Decomposing a Sullivan expectancy: An indirect incidence-based approach

Tim Riffe[1,2*†] and Rustam Tursun-zade[1†]

[1*]OPIK, Department of Sociology and Social Work, University of the Basque Country (UPV/EHU), Barrio de Sarriena s/n , Leioa, 48940, Spain.

[2]Ikerbasque (Basque Foundation for Science, Plaza Euskadi 5, Bilbao, 48009, Spain.

*Corresponding author(s). E-mail(s): tim.riffe@ehu.eus;
Contributing authors: rustam.tursunzade@ehu.eus ;
†These authors contributed equally to this work.

Abstract

Background: A lifetable and the prevalence of some health condition are sufficient to calculate a health expectancy (HLE) using the Sullivan method (Sullivan, 1971), and there are methods available to decompose differences in HLE (Nusselder & Looman, 2004; Shkolnikov, Andreev, et al., 2017). But you still might not be satisfied because observed prevalence could be driven by unobserved mortality differences between health states, and because the lifetable is itself a prevalence-weighted average of the mortality by different health states. Thus these decomposition methods are not guaranteed to isolate health and mortality effects.

Objectives: We aim to (i) transform Sullivan inputs into state-specific mortality schedules consistent with a given mortality rate ratio, and to (ii) convert prevalence into incidence for the case of health deterioration without possibility of recovery. These rates can be considered independent of one another, allowing for rate-based decomposition and a cleaner separation of health and mortality effects.

Methods: Our approach is premised on an algebraic transformation of mortality, prevalence, and an imposed all-cause mortality rate-ratio between states (which might come from the literature). Transition probabilities representing health deterioration are then inferred following the logic of multistate lifetable accounting. We first derive discrete-time transition probabilities, then we reframe the HLE calculation in terms of these. All resulting expectancies (total and for states)

are identical to the traditional Sullivan ones, but corresponding decomposition results are now based on incidence parameters.
Results: Both marginal sums and age patterns attributable to health and mortality differences are different when we decompose using the indirectly derived multistate expectancies versus a standard Sullivan decomposition. Interestingly, this method also enables a new decomposition of total life expectancy that includes a health component.
Conclusions: We give an indirect method to transform Sullivan-style inputs into a simple multistate model with an irreversible health state, and then show how to decompose expectancies in terms of these new parameters. This might seem hypothetical because we do not observe a mortality rate ratio, but we think even so it's more informative than previously proposed Sullivan decompositions.

Keywords: Healthy, Mortality, Prevalence, Incidence, Decomposition

1 Introduction

Healthy life expectancy (HLE, also called healthy life years) is most often calculated and reported using the Sullivan method (Sullivan, 1971), and often the resulting estimate is very close to what would be derived from a multistate model. It's only natural for a demographer to want to decompose a Sullivan-derived expectancy. Decomposition tells us what drives the differences, and has the potential to better guide efforts to improve health and longevity. The parameters of a Sullivan estimate are a single lifetable and information on the prevalence of a health condition. Decomposition approaches that have thus far been proposed include the widely-used Nusselder and Looman (2004) approach, which frames mortality in terms of survival, and the modification proposed by Shkolnikov et al. (2017), which frames the mortality in terms of mortality rates or probabilities. The latter is meant to more appropriately assign mortality effects to ages, since mortality rates can be treated as independent between ages, whereas survival is not. Both of these approaches have a single mortality schedule, however, and both keep health in terms of prevalence rather than incidence (Luy, Di Giulio, Di Lego, Lazarevič, & Sauerberg, 2020).

If the prevalent health condition in question has any lethality penalty, then it's easy to accept that the aggregate mortality over health states must be the prevalence-weighted average of two state-specific mortality schedules, one relatively high, the other relatively low. Prevalence is also in part determined by mortality dynamics. A very lethal health characteristic will not usually have high prevalence, and so on. Since the Sullivan data inputs are endogenous to one another, their use as decomposition parameters does not isolate the effects of health dynamics and mortality. There is not much we can directly do about this except try to find data that yields health transitions and state-specific mortality, but this is often simply not available. In this paper we propose an indirect transformation of Sullivan data inputs into incidence-based multistate transitions: two state-specific mortality schedules, and disease onset. Unless further information is given, this transformation is only possible for the case of irreversible health deterioration. Once the Sullivan data inputs are transformed to

incidence parameters, then you can calculate HLE using a multistate approach (see e.g., Caswell & van Daalen, 2021) and resulting decompositions should properly isolate health dynamics and mortality components (Moretti, Riffe, & Lorenti, 2023; Shen, Riffe, Payne, & Romo, 2023).

There's just one catch: we need to assume a mortality rate ratio between health states, as suggested by Brinks, Landwehr, Icks, Koch, and Giani (2013) or implied by Murray and Lopez (1994). Such information can come from the vast literature reporting such mortality rate ratios, which nowadays includes systematic reviews summarizing reports on the all-cause mortality risk ratio of people with and without specific health conditions. But transition parameter and corresponding decomposition results will naturally vary depending on the rate ratio value chosen. For our example, we apply this method to the case of sex differences in Alzheimer's disease including other dementias around 2018 in Germany and the United States. We show the advantages of the proposed transformation-decomposition procedure, and we promise to report thorough sensitivity testing for different risk ratio assumptions and health prevalence patterns. That future work will allow us to explore the trade-off between isolated but uncertain components versus observed but endogenous components, and other possible limitations of our method.

2 Methods

Let's define some variables

a will index age, we assume discrete single ages through, please pardon the following notation which might seem continuous.
$m(a)$ observed all-cause aggregate mortality rates for age a.
$q(a)$ conditional mortality probabilities derived from the (single age) mortality rates as $q(a) = 1 - e^{-m(a)}$
$l(a)$ lifetable survivorship
$L(a)$ lifetable exposure
$\pi(a)$ the prevalence of a health condition expressed as a probability
$R(a)$ the mortality rate ratio between health states. In practice we may or may not have this by age. We assume this.
$m^u(a)$and $m^h(a)$ the mortality rates of people with and without the health condition, respectively. We infer these.
$p^{h\to u}(a)$ transition probabilities from good to poor health. We infer this using some other intermediate steps described later.

For brevity, and at loss of demographic precision, let's use the following trick to convert rates $m(a)$ to survival probabilities (from birth) $l(a)$:

$$l(a+1) = e^{-\sum_{x=0}^{a} m(x)} \quad , \tag{1}$$

where age 0 is simply 1. From this we can directly transform to conditional death probabilities when needed, following standard lifetable calculations. We approximate

$L(a)$ using linear interpolation:

$$L(a) = \frac{l(a) + l(a+1)}{2} \quad . \tag{2}$$

Let's say that mortality $m(a)$ is the prevalence-weighted average of state-specific mortality rates that we don't observe:

$$m(a) = (1 - \pi(a))m^h(a) + \pi(a)m^u(a) \quad . \tag{3}$$

We do not see $m(a)^h$ ($m(a)^u$) directly, but we are comfortable importing a rate ratio $R(a)$ for the condition from some other epidemiological study or population, such that:

$$m^u(a) = R(a)m^h(a) \quad . \tag{4}$$

If you don't have the rate ratio $R(a)$ by age, then you might use a constant R, and the rest will be the same. In this case we can re-express $m(a)^h$ in the above two equations in terms of $m(a)$, $\pi(a)$, and $R(a)$:

$$m^h(a) = \frac{m(a)}{1 - \pi(a) + \pi(a)R(a)} \quad , \tag{5}$$

And now we have two health-specific mortality schedules that are consistent with observed mortality, prevalence, and an assumed rate ratio. Next we need to re-express prevalence in terms of transition probabilities from good to poor health. First, derive $l(a)$ per equation (1) (or your favorite lifetable method), then split it into healthy and unhealthy parts using $\pi(a)$ in the usual way:

$$\begin{aligned} l^h(a) &= l(a)(1 - \pi(a)) \\ l^u(a) &= l(a)\pi(a) \end{aligned} \tag{6}$$

The change in $l^u(a)$ is a net change $n^u(a)$, where we can now account for the decrement due to mortality, and the remainder must be transitions into poor health. To back out deaths in poor health $d^u(a)$, you can either convert $m(a)^u$ to a probability $q^u(a)$ and multiply with $l^u(a)$, or convert $l^u(a)$ to lifetable exposure per equation (2) (or again your favorite lifetable method), and multiply directly by the rates. The net change $n^u(a)$ plus the deaths $d^u(a)$ sum to the transitions into poor health $t^{h\to u}(a)$, and these can be converted to a transition probability $p^{h\to u}(a)$ by dividing out the healthy survivors $l^h(a)$

$$\begin{aligned} n^u(a) &= l^u(a+1) - l^u(a) \\ d^u(a) &= L^u(a)m^u(a) \\ t^{h\to u}(a) &= n^u(a) + d^u(a) \\ p^{h\to u}(a) &= \frac{t^{h\to u}(a)}{l^h(a)} \quad . \end{aligned} \tag{7}$$

Now we have parameters necessary to calculate HLE in terms of pure incidence, $m^h(a)$, $m^u(a)$, and $p^{h\to u}(a)$. We may also wish to specify an initial composition to

the radix based on $\pi(0)$. Recall the purpose of doing this is to give us conceivably independent parameters, so as to properly isolate effects when we decompose. So our expectancy function needs to be based on just these new parameters. You can set things up using matrix algebra as described by Caswell and van Daalen (2021), or simply iterate up like so to derive $l^h(a)$ and $l^u(a)$. You could do so additively like so

$$\begin{aligned} l^h(a+1) &= l^h(a) - d^h(a) - t^{h\to u}(a) \\ l^u(a+1) &= l^u(a) - d^u(a) + t^{h\to u}(a) \end{aligned} \tag{8}$$

Or multiplicatively like so

$$\begin{aligned} l^h(a+1) &= l^h(a)\left(1 - q^h(a) - p^{h\to u}(a)\right) \\ l^u(a+1) &= l^u(a)\left(1 - q^u(a)\right) + l^h(a)p^{h\to u}(a) \end{aligned} \tag{9}$$

Either way, the initial values are set by $\pi(0)$:

$$\begin{aligned} l^h(0) &= 1 - \pi(0) \\ l^u(0) &= \pi(0) \end{aligned} \tag{10}$$

Now we are ready to decompose! For this, we will recommend a lifetable response experiment approach to decomposition (Caswell, 1989), and that will simply require citing some other work giving the sensitivity calculation for this setup of a multistate model (in preparation, although Shen et al. (2023) takes this general approach). For the present, we will set up the decomposition using the linear integral approach (Horiuchi, Wilmoth, & Pletcher, 2008), using the Riffe (2023) implementation, as in Moretti et al. (2023), which yields virtually identical results. Either way, no new innovation is required to proceed to the decomposition now that expectancies can be equivalently redefined in terms of incidence.

We compare the incidence-based decomposition results with the standard Sullivan decomposition results. To be clear, the two expectancy functions take the following form:

$$HLE = f(m, \pi) \tag{11}$$

for the Sullivan case, versus

$$HLE = g(q^h, q^u, p^{h\to u}, \pi(0)) \tag{12}$$

for the indirect incidence case we propose.

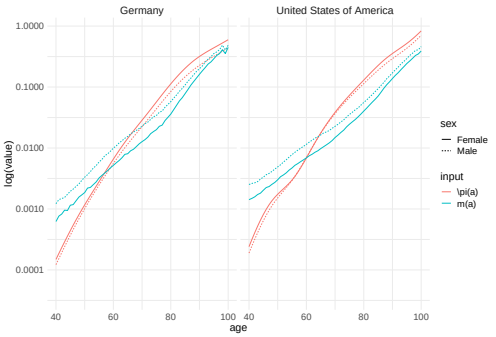
3 Application

3.1 Data

To demonstrate the method, we use data on the prevalence of Alzheimer and other dementia from the GBD (Global Burden of Disease Collaborative Network, 2020) and

mortality data from the Human Mortality Database (Barbieri et al., 2015) for men and women in Germany and the United States as the basic data inputs. Prevalence data from the GBD is delivered in abridged age groups up to 95+. We smooth and graduate this data to single ages using the PCLM algorithm of Rizzi, Gampe, and Eilers (2015) as implemented in the R package `ungroup` (Pascariu, Danko, Schoeley, & Rizzi, 2018), using HMD population counts as offsets. All analyses are truncated at age 40 because prevalence of this condition is nearly zero in younger ages in this data.

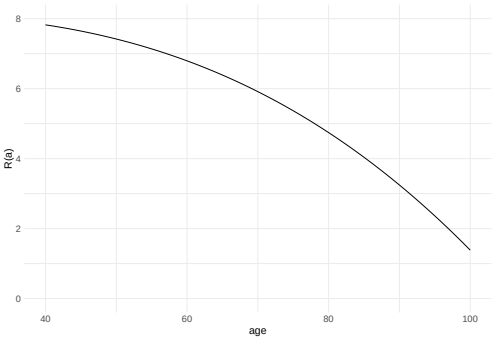
Fig. 1 Sullivan data inputs used: mortality $m(a)$ (rate scale) from HMD and prevalence of Alzheimer's disease and other dementias $\pi(a)$ (probability scale) from GBD



We construct a hypothetical single-age pattern of the risk ratio of all-cause mortality of those with and without this condition. Our hypothetical age pattern of the risk ratio is based on the systematic review of Liang et al. (2021), where a point estimate of 5.9 is stated as the average risk ratio, and studies such as James et al. (2014) and Garre-Olmo et al. (2019), which show that the risk ratio should diminish after age 70. We use a risk ratio pattern defined by the following polynomial equation over ages 0 to 100, shown also in Fig 2:

$$R(a) = 7.964 + .01411a - .0002017a^2 + -.000005974a^3 \quad . \tag{13}$$

Fig. 2 Risk ratio pattern $R(a)$ assumed

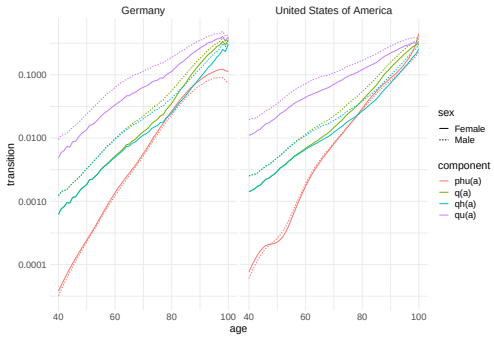


This pattern will be the primary object of sensitivity analysis as this paper develops. As it stands, assuming a flat value of 5.9 for R makes no obvious difference in the visualized decomposition results.

3.2 Results

Following equations (4) and (5), we derive the following two conditional death probabilities $q^h(a)$ and $q^u(a)$, compared with the original lifetable death probabilities $q(a)$ (converted to be consistent with eq (1)) in Fig 3. We also include onset $p^{h\to u}(a)$ (red line) derived from eq (7).

Fig. 3 Derived all-cause conditional mortality probabilities ($q^h(a)$,$q^u(a)$) by dementia status compared with original HMD lifetable ($q(a)$), as well as derived onset $p^{h\to u}(a)$.



Both the transitions in Fig 3 and the Sullivan inputs in Fig 1 give the same expectancies at age 40, displayed in Table 1.

Table 1 Remaining life expectancy at age 40 without (HLE) and with (ULE) dementia including Alzheimer's disease.

country	sex	HLE	ULE
Germany	Female	41.78	2.15
Germany	Male	38.47	1.11
United States of America	Female	40.30	2.49
United States of America	Male	37.27	1.57

Finally, we show the results of incidence-based decomposition of sex differences in HLE, ULE, and total life expectancy, comparing results obtained when we calculate HLE using the Sullivan parameters of Fig 1 versus the incidence parameters of Fig 3.

The main findings are that (i) initial conditions at age 40 play a negligible role in this data, (ii) males' lower onset of dementia in Germany reduced the sex-gap in HLE by around 4 months, and the overall life expectancy gap by 2 months, whereas in the USA sex differences in onset were negligible, (iii) most of the female advantage in all expectancies (HLE, ULE, Total LE) are due to lower mortality rates among those without dementia, accounting for over 3 years of the HLE and Total LE gaps in

Table 2 Comparison of decomposition approaches for sex differences in remaining life expectancy at age 40 with and without Dementia.

		Incidence				Sullivan	
Country	Expectancy	$\pi(40)$	$p^{h\to u}$	q^h	q^u	m	π
Germany	HLE	-0.00	-0.34	3.62	0.03	3.78	-0.46
Germany	ULE	0.00	0.16	0.41	0.49	0.59	0.46
Germany	Total	-0.00	-0.19	4.03	0.51	4.36	0.00
United States of America	HLE	-0.00	-0.04	3.04	0.02	3.31	-0.28
United States of America	ULE	0.00	0.03	0.41	0.48	0.64	0.28
United States of America	Total	-0.00	-0.02	3.45	0.51	3.95	0.00

both countries, and almost 5 months of the ULE gap, (iv) higher male mortality rates among those with dementia increased the sex gap in ULE and Total LE by around half a year in both countries.

The Sullivan results are different in a few key ways: (i) with the exception of ULE in the USA, more weight is given to mortality for ULE and HLE (ii) the health effect for HLE (π) is symmetrical to that of ULE, whereas for the incidence-based decomposition onset ($p^{h\to u}$) is not, (iii) the prevalence effect is of greater magnitude than the onset transition effect for both expectancies and countries, (iv) 100% of the Total LE difference is attributed to mortality differences by necessity, in essence because the Sullivan approach gives the same mortality to both health states.

4 Discussion

These preliminary results highlight how incidence-based decomposition can yield more specific insights into what explains differences in HLE, ULE and Total LE. In this current proposal, we include a health condition with a very large lethality penalty. We will add another example of a non-reversible health condition with a different prevalence pattern and different likely rate ratio that might highlight other ways that this method gives more information. For example, in the current dementia example, initial conditions at age 40 were unimportant, but in general we'd expect initial conditions to be very important for multistate models with non-reversible states of decreased health.

We think that leveraging the rate ratio as we do here could be a useful trick for other empirical exercises. For example, when estimating multistate transitions from surveys like HRS, one can adjust for mortality underestimation using the directly derived rate ratio and an a higher quality lifetable based on vital statistics. We have not yet reflected on what further minimal assumption would be required in order to extend the method to include recovery transitions.

Funding This work was supported by la Caixa Foundation grant number SR22-00502.

Availability of data and materials This work was based on publicly available datasets. We will give exact instructions for downloading these data, and archive our version for posterity.

Code availability We will share the code used to generate our results upon submission to a journal, as well as an Excel spreadsheet implementing the prevalence-incidence transformations.

Authors' contributions TR conceived of this work. TR and RTZ contributed equally to the code, drafting, revising, and data management.

References

Barbieri, M., Wilmoth, J.R., Shkolnikov, V.M., Glei, D., Jasilionis, D., Jdanov, D., ... Winant, C. (2015). Data resource profile: the human mortality database (hmd). *International journal of epidemiology*, *44*(5), 1549–1556,

Brinks, R., Landwehr, S., Icks, A., Koch, M., Giani, G. (2013). Deriving age-specific incidence from prevalence with an ordinary differential equation. *Statistics in Medicine*, *32*(12), 2070–2078,

Caswell, H. (1989). Analysis of life table response experiments i. decomposition of effects on population growth rate. *Ecological Modelling*, *46*(3-4), 221–237,

Caswell, H., & van Daalen, S. (2021). Healthy longevity from incidence-based models. *Demographic Research*, *45*, 397–452,

Garre-Olmo, J., Ponjoan, A., Inoriza, J.M., Blanch, J., Sánchez-Pérez, I., Cubıı, R., ... Vilalta-Franch, J. (2019). Survival, effect measures, and impact numbers after dementia diagnosis: a matched cohort study. *Clinical epidemiology*, 525–542,

Global Burden of Disease Collaborative Network (2020). *Global burden of disease study 2019 (gbd 2019) results.* Institute for Health Metrics and Evaluation (IHME). Retrieved from https://vizhub.healthdata.org/gbd-results/

Horiuchi, S., Wilmoth, J.R., Pletcher, S.D. (2008). A decomposition method based on a model of continuous change. *Demography*, *45*(4), 785–801,

James, B.D., Leurgans, S.E., Hebert, L.E., Scherr, P.A., Yaffe, K., Bennett, D.A. (2014). Contribution of alzheimer disease to mortality in the united states. *Neurology*, *82*(12), 1045–1050,

Liang, C.-S., Li, D.-J., Yang, F.-C., Tseng, P.-T., Carvalho, A.F., Stubbs, B., ... others (2021). Mortality rates in alzheimer's disease and non-alzheimer's dementias:

a systematic review and meta-analysis. *The Lancet Healthy Longevity*, *2*(8), e479–e488,

Luy, M., Di Giulio, P., Di Lego, V., Lazarevič, P., Sauerberg, M. (2020). Life expectancy: frequently used, but hardly understood. *Gerontology*, *66*(1), 95–104,

Moretti, M., Riffe, T., Lorenti, A. (2023). *Multistate analysis and decomposition of disability-free life expectancy trends in italy 2004-2019* (Tech. Rep.). Max Planck Institute for Demographic Research, Rostock, Germany.

Murray, C., & Lopez, A.D. (1994). Quantifying disability: data, methods and results. *Bulletin of the World Health Organization*, *72*(3), 481,

Nusselder, W.J., & Looman, C.W. (2004). Decomposition of differences in health expectancy by cause. *Demography*, *41*, 315–334,

Pascariu, M.D., Danko, M.J., Schoeley, 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, https://doi.org/10.21105/joss.00937

Riffe, T. (2023). Demodecomp: Decompose demographic functions [Computer software manual]. Retrieved from https://cran.r-project.org/web/packages/DemoDecomp (R package version 1.11)

Rizzi, S., Gampe, J., Eilers, P.H. (2015). Efficient estimation of smooth distributions from coarsely grouped data. *American Journal of Epidemiology*, *182*(2), 138–147,

Shen, T., Riffe, T., Payne, C., Romo, V.C. (2023). Decomposition of differences in health expectancies from multistate life table. https://doi.org/10.31235/osf.io/xcp2f

Shkolnikov, V.M., Andreev, E.M., et al. (2017). The decomposition of the difference between two healthy life expectancies. which formula is right. *Max Planck Institute for Demographic Research*, ,

Sullivan, D.F. (1971). A single index of mortality and morbidity. *HSMHA health reports*, *86*(4), 347,