Life lost, lifesaving, and causes of death.

Tim Riffe*[1] and Aïda Solé Auró[2]

[1]Department of Demography, University of California, Berkeley
[2]Universitat Pompeu Fabra

March 31, 2015

Abstract

The lives and potential years of life lost due to death are presented as a metric for describing the population impacts of death and for comparing causes of death. Lost lives and years of life may be classified by the ages in which deaths occurred, by the ages to which deaths would be postponed were they saved, by the ages through which the lost years would have been lived, or by the distribution of lost remaining lifespans. These temporal perspectives define the potential impacts of death and causes of death on population size and structure, and on the distribution of lifespans within populations. We illustrate these concepts using 2010 all-cause and cause-specific death data for the USA from the Human Mortality Database.

Introduction

A core task of demography is to account for and predict the population pyramid and the forces that shape it. The pyramid represents population size and age-sex structure, and it is shaped by the flows of births, deaths, and migrations[1]. Of these flows, births are usually regarded as the primary driver of variation in the profile of the pyramid,[2] whereas the pattern of survival exerts a gentler influence on the overall tapering and height of the pyramid. Year to year variation in the number of deaths tends to deduct smoothly from a wide range of ages, making

*triffe@demog.berkeley.edu

Research reported in this manuscript was supported by the U.S. National Institute On Aging of the National Institutes of Health under award numbers R01-AG011552 and R01-AG040245 and by the Spanish Ministry of Science and Innovation under award number ECO2013-48326-C2-1-P. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

[1]We are guilty of omitting migration from the following exposition, although some of the methods presented here would translate cleanly to emigration.

[2]This is true to the extent that wars, epidemics, and other mortality shocks effect broad age ranges rather than abrupt age groups.

all but the most severe mortality shocks illegible in the pyramid. This could be one of the reasons why relationships between mortality and age structure have been less charted than those between births and age structure.

Demographers most often quantify death in terms of age-specific rates via the lifetable and its summary indices because these are considered purged of accidental distortions from population age structure. Moreover, public health institutions and news media often also report trends in absolute numbers of deaths from particular causes and the total potential years of life lost (YPLL) due to these deaths.[3] The notion of YPLL dovetails with population size over time. In either of these treatments, mortality and deaths are isolated from age structure.

The point of departure for this article is to treat deaths analytically as population stocks through the notion of potentially saveable life. A potentially saveable life is a life that has been lost, a death, and a set of lost lives is a kind of population. The population of lost lives has static characteristics observed at the moment of death, such as age and sex structure. Were the population still living, it would still be subject to a continued force of mortality and therefore retain a structure of remaining lifespans. Observed mortality patterns may be used to project remaining lifespan structure onto the lost population as an approximation of what would happen were the population to be resurrected simultaneously. This step is a counterfactual exercise, in line with the formal treatment given in Vaupel and Yashin (1987).[4] Saveable life and lifetime can be quantified under various demographic perspectives on age and lifespan. These perspectives refine and supplement YPLL when assessing the population impacts of causes of death, and we think that they would provide useful information for the targetting and planning of public health interventions and the comparison of mortality burdens between populations and subpopulations.

We first formally define what is meant by age and lifespan perspectives, illustrating on the example of all-cause mortality before proposing an extention to causes of death. Concepts are illustrated based on the population of the United States in 2010 based on pre-release data newly collected by the Human Mortality Database (HMD). We propose a selection of strategies for visualizing and arranging results for purposes of reporting or making comparisons. Finally we discuss the limits of these methods and the utility of the information gained by them. All mortality data used in the following comes from the HMD.

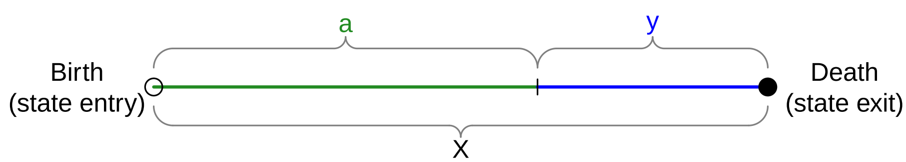
[3] Gardner and Sanborn (1990) review commonly used methods of calculating YPLL. The Global Burden of Disease reports refer to YPLL as YLL. As a news media example, in 2013 the Guardian ran a data blog entry visualizing the years of life lost in the USA due to gun violence (Rogers 2013).

[4] Vaupel (2008) follows in a similar vein, but aims at the effects of lifesaving on period distortions.

Temporal relationships: Age and lifespan perspectives

Population stock in a given year, t, can be structured by birth cohorts or age, the way we typically make population pyramids. If an entire lifespan is denoted by the random variable X, then the remaining lifespan, y, of a still-alive person aged a, $y = X - a$. Figure 1 gives a schematic representation of this simple

Figure 1: A lifeline, where chronological age (years lived) is indexed by a and thanatological age (years left) is indexed by y.



relationship between age and the lifespan for a single person-life. The lived part we call "age" and the yet-unlived part has no common name. Both a and y are placemarkers on the lifeline and could therefore be called "age". We refer to these indices as *chronological* and *thanatological* age, respectively.[5] The chronological and thanatological age perspectives are applicable to state durations in general, but in this paper we focus on the full lifespan.

For a cohort, the distribution of X is given by $f(X)$, which is equal to the lifetable death distribution, $d(a)$, for $a = X$, when the lifetable is specified with a radix equal to unity ($l(0) = 1$). The definition of the survival-conditioned distribution of *remaining* lifetime $f(y|a)$ can be summarized in words as the probability of surviving y years in the future given survival to current age a, and then dying at the exact age $a + y$.[6] Figure 2 shows selected cross-sections of the $f(y|a)$ surface calculated from the 2010 US male period lifetable (HMD). The area under each chronological age-conditioned curve is equal to one. In all cases where the underlying mortality pattern is fixed, the central mass of the curve approaches zero, moving one year down per year lived. In these data, the shape of the center of the curve does not change much until after chronological age 60, where conditional rescaling drives up death probabilities more and more. Upward scaling continues beyond those ages shown here, with $y = 0$ becoming the greatest single value in all ages beyond the modal chronological age at death.

$f(y|a)$ can be used to calculate the population having survived to age a and with y remaining years of life as $P(a, y) = P(a)f(y|a)$, where the total population with y remaining years of life, $P(y)$, is simply $\int_{a=0}^{\infty} P(a, y)\, \mathrm{d}a$ (Brouard 1986; 1989), a single death cohort with members from many birth cohorts. This decomposition sorts the lifeline segments of a living population by the part

[5] Thanatos was the Greek god of death, which marks the end of the lifeline to which y relates. By this token, one could just call chronological age *aphrodesian* age, but this would probably confuse things.

[6] A more explicit defintion is provided in the appendices.

Figure 2: Probability of surviving y years given survival to current age a, US males, 2010

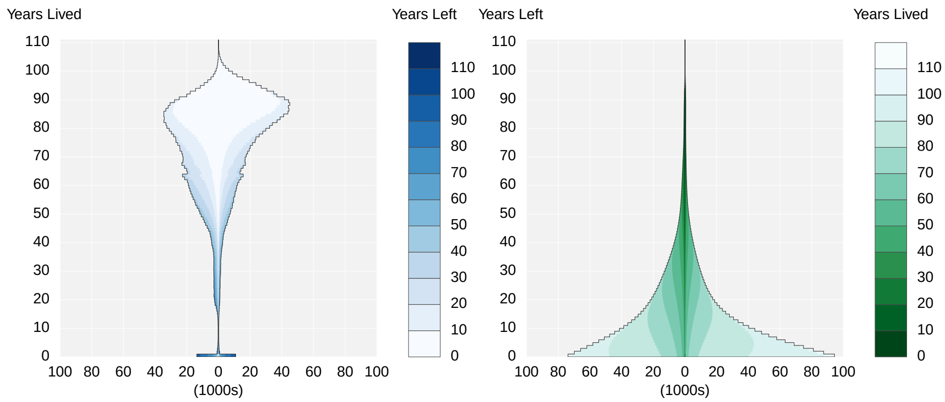
yet-unlived (years left, y) rather than by the part lived. The indices a and y differentiate between the past and future parts of a lifeline, respectively, and by extension of populations when so structured. As Brouard (1986), when comparing the lived and to-be-lived part of a population, we refer to $P(a)$ and $P(y)$. Is it so clear that the dead are no longer part of the population? If a life is completely saved, this life stays in the living population and is not counted as a death, but we (in common thinking) often imagine saved lives as a transient state classification. For demographers, however, among the living there are no saved lives but only lived lives. Still we can quantify hypothetically saveable life, and for this we must look to deaths. It is nice, and often realistic, to think that many of the lives taken by death are or will one day be saveable, but it is difficult to know what mortality rates would apply to a population of saved individuals. Consider the hypothetical population of lives saved a single time from death and subject to the same mortality as the population at large.

Figure 3a (left) shows US 2010 period deaths (the universe of lives that we counterfactually save) by age and sex (males on the left, females on the right). Over 1.23 million deaths each were recorded for US males and females in 2010. Deaths have been decomposed into discrete categories of remaining years of life (see appendix equation (1)), under the assumption that saved lives are subject to the same mortality schedule as the rest of the population and that all 2010 deaths get saved (just once). The results of this decompositon are represented by color bands in Figure 3a. The average chronological age at death observed

Figure 3: Potentially saveable lives (deaths) in the US by sex, 2010

(a) Classified by age (years lived) and sex, and decomposed by hypothetical remaining years of life (years left).

(b) Classified by hypothetical remaining years of life (years left) and sex, and decomposed by age (years lived).



$D^s(a)$ from appendix equation (2)

$D^s(y)$ from appendix equation (5)

for males was a full seven years lower than that for females: 69.9 versus, 76.9, respectively.[7] Figure 3b (right) displays the same decomposition after swapping the y axis and color gradient from Figure 3a. Now thanatological age (years left) of hypothetically saved lives are the primary y axis, while chronological age groups (years lived) are displayed with color. Figure 3b communicates that most saveable lives would live short remaining lifespans once saved and granted the same lifetable mortality. This is so in this data because most saveable lives are already in chronological ages subject to high mortality rates. In general, the only saveable lives that might live very long remaining lifespans are the few deaths that occur in young ages.

Randomly selected saveable males from this population would have on average longer remaining lifespans than randomly selected saveable females (16.7 versus 13.3 years, respectively). This is a paradox because females have lower mortality rates in nearly all ages, and have longer remaining life expectancies in all ages. Female mortality advantage is in this case more than offset by the relative youth of male deaths. Untangling the paradox further becomes a recursive exercise, since the relative youth of male deaths is due to an interaction between mortality schedules and population structure, itself a result of past vital forces.

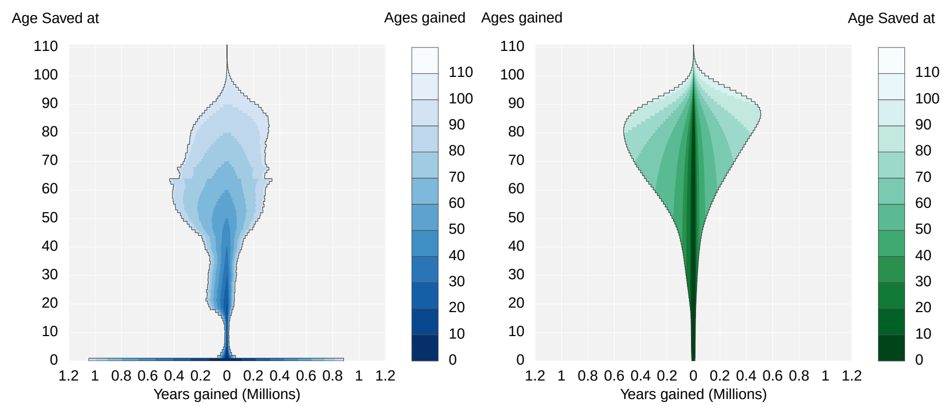
Figure 4a shows the person years of life potentially won by saving all the deaths in each age (see appendix equation (3)) for the same US data, which is essentially a reweighting of Figure 3a by the standard age-pattern of remaining life expectancy that these lives would hypothetically be subject to. Color bands

[7]This differs from period life expectancy (76.4 versus 81.2, respectively) because the population structure is not stationary.

Figure 4: Potentially person years of life won in the U.S. by sex, 2010*

(a) Classified by age at hypothetical saving and sex, $W^s(a)$, and decomposed by future ages to be lived.

(b) Classified by cumulative ages to be lived through and sex, and decomposed by age at saving.



$W^s(a)$ from appendix equation 3

$W^s(a+y|a)$ from appendix equation 6

*Note different x scale from Figure 3.

are assigned by decomposing the total life to be lived into the ages through which it will be lived. For example, if we save all 11700 of the 50-year-old US males that died in 2010, they would live a total of 349000 combined years (under a fixed 2010 mortality schedule), spread out over ages 50 and higher according to $\frac{l(50+y)}{l(50)}$. In Figure 4a we decompose these gained years of life according to the survival-conditioned distribution of remaining lifetime (see appendix equation (5)) and highlight this decomposition with color, while in Figure 4b, *gained* ages become the primary y axis, and color bands represent the ages in which populations in each age group were saved.

Figure 4b represents the cumulative contribution to the population pyramid that would result from saving all lives in 2010 and then surviving them forward according to 2010 mortality conditions. The chronological age axis indexes ages that are lived through at some point in the future, in sequence rather than simultaneously. Under the same assumption of fixed schedules, one could multiply this cumulative age structure with other age-schedules, such as age-specific fertility rates, or the economic age profiles produced by the NTA project, to benchmark the cumulative impacts of mortality on other quantities of interest. For instance, the females who died in 2010 would have given birth to 54122 babies cumulatively over their remaining lifetime assuming they were subject to constant 2010 period fertility (HFD) and mortality. In the present work, we only treat age structure, and we do not examine secondary consequences of this kind.

Causes of death

These basic relationships carry over when deaths and survival are adjusted to account for the hypothetical elimination of particular causes (in the case of independence). In this case, the total number of deaths observed, D, is the sum of the deaths from n causes. To speak of eliminating cause c (for instance, deaths from pancreatic cancer) from the lifetable is to speak of saving D^c lives and then subjecting them to mortality after having deducted cause c from the lifetable. This is problematic in that causes are not independent, and in that reductions in cause-specific mortality are not so thorough and immediate, but it serves as a basis for comparing the relative impacts of different causes on a given population structure. Humans have succeeded in eradicating certain causes of death in the past, and it is not so audacious to imagine that we may do so yet. While these eliminations may not deduct 100% of their magnitude from all-cause mortality due to substitution, there is an undeniable all-cause benefit, and at least we know its bounds. For some causes of death, independence is easier to imagine, such as deaths due to needless violence or particular kinds of accidents. We demonstrate concepts using large cause groupings.

Table 1: Major causes of death in the U.S. by sex, 2010 (HMD)

Cause	Female		Male		Total	
	count	%	count	%	count	%
Cardiovascular	395200	32.0	383077	31.1	778277	31.5
Cancer	369411	29.9	366373	29.7	735785	29.8
Infectious	156086	12.6	149139	12.1	305226	12.4
External	60673	4.9	125733	10.2	186405	7.6
Mental	76270	6.2	41413	3.4	117683	4.8
Infant	10368	0.8	12119	1.0	22487	0.9
Other	167995	13.6	154577	12.5	322572	13.1
Total	1236003	100.0	1232432	100.0	2468435	100.0

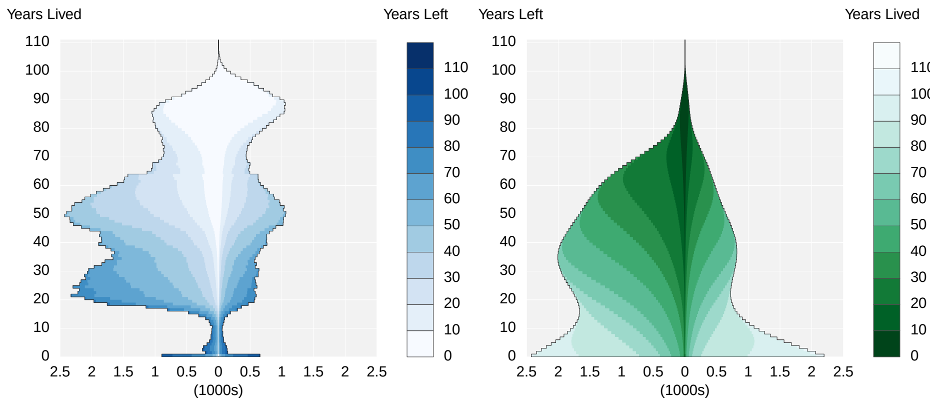
Table 1 lists a selection of grouped causes of death for the United States in 2010.

Now Figures 3 and 4 can be repeated for any particular cause of death, and the profile of each of the four perspectives characterizes the population impact of the given cause of death. Figures 5 and 6 depict the same temporal viewpoints, respectively, but now the deaths decomposed are only deaths to external causes, and external causes have been eliminated from the lifetable functions used for decomposition and redistribution. Causes differ in their impact profiles, and this forms a basis for comparison. As with standard population pyramids, one

Figure 5: Deaths from external causes in the USA, 2010

(a) Classified by age (years lived) and sex, and decomposed by hypothetical remaining years of life (years left).

(b) Classified by hypothetical remaining years of life (years left) and sex, and decomposed by age (years lived).



$D^{sc}(a)$ $D^{sc}(y)$

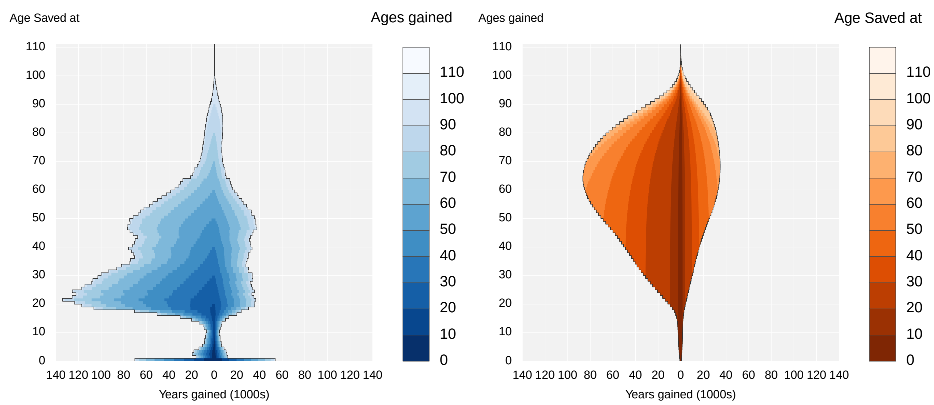
may prefer the use of percent scales to facilitate comparisons between causes or countries. These figures for external causes give visual form to the intuition that demographers have about external causes of death. External causes are important in young ages, and affect males more than females. Saving a randomly selected death from an external cause will on average have a higher payoff in terms of expected years of life gained than does preventing a death in general. Further, a typical life saved will traverse many working ages, and reach well into old ages. The group with the most to gain by eliminating external causes are males in their 20s (clearest in Figure 6).

The average chronological age of deaths to external causes in the USA in 2010 was 47.9 for males and 57.1 for females. Their cause-deleted mean remaining lifetimes were 34.0 and 29.8 years, respectively. The mean age-at-saving of all the person-years hypothetically won under these same conditions becomes 36.7 for males and 40.3 for females, whereas the mean of the ages *enjoyed* by these hypothetically saved people are 59.7 and 63.3, respectively. Means do not tell the story as well as images.

Figure 6: USA, 2010 Deaths from external causes, years of life potentially won*

(a) Classified by age at hypothetical saving and sex, $W^s(a)$, and decomposed by future ages to be lived.

(b) Classified by cumulative ages to be lived through and sex, and decomposed by age at saving.



$W^s(a)$ from equation 3

$W^s(a+y|a)$ from equation 6

*Note different x scale from Figure 5.

Observed patterns

In this section we summarize cause of death impacts from eight major cause groups: cancers, cardiovascular disease, external causes, ill-defined casues, infant and congenital causes, infectious disease, mental debilitation, and other causes. These results will be summarized for the USA, Canada, England and Wales, Norway, and Sweden based on a soon-to-be released collection from the HMD. This section may have a supplemental appendix of results.

Discussion

There are many perspectives under which the demography can account for the relationships between stocks and flows, but not all form part of the collective practice of demography. Our objective in these exercises has been to offer a novel quantitative and visual basis to assess the impact of mortality on population stocks. This is done by calculating the lifespan distributions foregone due to death and indexing the results based on various aspects of the lifespan. Practical suggestions have included both chronological and thanatological age perspectives, as well as two ways of accounting for years of lifespan gained: (i) The years would be won by saving the deaths in age a, and (ii) the ages saved individuals would pass through if survived forward.

The reader may choose to interpret this exercise as we have narrated it: "what would have happened if these lives had been saved?". We wish to point out that believing this statement is not necessary in order for the measures to be useful, just as common uses of period life expectancy require a certain degree of suspended disbelief. For clarity, we list the most important assumptions to be aware of when treating data as we have.

1. Unchanging period rate schedules. Note that the same formulas apply when data in the cohort perspective are used, but some care must be taken to allocate past deaths according to observed mortality within the cohort, and then complete non-extinct cohorts' mortality experience according to projection. Brouard (1986) combined history and projection in this way in his original study of population structure. In any case, period measures are the best barometer of the present that we have, and all calculations done in this paper fall under the period umbrella. The researcher is not limited to the use of static age schedules, and $f(y|a)$ could be calculated for age schedules that vary by birth cohort.

2. Homogenous populations. In using rate schedules derived from the population at large in order to describe a hypothetical population of saved individuals, we may neglect that saved individuals may be a frailer than the general population, and so subject to higher mortality rates going forward. We offer to remedy for this shortcoming, except to note that this possibility may hold truer for some causes than for others, and in general the degree of bias is unknown.

3. Independence of causes. As discussed in the text, all causes of death compete to be first, and removing the first cause may not reduce the all cause rate by the same amount we have partitioned, μ^c. The final reduction will lie somewhere between 0 and μ^c, and may depend on the cause and overall level of mortality. We think that this possible unsolvable limitation ought not keep the researcher from exploring in this direction.

References

Nicolas Brouard. Structure et dynamique des populations. la pyramide des années à vivre, aspects nationaux et exemples régionaux. *Espace, populations, sociétés*, 4(2):157–168, 1986.

Nicolas Brouard. *Mouvements et modèles de population.* Institut de formation et de recherche démographiques, 1989.

John W. Gardner and Jill S. Sanborn. Years of potential life lost (YPLL): What does it measure? *Epidemiology*, 1(4):pp. 322–329, 1990. ISSN 10443983. URL `http://www.jstor.org/stable/25759821`.

Human Fertility Database. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). Available at www.humanfertility.org (data downloaded on January 21, 2015).

Human Mortality Database. University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on January 20, 2015).

Arni SR Srinivasa Rao and James R Carey. Generalization of careys equality and a theorem on stationary population. *Journal of mathematical biology*, pages 1–12, 2014.

Simon Rogers. Gun violence in america: How many years of life have been lost? The Guardian, online, 2013. URL `http://www.theguardian.com/news/datablog/interactive/2013/feb/04/us-gun-violence-deaths-years-lost-periscopic`. [posted 4-February-2013].

James W Vaupel. Lifesaving, lifetimes and lifetables. In *How Long Do We Live?*, pages 93–107. Springer, 2008.

James W Vaupel. Life lived and left: Carey's equality. *Demographic Research*, 20(3):7–10, 2009.

James W Vaupel and Anatoli I Yashin. Repeated resuscitation: How lifesaving alters life tables. *Demography*, 24(1):123–135, 1987.

A Formulas

The probability of surviving exactly y years into the future given survial to age a, $f(y|a)$ is given by:[8]

$$f(y|a) = \mu(a+y)\frac{l(a+y)}{l(a)} \qquad , \tag{1}$$

where $\mu(a)$ is the force of mortality at exact age a, and $l(a)$ is the value of the survival function at exact age a, proportional to the probability of surviving from birth to age a.

Assume that all the deaths recorded in a year are saved and brought back to life. One may ask much more than the number and age structure of these saved lives, $D^s(a)$,

$$D^s(a) = \mu(a)P(a) \qquad , \tag{2}$$

[8] This definition is identical to that used in Brouard (1989), Vaupel (2009) or Rao and Carey (2014) in proving the equality of chronological and thanatological age structure in stationary populations. Brouard apparently had proven this earlier than 1989, since he cites the relationship in Brouard (1986). Equation (1) is easily modifiable to account for mortality schedules that change over time.

but also how many years the people saved at age a would live, $W^s(a)$[9]? The simplest calculation is to multiply the number of gained survivors by remaining life expectancy at each age, $e(a)$:

$$W^s(a) = D^s(a)e(a) = P(a)\mu(a)\frac{1}{l(a)}\int l(a+y)\ \mathrm{d}y \qquad . \tag{3}$$

$D^s(a)e(a)$ classifies potentially saved person-years by the ages in which they were saved. One may also wish to know the distribution of remaining lifespans of saved lives, which is quite different from (3):

$$D^s(a)f(y|a) = P(a)\mu(a)\mu(a+y)\frac{l(a+y)}{l(a)} \qquad . \tag{4}$$

Equation (4) aggregates up to the thanatological age distribution of saved lives, $D^s(y)$:

$$D^s(y) = \int D^s(a)f(y|a)\ \mathrm{d}a \qquad . \tag{5}$$

Or one might ask through which chronological ages the gained years of life would be lived, $W^s(a+y|a)$,

$$W^s(a+y|a) = D^s(a)\frac{l(a+y)}{l(a)} \qquad . \tag{6}$$

Define the force of mortality, $\mu(a) = \sum_{c=1}^{n} \mu^c(a)$, as the sum of n categorically separable causes. The people that will die from cause c are:

$$\begin{aligned} D^c &= \int_0^\infty D^c(a)\ \mathrm{d}a & (7)\\ &= \int_0^\infty \mu^c(a)P(a)\ \mathrm{d}a \qquad , & (8) \end{aligned}$$

and to hypothetically save all these people is to remove cause c from mortality, retaining D^c lives in the population. It makes sense to calculate the distribution of remaining lifespans of the D^c people that would have died of this cause using $l(a)$ removed of the cause in question, so we define $l^{-c}(a)$,

$$l^{-c}(a) = e^{-\int_0^a \mu(a)-\mu^c(a)\ \mathrm{d}a} \qquad , \tag{9}$$

which is hopefully more legible to render as

$$l^{-c}(a) = e^{-\int_0^a \mu^{-c}(a)\ \mathrm{d}a} \qquad . \tag{10}$$

[9] A mnemonic for W could be *won* years. This is essentially an age at death breakdown of YPLL.

This is a stronger supposition than the idea of repeated resuscitation from Vaupel and Yashin (1987), but the idea is to separate out the impacts of particular causes. Let us continue with the same notational concept of $^{-c}$ to define remaining life expectancy assuming survival to age a and no more death from cause c after age a, $e^{-c}(a)$:

$$e^{-c}(a) = \frac{1}{l^{-c}(a)} \int_{y=0}^{\infty} l^{-c}(a+y)\, \mathrm{d}y \qquad . \tag{11}$$

and so on, repeating equations (3) and (5) for the case of cause-specific saveable lives and their cause-deleted remaining lifespans.