

Healthy life expectancy, mortality, and age prevalence of morbidity

[Author(s) redacted]

March 12, 2018

Abstract

In calculating period healthy life expectancy, the use of age-specific morbidity prevalence patterns assumes that age captures the important time-variation in the given health condition, i.e. that the disabling process is related to how long an individual has lived. However, many morbidity patterns are better classified by time-to-death. At advanced ages the conflation of an increasing chronological-age mortality pattern and a time-to-death morbidity pattern produces an apparent morbidity pattern that increases with advancing age. Differences in period healthy life expectancy over time or between populations cannot easily be partitioned into morbidity and mortality components because the period morbidity pattern may depend on an unknown future time-to-death process not captured by period mortality. We illustrate these concepts formally and empirically, using morbidity data from the U.S. Health and Retirement Study. While holding the time-to-death morbidity pattern fixed, we show that mortality reduction alone reduces the total life years with disability. We estimate an upper bound of bias in estimates of disability life years that are based on age patterns of prevalence derived from different realistic time-to-death morbidity patterns. Our findings have implications for any between- or within-population comparisons of period healthy life expectancy conditioned on different age patterns of mortality.

1 Introduction

Healthy life expectancy (HLE) is among the most widely used metrics of population health. It combines information on mortality and morbidity (or disability) to summarize the expected years of life lived in good health, however measured.¹ If healthy life expectancy increases faster than life expectancy, morbidity is compressed into a smaller proportion of life. HLE can change because of changes in mortality, morbidity, or both.

Most often, HLE is calculated on the basis of morbidity prevalence and a life table. In calculating HLE, the use of age-specific morbidity prevalence data implicitly assumes that a chronological age pattern best characterizes variation in the health characteristic over the lifespan. However research has shown that the prevalence of many health characteristics in old ages is better measured by a pattern over time-to-death (TTD) or by both age and TTD (Klijs et al. 2010, Riffe et al. 2017, Rehkopf et al. 2017). Most morbidity

¹In this manuscript we use the terms morbidity and disability interchangeably when referring to prevalence patterns.

patterns increase with age in the aggregate, and can claim empirical regularity in this regard. However, we explain how the observed level and shape of a morbidity prevalence age-curve could change due to changes in mortality even with no change in underlying morbidity. At advanced ages the conflation of an increasing chronological-age mortality pattern and a TTD morbidity pattern produces a morbidity pattern that appears to increase with advancing age.

In the cohort perspective the TTD and age prevalence patterns imply the same HLE. Problems arise in the period perspective. Differences in period HLE over time or between two populations cannot easily be partitioned into morbidity and mortality components because period morbidity patterns may depend on a TTD *countdown* process already in place (Wolf et al. 2015).² For the same reason, comparisons of disability prevalence by age between populations with different underlying mortality may not be cut-and-dry.

In this paper we illustrate these concepts formally and empirically, using morbidity data from the U.S. Health and Retirement Study (RAND 2017, HRS 2017). Assuming a fixed TTD prevalence function, we show that mortality reduction alone can lead to a lower fraction of years lived in poor health, and consequently morbidity compression. We estimate the magnitude of potential biases for different empirical prevalence patterns, given different levels of mortality extracted from the Human Mortality Database (HMD 2018). We first explain how the age pattern of morbidity prevalence may partly be a function of mortality using both formulas and a schematic illustration.

2 Morbidity as a function of time to death

A joint pattern of morbidity prevalence by age and time to death could in principle obtain any shape. Limited research suggests that such joint patterns fall into a small set of basic types, which typically slope monotonically in a single direction over these two dimensions (Riffe et al. 2017). For example, Figure 1 shows US female poor self-reported health (SRH) prevalence after age 70 broken down by age and TTD under three views. Panel 1a shows these data as a surface, whose more horizontally running prevalence contour lines indicate that time until death is a more important time measure to describe prevalence of poor SRH than age. Figures 1b and 1c, show the same prevalence data plotted with respect to the respective TTD and age margins. Of these, the TTD lines (within age) in 1b yield a more compact set of patterns, which one could imaginably summarize with a single line without much loss of precision. Age lines (within TTD) in 1c are less articulated and more dispersed. Prevalence patterns such as this are inefficiently characterized by a single age pattern.

Thus an assumption of a single fixed morbidity prevalence pattern by time-to-death is simple, but not unrealistic, at least for characteristics with this joint prevalence pattern. This pattern holds for many, but not all, indicators of health and disability (Riffe et al. 2017). Klijs et al. (2010) has similarly shown that, compared to age, TTD was a stronger predictor of both incidence and prevalence of activities of daily living disabilities, but that functional limitations were better predicted by age. Overall, there is not a broad evidence base on joint age and TTD patterns. While our simplifying assumption of fixed TTD patterns is an acceptable abstraction of the particular joint patterns from which they were derived, these patterns only serve to demonstrate an arithmetic point in the

²TTD prevalence patterns may also arise from age-structured health and mortality transition rates, but this is something we do not treat in detail.

extreme. The degree to which prevalence is better measured by age or time until death for different disabling conditions remains an open research question.

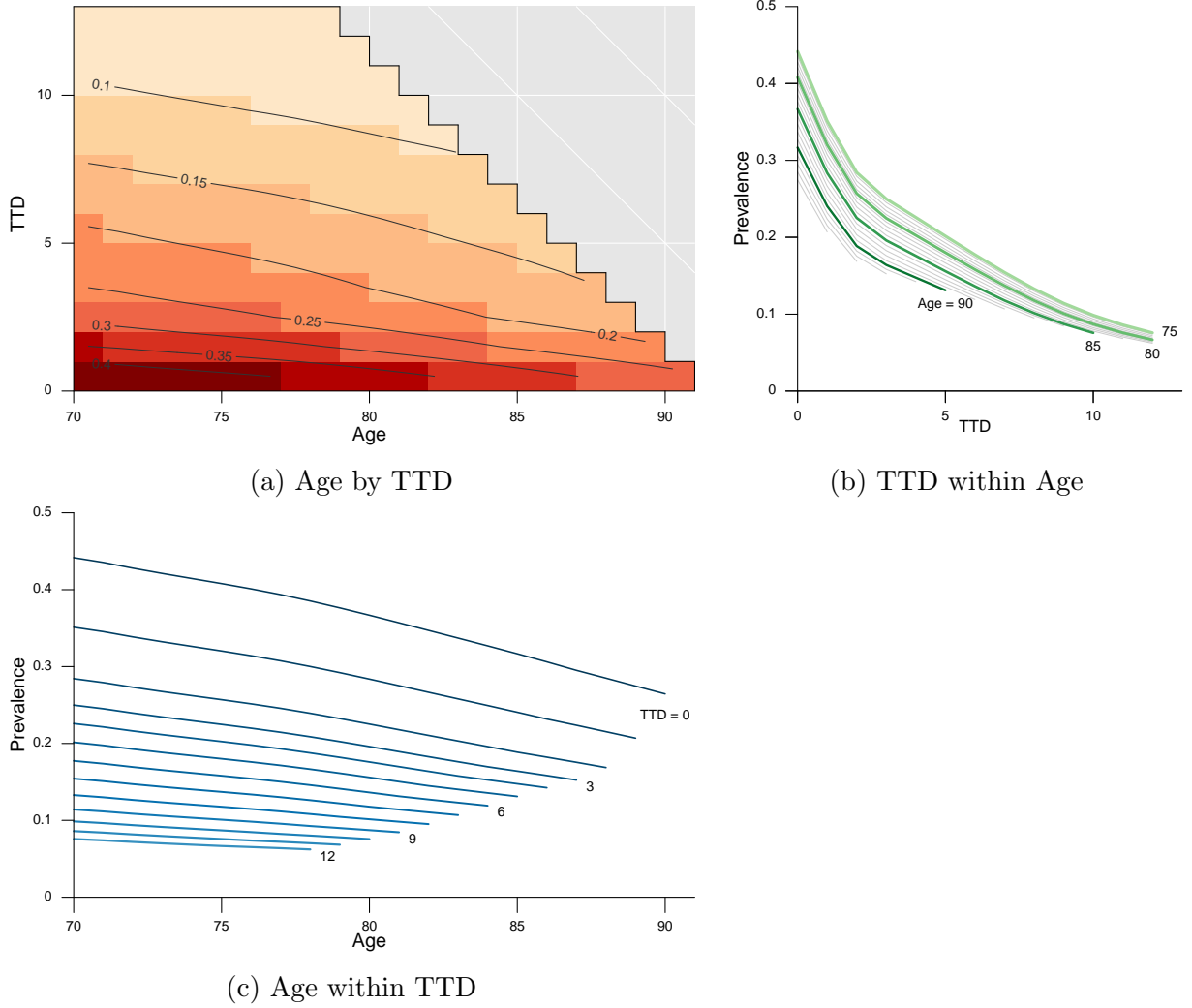


Figure 1: Prevalence of females from the 1920-1924 cohort with self-reported poor health, by time-to-death and age. The same data are shown by age over the time-to-death margin in (b) and by time-to-death over the age margin in (c). Comparing (b) and (c) we conclude that a single time-to-death pattern would better approximate the surface (a) than would a single age pattern.

Still we wager that TTD-dominant prevalence patterns are widespread in old ages, and an understanding of their mechanical role in lifetable-based demographic methods bears lessons for the practices of comparing, decomposing, or analyzing trends in HLE. Imagine a bad health condition, G , with prevalence that varies only as a function of time to death, y , and not as a function of chronological age, a . Since the TTD prevalence distribution is very closely concentrated at the end of life, there will still be an apparent age function, $g^*(a)$. In this case $g^*(a)$ is a heterogeneous aggregate based on both mortality and the underlying TTD process:

$$g^*(a) = \frac{\int_0^\omega g(y)N(a, y) dy}{N(a)} \quad (1)$$

$$= \frac{\int_0^\omega g(y)N(a)\mu(a+y)\frac{\ell(a+y)}{\ell(a)} dy}{N(a)} \quad (2)$$

$$= \int_0^\omega g(y)f(y|a) dy \quad , \quad (3)$$

where $N(a)$ is the population aged a , $\ell(a)$ is life-table survivorship, and $\mu(a)$ is the force of mortality. $f(y|a)$ is the conditional remaining-years distribution, which gives the probability of dying in y years given survival to age a . The expression (3) says that the proportion of those in age a that has condition G does not depend on population structure at all, but only on future mortality rates and the TTD pattern of G , $g(y)$.

A function such as $g(y)$ would have implications for the interpretation of period age patterns of morbidity, and by extension, HLE. If a function such as $g(y)$ holds, it is tautologically true that the measurement of HLE in completed cohorts (or stationary populations) will be identical whether calculated on the basis of $g^*(a)$ or the underlying $g(y)$ pattern. Distortions arise in the interpretation of period HLE under changing mortality, or with period HLE comparisons between populations with different mortality.

Since morbidity prevalence in this scenario is partly a function of mortality, the age patterns of morbidity for populations with different mortality levels or patterns cannot be compared without additional information. Under these circumstances, it is also deceptively tricky to partition period HLE differences into underlying morbidity and mortality components, because the morbidity component is (arithmetically) a function of an uncertain future mortality pattern that accompanies the apparent age pattern of morbidity. Although cohort HLE (a gold standard) is theoretically unbiased (Imai and Soneji 2007)³, and therefore comparable, this quantity cannot be faithfully decomposed into morbidity and mortality components based on age patterns of morbidity and mortality alone if the underlying morbidity pattern is a function of time until death.

A toy example serves to illustrate these concepts. Figure 2 provides a schematic overview of two stationary populations. The underlying survival pattern of these two populations is based on period survival curves from Japanese males in 1970 (a) and 2010 (b) (HMD 2018), but the reader may imagine these as two hypothetical populations. Population (a) has a life expectancy of 69.3, while population (b) has a life expectancy of 79.5, slightly more than 10 years higher. For demonstration, we partition each survival curve into 10 lifespan quantiles, represented with horizontal bars. Our simple TTD prevalence, $g(y)$, is drawn with identical yellow triangles at the end of each lifespan bar. Onset begins 5 years before death and culminates with 80% prevalence. The chronological prevalence function is drawn with blue triangles, with onset at age 50 reaching a maximum prevalence of 50% at hypothetical age 111.5. Both prevalence functions are identical for populations (a) and (b).

The resulting disabled life expectancy (DLE) is shown with barplots next to each stationary population. In population (a) the age and TTD prevalence functions yield the same DLE (and HLE). In population (b) the time-to-death DLE is identical to population (a), but the age DLE is nearly twice as high. For the age prevalence function, it is correct

³We confirm that this remains so even if morbidity prevalence is strongly patterned by time until death.

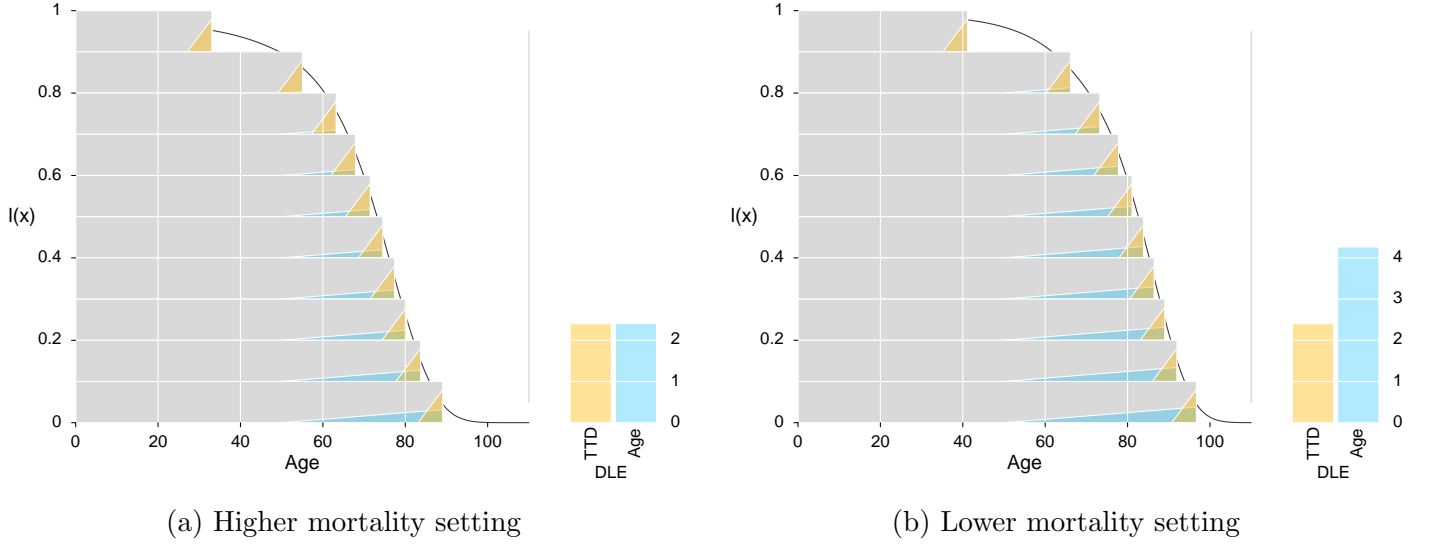


Figure 2: Schematic survival curves from higher (2a) and lower (2b) mortality populations. Each population is subjected to the same chronological age (blue) and time-to-death (yellow) morbidity prevalence patterns. The total prevalence of each sums to disabled life expectancy (DLE), drawn on the right of each survival curve. In 2a the age and time-to-death prevalences imply the same DLE. In 2b the time-to-death DLE is identical to 2a, but the age DLE is two years higher, due entirely to improved longevity.

to conclude that increased longevity leads to increases in prevalence, but for the TTD prevalence function there is no morbidity-mortality trade-off. Instead, improved longevity leads to increased proportions of life lived disability-free, albeit with no change in the absolute concentration of morbidity in the final years of life. Analyses based on the standard Sullivan method (Sullivan 1971) are only capable of predicting increased DLE when projecting from the mortality of (a) to (b). This is so for both kinds of morbidity because the TTD prevalence pattern is erased when the same condition is measured over age. The same Sullivan method can also only conclude that the morbidity of the TTD process is more compressed in (b) than in (a), even though its essential character is unchanged. Prevalence functions are in fact more nuanced than those presented here, often varying by both age and TTD, but this example provides a useful heuristic to understand this source of bias in common applications of the Sullivan method.

3 Data and methods

For our exercises we require lifetables and morbidity or disability prevalence patterns structured by time to death. Lifetables are taken from the Human Mortality Database (HMD 2018). Estimating disability prevalence by time until death is straightforward for health surveys with mortality follow-up modules. We use the RAND version P of the U.S. Health and Retirement Study (HRS),⁴ which consists of 12 survey waves from 1992 through 2014 and an extensive set of measures for various aspects of morbidity and disability. Time to death is calculated directly for deceased respondents as the difference between the date of death and interview date. Within-individual time-to-death

⁴The HRS is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

trajectories of health measures are therefore directly available as such in the data. For a given binary health measure on a single age by TTD grid, the specific prevalence ($g_{a,y}$) is defined as the average value of that respective health measure in that age by TTD interval. We estimate patterns for variables including counts of disabilities in Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL), living in a nursing home, having poor self-rated health, and being unable to name the month of the year.

To ease interpretation, for each sex separately we produced age by TTD prevalence surfaces in which random variation is filtered out by applying logistic regression models with a given binary health measure as the outcome and natural splines over birth cohort, TTD, and age as covariates. We used the predictions from these regression models to produce a series of prevalence surfaces for each variable. To account for the longitudinal nature of the HRS, potentially resulting in correlated error terms, we reproduced the HRS sampling structure in a bootstrap procedure (Efron and Tibshirani 1994). The bootstrap procedure is described in a technical appendix. In a sensitivity analysis, we investigated the influence on our results of the underlying assumption that morbidity prevalence is a smooth function over birth cohort, TTD, and age; both natural splines and LOESS curves over these time measures (Riffe et al. 2017) produced highly similar results. Annotated R-code for all empirical results in this analysis is available in an open online repository.

4 Results

We follow the logic of our examples from section 4.1 to design two empirical exercises that help develop intuition on how time-to-death prevalence patterns create age patterns of prevalence, and how they may lead to bias in common decomposition approaches. We first develop an empirical sense of how different time-to-death morbidity prevalence curves interact with mortality in late life to change the shape and level of age patterns of prevalence. In section 4.2 we examine the degree to which standard decompositions of disabled life expectancy into disability and mortality components could be biased if these relationships hold.

4.1 Time-to-death morbidity interacted with mortality

The age pattern of morbidity prevalence observed in the cross-section (in older ages) depends on the extent to which prevalence is principally described by age versus time-to-death and on the underlying mortality level. If the prevalence is principally a function of TTD, the specifics of its shape are also important. In Figure 3 we show the age-translations of different schematic TTD prevalence patterns when interacted with fictitious stationary populations under different mortality levels.

The first column contains three different types of disability, all of which are experienced by half of the population at the time of death, but which differ in the timing of onset prior to death and in the steepness of the curve with the approach to death. The first type of disability is virtually nonexistent 5 years prior to death, but then increases very rapidly as death approaches. The middle variant of disability is rare 15 years before death, but increases to about 20 percent of the population 5 years before death and rises sharply thereafter. The bottom figure depicts a disabling process that although still strictly determined by time-to-death, is common and accumulates very slowly starting from about 50 years before death.

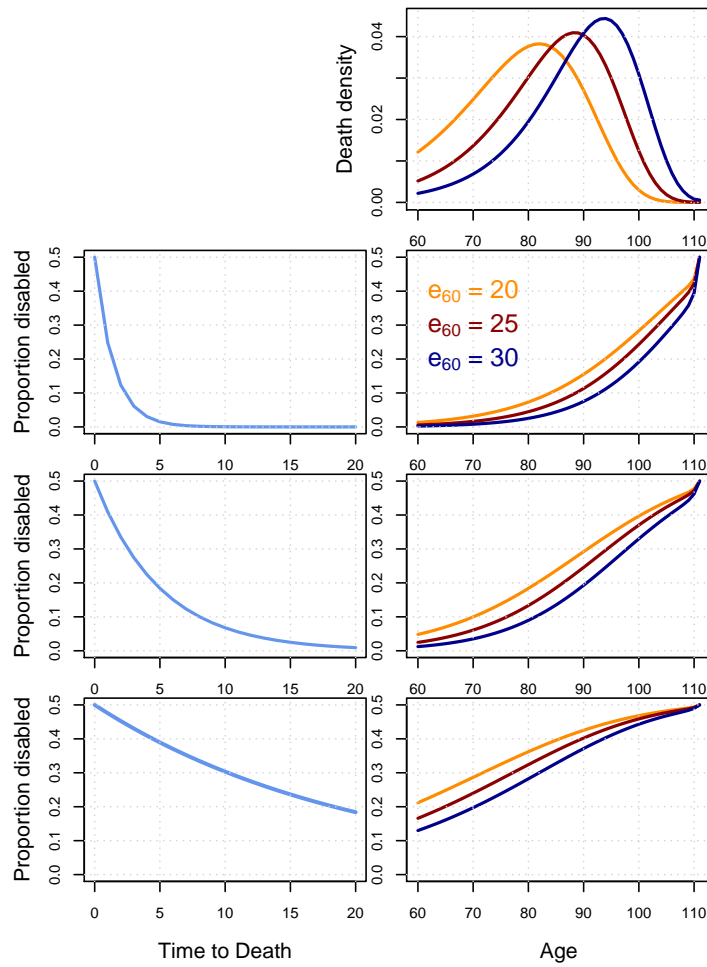


Figure 3: The age pattern of morbidity prevalence in ages 60+ derived from interacting different TTD prevalence patterns (left column) with fictitious stationary populations subject to the color-coded death distributions depicted in the top right figure.

The second column translates the time-to-death disability prevalence curves into the “apparent” chronological age prevalence of disability for different mortality levels depicted by the death density curves above. These mortality levels roughly correspond to USA males in 2002 $e_{60} = 20.0$ years, Canadian females in 2004 $e_{60} = 25.0$ years, and projected Japanese females a decade or so from now $e_{60} = 30.0$ years (latest observed level in 2012 was $e_{60} = 28.3$ years (HMD 2018)). In all cases increasing remaining life expectancy results in decreasing age-specific disability prevalence by chronological age. With steeply increasing disability prior to death (first row), the differences in disability prevalence are largest above age 80, where the bulk of mortality occurs, while with more gently increasing disability (second and third rows) the greatest differences in disability age-prevalence curves appear at younger ages.

These mortality-induced differences are large and also qualitatively different. In the middle variant, which closely resembles the TTD prevalence of disability in being able to bathe oneself, a 10-year increase in e_{60} results in a 50 percent drop in disability prevalence at age 80 from around 20 to 10 percent. Meanwhile, the age at which a quarter of the population were considered disabled in this scenario differed by about 5 years with a 5-year improvement in e_{60} from 20 to 25. The sharply increasing TTD prevalence pattern (top row) translates to a concave age pattern of prevalence, whereas the high and gradually increasing TTD pattern in the bottom row translates to a convex age pattern. This sheds light on some limiting conditions: As a TTD prevalence pattern tends to concentrate in the final moment of life, its corresponding age pattern will tend to be proportional to the force of mortality; as a TTD prevalence pattern tends to uniformity, its corresponding age pattern will also tend to uniformity.

4.2 The impact of mortality differences on comparisons of disabled life expectancy

The Sullivan method is the most commonly used method to partition life expectancy into estimates of the average life years lived in a state of good health (HLE) or disability (DLE). Its popularity owes to its minimal data requirements. Only current age-specific disability prevalence rates are needed in addition to a life table. Specifically, the stationary number of person-years with disability in age-group x to $x + n$ is the product of the person-years lived from the life table ${}_nL_x$ and the proportion disabled ${}_n\pi_x$. The total DLE is the sum of ${}_n\pi_x \times {}_nL_x$ over all age groups (Sullivan 1971).

Comparing populations on the basis of HLE or DLE is standard practice in macro studies of population health. The difference in either metric, either a within-population difference over two time periods or a between-population difference in the same time period, is often decomposed into mortality and morbidity components on the basis of differences in mortality rates and morbidity age-prevalence respectively. Andreev et al. (2002) propose an analytic decomposition of HLE based on age patterns of mortality incidence and morbidity prevalence, which partitions a difference in HLE into two additive components for mortality and morbidity.

If we take the interaction between TTD-prevalence and mortality from section 4.1 at face value, then there is reason to suspect that decompositions based on an age-pattern of prevalence might be biased. Specifically, a marginal age pattern of prevalence can itself be partitioned into morbidity and mortality components, at least for the TTD-type prevalence patterns considered here, which appear to be common. This relationship is not accounted for in standard decompositions, which here constitutes a compositional bias in

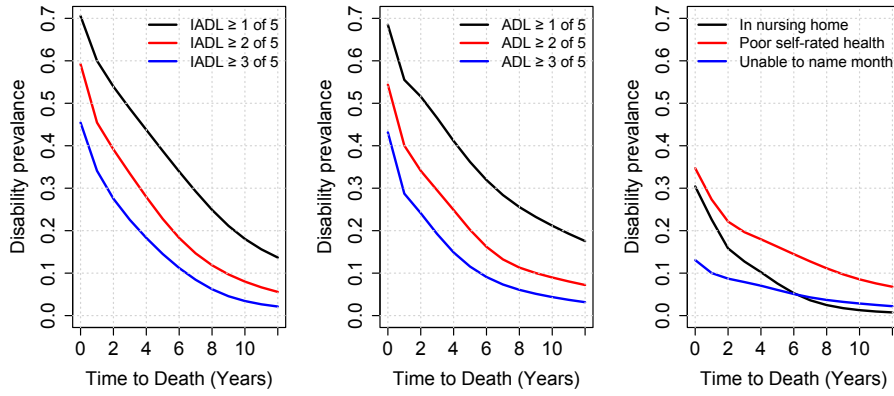


Figure 4: The disability prevalence by time to death for various disability types, based on U.S. HRS data (females).

the decomposition. To get a sense of the upper bound in the magnitude of this bias, we test how the disability/morbidity component of the Andreev et al. (2002) decomposition method would change when a fixed time-to-death prevalence pattern is applied to different mortality regimes using empirical data.⁵

Specifically, we consider the age prevalence of difficulties in carrying out at least 1, 2, or 3 (out of 5) functional Activities of Daily Living (ADL), difficulties in carrying out at least 1, 2, or 3 (out of 5) instrumental Activities of Daily Living (IADL), living in a nursing home, having poor self-rated health, and being unable to name the month of the year. We estimate a linear time-to-death prevalence profile for each disability by averaging over the predicted time-to-death profiles for each lifespan and birth cohort. The TTD prevalence of morbidity for each disability type is shown for females in Figure 4.

We calculate the apparent period chronological age prevalence of morbidity, $g^*(a)$ for all medium to large populations of the Human Mortality Database, had they experienced the US time-to-death profile of morbidity. Eastern European countries are excluded from this exercise due to widely varying age patterns of mortality, particularly in the years surrounding political transition. To calculate the age-pattern of morbidity, we assume the survival pattern of each lifetable to be a stationary population and apply a discretization of equation (3).

We then make pair-wise comparisons of DLE between each population in the same year for the years 1980, 1990, and 2000. For within-country comparisons, we compare each population in 10-year jumps, for all years starting from 1950. Altogether this leads to 187 within-population comparisons and 1785 between-population comparisons for each sex. Finally we decompose the change or difference in DLE between the population pairs into mortality and morbidity components using the method of Andreev et al. (2002). The true value of the change or difference in DLE and the true value of the disability component are both zero by design. Thus, the estimated disability component from this decomposition gauges bias.

We compare the association between the change in the estimated disability component and the difference in remaining life expectancy at age 60 for each population pair

⁵Results from this exercise would be nearly identical if decomposing with the method of Nusselder and Looman (2004), since we do not examine the age pattern of bias (Shkolnikov and Andreev 2017).

in Figure 5. This provides a rough empirically-based estimate of the upper bound of the change in the disability component that is attributable to the different underlying mortality levels of any two populations being compared. Thus for 2+ ADLs, if female e_{60} increases in a country by 5 years, up to about 1 year of the reduction in DLE that is attributed to the disability component could be solely arising from the decrease in mortality.

Overall, the relationship between the change in disability component and the increase in e_{60} is strikingly linear, although the slopes differ for males and females and by disability type. Two factors apparently account for this variation: The degree to which prevalence is concentrated close to the moment of death, and the overall level of the TTD prevalence curve. The second correlates closely with the first: A TTD prevalence curve is higher on average if it increases gradually with the approach to death, which implies a higher proportion of prevalence spread out over a wider range of ages before each age at death. “Real” departures from this upper bound depend on how well a 1-dimensional TTD pattern describes its underlying joint age by TTD prevalence pattern, the extent to which the US average TTD prevalence is representative, and the departure from the stationary population assumption.

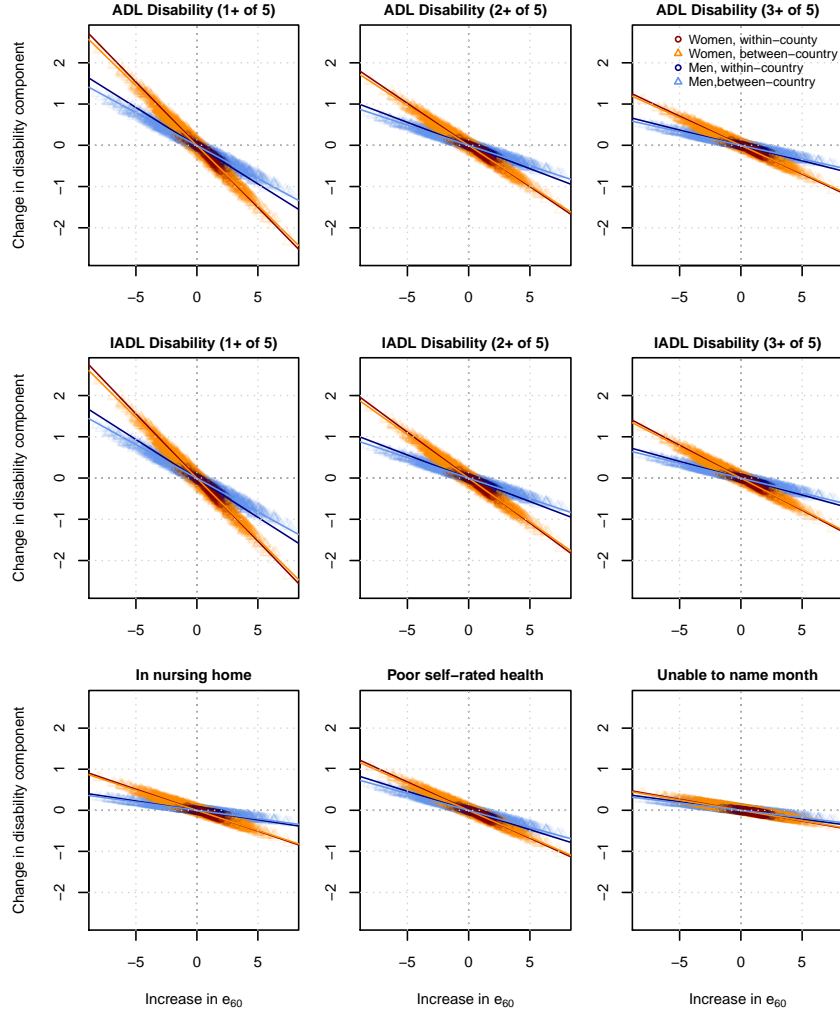


Figure 5: Results of the hypothetical decomposition exercise. The size of the morbidity component using a standard decomposition method is plotted against the difference in remaining life expectancy at age 60 (e_{60}) in each pair of populations. Linear trend lines are also provided for each sex and decomposition type.

5 Discussion

Healthy life expectancy is a popular measure of population health, forming a snapshot of the hypothetical life years lived in good or poor health. This information is well-captured in the cohort perspective, irrespective of whether morbidity prevalence is a function of chronological age or time to death, or whether morbidity prevalence is estimated directly or derived from a multistate incidence-based model. However, difficulties arise in the interpretation of period differences in HLE. The age pattern of disability prevalence can increase or decrease solely as a function of mortality change when the underlying TTD morbidity prevalence is held constant. Thus, for instance, observed widening ratios in the age profiles of disability prevalence between subgroups (Crimmins and Saito 2001) cannot be attributed to changes in the disabling process without taking into account changing mortality profiles. This observation should inform the practice of forecasting observed age-specific rates of decline in disability (Manton et al. 2006, Khaw 1999), and it especially calls into question the more common practice of holding age-patterns of disability prevalence fixed in morbidity projections. Health economists refer to a similar ‘red herring’ argument, namely that medical costs are more closely associated with time to death than with chronological age. As a result, health care cost projections based on a chronological age rather than time-to-death pattern of expenditure are artificially inflated when coupled with forecasted improvements in mortality (Zweifel et al. 1999, Lee and Miller 2002, Geue et al. 2014).

It is preferable to measure changes in health or disability from a cohort perspective (Manton and Land 2000, Manton et al. 2008, Christensen et al. 2013). Manton and Land (2000), for instance, found large differences between period and cohort estimates of active life expectancy (ALE). ALE at ages 65 and 85 was between 1.6 and 2.6 times larger in the cohort perspective than for similar period estimates, and the expected years of life disabled were smaller in the cohort perspective. Additionally, they uncovered larger differences between the cohort and period perspectives for men than women, which they attribute to sex differences in disability transition rates. The cohort-period direction of difference in ALE is also consistent with a fixed TTD pattern combined with improving mortality, as of our toy example in Section 2. We hypothesize that some of these larger differences might also be attributable to larger mortality reduction among men. Further, although cohort HLE estimates are unproblematic as an index, decompositions of prevalence-derived HLE differences between cohorts (with different mortality) into morbidity and mortality components is still potentially biased. This is because the age pattern of morbidity is itself decomposable into morbidity and mortality components.

As our example illustrates, differences in underlying mortality can imply differences in the age profile of disability. It is of course more intuitive to reframe this statement to say that differences in health transitions can drive aggregate mortality change. That mortality levels and disability prevalence are related is perhaps not surprising. Several studies have looked at the macro relationship between overall mortality levels and sex differences in HLE. At higher levels of life expectancy, female advantage in healthy life expectancy diminishes, or even reverses into male advantage (Van Oyen et al. 2013). As well, the larger the proportional female advantage in longevity, the larger the female excess in the proportion of life in poor health (Luy and Minagawa 2014). Additionally, although the association between the severity of chronic conditions and poor health was found to be similar for men and women, morbidity prevalence rates are generally higher among women (Case and Paxson 2005). It would be worthwhile to investigate whether

there might not only be differences in the composition of chronic conditions between the sexes, but whether the underlying morbidity pattern itself might differ between the sexes in its joint age and time-to-death pattern (Riffe et al. 2017).

Empirically, the prevalence of health conditions in the final decades of life is in most cases not an exclusive function of time to death; rather, morbidity often varies as a function of both age and time until death, and expressing morbidity prevalence as a function of both can increase precision (Klijs et al. 2011, Riffe et al. 2017). There is great variety in the temporal variation of the prevalence of late-life health conditions. There is also great variety in individual trajectories with the approach to death (Lunney et al. 2003).

That morbidity prevalence may for certain health conditions be patterned by time to death does not require that morbidity incidence is a function of time to death. First, a sequence of health states wherein mortality risk increases in each successive state could produce a time-to-death prevalence pattern.⁶ Second, it is also plausible that some morbidity conditions are linked to a more general process of dying, thereby linking morbidity to a process that ends with death and consequently producing a time-to-death prevalence pattern. For example, certain conditions may manifest themselves that are not primary causes of the impending death but consequences of nearness to death caused by some other primary factor. Neither of these explanations conflicts with the reality that causes must precede effects, and that therefore death cannot cause the morbidity that precedes it (Lynch 2015). A third possibility, also consistent with the second, is that certain health transition incidence rates are indeed patterned by time to death, as reported by Klijs et al. (2010).

To model prevalence as a function of time to death requires no surreal understanding of how things work, but is rather a modelling choice (Wolf et al. 2015). When modeling for descriptive or exploratory purposes (as we have done to produce e.g. Figure 3), and possibly for projections, one can safely use TTD as a predictive variable. However, using time to death in models intended for causal interpretation is more perilous; Some argue that TTD may function as a proxy for unobserved variables such as biomarkers for impending mortality (Wolf et al. 2015) although a limited exploration of this possibility yielded mixed results (Rehkopf et al. 2017). Others argue that including TTD as a proxy in models will introduce omitted variable bias (Lynch 2015). This does not preclude use of TTD patterns in classic aggregate demography settings, such as lifetable-based methods as here. Much more empirical work is needed in order to determine whether modeling morbidity prevalence as a function of time to death is more widely applicable to other health conditions, younger ages, more recent birth cohorts, and other populations in different stages of epidemiological transition. Further, even joint age-TTD prevalence patterns, such as that of Figure 1 may change over time, and are not fixed as in our examples. Nevertheless, the distortions demonstrated are likely to arise in everyday practice when comparing health trends over age between populations or over time, since many health conditions appear to show strong time-to-death components. To further test the bias we hypothesize and to design more robust synthetic measures of life lived in good or poor health, more effort should first be given to data collection, mortality followup, and measurement practices.

Increasingly, data on mortality incidence and disability onset and recovery are available from multiple waves of survey panels such as the HRS, allowing researchers to cal-

⁶Even a minimal state model with no age patterns for state transitions or mortality risk within states can produce both an overall age pattern of mortality and a time-to-death pattern of morbidity prevalence.

376 culate healthy life expectancy using sophisticated multistate models. Unfortunately such
 377 data is not available in all countries, or if it is, time trends are limited to the recent
 378 past. We are not arguing that a TTD approach should replace multistate models when
 379 such data is available. Our aim is rather to expose the implications of comparing healthy
 380 life expectancy from age-structured prevalence-based models with different underlying
 381 mortality regimes.

382 6 Conclusions

383 We describe a likely source of bias in comparisons of healthy life expectancy or life
 384 years lived in poor health. This bias derives from a failure of age structure to capture
 385 the principal pattern of time variation in morbidity. In short, age standardization does
 386 not guarantee comparability of morbidity levels. Comparisons of period healthy life
 387 expectancy calculated from marginal age patterns of morbidity prevalence cannot be
 388 directly partitioned into morbidity and mortality effects if the joint age by time to death
 389 morbidity prevalence pattern varies at all by time until death.

390 In Section 4 we provide an example of a variable whose joint prevalence pattern over
 391 age and time-to-death is better summarized by a single time-to-death pattern than by an
 392 age pattern. We show how simplistic time-to-death and age-patterns of morbidity interact
 393 with lifetable survivorship to yield opposite conclusions on morbidity compression under
 394 improved mortality. A fixed increasing age pattern of morbidity predicts a greater burden
 395 of morbidity under improved survival, but a fixed time-to-death pattern of morbidity
 396 predicts no change in the average time spent in poor health, and therefore morbidity
 397 compression. In Section 4.1 we describe in greater detail how a range of schematic time-to-
 398 death morbidity patterns interact with different stationary mortality levels to produce age
 399 patterns of morbidity. We describe how concave age patterns of prevalence are suggestive
 400 of strong time-to-death prevalence patterns, whereas convex age patterns suggest gently
 401 sloping time-to-death patterns, or even pure age functions. We demonstrate in Section 4.2
 402 that the degree of bias in partitioning effects depends on the shape and level of time-to-
 403 death variation.

References

- Evgueni M. Andreev, V.M. Shkolnikov, and Alexander Z. Begun. Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates. *Demographic Research*, 7(14):499–522, 2002.
- Peter Bühlmann and Hans R Künsch. The blockwise bootstrap for general parameters of a stationary time series. *Scandinavian Journal of Statistics*, pages 35–54, 1995.
- Bendix Carstensen. Age–period–cohort models for the lexis diagram. *Statistics in medicine*, 26(15):3018–3045, 2007.
- Anne Case and Christina Paxson. Sex differences in morbidity and mortality. *Demography*, 42(2):189–214, 2005.
- Kaare Christensen, Mikael Thinggaard, Anna Oksuzyan, Troels Steenstrup, Karen Andersen-Ranberg, Bernard Jeune, Matt McGue, and James W Vaupel. Physical and cognitive functioning of people older than 90 years: a comparison of two Danish cohorts born 10 years apart. *The Lancet*, 382(9903):1507–1513, 2013.
- D Clayton and E Schifflers. Models for temporal variation in cancer rates. ii: age–period–cohort models. *Statistics in medicine*, 6(4):469–481, 1987.
- Eileen M Crimmins and Yasuhiko Saito. Trends in healthy life expectancy in the United States, 1970–1990: gender, racial, and educational differences. *Social Science & Medicine*, 52(11):1629–1641, 2001.
- Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.
- Claudia Geue, Andrew Briggs, James Lewsey, and Paula Lorgelly. Population ageing and healthcare expenditure projections: new evidence from a time to death approach. *The European Journal of Health Economics*, 15(8):885–896, 2014.
- HMD. Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org (data downloaded February, 2018). 2018.
- HRS. Health and retirement study public use dataset, 2017. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740).
- Kosuke Imai and Samir Soneji. On the estimation of disability-free life expectancy: Sullivan’s method and its extension. *Journal of the American Statistical Association*, 102(480):1199–1211, 2007.
- Kay-Tee Khaw. How many, how old, how soon? *BMJ: British Medical Journal*, 319(7221):1350, 1999.
- Bart Klijs, Johan P Mackenbach, and Anton E Kunst. Disability occurrence and proximity to death. *Disability and rehabilitation*, 32(21):1733–1741, 2010.

- Bart Klijs, Johan P Mackenbach, and Anton E Kunst. Future disability projections could be improved by connecting to the theory of a dynamic equilibrium. *Journal of Clinical Epidemiology*, 64(4):436–443, 2011.
- Ronald Lee and Timothy Miller. An approach to forecasting health expenditures, with application to the us medicare system. *Health Services Research*, 37(5):1365–1386, 2002.
- June R Lunney, Joanne Lynn, Daniel J Foley, Steven Lipson, and Jack M Guralnik. Patterns of functional decline at the end of life. *Jama*, 289(18):2387–2392, 2003.
- Marc Luy and Yuka Minagawa. Gender gaps - life expectancy and proportion of life in poor health. *Health Reports*, 25(12):12–19, 2014.
- Scott M Lynch. Commentary on disability trajectories at the end of life: A countdown model the problems with time-to-death as a predictor of disability. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 70(5):753–756, 2015.
- Kenneth G Manton and Kenneth C Land. Active life expectancy estimates for the US elderly population: a multidimensional continuous-mixture model of functional change applied to completed cohorts, 1982–1996. *Demography*, 37(3):253–265, 2000.
- Kenneth G Manton, XiLiang Gu, and Vicki L Lamb. Long-Term Trends in Life Expectancy and Active Life Expectancy in the United States. *Population and Development Review*, 32(1):81–105, 2006.
- Kenneth G. Manton, XiLiang Gu, and Gene R. Lowrimore. Cohort Changes in Active Life Expectancy in the U.S. Elderly Population: Experience From the 19822004 National Long-Term Care Survey. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 63(5):S269–S281, 2008. doi: 10.1093/geronb/63.5.S269. URL <http://psychsocgerontology.oxfordjournals.org/content/63/5/S269.abstract>.
- WJ Nusselder and CWN Looman. Decomposition of differences in health expectancy by cause. *Demography*, 41:315–334, 2004.
- RAND. RAND HRS Data, Version M, 2017. Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration.
- David H Rehkopf, Luis Rosero-Bixby, and William H Dow. A cross-national comparison of 12 biomarkers finds no universal biomarkers of aging among individuals aged 60 and older. *Vienna Yearbook of Population Research 2016*, 14:255–277, 2017.
- T. Riffe, P. H. Chung, J. Spijker, and J. MacInnes. Time-to-death patterns in markers of age and dependency. *Vienna Yearbook of Population Research 2016*, 14:229–254, 2017. doi: 10.1553/populationyearbook2016s229.
- V. M. Shkolnikov and E. M. Andreev. The decomposition of the difference between two healthy life expectancies. which formula is right? *MPIDR Working Paper*, WP-2017 (015), 2017.

- 481 DF Sullivan. A single index of mortality and morbidity. *HSMHA Health Reports*, 86:
482 347–354, 1971.
- 483 Herman Van Oyen, Wilma Nusselder, Carol Jagger, Petra Kolip, Emmanuelle Cambois,
484 and Jean-Marie Robine. Gender differences in healthy life years within the EU: an
485 exploration of the health–survival paradox. *International Journal of Public Health*, 58
486 (1):143–155, 2013.
- 487 Douglas A Wolf, Vicki A Freedman, Jan I Ondrich, Christopher L Seplaki, and Brenda C
488 Spillman. Disability trajectories at the end of life: A countdown model. *The Journals*
489 *of Gerontology Series B: Psychological Sciences and Social Sciences*, 70(5):745–752,
490 2015.
- 491 Peter Zweifel, Stefan Felder, and Markus Meiers. Ageing of population and health care
492 expenditure: a red herring? *Health Economics*, 8(6):485–496, 1999.

493 A Description of bootstrapping technique

494 The Health and Retirement Survey (HRS) data is a longitudinal dataset, and hence
495 observations are not independent but correlated within individuals. To account for the
496 influence of the correlated error structure on the estimation of age by time-to-death
497 (TAL) surfaces, we applied a bootstrap (Efron and Tibshirani 1994) procedure in which
498 we approximated the sampling structure of the HRS. In our application, a bootstrap with
499 999 iterations provided a good approximation of the distribution of our estimator; if our
500 method is adapted to other datasets, we recommend the use of bootstrap diagnostics
501 applied to various locations on the TAL surface (with special attention to the “edges” of
502 the TAL surface) to determine bootstrap size.

503 A.1 Bootstrap steps

504 An R-code application of our method is available as Supplemental Material. The method,
505 and therefore the code, works through the following 9 steps:

- 506 1. Choose an outcome variable (i.e. BMI, hospital stay duration, etc.).
- 507 2. Create an empty dataset with time dimensions of interest (in our application: birth
508 year, time to death, and chronological age, but this could be more dimensions if
509 desired), which we call the Lexis space
- 510 3. Resample individuals with replacement.
- 511 4. Create natural cubic splines for the dimensions of the Lexis space.
- 512 5. Fit regression model of a type appropriate to the outcome variable to the resampled
513 data with the natural cubic splines (step 4) as covariates.
- 514 6. Use the regression fit to predict the Lexis space and save these predictions.
- 515 7. Repeat steps 3 to 6 until desired bootstrap size is achieved.

8. Take the median or mean over the bootstrap iterations for each Lexis cell to create a Lexis space with median or mean estimates.

- Optional: take $\alpha/2$ and $1 - \alpha/2$ quantiles over the bootstrap iterations for each Lexis cell in order to create Lexis spaces containing the lower and upper $1 - \alpha$ confidence bounds of the estimates.

9. Take a cross-section of the Lexis space with median or mean predictions in order to plot the Lexis surface.

The rest of this appendix sheds more light on these steps. See also the supplementary R-code for the implementation of the method.

A.2 Resampling with replacement and probability weights

The HRS samples individuals each year. Sampled individuals stay in the sample until either death or drop-out. To account for the changing population structure of the United States, probability weights for an individual can differ between observations (i.e. they change over time).

In order to account for the longitudinal nature of the data, we resample individuals following the sampling structure of the HRS. A (re)sampled individual contributes all of his or her observations to that respective (re)sample. This type of resampling is also known as blockwise bootstrapping in a time series context (Bühlmann and Künsch 1995). The probability for an individual to be sampled is equal to the HRS probability weight of the first observation of that individual (RAND variable `wtresp`). We resample individuals with replacement. The number of individuals resampled is equal to the number of individuals in the original sample.

The information contained in the probability weights of subsequent observations of individuals is also used: We create new weights termed rescale weights, which are each individuals probability weights divided by that individuals respective sampling probability (and hence, the rescale weight on the first observation for each individual is always 1). These rescaled weights are used in the regression procedure (see section A.3), where the sampling probability represents the representativeness of the individual in their respective demographic category, the rescale weights capture to what extent this weight changes as a consequence of changes in population structure, and therefore captures to what extent an individuals representativeness changes. In our application of the method, the vast majority of resampling weights did not exceed the 0.90 to 1.10 range, i.e. the sampling probability was never inflated or deflated by more than 10 percent.

A.3 Regression natural cubic splines

To produce smooth age by TAL surfaces we applied regression models to the outcome of interest (BMI, hospital stays, etc.). Covariates in these regression models were natural cubic splines of the dimensions age, birth year, and time to death. The cubic spline is defined as follows:

$$S(x) = \begin{cases} S_0(x) = a_0x^3 + b_0x^2 + c_0x + d_0 & t_0 \leq x \leq t_1 \\ \vdots & \\ S_{k-1}(x) = a_{k-1}x^3 + b_{k-1}x^2 + c_{k-1}x + d_{k-1} & t_{k-1} \leq x \leq t_k \end{cases} , \quad (4)$$

where x represents points on the dimension (in our application, a time dimension) of interest, and t_0 and t_k denote the two extremes (endpoints) of that part of the dimension of interest for which we wish to create this spline. Furthermore, the constraints

$$S_{k-1}(x_i) = S_k(x_i) \quad (5)$$

$$S'_{k-1}(x_i) = S'_k(x_i) \quad (6)$$

$$S''_{k-1}(x_i) = S''_k(x_i) \quad (7)$$

with $i = 1, 2, \dots, k-1$ must be satisfied. This ensures that the segments of the spline meet at the boundaries (first constraint) in a smooth manner (second and third constraint). The additional constraint

$$S''_{k-1}(x_i) = S''_k(x_i) \quad (8)$$

makes it into a natural cubic spline, i.e the second derivative of each polynomial is set to 0 at the endpoints of our part of the time dimension of interest.

The part of the dimension of interest that can be found between t_0 and t_k is subdivided into $k - 1$ segments, the connection of these segments are known as knots. In our application, knots for birth year were placed at the year 1902.5 to 1925.5 at 5-year intervals. Knots for time to death were placed at 0.5, 1, 2, 4, 7.5 and 10 years to death. Knots for chronological age were placed at 72.5 to 97.5 years of age at 5 year intervals. We denote the natural cubic spline for age as $A(x)$, for birth cohort as $C(x)$ and for time to death as $T(x)$.

For continuous outcomes, we used these splines in linear regression models, i.e.

$$\mathbb{E}[Y|x_a, x_c, x_t] = A(x_a) + C(x_c) + T(x_t) \quad , \quad (9)$$

where Y is the outcome variable of interest, and x_a, x_c, x_t represent (observed) points on the age, birth year and time to death axes, respectively. For binary outcomes, we used logistic regression, i.e.

$$\text{logit}\{\mathbb{E}[Y|x_a, x_c, x_t]\} = A(x_a) + C(x_c) + T(x_t) \quad . \quad (10)$$

Ordinal outcomes were dichotomized during data handling and therefore also modeled using logistic regression.

For count outcomes we used zero-inflated Poisson regression, which is a mixture model, combining logistic and Poisson regression. This type of model separates modeling the probability that an observation is an “excess” 0 (or not) from the “count” component (which is modeled if the observation is not an excess 0). The count component was therefore modeled as:

$$\ln\{\mathbb{E}[Y|x_a, x_c, x_t]\} = A(x_a) + C(x_c) + T(x_t) \quad . \quad (11)$$

The 0 or not component was modeled with logistic regression as denoted above. We used this type of model because most of our observations contained a very large number of zeros relative to observed counts and therefore appeared to come from a mixture distribution. Ordinary Poisson models can be compared with zero-inflated Poisson models using the Vuong test, but this was omitted in our application since most outcomes were very clearly zero-inflated and furthermore zero-inflated Poisson models can also adequately model non-zero inflated count outcomes (with some loss of statistical precision, but significance or confidence intervals are not the focus of this work). Most count variables in

our dataset were bounded, e.g. the number of nights that can be spent in the hospital is bounded by 365 (maximum number of days in a year). In those instances, we included those bounds in our Poisson model through an offset term:

$$\ln\{\mathbb{E}\left[\frac{Y}{\text{max bound}}|x_a, x_c, x_t\right]\} = A(x_a) + C(x_c) + T(x_t) \quad (12)$$

\Rightarrow

$$\ln\{\mathbb{E}[Y|x_a, x_c, x_t]\} = A(x_a) + C(x_c) + T(x_t) + \ln(\text{max bound}) \quad .$$

Since the functions $A(x)$, $C(x)$ and $T(x)$ are cubic natural splines, they represent a vector of coefficients corresponding to $S(x)$ as defined earlier. Specification of our models in this manner with natural splines was inspired by work of Carstensen (2007).

Because the three time dimensions included in our models are not linearly dependent, additional constraints on one dimension (such as required in APC modeling, see e.g. Clayton and Schifflers (1987)) were not required. We did not exhaustively investigate the application of splines along different time dimensions, or such splines in addition to splines of the time dimensions already present. However, exploratory results indicated that this would not have substantially affected our conclusions. Similarly, a higher number of knots had only minor effects on the smoothed TAL surfaces. Finally, we also compared our natural cubic splines to other flexible curve fitting methods (such as LOESS and GAM), and found natural cubic splines to provide the best fit overall to the empirical data, as expressed in distance to squared and absolute residuals and propensity for systematic under- or over-estimation of the data along relevant time dimensions (including edge effects).

Splines are known to function less well near the endpoints (t_0 and t_k); to determine if this was present in our data, we investigated the residuals of regression models. We found only weak evidence for the existence of edge effects, and for most outcomes edge effects were absent. For those outcomes where edge effects were potentially present, as a sensitivity analysis, in our resampling procedure we purposely oversampled observations near the edges and estimated TAL surfaces under this regime. We did not find this to substantially affect our conclusions.

A.4 Predicting the age by time-to-death surface

To produce age by time-to-death surfaces, we created a dataset with birth year, time to death, and chronological age as columns. The minimum and maximum values of birth year, time to death and chronological age were equal to those found in the empirical data, and intermediate values were spaced at 1-year intervals. This dataset therefore represents an empty Lexis space. Using the fitted regression models, we then predicted outcomes for this Lexis space; i.e. we created a column with an expected value for each birth year, time to death and chronological age combination (henceforth: Lexis cell) in the Lexis space. Each iteration of the bootstrap added a new column of such predictions. Once all iterations are finished, to produce the final Lexis surface we then plot the median prediction for each Lexis cell. Since three dimensions are more difficult to visualize due to overplotting, in our paper we chose the predictions of the 1915-1919 birth cohort and plotted along the time to death and age dimensions.

611 We found that, overall, predictions for each Lexis cell were normally distributed,
612 and TAL surfaces based on Lexis cell averages were nearly identical to those for TAL
613 surfaces based on median values. However, since some cells had more skewed predictions,
614 namely when the outcome was binomial and predictions were close to 0 or 1 (the edge
615 of the parameter space), we chose the median as the appropriate measure of centrality.
616 Additionally, the median, being the 50% quantile, is also consistent with 95% quantile
617 bootstrap confidence intervals. As their name implies, 95% quantile bootstrap confidence
618 intervals can be calculated by taking the 2.5% and 97.5% quantile for each Lexis cell and
619 plotting those surfaces. The value of the median will, by definition, not exceed the 2.5%
620 and 97.5% quantiles, whereas in theory the mean could exceed these confidence bounds
621 in highly skewed distributions.