



MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

Konrad-Zuse-Strasse 1 · D-18057 Rostock · Germany · Tel +49 (0) 3 81 20 81 - 0 · Fax +49 (0) 3 81 20 81 - 202 · www.demogr.mpg.de

MPIDR Working Paper WP 2023-046 | November 2023
<https://doi.org/10.4054/MPIDR-WP-2023-046>

An Overlapping Cohorts Perspective of Lifespan Inequality

Héctor Pifarré i Arolas

José Andrade | andrade@demogr.mpg.de

Mikko Myrskylä | myrskylä@demogr.mpg.de

© Copyright is held by the authors.

Working papers of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in working papers are attributable to the authors and do not necessarily reflect those of the Institute.

An Overlapping Cohorts Perspective of Lifespan

Inequality

¹ Héctor Pifarré i Arolas ^{1,2}, José Andrade ³, Mikko Myrskylä ^{3,4,5}

¹ La Follette School of Public Affairs, University of Wisconsin—Madison, Madison, WI, United States of America; hparolas@lafollette.wisc.edu. ² Center for Demography and Ecology, University of Wisconsin—Madison, Madison, WI 53706. ³ Max Planck Institute for Demographic Research, Rostock, Germany; andrade@demogr.mpg.de. ⁴ Population Research Unit, University of Helsinki, Helsinki, Finland, ⁵ Max Planck – University of Helsinki Center for Social Inequalities in Population Health, Rostock, Germany and Helsinki, Finland; myrskylä@demogr.mpg.de.

November 2023

¹ Funding

MM was supported by the Strategic Research Council (SRC), FLUX consortium, decision numbers 345130 and 345131; by the National Institute on Aging (R01AG075208); by grants to the Max Planck – University of Helsinki Center from the Max Planck Society, Jane and Aatos Erkko Foundation, Faculty of Social Sciences at the University of Helsinki, and Cities of Helsinki, Vantaa and Espoo; and the European Union (ERC Synergy, BIOSFER, 101071773). Views and opinions expressed are, however, those of the author only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Support for this research was provided to HPA by the University of Wisconsin-Madison, Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation; and by the Herb Kohl Public Service Research Competition through the La Follette School of Public Affairs, University of Wisconsin-Madison.

Abstract

A growing applied literature investigates the levels, trends, causes, and effects of lifespan variation. This work is typically based on measures that combine partial cohort histories into a synthetic cohort, most frequently in a period-life table, or focus on single (completed) cohort analysis. We introduce a new cohort-based method, the overlapping cohorts perspective, that preserves individual cohort histories and aggregates them in a population level measure. We apply these new methods to describe levels and trends in lifespan variation, and to the assessment of temporary and permanent mortality changes in several case studies, including the surge of violent deaths in Colombia in the 1990s and 2000s, and cause-deleted exercises for top mortality causes such as cardiovascular diseases and cancer. The results from our approach differ from those of existing methods in the timing, trends, and levels of the impact of these mortality developments on lifespan variability, bringing new insights to applied work.

Keywords: Lifespan variation, age-at-death distribution, overlapping cohorts, life table.

Data availability: All data and code to fully reproduce the analyses are available at the OSF: <https://osf.io/t6qns/>

Acknowledgements: We thank Jason Fletcher, Alberto Palloni, Marcus Ebeling and Alyson van Raalte for their insightful comments and suggestions in previous versions.

Introduction

Lifespan variability, or the variation in ages at death, has emerged as a central demographic measure of population mortality, together with the widely used life expectancy. Interest in the levels and trends of lifespan variability is based on its behavioral (Barro and Friedman 1997; Picone et al. 2004), biological (Aburto et al. 2020), and ethical (Seligman et al. 2016) implications, with the potential to inform policy recommendations. Alongside the applied literature on lifespan variability, a growing body of work has continued to develop its theoretical underpinnings (Aburto et al. 2019; Gillespie et al. 2014; Nau and Firebaugh 2012; van Raalte and Caswell 2013); this piece is contribution to the latter.

Most current applications use period measures (Aburto et al. 2023; Seaman et al. 2016; Xu et al. 2021), with few exceptions focusing on cohort measures (e.g. Myrskylä 2010). Interpretation of the cohort measure is straightforward, but the typical single-cohort approach may not represent the entire population. The dominant period approach has several advantages. Chief among these is that often the applied interest lies in what is happening now, or in a specific narrow time window – and to this the period provides a seemingly reasonable answer. In addition, the data requirements for the period approach are modest compared to the cohort approach, which may require data over decades or centuries.

The key advantage of the period indicators – they answer to the question what is happening now – is however also a key disadvantage. A prominent criticism of period indicators is their sensitivity to temporary variations in rates. This is widely recognized in the context of the analysis of fertility rates, as fertility delays may create a wedge between period and cohort measures (Bongaarts and Feeney 1998), and it also plays a role in the study of mortality patterns (Bongaarts 2005; Guillot 2006). The key assumption for period indicators is that a hypothetical (or synthetic) cohort will experience the period rates at all ages. For example, a period measure

calculated during a pandemic will capture the lifespan variability of the hypothetical cohort subject to the unusually high mortality patterns throughout their entire life, which is not accurate of any cohort's experience.

Thus, the (period) synthetic cohort representation of population level patterns is founded on stringent assumptions, which can lead to misleading assessments when they are not met. In a stationary population, all cohorts share the same age-specific mortality rates, hence the period and cohort-based perspectives coincide. In this context, measures derived from a period life table are identical to their cohort equivalents. For this reason, the period-based synthetic cohort provides a sufficient single cohort representation of the population's mortality patterns. That is, lifespan inequality measured from the period life table coincides with that of all of the single cohorts present in the population. The reality is, though, that actual populations are composed of cohorts that have experienced a diversity of living conditions, leading to differences in their mortality profiles. Hence, this challenges the interpretation of single cohort-based measures as a sufficient statistic for those of the aggregate population, or of any of the individual cohorts that compose it. While this is well understood, the implications on how the period measures should be interpreted are not always thought through. Recent work by Nepomuceno et al. (2022) has introduced a lifespan indicator, the cross-sectional average inequality in lifespan (CAL^\dagger), which incorporates some of this cohort diversity. The authors justify their measure on behavioral grounds, as they posit it might better approximate how individuals form their subjective expectations. However, CAL^\dagger can also be interpreted as an alternative synthetic cohort, built with partial cohort information, that also relies on the same strict assumptions to reflect the real population level mortality.

We propose an overlapping cohorts perspective of lifespan variation, fully incorporating cohort trajectories. To that effect, we begin by considering the mortality experience of all the

cohorts alive in the period of interest. To do so, in applications, we must rely on a combination of historical data and mortality projections. Then, the researcher must consider which elements of each cohort's mortality to include, based on the goal of the analysis. The cohorts' remaining lifespan variation (from its current age forward) would be suitable for the study of the effects of short-term mortality fluctuations. Long run or permanent mortality changes, such as the analysis of the importance of specific causes of death, would be better studied by considering the full cohort mortality experience. Finally, individual cohorts' mortality are aggregated based on their population weight. Our method coincides with the two existing synthetic cohort approaches, as well as with the standard real cohort approach under the stationary population assumption. However, it also illustrates the effects of departing from them and provides an alternative more reasonable measurement on those occasions.

We develop the overlapping cohorts perspective on lifespan inequality and demonstrate how results differ from those from current methods across several applications. First, we report the trend and level differences across methods for a historical population covering the period 1901 through 2015. Then, we study the impact of temporary and permanent mortality changes on lifespan variability through historical cases. We analyze French war and flu related mortality spikes in the first half of the nineteenth century, and the more recent surge in cartel violence-related deaths in Colombia (1984-2015). An additional counterfactual exercise investigates the effect of permanent mortality changes in 1951 to 2010 in Sweden. We evaluate the impact of cancer and cardiovascular diseases on lifespan variability through a cause-deleted exercise. Finally, we address data requirement considerations, among other issues, in the discussion.

Methods

Synthetic Life Tables

We denote *synthetic* the life tables that are built by combining part of the mortality rates from multiple cohorts into a single life table. While the most common synthetic approach is the period life table, more recently alternative methods that use more comprehensive cohort mortality trajectories have been suggested, inspired by the cross-sectional average approach pioneered by Brouard (1986). These methods tie together age-specific mortality rates profiles from different cohorts. These approaches aim to provide an adaptation of the life table, a method designed to analyze a single cohort, to population level analysis that encompasses multiple cohorts.

Our aim also is to provide a method that encompasses multiple cohorts. In contrast to existing methods, our approach takes into account cohorts' mortality both before (as in CAL) and after the period that is in focus. Instead of creating a synthetic cohort combining pieces of the mortality profiles of different cohorts, we preserve the overlapping cohort structure of the life table by aggregating full cohort profiles.

As we compare existing and our novel method, we use the variance of the age-at-death distribution as a measure of lifespan variation, frequently employed in applications (Aburto et al. 2023; Xu et al. 2021). Though other measures have been proposed (van Raalte and Caswell 2013; Wilmoth and Horiuchi 1999), the variance can be readily decomposed into its components (Shorrocks 1982), which is particularly useful for our cross-cohort approach.

Period Life Table

The variance of the ages-at-death distribution is expressed as:

$$V_t(a) = \frac{1}{l_a} \sum_{x=a}^{\omega} [d_x(x - M_a)^2], \quad (1)$$

where t refers to a particular period, a denotes the earliest age recorded in the life table, l_a is the radix of the population at age a . M_a is the average age at death after age a (i.e. $M_a = e(a) + a$). d_x refers to the number of deaths of the life table, x and ω are the age at death and the terminal age in the life table respectively. In its application to the period approach, this measure is based on the age-at-death distribution of the period life table, which is built from the period age-specific mortality rates. The same equation applies for the single-cohort approach, as well as for the overlapping cohorts approach.

Cross-sectional Average Life Table

The *cross-sectional average* (CAL) family of measures is an alternative to period life tables indicators that is based on the past mortality experience of all the birth cohorts in a population. That is, unlike conventional period measures, it is based on the survival up to the period under analysis, and not exclusively period mortality rates. Typically, life table-based measures are then constructed as a function of the survival curve of these cohorts (Guillot 2003). For the case of lifespan variation as measured by e^\dagger , the cross-sectional average version (Nepomuceno et al. 2022), CAL_t^\dagger , is defined as follows:

$$CAL_t^\dagger = \sum_{x=a}^{\omega} \left[-\ln \left(\frac{1}{l_a} l_{x,t-x}^i \right) \frac{1}{l_a} l_{x,t-x}^i \right], \quad (2)$$

where t is the period at which the measure is calculated, l_a represents the radix of the population at age a . $l_{x,t-x}^i$ is the number of survivors recorded in the cohort life table at age x for the cohort i born in the year $t - x$.

Alternatively, it is possible to interpret this approach as defining a new type of synthetic life table. That is, the life table such that $l_x = l_i$, which we call the cross-sectional average life table. This interpretation is useful because it allows us to compute other lifespan measures from its age-at-death distribution. For comparability, we will use the variance of the CAL-based life

table instead of CAL^\dagger .² Note, however, that, beyond facilitating the implementation of alternative measures, an interpretation in terms of an individual cohort in a strict sense is not possible (Guillot 2003). The reason being that combining past cohort survival in this manner may imply negative age-specific mortality rates.³ Nonetheless, this will not play a role in our analysis.

Overlapping Cohorts

We begin with the premise that, in any given period, the lifespan variation of a population is an aggregation of this measure for each of the overlapping cohorts (OC) that compose it. In a cross-section, each cohort is at a different current age, and thus we must decide what part of their age-at-death distribution will be considered in the lifespan variation calculation for the current period. As we discuss below, depending on the goal of the analysis, we may consider their past, future, or entire mortality history. Finally, an aggregation method is required to condense all the cohort-specific information for a population level measure. This section fully develops this intuition.

Perspectives

In any period, for each of the multiple cohorts in a population, we may consider lifespan variability according to three perspectives: i) from birth to its maximal age (full), ii) from its current age onward (forward), or iii) up to its current age (backward). This is similar but not the same as conditional lifespan variation (i.e. above a threshold age); conditional measures are frequently used, for example, in studies that study lifespan variation trends net of the influence of infant mortality (Edwards and Tuljapurkar 2005). Instead, our approach considers each of

² Nepomuceno et al. (2022) report a strong correlation between the standard deviation and CAL^\dagger using the CAL approach.

³ For example, mortality rates calculated for Sweden based on the cross-sectional average life table for the cohorts born in 1800 to 1885 provide negative values for several ages (e.g., 10, 18, 24, 84). More details can be found in the OSM-2.

the three perspectives from a specific period, thus forward and backward measures capture the lifespan variation of every cohort for different age ranges.

Each perspective is best suited for specific analysis, as we illustrate with our applications. We may be interested in full lifespan variation when assessing permanent mortality changes that will eventually affect all cohorts. In turn, the other two perspectives may be better suited for temporary mortality changes. A forward measure captures the effects on the lifespan variability that will affect the remaining cohort members. Thus, it is well-suited to assess the effects of short term and unexpected mortality fluctuations, for which we may care about the effects on the survivors that will actually experience the event. Finally, backward looking perspectives capture the past lifespan variability experienced by cohorts. While we do not explore this in the current work, present cohorts may look back to inform their subjective survival expectations and possible lifespan variability (Nepomuceno et al. 2022).

Aggregation

Having settled on a particular perspective, the basic object of interest is the (cross-sectional) age-at-death distribution of a given population. Here, instead of using the age-at-death distribution of a single (synthetic or true) cohort, our approach is based on the mixture of the individual cohort's distribution. The idea being that this is a reflection of the lifespan variation of the cohorts present in the population. The population's age-at-death distribution, of all cohorts i alive in period t , is constructed as follows:

$$d_t = \sum_{i=1}^n [s_0^i d^i], \quad (3)$$

where d_t and d^i are the density function for the entire population and for the cohort of age i at t , respectively. n is the total number of cohorts present in t . s_0^i is the weight in the period's population of cohort i at birth. Note that while we will work with actual cohort weights, it is

possible to age-standardize the distribution in applications where age structures are a confounder, and not a key element.

Once the age-at-death distribution is constructed, we simply compute the variance $V_t(a)$ of this distribution to measure lifespan variation. Alternatively, this may be calculated from cohort life tables using the well-known decomposition of the variance (Shorrocks 1982):

$$V_t(a) = \underbrace{\sum_{i=1}^n [s_a^i V_a^i]}_{\text{Within}} + \underbrace{\sum_{i=1}^n [s_a^i (M_a^i - M_t)^2]}_{\text{Between}}, \quad (4)$$

where n is the total alive cohorts present in time period t . s_a^i represents the share of the population who survived to age a for each cohort. M_a^i is the average age at death after age a in cohort i , M_t denotes the average age at death of individuals in all cohorts who survived to age a in time period t and V_a^i is the variance of age at death of individuals present in a cohort i that survived to age a . That is, the overall lifespan in the population is composed of the weighted sum of the individual cohort variances, and the between-cohort component. In applications, however, the within component accounts for nearly all of the total variance (see OSM-1).

Comparison Between Approaches

Figure 1 depicts the lexis diagram of the cohorts present in the period $t = 1915$. Each approach, CAL, period, and OC, uses parts of the mortality rates located within that parallelogram to compute lifespan variation. Thus, all approaches use some of the mortality experience of the cohorts alive in 1915. Period life table lifespan variability is based on the aggregation of the contemporary mortality rates (the vertical line). CAL, instead, incorporates the mortality history up to 1915, represented by the triangle to the left of (and up to) 1915. In other words, CAL accounts for the mortality conditions of the surviving cohorts in 1915. Finally, our measure uses different parts of the parallelogram depending on the perspective. Both past (left

hand triangle), and future (right hand side triangle) mortality rates for all cohorts alive in 1915 are included in the full perspective. In turn, the left-hand side is used in the backwards based calculation, and the right hand side for forward measures. Note, that while the backwards perspective is based on the cohort histories as CAL, there are differences in their respective calculations.

The differences in the information comprised in each approach only come into play when evaluating real-world population experiences. While there are analytical differences across methods, on top of the different mortality rates considered, in a constant mortality steady state all approaches yield the same result. That is, the mortality rates in the vertical (cross-section) cut of the Lexis diagram coincide with those in the diagonals (both right and left sides). It is when we recognize the diversity in mortality experiences across real cohorts that differences between approaches emerge. CAL-based lifespan variability measures acknowledge this and incorporate the past mortality histories of current cohorts. However, CAL does not attempt to reflect the actual lifespan variation of the current population. Instead, it has been justified from a behavioral perspective. The premise is that current cohorts may inform their subjective expectations on their own future mortality trajectories based on the past experiences of their kin or those socially close to them (Nepomuceno et al. 2022). Thus, the gap we intend to fill with this approach is that of a measure that preserves true cohort mortality experiences, but that also reflects the overall lifespan variability of the population.

Panel a in Figure 2 is the empirical counterpart of Figure 1 for the year 1915 in Sweden, displaying age-specific mortality rates. These are aggregated into their respective implied mortality rate distributions in panel b. The main differences across methods concentrate in both ends of the age distribution, with particularly substantial discrepancies between CAL and the rest of the approaches at older ages. Note the implicit negative mortality rates for the CAL-based distribution, for example in the range 40 through 50, and at older ages. These differences

manifest in substantial disparities in the respective age-at-death distributions, as shown in panel c. Specifically, panel c reports the disparities in survivors (l_x), as a way to deal with negative mortality rates from CAL life tables. In the later sections, we will illustrate how the large differences in ages-at-death influence lifespan variability calculations in different contexts.

Counterfactuals

Counterfactual calculations for lifespan variability may be used to study the effect of short-term mortality fluctuations (Aburto et al. 2021; García and Aburto 2019; Vigezzi et al. 2022), permanent mortality changes, or in exercises that assess the contribution of specific causes of death to lifespan variation (Aburto et al. 2023; Seligman et al. 2016). Typically, these calculations for mortality changes are based on period lifespan variability measures. While there exist differences in the analytical strategies, at their core, these approaches evaluate the effect of mortality variations based on stationary population assumption. There are two fundamental assumptions in these exercises, which also disregard considerations related to the transition period, i.e. until the entire population has been born under the new mortality regime. First, any mortality change is presumed to be permanent, such that current period mortality rates will be constant and thus match, in the long run, cohort mortality rates. Second, the current shares of deaths by cause will be constant and shared across cohorts. Related to this, in addition, decomposition approaches often (Aburto et al. 2023), but not always (Seligman et al. 2016), rely on the independence of mortality across causes.

Temporary Mortality Fluctuations

What happens when these assumptions are challenged? We illustrate the consequences of replacing some of these assumptions with real data in an effort to better match the actual mortality developments of the population in applications. We use case studies that cover a range of applications. In the first two applications, we analyze the impact of short-lived periods of

elevated mortality. We drop the permanent mortality change assumption and allow for mortality to decline after the mortality increasing shock, following its actual (or forecasted) patterns.

To create the counterfactual mortality rates (absent of the shock), we interpolate mortality for those periods. The first set of historical mortality shocks are the relatively short peaks of mortality in 1914-1918 and 1940-1945 in France, due to a combination of the Spanish flu epidemic and world war casualties. Thus, we replace mortality rates during the peak mortality years with a cubic spline interpolation that fits a third-degree polynomial to the surrounding periods' age-specific mortality rates. In this case, we work with all-cause mortality rates, since we lack data on cause specific mortality. Then, we explore the implications of our approach in the context of a longer-lived mortality crisis, the surge of violent deaths in Colombia in the period 1984-2015. Here, we generate a counterfactual eliminating specifically the surge of violent deaths. In this case, a linear interpolation performs the best, given the gaps in the data for some of the adjacent years.

Permanent Mortality Changes

We also investigate the trends in the contribution of cardiovascular and neoplasm mortality to lifespan variability in Sweden for the period 1951-2018. In this exercise, we deviate minimally from existing period-based approaches in the implementation of the OC approach. Cohort-specific counterfactual mortality profiles are created using standard cause-deleted approaches (Preston et al. 2000), and cause-specific death shares are assumed to remain constant at the period's level for all cohorts.

The main difference with respect to the traditional period analysis is that we do not assume equal mortality profiles across cohorts. Instead, we keep existing age-specific mortality rate differences between cohorts constant in relative terms. In other words, our alternate assumption

is that the rate of mortality improvements over time will keep their pace, and thus so will cohort relative mortality positions. This assumption has been chosen for simplicity; alternative (and perhaps empirically founded) scenarios for future cohort mortality differences are left for future exercises.

Based on this idea, we proceed as follows. First, using the last-born cohort (age 0 in year under study) as a baseline, we calculate relative age-specific relative mortality rates across cohorts. For example, in 2000, the age 0 mortality rate for the oldest cohort (born in 1910) is 54 times that of the youngest cohort. Once relative cohort mortality is established, cohort-specific mortality rates depend on the level of mortality to which relative mortality is anchored. We use the period's mortality rates to set the mortality level. For each period, contemporary mortality rates are the assumed mortality for the last-born cohort, and all other cohorts are scaled accordingly based on their relative mortality.

Data

Mortality rates and population exposures by single year of age are retrieved from the Human Mortality Database (HMD). To increase the range of our period of analysis, HMD data is complemented with World Population Prospects (2022) mortality and population projections (medium mortality scenario) from the last year available in the HMD and up to 2100. Period and cohort life tables are constructed following standard demographic techniques (extrapolation), considering the 85+ and 90+ open-ended age intervals.

Causes of death data for Colombia and Sweden were obtained from the World Health Organization (WHO, 2020) Mortality Database. For Colombia, all causes of death associated to violence are considered (ICD10: X85-Y09). In the case of Sweden, the following ICD codes CD10 I00-I99 are used to capture cardiovascular diseases (CVD), and ICD 10 C00-C97 and D00-D48 for malignant neoplasms. Causes of death categories were harmonized across

versions of the International Classification of Diseases in analysis that extend across different ICD versions. Age-specific deaths are ungrouped into single ages using a univariate penalized composite link model (Pascariu et al. 2018).

Results

Trends

An examination of lifespan variation across both synthetic approaches and the OC perspective reveals differences in both levels and trends. Panel a in Figure 3 displays the trends in lifespan variation for France by method, as well as the underlying cohort-specific lifespan variation. Given their data requirements, CAL and (full) OC measures are available from 1901 (vertical dashed line) to 2015. Throughout the years considered, period measures report the lowest lifespan variation, followed by CAL and OC; the exception being the surge in mortality during the second world war (which we will revisit in the next section). In addition, while period and CAL lifespan variation trends downward after the end of the 1800s, OC measures remain higher (growing moderately) up to the 1950s, when they rapidly decline.

The reason for this divergence across methods is that full OC indicators encompass more past cohort information than period methods, and also future developments (contrary to CAL indicators). In the case of France, the cohorts in the early 1900s will experience the two world wars and the flu pandemic. Thus, OC measures do not follow the contemporary declines in mortality rates, which dictate period trends, and are also pushed upward by the future mortality spikes. As individual cohorts' lifespan variability starts declining monotonically in the 1950s, so do OC indicators. We will examine in detail these effects in a case study covering the upsurge in violent deaths in Colombia starting in the mid-1980s. The extent to which CAL and period approaches deviate from the experience of contemporary cohorts can be seen more clearly in panel b. The figure displays the lifespan variation of the cohorts alive in 1901 (blue

dots), as well as the population level indicator according to the three methods. CAL and period indicators are below the experience of any one cohort in 1901, whereas OC is a (cohort size) weighted average of the individual lifespan variability.

Temporary Mortality Changes

High Mortality Episodes — France War Mortality

Throughout the first half of the twentieth century, France experienced two temporary large high mortality episodes due to the two world wars, combined, in 1918, with the additional mortality from the Spanish flu epidemic. In this exercise, we assess the impact that these events had on lifespan variability from the perspectives of the cohorts in the year 1940 in a counterfactual exercise.

Panel a in Figure 4 displays both series (real and counterfactual) of aggregate mortality rates in France from 1855 to 2025, with elevated mortality in the years 1914 through 1918, and 1940 through 1945. The age pattern of elevated mortality is shown in panel b for all cohorts present in 1940. We can see that while mortality was relatively higher for most age groups during these peaks, the age group from 15 to 40 experienced the worst effects in both occasions.

Panel c displays the overall effects of these events on the cohorts in 1940 according to each method, as well as the cohort specific effects. Three noteworthy differences become apparent. First, period-based measures have a higher magnitude than either of the cohort-based approaches. As we know, period methods assume that the unusually elevated mortality will remain constant, over-estimating the mortality effect of the event.

Second, we obtain opposite messages according to forward and full perspectives. While according to the forward perspective lifespan variation increased, the full perspective indicates a small decline in it. That is, for the survivors of the cohorts alive in 1940, the lifespan variation

for the remaining part of their lives increased as a result of these high mortality events, and instead it declined when considered for these cohorts from birth. Third, there is substantial heterogeneity across the effects it has on cohorts, depending on their age in 1940. In the case of the forward measure the extent to which cohorts experience increases in lifespan variation varies, peaking for those aged 21 in 1940. These will be the cohorts that reach the ages with highest mortality throughout those years. In turn, the full measure shows a cross-over around age 45. Those older than 45 will experience a decline in lifespan variation as a result of these high mortality events, whereas the later born cohorts will face an increase. Overall, full lifespan variation declines, as the effect for older cohorts dominates.

High Mortality Episodes — Colombia's Cartel Violence

During the last years of the 20th century and the beginning of the 21st, Colombia experienced one of the most violent episodes in South America, with high rates of homicide. This was mainly driven by the Drug Wars in the 1980s and guerrilla warfare in the 1990s (Coatsworth 2003; Luna 2019). The war between the Colombian Armed Forces, paramilitary groups, and guerrilla groups lasted until 2016 when Colombia approved the peace agreement referendum, leading to the termination of hostilities between the Colombian government and guerrilla groups (Luna 2019). This exercise illustrates the extent of the impact of high and large mortality episodes on lifespan variability, based on a counterfactual exercise.

Panel a in Figure 5 depicts the trends of the real and counterfactual mortality rates in Colombia from 1953 to 2018, with high mortality occurring during the period from 1984 to 2015 due to the violent episode in the country. The age-specific excess mortality resulting from it is shown in panel b, as the ratio of the real and counterfactual mortality. This high mortality episode disproportionately affected individuals between the ages of 15 and 40. The resulting cohort-specific increase in forward lifespan variability is reported in panel c. Cohorts born

between 1970 and 1985 experienced the violent episode at the ages worst affected (15 through 40) and, as a result, had the greatest increase in their lifespan variability. Older and younger cohorts progressively lived fewer of their more vulnerable years during the period of elevated mortality and hence suffered lower lifespan variability increases.

Panel d displays the overall effect of the violent episode on lifespan variability over time and across different methods. The most striking difference is in the peak years in which the population's lifespan variation is most affected. While the impact peaks in years when mortality was the highest due to violence (e.g. 2002, 1996, and 1992) for period measures, the cohort-based forward perspective peaks at the onset of the violent episode and gradually decreases over time. This result builds on the intuition from panel c: the beginning of the violence spike concentrates the largest number of highly affected cohorts, i.e. cohorts that will experience the pandemic in their most vulnerable ages. In contrast, period-based measures track contemporary mortality rates.

Permanent Mortality Changes

The next case involves permanent mortality changes, in the form of cause-specific assessments, for which interpret the sign and magnitude of the influence on lifespan variability. From an analytical standpoint, the direction of the effect is dictated by whether age-specific mortality changes fall below or above a specific threshold age. When mortality declines below the threshold age, it results in a decrease in lifespan variability. Conversely, if mortality decreases above the threshold age, it leads to an increase in lifespan variability (Seligman et al. 2016; van Raalte and Caswell 2013). This will play an important role in the cause-deleted exercise, since the sign of the effect on lifespan variability will depend on the age distribution of the cause of death in relation to the method-specific threshold ages. The magnitude of the effect is

influenced by the distance from the threshold for the ages impacted by the mortality change, as well as the overall size of the change.

We examine the contribution of top causes of death to lifespan variability. In particular, we investigate the contribution of cardiovascular diseases (CVD) and neoplasms in Sweden for the period 1951 to 2010. To identify the cause-specific impact, we generate a counterfactual through a standard cause-deleted life table analysis.⁴ Panel a and d in Figure 6 displays the age-specific share of deaths for cardiovascular diseases (CVD) and neoplasms by age in the year 1982. In the case of CVD, the share grows steadily with age, concentrating the majority of CVD-related deaths among the elderly population. On the other hand, the proportions of neoplasms are higher within the age range of 30 to 80. These patterns remain stable throughout the entire analyzed period.

Panel b and e in Figure 6 depict the baseline and cause-deleted mortality rates, and corresponding threshold ages (vertical lines) for both period-based and cohort-based methods, also for 1982. Remarkably, the threshold age for the overlapping cohorts method is about half that of the period mortality distribution (33.8 v. 61.8). This will influence the contribution these causes have to their respective lifespan variations, given that, for the period method, a higher proportion of the mortality declines will occur above the threshold age.

These differences in the threshold age result in substantial discrepancies across methods, as illustrated by panel c and f, which displays the impact of each cause across the period (the ratio of real and counterfactual lifespan variation). We find that, for most of the period, CVD are positive contributors to lifespans variation in both methods (panel c). However, practically all mortality changes are located well above threshold in our approach, but less so for the period-

⁴ Life tables are closed at 90+ to better represent the mortality of causes such as CVD that are heavily concentrated at older ages.

based measure. This explains the differences in levels in the contribution, which are up to 1.2 times larger for our measure (in 1981). However, as mortality declines, increasing the threshold age, the sign of the contribution reverses for the period measure, and CVD reduces lifespan variation from the 1990s onward. Threshold differences manifest in a more extreme manner for cancer (panel f). According to the period measure, across all years, neoplasms have a negative contribution to lifespan variation. In contrast, up to the early 1990s, they lower lifespan variability in our approach.

Discussion

We introduced a new lifespan variability indicator that fully incorporates cohort trajectories of all cohorts alive in the period of consideration, in contrast with current measures that only utilize partial cohort histories to create a synthetic cohort, or focus on single cohorts. We applied it to several case studies to illustrate its implications for the assessment of temporary and permanent mortality developments. Relatively short-lived high mortality episodes were studied in the context of the early to mid-1900s mortality developments in France, and the surge in violent deaths in Colombia (1984-2015). Our method differs from existing approaches in the magnitude, timing, and direction of the effects of these episodes on lifespan variation. Finally, we perform a counterfactual exercise to assess the impact of permanent mortality variations; specifically, cause-specific contributions to lifespan variability. We demonstrated that the extent to which our results differ from those of current methods depends on the role played by the underlying discrepancies in the age-at-death distributions across approaches. These mortality differences affect the position of the critical threshold ages, with consequential results on the findings by each method. We find that CVD have a substantially larger impact on lifespan variability compared to existing methodologies (up to 1.2 times), and that cancer

diminishes lifespan variability from a period perspective, but increases it, for most periods, based on our approach.

It might appear that the requirements to deploy our approach to measure lifespan variability go beyond the commonly available mortality data, while period methods do not suffer from this. For the most part, we were able to apply our method by relying on countries with high quality historical data, and on long term mortality projections. Strong data requirements, however, are inherent to the measurement of lifespan variation, a longitudinal measure that is highly sensitive to the shape of future mortality. Period measures address these inherent data requirements with strong assumptions on future mortality developments. Specifically, a critical assumption is that of equal mortality across cohorts. In this work, we have shown how failure to meet this and other assumptions can result in potentially misleading assessments. In other words, either data and projections or strong assumptions will always be needed to assess lifespan variability; there is no such thing as a free lunch.

Beyond data-related considerations, there remain challenges to the rationale behind the period perspective on longitudinal measures. In the case of measures such as crude mortality rates, the period justification is clear: they aim to reflect purely contemporary mortality conditions. However, for inherently longitudinal measures, like lifespan variability, they are more difficult to interpret outside of the stationary population equivalence framework. As we know, if mortality is not equal across cohorts, this is not representative of the population's experience. A common approach is to refer to them as measuring the experience of an hypothetical cohort. However, given that period lifespan variation measures can differ substantially from those of any one cohort, this may limit the interpretability of the synthetic cohort results. Thus, it is useful to consider alternatives. We believe that the overlapping cohorts approach presented in this paper is one such useful alternative.

We are not alone in recognizing the limitations of period measures. A well-established body of work challenges the notion that period mortality rates truly reflect contemporary conditions. The argument being that past behavioral and contextual influences shape cohort mortality (Barker 2007; Myrskylä 2010; Wang and Preston 2009). A strand of work that is closer to our contribution also recognizes that, with non-constant mortality, period measures are hard to interpret given the resulting heterogeneous cohorts. CAL and related measures provide an alternative, based on single cohort experiences (Boongarts 2005), on combinations of cohorts (Guillot 2003; Nepomuceno et al. 2022), and using historical and projection-based mortality data for cohorts (Guillot and Payne 2019). Our proposed measures follow in this tradition.

We believe that a new mixture of revised assumptions, together with a larger role for mortality projections, is needed moving forward. In our cause-deleted exercise we have introduced a new assumption that preserves the overlapping cohort perspective and alleviates data requirements. At the same time, promising new efforts to complete cohort fertility and mortality open the door for more accurate forecasts (Basellini et al. 2020; Bohk-Ewald et al. 2018; Goldstein et al. 2023). Thus, there exist promising alternatives to the equal cohort mortality assumption. We view our work as a methodological innovation that provides a benchmark for lifespan variability measurement, in a data rich context. Ultimately, the best mixture between assumptions, data, and projections is likely to continue shifting in the future as new data becomes available, and further developments in mortality forecasting arrive.

Conclusion

We have introduced a new methodological framework to measure lifespan variability based on an overlapping cohorts conception of the population. In our view, our contribution is twofold. We propose a different approach to aggregate cross-cohort information that does not rely on a

synthetic cohort. In addition, we discuss which fragments of cohort mortality to utilize in applications, beyond conditional lifespan variability considerations. There remain significant challenges to apply this new method, particularly in low data contexts. However, we also believe valuable new insights can be gained by applying this framework to other life tables and adjacent demographic measures.

References

- Aburto, J. M., Alvarez, J. A., Villavicencio, F., & Vaupel, J. W. (2019). The threshold age of the lifetable entropy. *Demographic Research*, *41*, 83-102.
- Aburto, J. M., Di Lego, V., Riffe, T., Kashyap, R., Van Raalte, A., & Torrissi, O. (2023). A global assessment of the impact of violence on lifetime uncertainty. *Science advances*, *9*(5), eadd9038.
- Aburto, J. M., Kashyap, R., Schöley, J., Angus, C., Ermisch, J., Mills, M. C., & Dowd, J. B. (2021). Estimating the burden of the COVID-19 pandemic on mortality, life expectancy and lifespan inequality in England and Wales: a population-level analysis. *J Epidemiol Community Health*, *75*(8), 735-740.
- Aburto, J. M., Villavicencio, F., Basellini, U., Kjærgaard, S., & Vaupel, J. W. (2020). Dynamics of life expectancy and life span equality. *Proceedings of the National Academy of Sciences*, *117*(10), 5250-5259.
- Barker, D. J. (2007). The origins of the developmental origins theory. *Journal of internal medicine*, *261*(5), 412-417.
- Barro, R. J., & Friedman, J. W. (1977). On uncertain lifetimes. *Journal of Political Economy*, *85*(4), 843-849.

- Basellini, U., Kjærgaard, S., & Camarda, C. G. (2020). An age-at-death distribution approach to forecast cohort mortality. *Insurance: Mathematics and Economics*, *91*, 129-143.
- Bohk-Ewald, C., Li, P., & Myrskylä, M. (2018). Forecast accuracy hardly improves with method complexity when completing cohort fertility. *Proceedings of the National Academy of Sciences*, *115*(37), 9187-9192.
- Bongaarts, J. (2005). Five period measures of longevity, *Demographic Research*, *13*(21), 547–558.
- Bongaarts, J., & Feeney, G. (1998). On the quantum and tempo of fertility. *Population and development review*, 271-291.
- Brouard, N. (1986). Structure et dynamique des populations. La pyramide des années à vivre, aspects nationaux et exemples régionaux. *Espace Populations Sociétés*, *4*(2), 157-168.
- Coatsworth, J. H. (2003). Roots of violence in Colombia: Armed actors and beyond. *ReVista: Harvard Review of Latin America*. Spring 2.
- Edwards, R. D., & Tuljapurkar, S. (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review*, *31*(4), 645-674.
- García, J., & Aburto, J. M. (2019). The impact of violence on Venezuelan life expectancy and lifespan inequality. *International journal of epidemiology*, *48*(5), 1593-1601.
- Gillespie, D. O., Trotter, M. V., & Tuljapurkar, S. D. (2014). Divergence in age patterns of mortality change drives international divergence in lifespan inequality. *Demography*, *51*(3), 1003-1017.

- Goldstein, J. R., Osborne, M., Atherwood, S., & Breen, C. F. (2023). Mortality Modeling of Partially Observed Cohorts Using Administrative Death Records. *Population Research and Policy Review*, 42(3), 36.
- Guillot, M. (2003). The cross-sectional average length of life (CAL): A cross-sectional mortality measure that reflects the experience of cohorts. *Population Studies*, 57(1), 41-54.
- Guillot, M., & Payne, C. F. (2019). Tracking progress in mean longevity: The lagged cohort life expectancy (LCLE) approach. *Population Studies*, 73(3), 405-421.
- Human Mortality Database*. (2021). Berkeley, CA (USA): University of California, Berkeley; Rostock, Germany: Max Planck Institute for Demographic Research. Available from www.mortality.org (6 May 2023, date last accessed).
- Luna, L. (2019). Colombian violent conflict: A historical perspective. *International Journal on World Peace*, 36(4), 53-84.
- Myrskylä, M. (2010). The effects of shocks in early life mortality on later life expectancy and mortality compression: A cohort analysis. *Demographic Research*, 22(12), 289–320.
- Nau, C., & Firebaugh, G. (2012). A new method for determining why length of life is more unequal in some populations than in others. *Demography*, 49(4), 1207-1230.
- Nepomuceno, M. R., Cui, Q., Van Raalte, A., Aburto, J. M., & Canudas-Romo, V. (2022). The cross-sectional average inequality in lifespan (CAL[†]): A lifespan variation measure that reflects the mortality histories of cohorts. *Demography*, 59(1), 187-206.
- Pascariu, M. D., Daňko, M. J., Schöley, J., & Rizzi, S. (2018). Ungroup: An R package for efficient estimation of smooth distributions from coarsely binned data. *Journal of Open Source Software*, 3(29), 937.

- Picone, G., Sloan, F., & Taylor, D. (2004). Effects of risk and time preference and expected longevity on demand for medical tests. *Journal of Risk and Uncertainty*, 28, 39-53.
- Preston, S., Heuveline, P., & Guillot, M. (2000). Demography: measuring and modeling population processes. 2001. *Malden, MA: Blackwell Publishers*.
- Seaman, R., Leyland, A. H., & Popham, F. (2016). Increasing inequality in age of death at shared levels of life expectancy: a comparative study of Scotland and England and Wales. *SSM-population health*, 2, 724-731.
- Seligman, B., Greenberg, G., & Tuljapurkar, S. (2016). Equity and length of lifespan are not the same. *Proceedings of the National Academy of Sciences*, 113(30), 8420-8423.
- Shorrocks, A. F. (1982). Inequality decomposition by factor components. *Econometrica: Journal of the Econometric Society*, 193-211.
- United Nations, Department of Economic and Social Affairs, Population Division. (2022). World Population Prospects 2022: Data Sources. <https://population.un.org/wpp/> (6 May 2023, date last accessed).
- Van Raalte, A. A., & Caswell, H. (2013). Perturbation analysis of indices of lifespan variability. *Demography*, 50(5), 1615-1640.
- Vigezzi, S., Aburto, J. M., Permanyer, I., & Zarulli, V. (2022). Divergent trends in lifespan variation during mortality crises. *Demographic Research*, 46, 291-336.
- Wang, H., & Preston, S. H. (2009). Forecasting United States mortality using cohort smoking histories. *Proceedings of the National Academy of Sciences*, 106(2), 393-398.
- Wilmoth, J. R., & Horiuchi, S. (1999). Rectangularization revisited: Variability of age at death within human populations. *Demography*, 36(4), 475-495.

World Health Organization. (2020). WHO methods and data sources for global burden of disease estimates 2000-2019. Geneva: Department of Data and Analytics. <https://www.who.int/data/data-collection-tools/who-mortality-database> (6 May 2023, date last accessed).

Xu, W., Engelman, M., & Fletcher, J. (2021). From convergence to divergence: Lifespan variation in US states, 1959–2017. *SSM-Population Health*, 16, 100987.

Figures

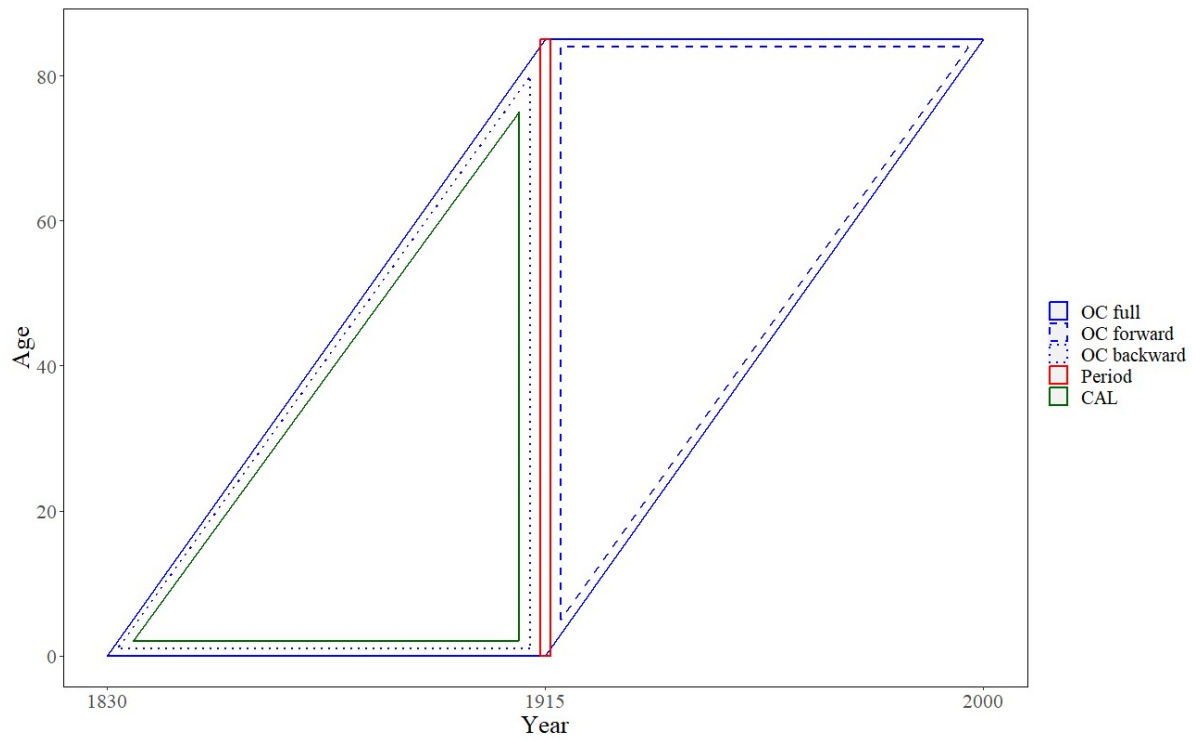


Fig. 1 Lexis diagram and life table methods. Lexis diagram showing the mortality rates used in the calculation of lifespan variability according to the period method (period), cross-sectional average approach (CAL), and the overlapping cohorts perspective (OC).

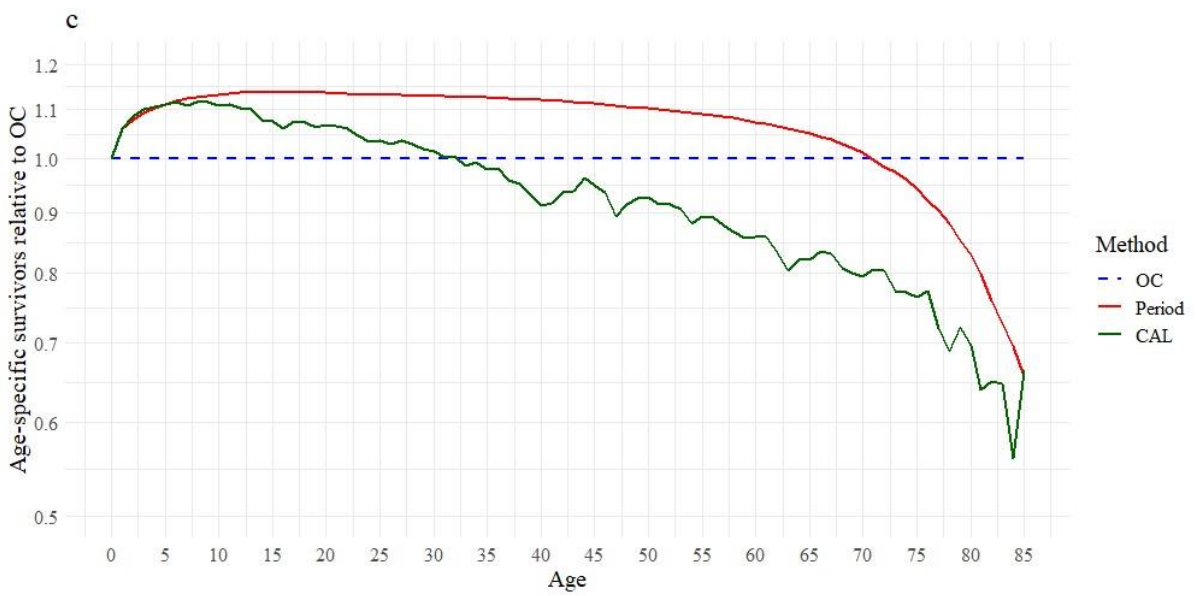
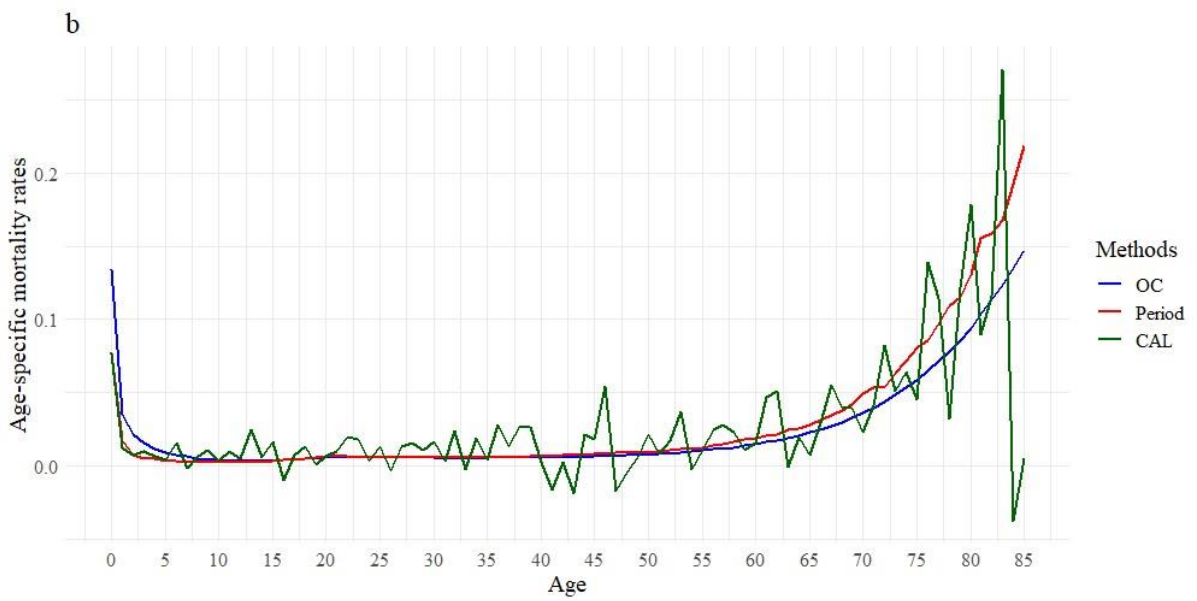
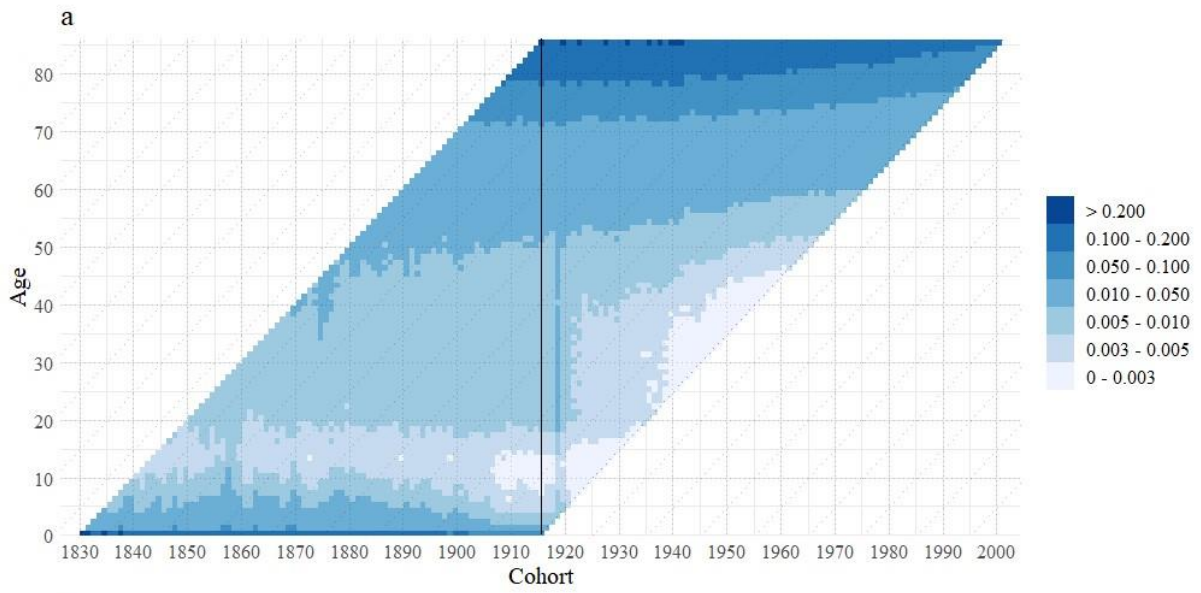


Fig. 2 Differences in mortality over life table methods in Sweden. (a) lexis diagram with age-specific mortality rates for cohorts born in the period 1830 to 1915 in Sweden. (b) age-specific mortality rates across life table methods for Sweden in 1915. (c) survivors (l_x) for CAL and period life table method relative to OC for Sweden in 1915.

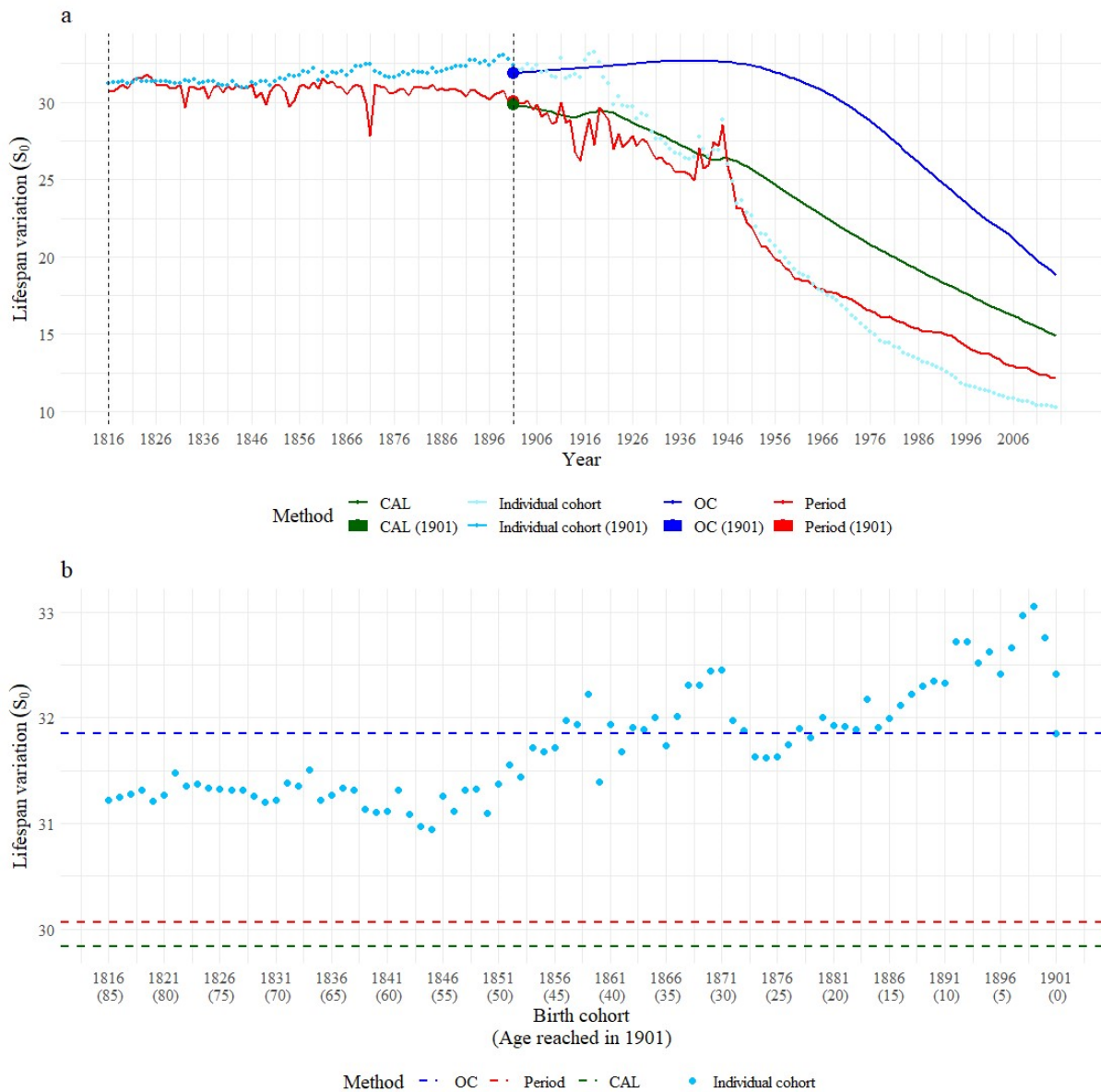


Fig. 3 Trends of lifespan variation over life table methods in France. (a) lifespan variation trends of CAL-based (CAL), period life table (period), and overlapping cohorts (OC) measured as standard deviation (S_0) from 1816 to 2015 in France. The dashed vertical line indicates the year 1910. (b) dashed lines are standard deviations according to CAL (CAL), period (period), and overlapping cohorts approaches (OC); dots are the lifespan variation (in standard deviations) of the cohorts alive in 1910.

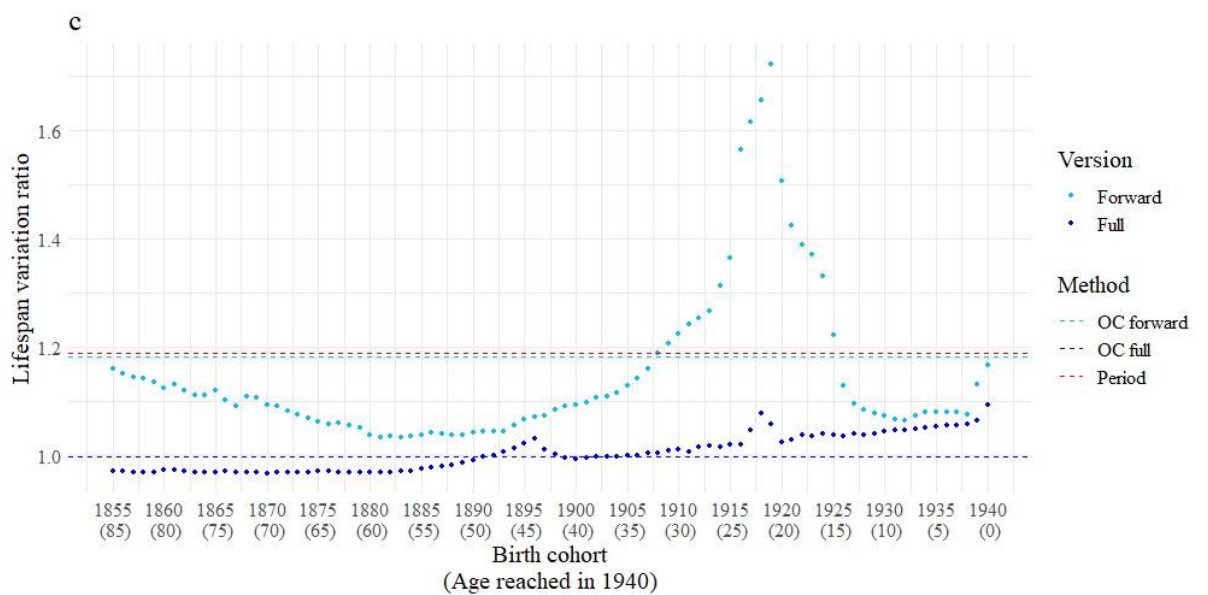
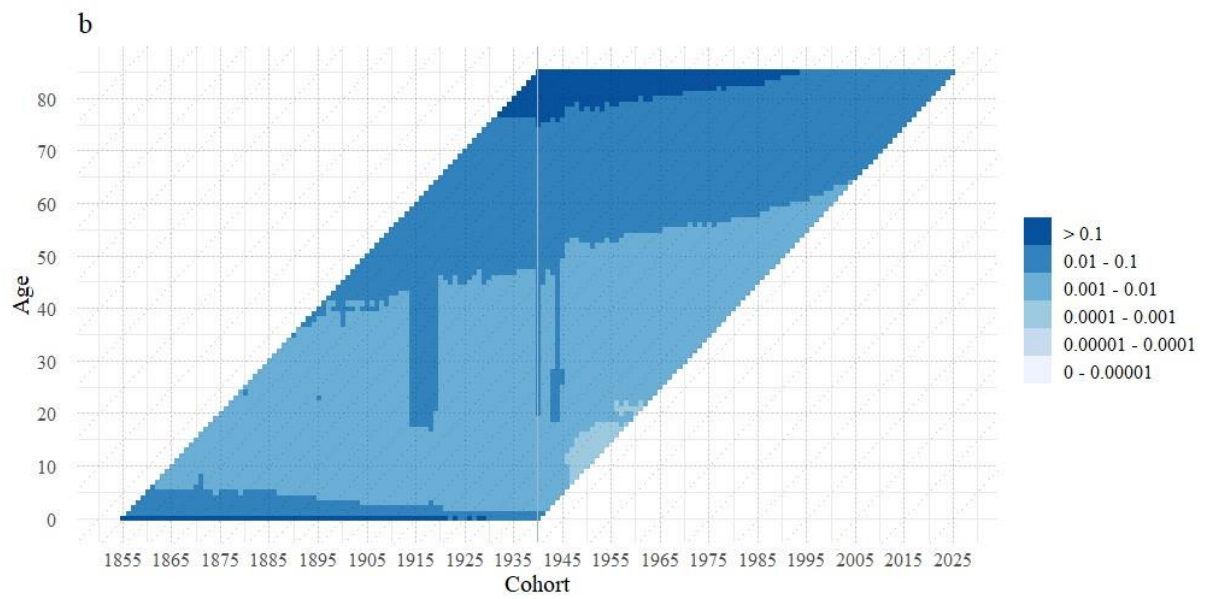
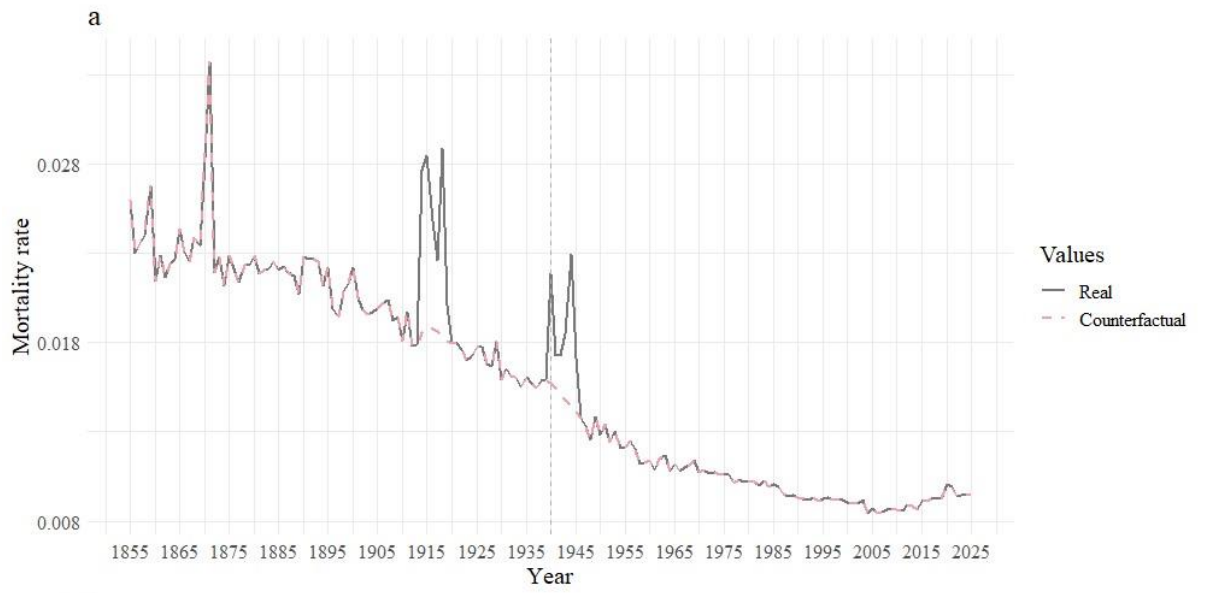


Fig. 4 Lifespan variation and mortality shocks in France. (a) for France from 1855 to 2025, the solid line represents the crude mortality rate (per 100,000), the counterfactual (interpolated crude mortality rates during WWI, WWII) is the dashed line. (b) lexis diagram of the age-specific mortality rates for cohorts born during 1855 to 1940 in France. (c) dots are single-cohort lifespan variation from a forward (light blue) and full (blue) perspective; population-level measures for lifespan variability according to the full (dashed blue) and forward (dashed light blue) perspectives in overlapping cohorts, period lifespan variability in dashed red. contemporary mortality rates.

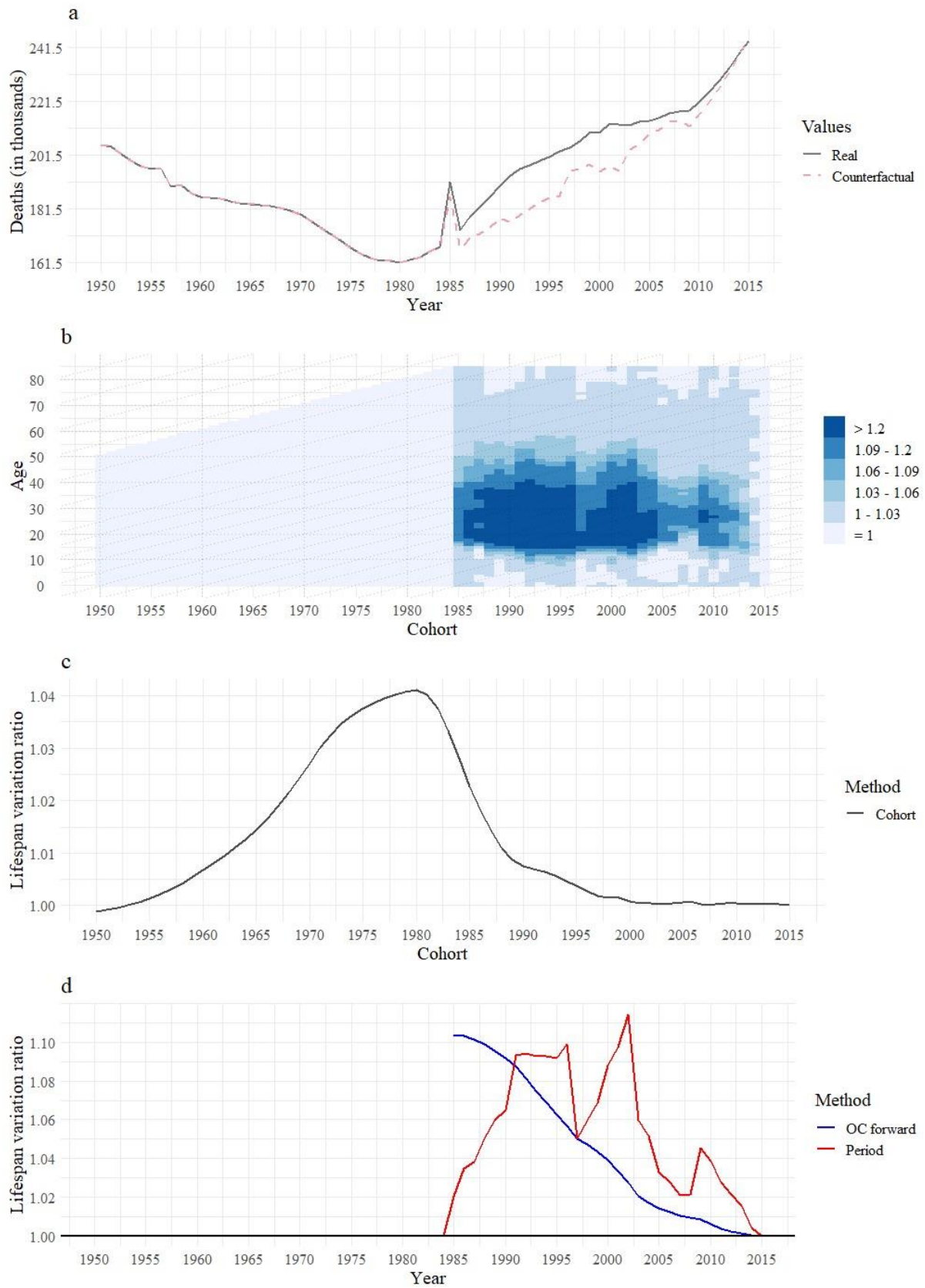


Fig. 5 Lifespan variation during violence episode in Colombia. (a) number of deaths in thousands and its counterfactual (interpolated deaths during the violence episode from 1984 to 2015) for Colombia

from 1951 to 2015. (b) the ratio real / counterfactual of age-specific mortality rates (1984 to 2015). (c) the ratio real / counterfactual of lifespan variation for all individual cohorts born during 1951 to 2015 in Colombia. (d) ratio real / counterfactual of the overlapping cohorts (blue) and period (red) lifespan variation in Colombia from 1984 to 2015.

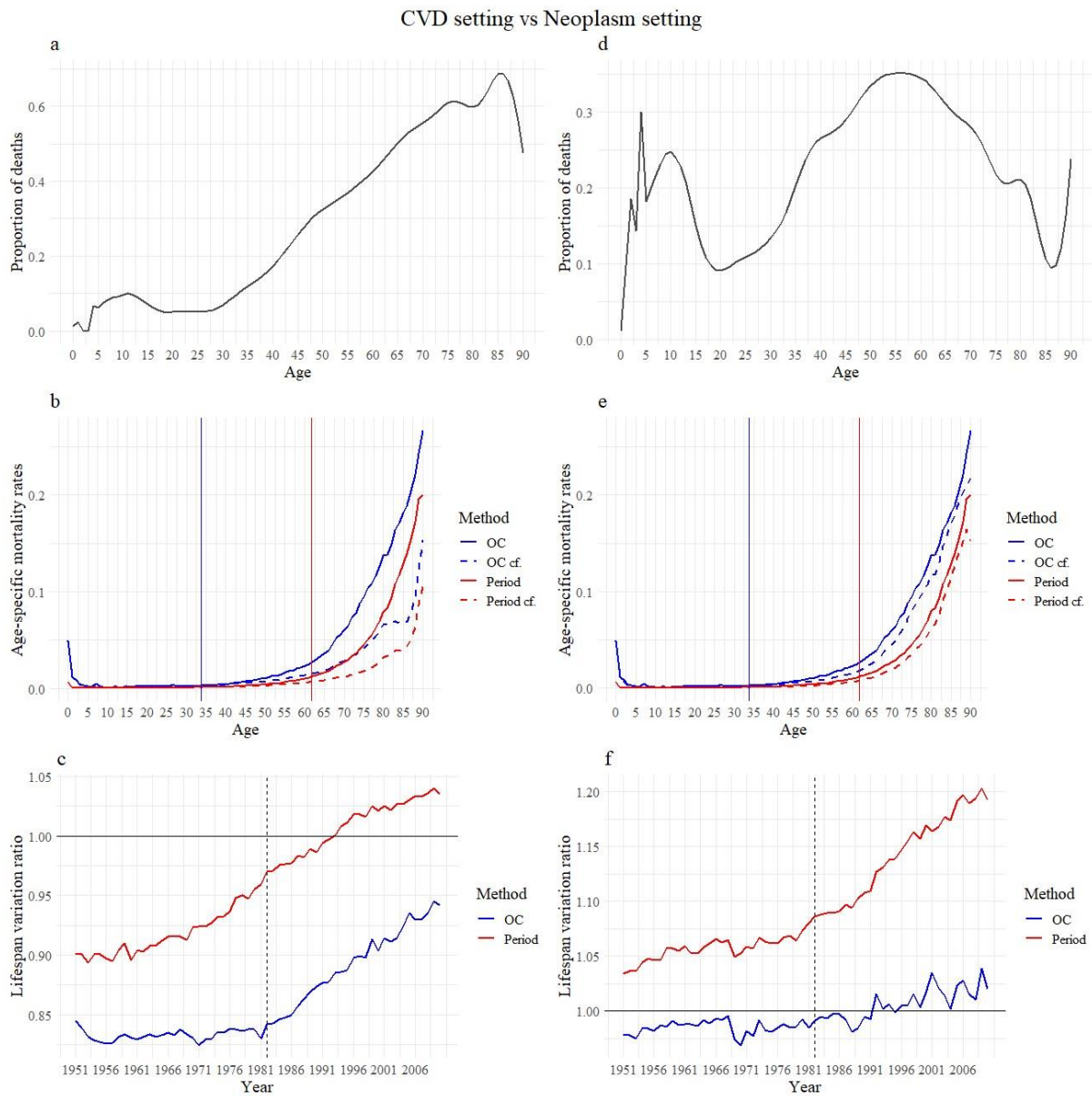


Fig. 6 Lifespan variation and permanent mortality changes in Sweden. (a) and (d), show the proportion of deaths related to cardiovascular diseases (CVD) and neoplasms accordingly for Sweden in 1982. (b) and (e), real and counterfactual age-specific mortality rates for the overlapping cohorts (OC) and period (period) methods for CVD and neoplasms, respectively, for Sweden in 1982; vertical solid lines illustrate the threshold ages associated with each method's mortality rates. (c) and (f), evolution of the

ratio real / counterfactual of the overlapping cohorts (blue) and period (red) lifespan variation for CVD (c) and neoplasm (f) in Sweden for the period 1951-2010.