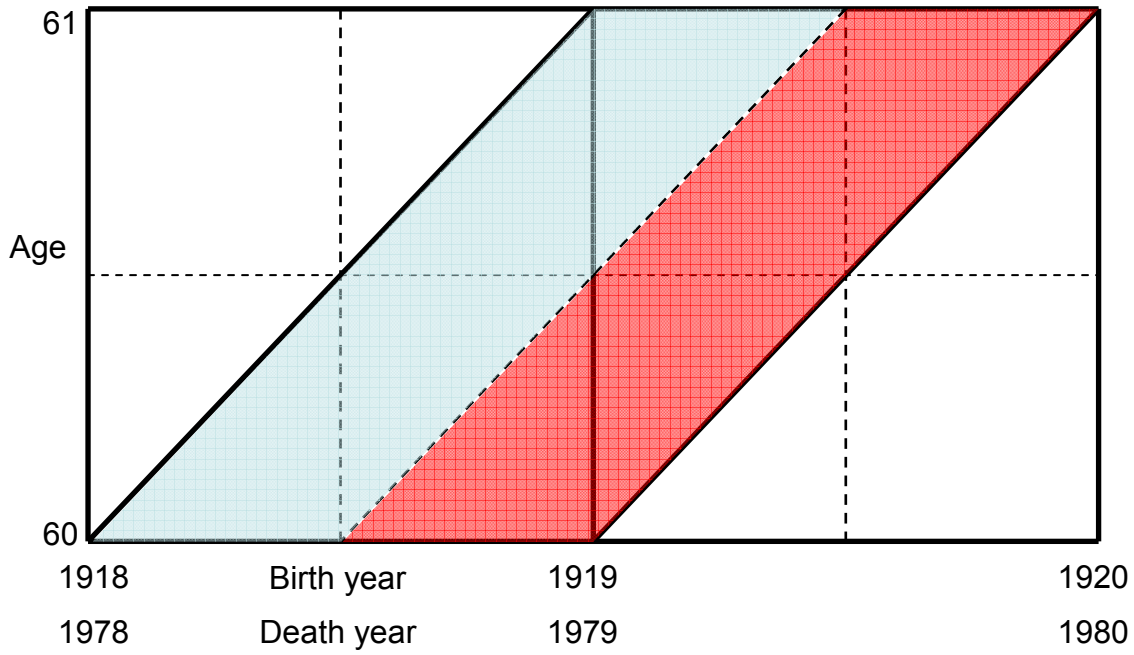


Fig. A2: A Lexis diagram of what we'd like to know: half-year cohorts

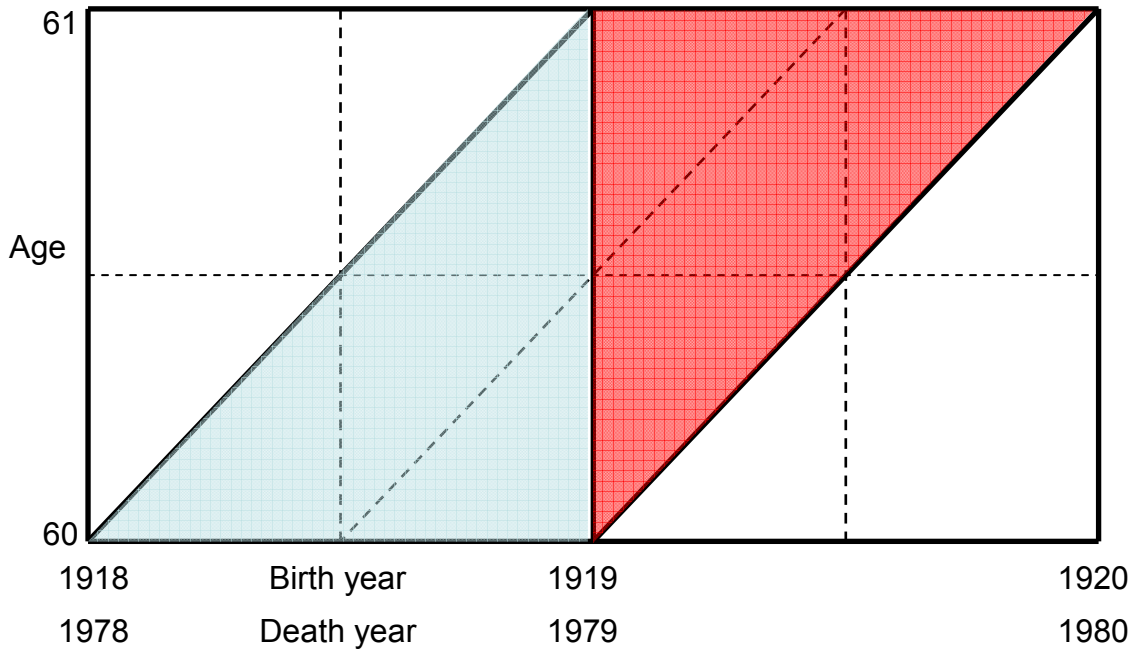


Fig. A3: Lexis diagram of what is known with cohort, year of death, and age at death ("Lexis data" from the Human Mortality Database).

At first glance, it is not immediately apparent how we can get more information. For example, our 1918-born 60-year-old can now be definitively assigned to, say, 1979 as a death year, but we cannot say with certainty when in 1918 she was born or when in 1979 she died. She could have been born early in 1918 (if she died even earlier in 1979) or have died late in 1979 (if she was born even later in 1918). So, unlike our source data,

we cannot use absolute knowledge of birth or death at finer scale than a year; at best, we can get probabilistic information.

However, the probabilistic information should be relatively accurate, especially if we make two assumptions: (1) that there are roughly equal numbers of births on each day of a year, and (2) that age-specific mortality rates are constant throughout an age-class (e.g., chance of dying at 60.0 is the same as at 60.9). Of course, neither assumption is strictly true, but they are both approximately true, and as we will see their violation should not undermine the validity of our analysis under most circumstances.

The basic insight that allows us to get probabilistic information is that a person born on Jan. 1 of a year is much more likely to die in the first than the second possible death year. If our example subject were born Jan. 1, 1918 the only way she could die in 1979 at age 60 would be to die on Jan. 1 earlier in the day than the time she was born – highly unlikely relative to the probability of dying sometime in 1978. Given that we only know her age at death to the year, we can calculate this probability. If we know that she was born on Jan. 1, 1918 at 11:59 pm and died at age 60, there is a 1/365 chance that she died in 1981 and a 364/365 chance that she died in 1978 (again, assuming constant death rates from age 60.00 to age 60.99). If she were born on Jan. 2 at 11:59 pm, these probabilities would shift to 2/365 and 363/365 respectively, and so on. The probabilities always sum to 1, of course, and they are linear functions of birth date within a year (Fig. A4).

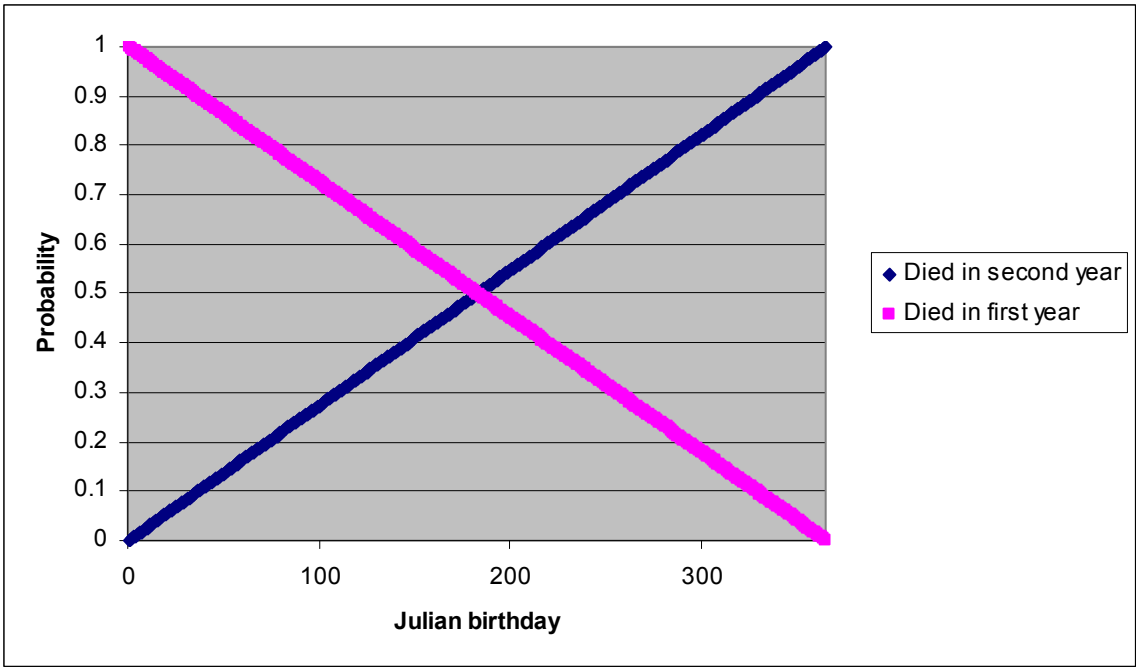


Figure A4: Probability of different Julian birthdays (i.e., days in the year numbered 1-365) for those dying in the earlier or later possible year for a given age and cohort

All we know from our Lexis data is that someone died in the earlier or later of the two possible years. However, even with just this knowledge, we can establish the probability distributions of date of birth for these two groups as the lines in Fig. A4. As can be seen, these distributions overlap but are markedly different. The mean date of birth for those who died in the first year is approximately May 1 (1/3 of the way through the year) and

for those who died in the second year is Sept. 1 (2/3 of the way through the year). We thus can effectively create two probabilistic "cohorts" with known mean birthdays, what we call "Lexis cohorts." If we conduct analyses on these Lexis cohorts, we can now see patterns that emerge at finer scales than one year (Fig. A5). Our precision will of course not be as good as if we had two distinct cohorts born with certainty in the first and second half of the year (the precision is 75%, to be precise, Fig. A5), but if the effects we are looking for are strong enough we should still be able to detect them.

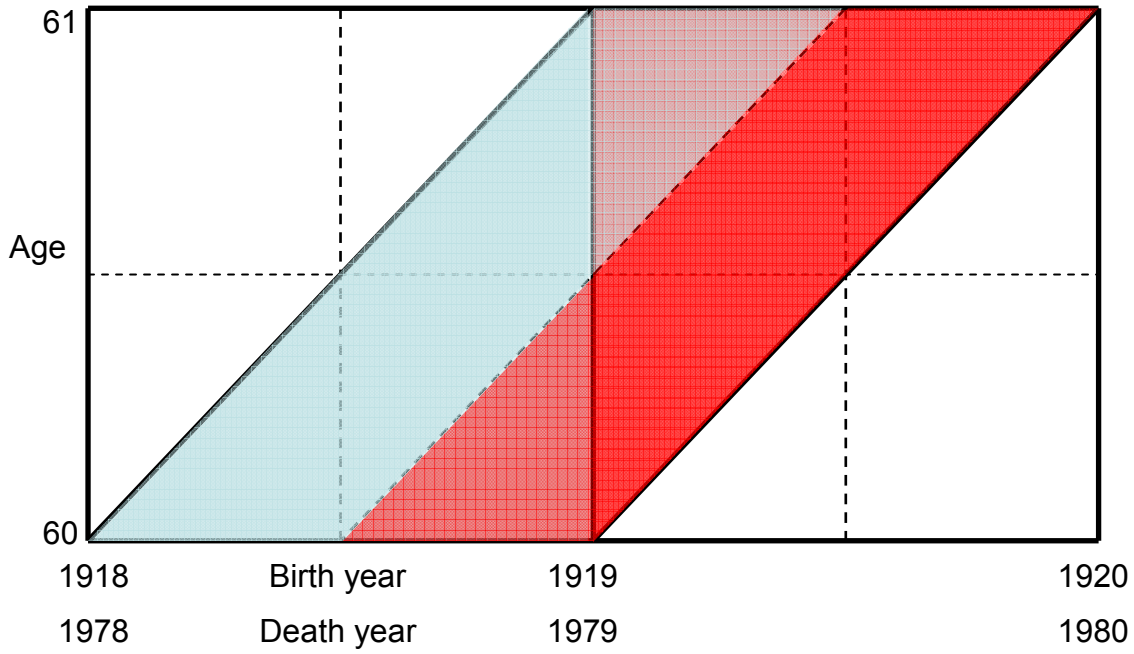


Fig. A5: A Lexis diagram of the 75% overlap between the "Lexis cohorts" and true half-year cohorts.

Relaxing the assumptions

One assumption of the probability distributions for our two cohorts is that there is no seasonality to birth rates. However, seasonality of births will not affect the probability of each year of death for a given birthday; it affects only the distribution of the birthdays. This is easily incorporated into a model. Fig. A6 uses a cosine function to illustrate the sort of effect seasonality could have on our probability distributions. A less symmetric seasonality could bias the original estimates somewhat (Fig. A7). In all cases, seasonality could affect estimates of mean birthday for the two cohorts, but should not be strong enough to change the fundamental difference between an earlier and a later cohort that are approximately 50% distinct.

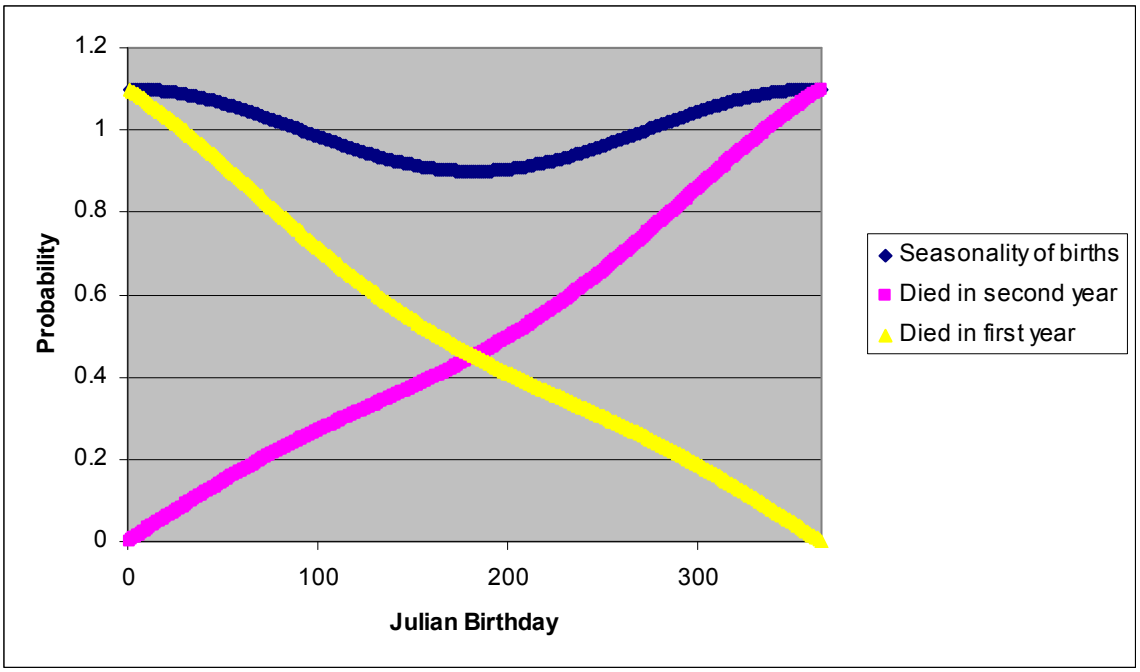


Figure A6: Probability of different Julian birthdays (i.e., days in the year numbered 1-365) for those dying in the earlier or later possible year for a given age and cohort assuming seasonality of birth following a cosine function symmetrical around mid-year.

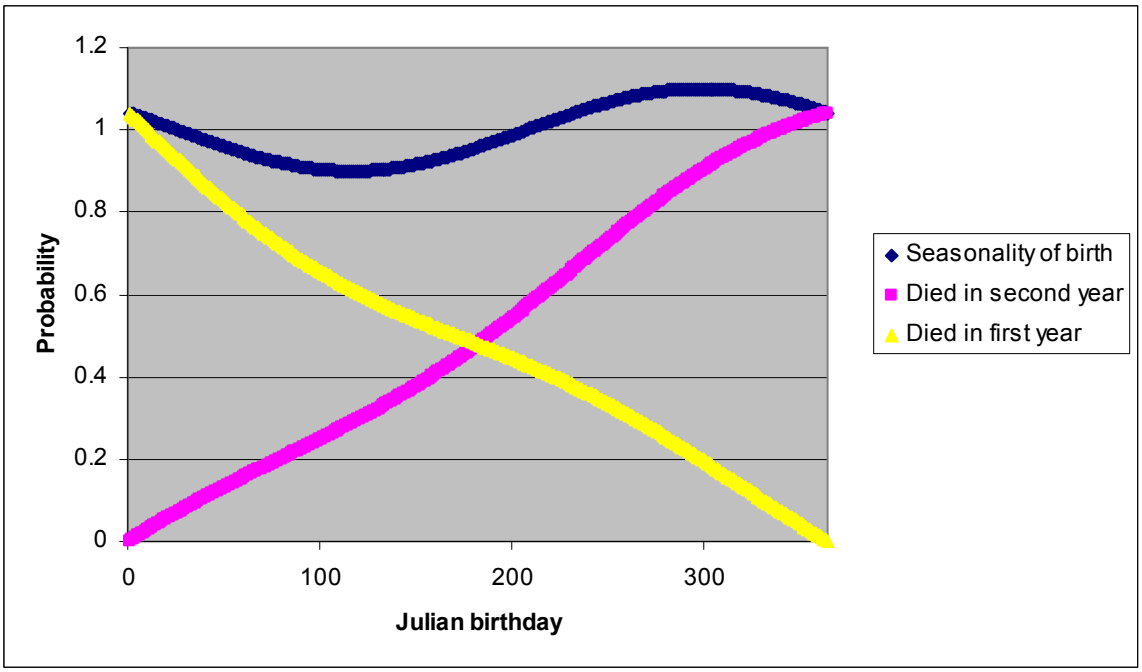


Figure A7: Probability of different Julian birthdays (i.e., days in the year numbered 1-365) for those dying in the earlier or later possible year for a given age and cohort assuming seasonality of birth following a sine function that is not symmetrical around mid-year.

If we allow death rates to vary within an age class, there should be relatively little effect on our model except for infants. Infants have much higher death rates in the first

few months of their first year than in the last few months, and this approach should not be used on infant data. However, at older ages death rates change little from year to year. For 60-year olds in Spain from 1916-1922, average death rate is 0.0108, and for 61-year-olds is 0.0113, a difference of under 5% - small relative to the 75% overlap assumption we've already built into the model. This should result in a systematic overestimation of birthday (since people are more likely to have died at an older age and thus to have been born earlier). The effect is thus to slightly up-weight early births and down-weight late births (Fig.A8). Again, we see that there should be no major effect on our ability to infer differences between the half-year cohorts based on differences between the Lexis cohorts.

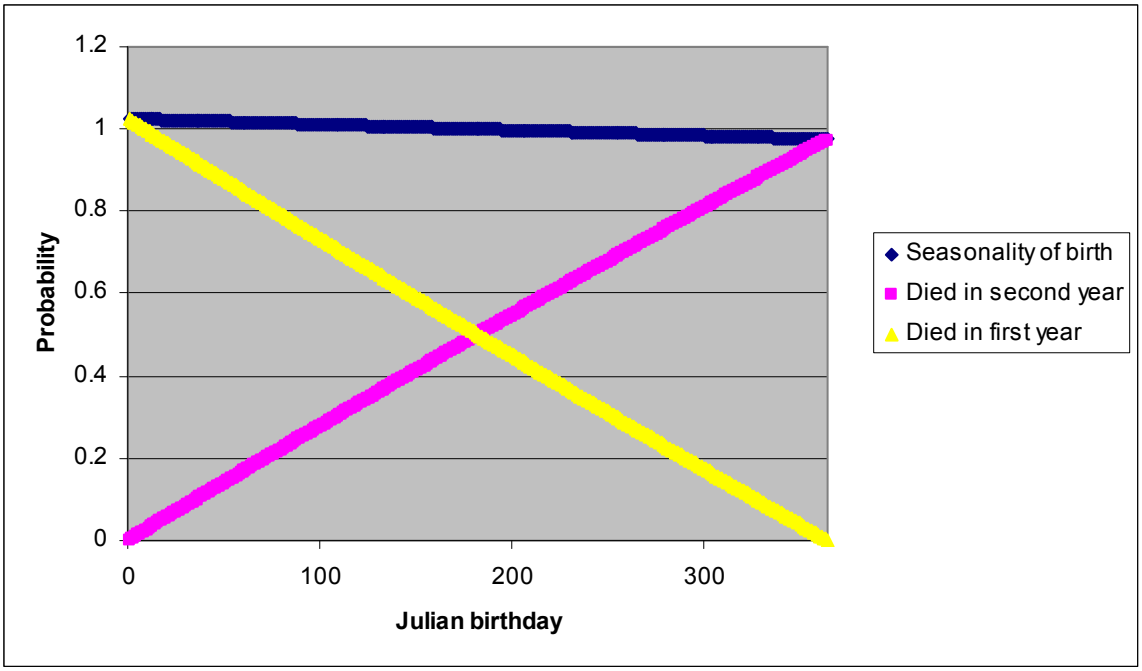


Fig A8: Probability of different Julian birthdays (i.e., days in the year numbered 1-365) for those dying in the earlier or later possible year for a given age and cohort assuming a 5% difference in allocation of births between the beginning and end of the year, based on the slightly higher probability of death at older ages.

Appendix B: A Selection of Sensitivity Analyses

Fig. B1a

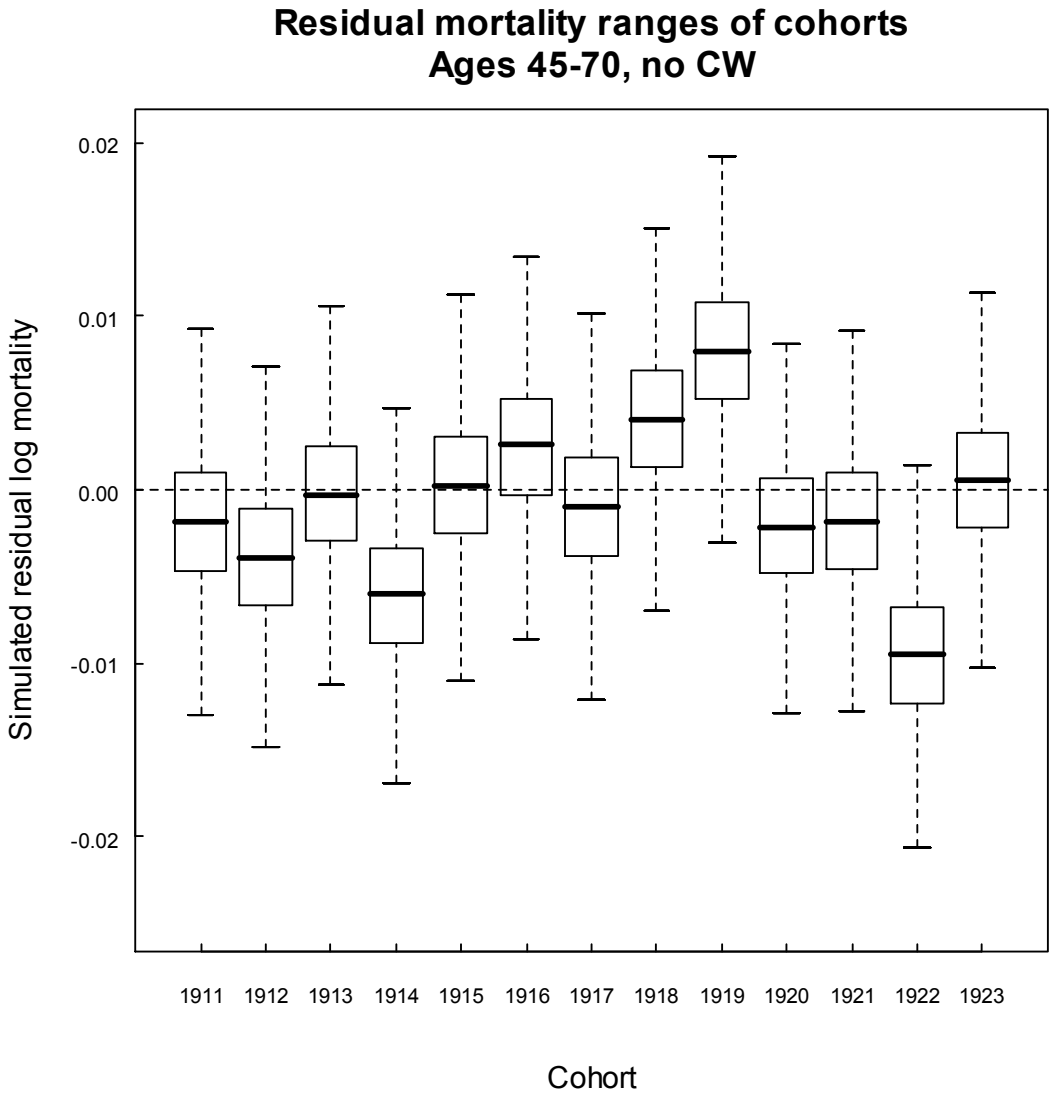

Residual mortality ranges of cohorts
Ages 45-70, no CW
Simulated residual log mortality
0.02
0.01
0.00
-0.01
-0.02
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
Cohort

Fig. B1b

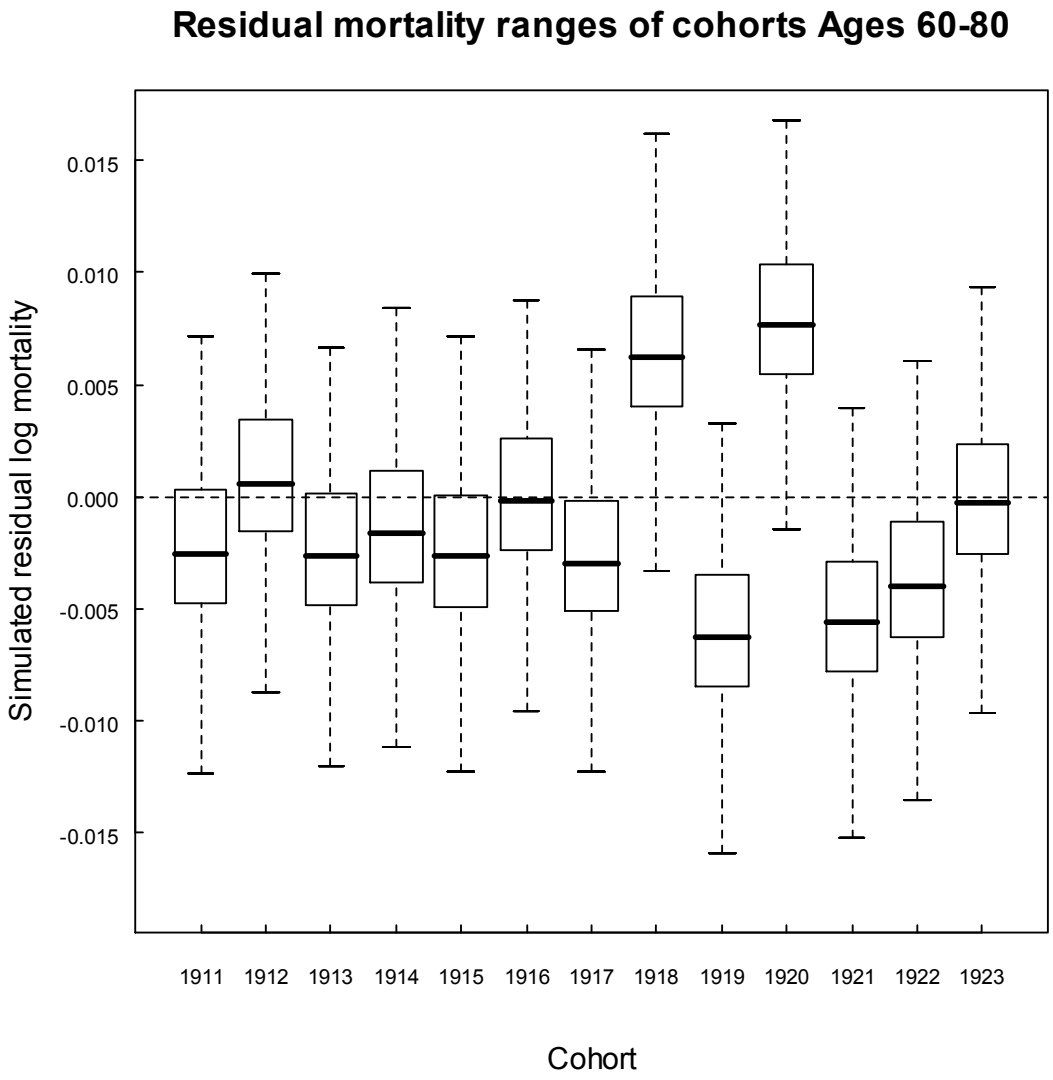

Residual mortality ranges of cohorts Ages 60-80
Simulated residual log mortality
0.015
0.010
0.005
0.000
-0.005
-0.010
-0.015
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
Cohort

Fig. B1c

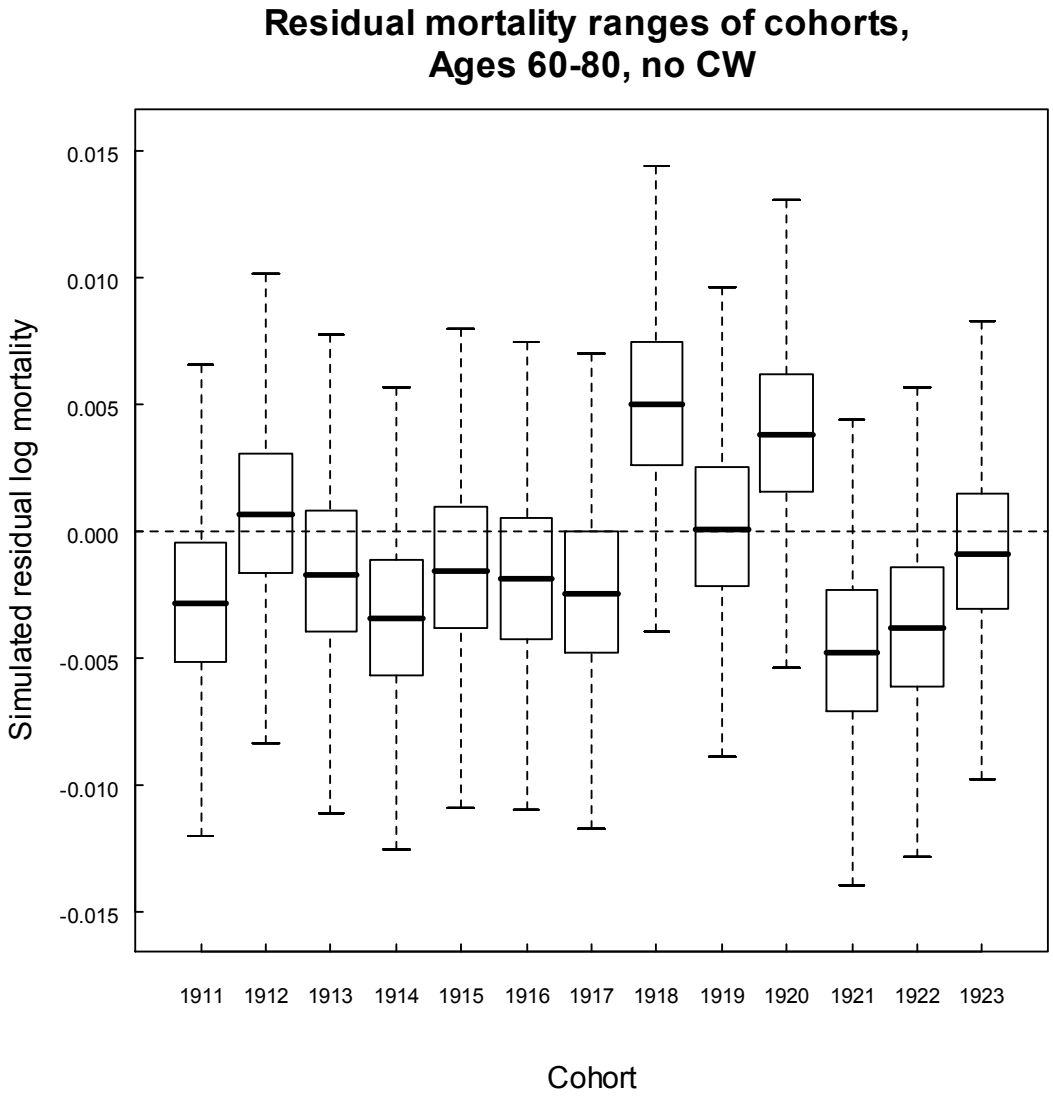


Fig. B1d

Figure B1: Estimated Bayesian deviations of residual mortality rates with 50% credible intervals (boxes) and 1.5 times interquartile range (whiskers) for the cohorts 1911-1923 a) for ages 45-70 for all 24 countries; b) for ages 45-70 for 21 countries, excluding England and Wales, New Zealand, and Canada; c) for ages 60-80 for all 24 countries; and d) for ages 60-80 for 21 countries, excluding England and Wales, New Zealand, and Canada.

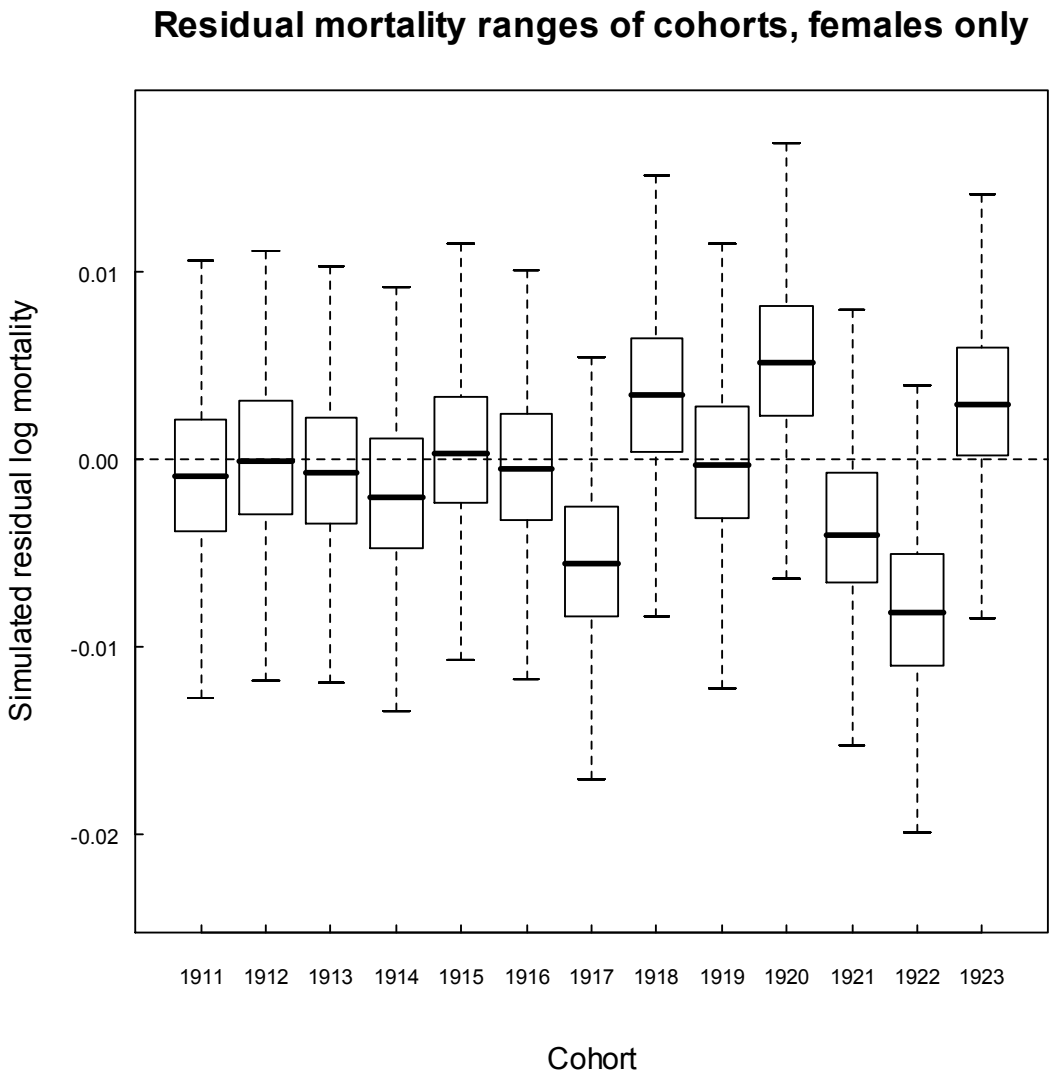

Residual mortality ranges of cohorts, females only
Simulated residual log mortality
0.01
0.00
-0.01
-0.02
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
Cohort

Fig B2a

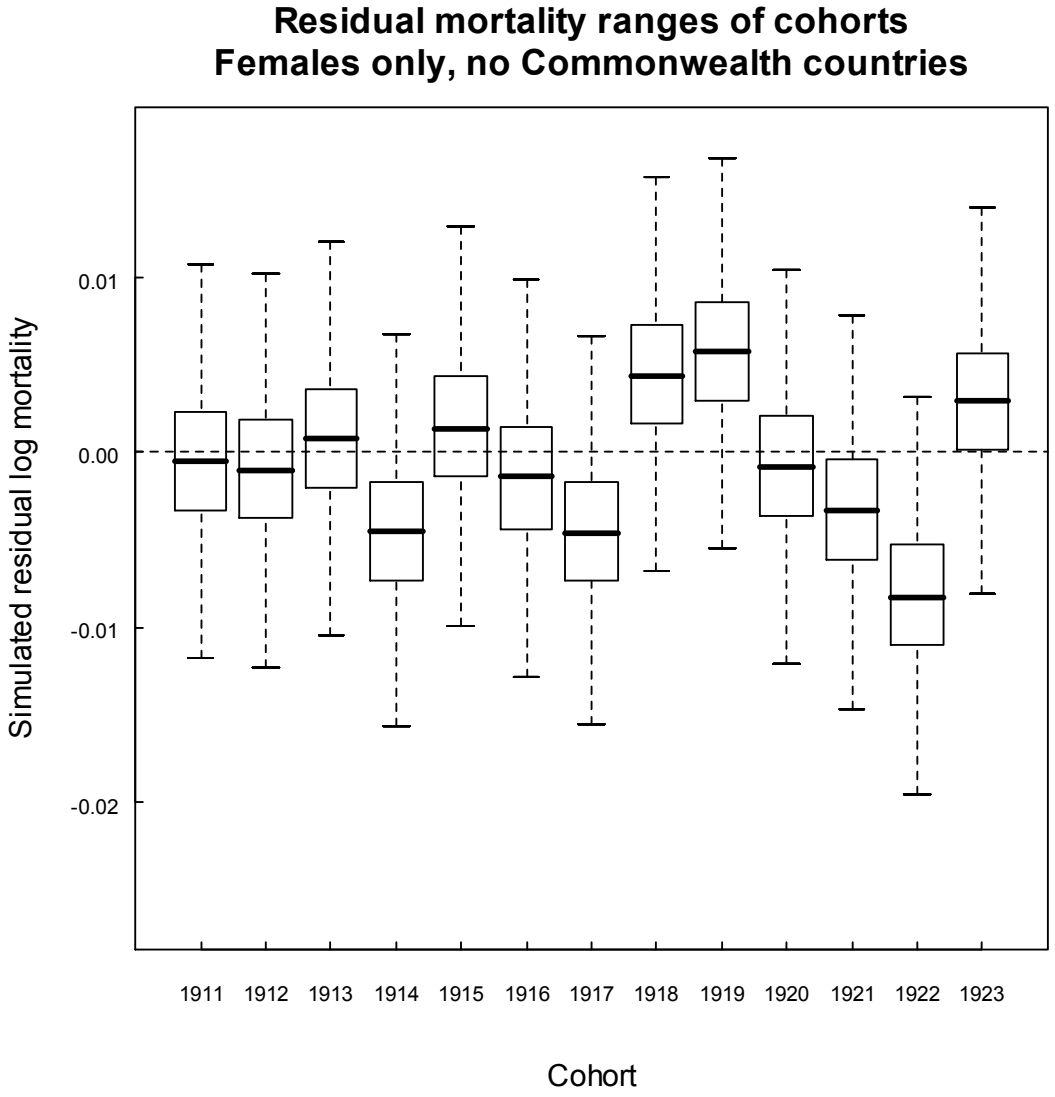


Fig B2b

Figure B2: Estimated Bayesian deviations of residual mortality rates with 50% credible intervals (boxes) and 1.5 times interquartile range (whiskers) for females only in the cohorts 1911-1923 a) for all 24 countries; and b) for 21 countries, excluding England and Wales, New Zealand, and Canada.

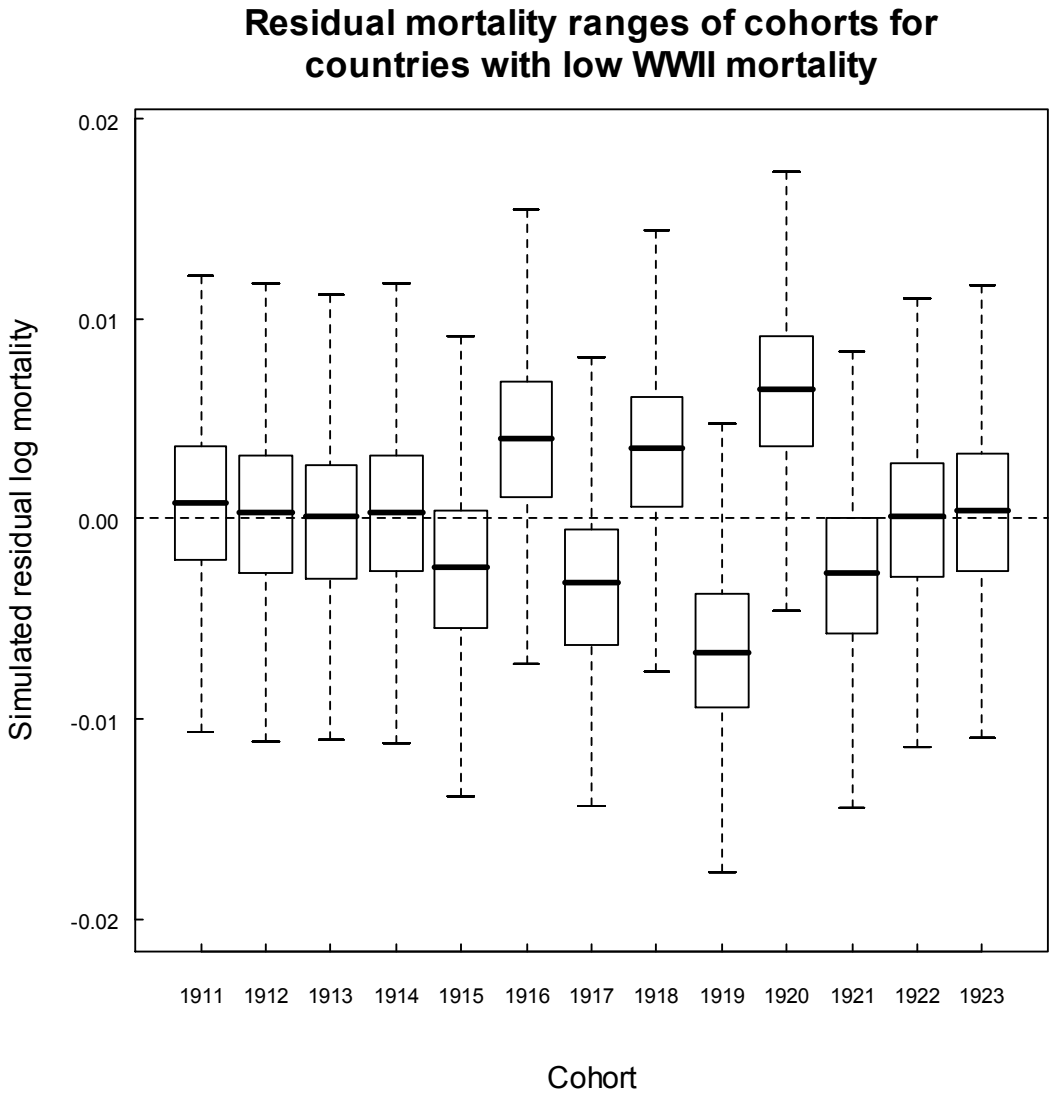


Figure B3: Estimated Bayesian deviations of residual mortality rates with 50% credible intervals (boxes) and 1.5 times interquartile range (whiskers) for the cohorts 1911-1923 in the ten countries in our sample with low (<0.5%) mortality during World War II: Australia, Bulgaria, Canada, Denmark, New Zealand, Norway, Portugal, Sweden, Switzerland, and USA.

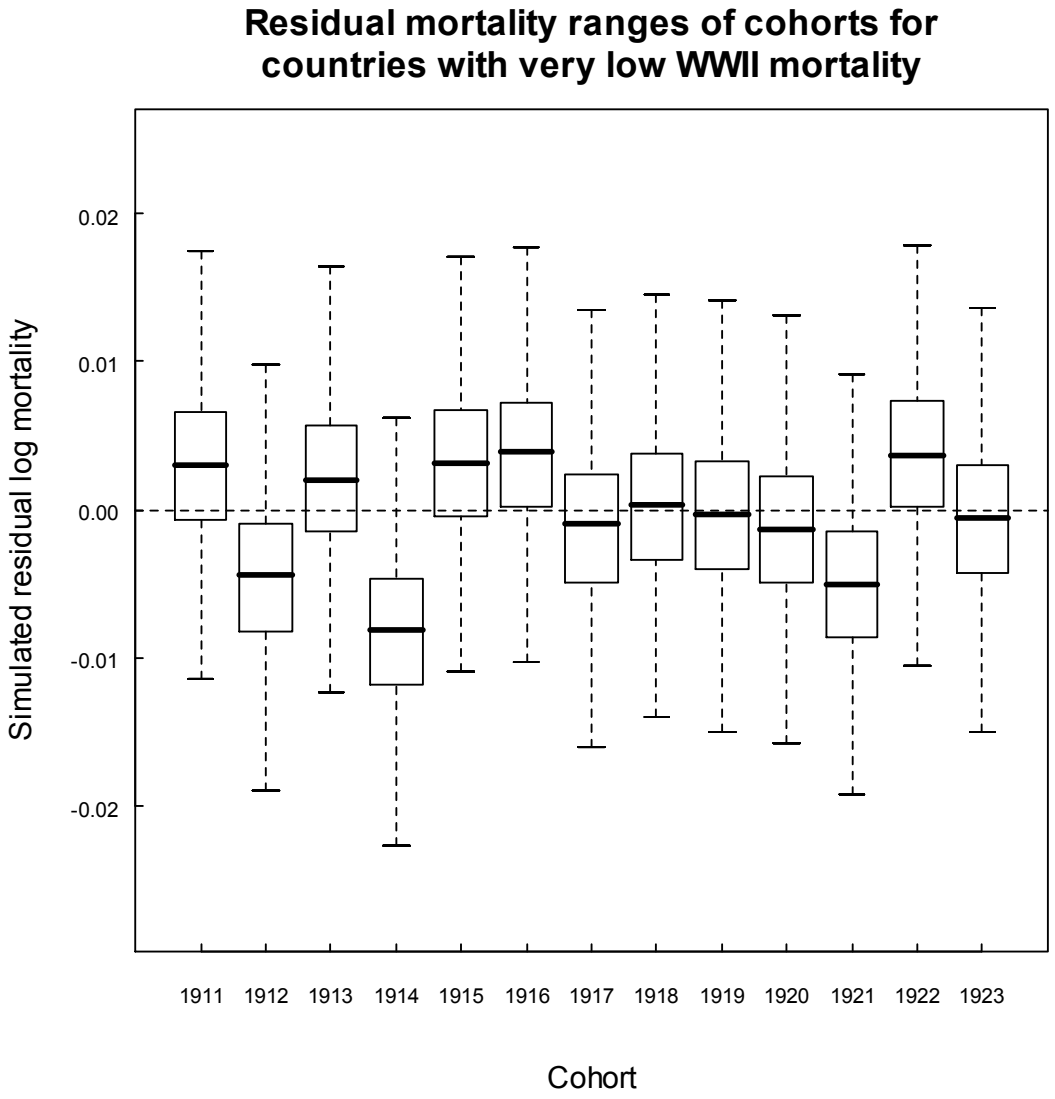


Figure B4: Estimated Bayesian deviations of residual mortality rates with 50% credible intervals (boxes) and 1.5 times interquartile range (whiskers) for the cohorts 1911-1923 in the four countries in our sample with very low (<0.1%) mortality during World War II: Denmark, Portugal, Sweden, and Switzerland.

Appendix C
Adjacent-bias artifacts

When we first started this project, we ran a series of preliminary analyses. One such analysis was to create pseudo-cohorts from period data in the Human Mortality Database (HMD) by subtracting age-at-death from year-of-death. While this is not a method we would seek to defend, it produced a striking result: in almost every country, mortality at older ages for the 1919 cohort was far lower than for any other cohort, and mortality for the 1920 cohort was far higher. Our pseudo-cohorts differed from actual cohorts in that they (a) had a mean birthday at Jan. 1, not halfway through the year, (b) had a range of birthdays that spans two years, not one, and (c) had a higher concentration of birthdays close to the center of the range rather than a uniform distribution. Based on this, we suspected that there was some protective effect of the flu on the very young in most cases (those born in the 1919 pseudo-cohort), but that those exposed very early in pregnancy (i.e., born in the second half of 1919 and thus likely in the 1920 pseudo-cohort) might have suffered severely. We then discovered that this is a known artifact in the HMD (John R. Wilmoth, *personal communication*).

It is easy for such an artifact to arise when there is some confusion as to the allocation of either deaths or exposures between adjacent cohorts. For example, if a portion of the 1920 population is mistakenly allocated to 1919 but the death counts remain the same, the 1919 cohort mortality rates will appear low while the 1920 rates appear high. The biases should be of similar magnitude but in opposite directions. While such patterns can also be due to real phenomena – for example, periods of low birth rates are often followed by a baby boom – in most cases it is unusual to have symmetrically opposed effects in adjacent cohorts. This can thus serve as a check on data quality and the accuracy of the analysis: symmetrically opposed effects in adjacent cohorts should be treated with suspicion. In the data presented here, we notice such effects in several cases. In the three Commonwealth countries England and Wales, New Zealand, and Canada, we see low mortality in the 1919 cohorts and high mortality in the 1920 cohort. In Switzerland, we see high mortality for the spring 1919 cohort and low mortality for the fall 1919 cohort. In Italy, we see high mortality in the spring 1921 cohort and low mortality in the fall 1921 cohort. In the case of Italy, there is good reason to suspect just such an allocation: the birth-rate calibration we used was with birth rates averaged across the years 1921-23, and a deviation from the average in 1921 would produce just this effect.

The trouble, of course, is that it is nearly impossible in such cases to demonstrate definitively whether the effect is real or an artifact. Nonetheless, worry about such artifacts should not undermine the overall analysis too much: in most cases, strong effects should be apparent without potential artifacts, since it is rare for adjacent cohorts to show truly opposite trends. The fact that we show few such effects can be considered substantial negative evidence.