University of Pennsylvania
ScholarlyCommons

PSC Working Paper Series

4-5-2007

A New Method for Attributing Changes in Life Expectancy to Various Causes of Death, with Application to the United States

Hiram Beltrán-Sánchez
University of Pennsylvania, hbeltra@sas.upenn.edu

Samuel Preston
University of Pennsylvania, spreston@sas.upenn.edu

A New Method for Attributing Changes in Life Expectancy to Various Causes of Death, with Application to the United States

Abstract

This article focuses on decomposition of changes in life expectancy by cause of death. We propose an alternative to Arriaga's (1984) method for performing such decompositions. We apply our method to changes in life expectancy in the United States between 1970 and 2000 and compare results to those produced using Arriaga's formulation. The major difference between the approaches pertains to diseases prominent at older ages such as cardiovascular disease. For applications where causes of death are the central focus, our technique appears to have a modest advantage because of its conceptual clarity and attractive byproducts in the form of cause-deleted life tables.

Disciplines

Demography, Population, and Ecology | Social and Behavioral Sciences | Sociology

A New Method for Attributing Changes in Life Expectancy to Various Causes of Death, with Application to the United States[*]

Hiram Beltrán-Sánchez

Samuel H. Preston

Abstract

This article focuses on decomposition of changes in life expectancy by cause of death. We propose an alternative to Arriaga's (1984) method for performing such decompositions. We apply our method to changes in life expectancy in the United States between 1970 and 2000 and compare results to those produced using Arriaga's formulation. The major difference between the approaches pertains to diseases prominent at older ages such as cardiovascular disease. For applications where causes of death are the central focus, our technique appears to have a modest advantage because of its conceptual clarity and attractive byproducts in the form of cause-deleted life tables.

[*] Hiram Beltrán-Sánchez, Population Studies Center, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6298, E-mail: hbeltra@sas.upenn.edu. Samuel H. Preston, Population Studies Center, University of Pennsylvania, E-mail: spreston@sas.upenn.edu.

Introduction

Demographic decomposition techniques were formalized and generalized by Kitagawa (1955) and elaborated by Das Gupta (1978, 1989, 1991, 1993). The first efforts to apply decompositional methods to life expectancy occurred in the 1980's (Arriaga 1982; Arriaga 1984; Arriaga 1989; Pollard 1982; Pollard 1988; Pressat 1985). Initially, these efforts attributed differences in life expectancy to mortality differences in various age groups. At a second stage, analysis was extended to various causes of death. Later developments applied decompositional methods to other demographic measures and introduced dynamic elements (Andreev, Shkolnikov, and Begun 2002; Vaupel and Romo 2002; Vaupel and Romo 2003).

This paper focuses on decomposition of changes in life expectancy by cause of death. We propose an alternative to Arriaga's (1984) initial formulation, one that is more directly aligned with the traditions of Kitagawa and Das Gupta and that has somewhat more attractive properties. We apply our method to changes in life expectancy in the United States between 1970 and 2000 and compare results to those produced using Arriaga's formulation.

Background

In the early 1980s a technique to decompose the change in life expectancy by age and cause of death was independently developed by Arriaga (1982, 1984), Pollard (1982), Pressat (1985) and Andreev (1982) (cited in Andreev et al. 2002). These versions are equivalent (Pollard 1988). However, Pollard (1988) pointed out a limitation of the cause-of-death formulation shortly after it was proposed: "a limitation of …[the formula] is that the weighting factor for a particular cause is itself a function of the level of mortality from the various causes of death in the two populations, *including the cause itself"* [italics added] (p.267). We suspect that this property is a result of the fact that age-decompositions preceded cause-decompositions and that the latter were designed along the lines of the former; had the initial target been cause-decompositions, a more straightforward procedure was available, which is developed below.

Methods

Arriaga's formulation of the change in life expectancy at birth between times 1 and 2 by causes of death can be written, in the continuous case, as (Pollard (1988:267)):

$$\overset{\circ}{e}_2(0) - \overset{\circ}{e}_1(0) = \sum_{i=1}^{n} \int_0^\infty \left(\mu_2^i(a) - \mu_1^i(a) \right) w_a \, da \,, \tag{1}$$

where $w_a = \dfrac{1}{2}\left(p_1(a)\overset{\circ}{e}_2(a) + p_2(a)\overset{\circ}{e}_1(a) \right)$,

Where $\overset{\circ}{e}_t(a)$ represents the life expectancy at age a at time t (t=1, 2), $\mu_t^i(a)$ is the force of mortality at time t for cause of death i at age a, and $p_t(a)$ is the probability of surviving from birth to age a in the all-cause mortality life table at time t. Clearly, the same formula can be used for decomposing the difference in life expectancy between any two populations.

Reflected in the weighting factor, w_a, are mortality rates from all causes of death, including the cause of death whose contribution one is trying to isolate (Pollard 1988). This procedure is akin to measuring change of mass on a scale whose calibration depends on the mass being measured. A cleaner formulation would use a weighting function that does not depend on the cause of death at issue.

The development of an alternative approach proceeds as follows. Suppose that there are two mutually exclusive and exhaustive causes of death operating in a population at times 1 and 2. The probability of surviving from birth to age a at time t if the only cause of death operating were cause i is $p_t^i(a) = e^{-\int_0^a \mu_t^i(s)ds}$. In the life table for all causes of death combined, the equivalent survival function is $p_t(a) = e^{-\int_0^a \mu_t(s)ds}$. Assuming independence of the causes of death, $p_1(a) = p_1^1(a) \cdot p_1^2(a)$ and $p_2(a) = p_2^1(a) \cdot p_2^2(a)$.

Applying Kitagawa's (1955) formula for two-factor decompositions without residual terms, the difference between life expectancies at birth at times 1 and 2 is:

$$e_2 - e_1 = \int_0^\infty p_2(a)da - \int_0^\infty p_1(a)da = \int_0^\infty p_2^1 p_2^2 da - \int_0^\infty p_1^1 p_1^2 da = \int_0^\infty \left(p_2^1 - p_1^1 \right)\left[\frac{p_1^2 + p_2^2}{2} \right]da + \int_0^\infty \left(p_2^2 - p_1^2 \right)\left[\frac{p_1^1 + p_2^1}{2} \right]da .$$

$$e_2 - e_1 = \int_0^\infty \left(p_2^1 - p_1^1 \right)\left[\frac{p_1^{-1} + p_2^{-1}}{2} \right]da + \int_0^\infty \left(p_2^2 - p_1^2 \right)\left[\frac{p_1^{-2} + p_2^{-2}}{2} \right]da . \tag{2}$$

Equation (2) is more straightforward than equation (1). It clearly isolates the contributions of the two causes of death, and the contribution of each cause is weighted by the

mean survival probability for the other cause. Since the two causes of death can always be construed as any one particular cause and "all other causes", equation (2) has broad utility. It is the basis of what follows.

Extensions to more than two causes of death

With three causes of death operating in the population, the difference in life expectancy at birth between times 1 and 2 is given by (see Appendix 1):

$$e_2 - e_1 = \int_0^\infty \left(p_2^1 - p_1^1 \right) \left[\frac{p_1^{-1} + p_2^{-1}}{2} \right] da + \int_0^\infty \left(p_2^2 - p_1^2 \right) \left[\frac{p_1^{-2} + p_2^{-2}}{2} \right] da + \int_0^\infty \left(p_2^3 - p_1^3 \right) \left[\frac{p_1^{-3} + p_2^{-3}}{2} \right] da$$

$$- \int_0^\infty \left(\frac{\left(p_2^1 - p_1^1 \right)\left(p_2^2 - p_1^2 \right)\left(p_2^3 - p_1^3 \right)}{2} \right) da \,,$$

where p_1^{-i} and p_2^{-i} are the product of survival probabilities for all causes except cause i in the population at times 1 and 2, respectively (e.g. $p_1^{-1} = p_1^2 p_1^3$). Note that the two terms appearing in equation (2) are also present in this equation (there are now three such terms because there are three causes of death) and that an additional term has been added that reflects the interactions among the three causes of death.

It can be shown that for n causes of death ($n>5$), $e_2 - e_1$ can be written as (see Appendix 1):

$$e_2 - e_1 = \sum_{i=1}^n \int_0^\infty \left(p_2^i - p_1^i \right) \left[\frac{p_2^{-i} + p_1^{-i}}{2} \right] da -$$

$$\sum_{i=1}^n \int_0^\infty \left(p_2^i - p_1^i \right) \left\{ \sum_{r=1}^{\delta} \frac{ \displaystyle\sum_{\substack{j=1 \\ j\neq i}}^{n-(r-1)} \sum_{\substack{k=j+1 \\ k\neq i,j}}^{n-(r-2)} \sum_{\substack{l=k+1 \\ l\neq i,j,k}}^{n-(r-3)} \cdots \sum_{\substack{z=y+1 \\ z\neq i,j,\ldots,y}}^{n} \left(p_2^j p_2^k p_2^l \ldots p_2^z - p_1^j p_1^k p_1^l \ldots p_1^z \right)\left(\prod_{\substack{\theta=1 \\ \theta\neq i,j,\ldots,z}}^{n} p_2^\theta - \prod_{\substack{\theta=1 \\ \theta\neq i,j,\ldots,z}}^{n} p_1^\theta \right) }{ n \dbinom{n-1}{r} } \right\} da$$

where $\delta = \begin{cases} \dfrac{n-1}{2} & \text{if } n \text{ is odd} \\[2mm] \dfrac{n-2}{2} & \text{if } n \text{ is even} \end{cases}$

(3)

3

As the number of causes of death increases, the number of interaction terms will also increase and the contribution of each element in the interaction term to the change in life expectancy will become very small. Although the formula for the set of interaction terms can be expressed mathematically, as in equation (3), in practice there will be little reason to calculate its value; it can be estimated as a residual once the value of the cause-specific terms in equation (2) are estimated for each cause of death of interest. Appendix 2 provides details about the discrete approximations that are employed to implement the approach.

Illustrations

A comparison of the weighting functions in equations (1) and (2) provides an intuition about how results should differ between the two approaches. Suppose that the cause of death whose contribution we are interested in is cardiovascular disease, a cause with a very old age incidence. In our equation (2), changes in cardiovascular disease at older ages will receive relatively heavy weight because they are being weighted by survivorship from *other* causes of death, survivorship that is relatively high at older ages. In contrast, cardiovascular survival is fully represented in the function weighting changes in cardiovascular disease in equation (1), so that developments at older ages receive somewhat less weight than in equation (2) (of course, different functions are being weighted in (1) and (2)). Thus, we would expect changes in cardiovascular disease to appear less important when applying Arriaga's equation (1) than when applying our equation (2). The choice between them is not entirely arbitrary; changes in cardiovascular disease are quite important precisely *because* other causes of death are of relatively mild import at older ages. The inclusion of the low survivorship from cardiovascular disease itself in the weighting function of (1) discounts the importance of changes therein.

We illustrate the difference with an example drawn from England and Wales during a period when cardiovascular death rates and those from other causes of death were moving in opposite directions. Over the period 1921-1951, life expectancy at birth for females in England and Wales rose from 59.9 years to 70.9. During this period, recorded cardiovascular death rates were rising while death rates from other causes of death were falling (Preston, Keyfitz, and Schoen 1972). Thus, cardiovascular disease contributed negatively to the change in life expectancy. We applied the two formulas to these data. According to our formula (2), cardiovascular diseases contributed -0.58 years to the gain in life expectancy; according to

Arriaga's formula (1), it contributed only -0.14 years, less than a quarter as much. As expected, changes in an "old" cause of death are discounted in formula (1) relative to formula (2).

Application to Recent Mortality Change in the United States

After several decades of slow improvement in mortality, the United States experienced substantial advances in longevity between 1970 and 2000. These are widely believed to reflect primarily reductions in death rates from cardiovascular diseases, facilitated by medical advances such as cardiovascular bypass surgery and expanded use of blood pressure reduction drugs and beta blockers, as well as by reductions in cigarette smoking (Cutler 2004; Ergin, Muntner, Sherwin, and He 2004). We apply the two methods of attributing mortality reductions to various causes of death in order to estimate their relative contributions during the period 1970-2000.

Data

Mortality data for 1970 are drawn from the publicly available multiple cause of death file, 1968-1973, obtained through the Inter-university Consortium for Political and Social Research (National Center for Health Statistics 2001). For the year 2000, we use the multiple cause of death public use file published by the National Center for Health Statistics (2002). The midyear population figures in 1970 and 2000 are drawn from the Census Bureau (U.S. Bureau of the Census 1971; U.S. Bureau of the Census 2005).[1] For blacks in 1970 we use the reconstructed size of the African-American population (Preston, Elo, Foster, and Fu 1998:9).[2] The terminal age category in our application is age 100+.[3]

We focus on the ten leading causes of death in the U.S. for the year 2000, and then construct comparable categories in 1970. The leading causes of death in 2000 are derived from the tabulation: Leading Causes of Death Reports 1999-2004 published by the Centers for Disease Control (CDC) website (Centers for Disease Control 2005). We combined several

---

[1] For the year 2000 we use the electronically available monthly postcensal estimation of the resident population.
[2] We use Preston et al. (1998) estimates for ages 0-84, and census estimates for ages 85-99 (U.S. Bureau of the Census 1971).
[3] The centenarian population in 1970 was improperly counted (Siegel 1974). Thus, we use the preferred estimates of centenarians by race and sex from Siegel and Passel (1976).

causes of death to produce the following groups: cardiovascular diseases (heart and cerebrovascular diseases); malignant neoplasms; chronic lower respiratory diseases; violence (accidents (unintentional injuries), intentional self-harm (suicide), and assault (homicide)); diabetes; influenza-pneumonia; nephritis (nephritis, nephrotic syndrome and nephrosis); septicemia; liver cirrhosis (chronic liver disease and cirrhosis); and hypertension (essential (primary) hypertension and hypertensive renal disease). Comparable codes for 1970 were derived from CDC (2001).

Results

We estimate life expectancy at birth for the total U.S. population in 1970 to be 70.70 years (slightly lower than the value of 70.9 estimated by the NCHS (1974)) and 76.96 years for the year 2000 (slightly higher than the estimation of Arias (2002)). Thus, life expectancy at birth increased by about 6.22 years between 1970 and 2000 for the total population of the U.S. Similar figures are shown in Table 1 by race and sex.

Table 1. Change in Life Expectancy at Birth for the U.S. Population by Sex and Race between 1970 and 2000

Life Expectancy at Birth	Total Population			Race		White		Black	
	All	Males	Females	White	Black	Males	Female	Male	Female
2000	76.96	74.24	79.56	77.43	71.74	74.79	79.99	68.15	75.06
1970	70.70	66.97	74.62	71.58	64.78	67.87	75.47	61.15	68.66
Difference	6.26	7.27	4.94	5.85	6.96	6.92	4.52	7.00	6.40

Source: Multiple cause of death file (1968-1973), multiple cause of death 2000, and census bureau estimates.

The ten causes of death selected in this study account for the majority of the deaths in both time periods. In 1970, about 84% of the total deaths were attributed to one of those ten causes for the total population, and in 2000, the percentage slightly decreased to about 80% (Table 2). Cardiovascular diseases (heart and cerebrovascular) and malignant neoplasms are responsible for most of the deaths in both years; together, they make up about two thirds of the total deaths.

Table 2. Percentage Distribution of Deaths by Cause for the Total U.S. Population, 1970- 2000

Cause of Death	1970	2000
Cardiovascular	49.06	36.55
Neoplasms	17.21	23.00
Respiratory	1.61	5.07
Violence	8.08	5.29
Diabetes	1.99	2.88
Influenza-Pneumonia	3.26	2.72
Nephritis	0.46	1.55
Septicemia	0.18	1.30
Liver cirrhosis	1.63	1.11
Hypertension	0.43	0.75
Residual	16.07	19.79
Sum	100.00	100.00

Source: see Table 1.

Our estimates of the contribution of each cause of death to the change in life expectancy at birth by race and sex between 1970 and 2000 are shown in Table 3. Cardiovascular diseases contributed the majority of the increase in life expectancy at birth during the 30-year period for all groups except black males, for whom large declines in death rates from accidents and violence contributed 1.94 years of life to the total gain of 7.00 years. This exceptionally large decline in violent mortality resulted in a larger gain in life expectancy for black males than for any other race/sex group during the period. Cancers made a positive contribution to life expectancy gains among white people, while for blacks, its contribution is negative. Diabetes also shows a negative contribution but among males only. The relative importance of violence is more pronounced among males, particularly among black males.

The difference in cause-of-death attributions between our method and Arriaga's is shown in Table 4. Except for cardiovascular diseases, the difference between the two sets of estimates is very small. For the total population, our method attributes a contribution of cardiovascular diseases of about 0.2 years higher than Arriaga's. This difference is of similar magnitude for the total population of males, and somewhat lower for females and whites; but it increases for the black population. Our approach estimates a contribution of cardiovascular diseases to be about 3%, 2.8% and 3.3% higher for the total black population, black males and black females, respectively, compared to Arriaga's formula (data not shown).

The contribution of the interaction terms to the change in life expectancy at birth is very small (Table 3). For most of the causes of death, the estimated contribution is in the range of one one-hundredth of a year. Apart from rounding error, direct estimation of the interaction terms using equation (3) gives identical results to those obtained indirectly in Table 3 (data not shown).

Table 3. Change in Life Expectancy at Birth, 1970-2000, Attributable to Various Causes of Death in the United States (in Years)

Cause of death	Total Population			Race		White		Black	
	All	Males	Females	White	Black	Males	Female	Male	Female
Cardiovascular	3.76	3.98	3.35	3.74	3.37	4.04	3.26	2.83	3.74
Neoplasms	0.27	0.35	0.17	0.26	-0.02	0.36	0.15	-0.06	-0.04
Respiratory	-0.21	-0.08	-0.36	-0.23	-0.14	-0.07	-0.39	-0.12	-0.16
Violence	0.90	1.23	0.53	0.77	1.42	1.04	0.47	1.94	0.74
Diabetes	0.01	-0.03	0.05	0.01	-0.05	-0.02	0.05	-0.10	0.03
Influenza-Pneumonia	0.38	0.42	0.32	0.32	0.67	0.36	0.27	0.73	0.57
Nephritis	-0.06	-0.05	-0.07	-0.05	-0.08	-0.05	-0.06	-0.06	-0.11
Septicemia	-0.08	-0.07	-0.10	-0.07	-0.16	-0.06	-0.09	-0.13	-0.18
Liver-Cirrhosis	0.17	0.20	0.14	0.14	0.32	0.17	0.12	0.33	0.28
Hypertension	0.00	0.01	-0.01	0.00	0.02	0.01	-0.01	0.03	0.00
Residual	1.12	1.30	0.91	0.97	1.63	1.15	0.75	1.61	1.54
SUM	6.26	7.26	4.93	5.86	6.98	6.93	4.52	7.00	6.41
Interaction	0.00	-0.01	-0.01	0.01	0.02	0.01	0.00	0.00	0.01
Total change in e(0)	6.26	7.27	4.94	5.85	6.96	6.92	4.52	7.00	6.40

Note: The interaction term is computed as the SUM of causes minus the total change in life expectancy at birth.
Source: See Table 1.

Table 4. Difference Between Present Method and Arriaga's Method in the Attribution to Various Causes of Death of the Change in Life Expectancy at Birth in the United States, 1970-2000 (Present Estimate Minus Arriaga's, in Years)

Cause of death	Total Population			Race	
	All	Males	Females	White	Black
Cardiovascular	0.16	0.18	0.13	0.14	0.22
Neoplasms	-0.03	-0.03	-0.02	-0.03	-0.05
Respiratory	-0.03	-0.03	-0.02	-0.03	-0.02
Violence	-0.01	-0.03	0.00	-0.01	-0.02
Diabetes	0.00	0.00	0.00	0.00	-0.02
Influenza-Pneumonia	0.00	0.01	0.00	0.01	0.00
Nephritis	-0.01	-0.01	-0.01	-0.01	-0.02
Septicemia	-0.01	-0.01	-0.01	-0.01	-0.02
Liver-Cirrhosis	0.00	0.00	0.00	0.00	-0.01
Hypertension	0.00	0.00	0.00	0.00	-0.01
Residual	-0.08	-0.06	-0.08	-0.07	-0.05

Note: The implementation of Arriaga's formula is taken from Preston, Heuveline, and Guillot (2001: 64, 84) in which the master life tables are closed using the approach described in Appendix 2.
Source: See Tables 1 and 3.

Discussion

A valuable byproduct of the method of cause-decomposition developed here is a set of associated single decrement life tables, including cause-deleted life tables. Cause-deleted life tables show the gain in life expectancy that would result from the complete elimination of a cause of death (Preston et al. 2001: 84). Such tables provide a clear measure of the public health significance of a particular disease or injury process.

The construction of the $_nL_x$ column of a cause-deleted life table is required in order to implement the decompositions proposed here. It is recommended that equation (2.1) be used for this purpose. Once computed, the elements of this column simply need to be added up and the sum compared to the original life expectancy in order to estimate the gain from the elimination of deaths associated with a particular disease. Table 5 presents the results of this exercise for the United States.

This table highlights the changing significance of cancer and cardiovascular disease. In both 1970 and 2000, these are the two causes of death whose elimination would produce the greatest advance in life expectancy at birth. But the relative importance of the two has changed dramatically: in 1970, cardiovascular disease for all groups combined was responsible for 3.75 times the loss of life years of neoplasms. In 2000, the ratio was 1.61. Even though the death rate from neoplasms was slowly declining between 1970 and 2000, more years of life were sacrificed to neoplasms in 2000 than in 1970.[4] The primary reason is the decline in death rates from cardiovascular diseases themselves. Because of this decline, those who died from cancer in 2000 would otherwise have lived longer, on average, than those who died from cancer in 1970; therefore, cancer is causing a greater loss of life in 2000 even though its death rates have declined. This result underscores the importance of interactions among diseases in determining longevity.

---

[4] Using the 2000 U.S. Standard Population (National Cancer Institute, 2007) the age-standardized crude death rate for neoplasms declined very little from 0.0020 in 1970 to about 0.0019 in 2000.

Table 5. Gain in Life Expectancy at Birth Resulting from the Elimination of a Cause of Death by Sex and Race for the U.S. Population, 1970-2000 (in Years)

| | 1970 | | | | | | | | |
| Cause of death | Total Population | | | Race | | White | | Black | |
	All	Males	Females	White	Black	Male	Female	Male	Female
Cardiovascular	9.53	9.03	9.60	9.48	9.94	9.08	9.39	8.42	11.33
Neoplasms	2.54	2.38	2.61	2.56	2.42	2.39	2.64	2.33	2.40
Respiratory	0.20	0.26	0.11	0.21	0.14	0.27	0.10	0.17	0.10
Violence	1.86	2.51	1.08	1.71	2.60	2.28	1.03	3.72	1.29
Diabetes	0.25	0.18	0.32	0.23	0.37	0.17	0.29	0.23	0.55
Influenza-Pneumonia	0.50	0.50	0.47	0.44	0.80	0.44	0.42	0.85	0.72
Nephritis	0.07	0.07	0.06	0.05	0.16	0.05	0.05	0.15	0.17
Septicemia	0.03	0.04	0.03	0.03	0.06	0.03	0.03	0.06	0.07
Liver-Cirrhosis	0.29	0.33	0.23	0.27	0.41	0.31	0.20	0.43	0.35
Hypertension	0.05	0.05	0.05	0.04	0.14	0.03	0.04	0.13	0.15

| | 2000 | | | | | | | | |
| Cause of death | Total Population | | | Race | | White | | Black | |
	All	Males	Females	White	Black	Male	Female	Male	Female
Cardiovascular	5.25	5.13	5.09	5.12	5.96	5.04	4.92	5.42	6.19
Neoplasms	3.27	3.32	3.17	3.27	3.44	3.29	3.19	3.59	3.22
Respiratory	0.55	0.54	0.56	0.58	0.36	0.56	0.60	0.38	0.31
Violence	1.20	1.66	0.66	1.17	1.49	1.60	0.66	2.20	0.69
Diabetes	0.34	0.32	0.36	0.31	0.58	0.30	0.31	0.45	0.68
Influenza-Pneumonia	0.25	0.24	0.25	0.24	0.26	0.23	0.25	0.26	0.24
Nephritis	0.16	0.16	0.16	0.14	0.32	0.14	0.13	0.28	0.35
Septicemia	0.14	0.13	0.15	0.12	0.27	0.12	0.13	0.24	0.30
Liver-Cirrhosis	0.17	0.22	0.11	0.17	0.15	0.22	0.11	0.18	0.12
Hypertension	0.07	0.06	0.07	0.05	0.17	0.05	0.06	0.14	0.19

Source: Table 3 using Appendix equation 2.1

Summing Up

This paper presents a straightforward and parsimonious formula (equation 2) for decomposing differences in life expectancy into various causes of death. It follows directly in the traditions of Kitagawa and Das Gupta and avoids the messy property of Arriaga's decomposition whereby the importance of a cause of death in accounting for mortality differences is determined by a weighting function that includes the cause itself.

Nevertheless, our results are quite similar to those derived by using the Arriaga formula except for cardiovascular disease, an important cause of death that is heavily concentrated at older ages. Both formulas are algebraically "correct". The main advantage of Arriaga's approach relative to ours is that it is consistent with the age-decompositions out of which it developed. An advantage of our approach relative to Arriaga's is that it automatically produces a cause-deleted life table that highlights the public health significance of a cause of death. For applications where causes of death are the central focus, it would appear to have a modest advantage because of its conceptual clarity and attractive byproducts. At a minimum, analysts should be aware that different methods are available to perform these important calculations.

Appendix 1

The development of the new decompositional formulas relies heavily on the formulation developed by Das Gupta (1993), in which he decomposed the difference between two rates into differences of the individual pieces that make up the rates and interaction terms.

Two causes of death

The simplest case is when we have two causes of death. Let $p_1(a)$ and $p_2(a)$ be the survival probabilities from birth to age a for a population at times 1 and 2 (or two populations at a particular time). Then, as a multiple decrement process, we have that $p_1(a) = p_1^1(a) \cdot p_1^2(a)$ and $p_2(a) = p_2^1(a) \cdot p_2^2(a)$, where p_1^i and p_2^i for $i=1,2$, are two causes of death for the population at times 1 and 2.

The difference in life expectancy at birth in this case is given by

$$e_2 - e_1 = \int_0^\infty p_2(a)\,da - \int_0^\infty p_1(a)\,da = \int_0^\infty p_2^1 p_2^2\,da - \int_0^\infty p_1^1 p_1^2\,da .$$

For simplicity, let $p_t^i(a) = p_t^i$, for $i=1,2$ and $t=1,2$. It is fairly easy to show (following Kitagawa, 1955) that

$$e_2 - e_1 = \int_0^\infty \left(p_2^1 - p_1^1\right)\left[\frac{p_1^2 + p_2^2}{2}\right]da + \int_0^\infty \left(p_2^2 - p_1^2\right)\left[\frac{p_1^1 + p_2^1}{2}\right]da .$$

Given that there are only two causes of death in this case, the above equation can be rewritten as

$$e_2 - e_1 = \int_0^\infty \left(p_2^1 - p_1^1\right)\left[\frac{p_1^{-1} + p_2^{-1}}{2}\right]da + \int_0^\infty \left(p_2^2 - p_1^2\right)\left[\frac{p_1^{-2} + p_2^{-2}}{2}\right]da ,$$

where p_t^{-i} represents the survival probability for all causes but cause i, at time t.

Three causes of death

Similarly, let $p_1(a)$ and $p_2(a)$ be the survival probabilities from birth to age a for a population at times 1 and 2, where $p_1(a) = p_1^1 \cdot p_1^2 \cdot p_1^3$ and $p_2(a) = p_2^1 \cdot p_2^2 \cdot p_2^3$ (assuming that $p_t^i(a) = p_t^i$, for $i=1,2,3$ and $t=1,2$).

The difference in life expectancy is then given by

$$e_2 - e_1 = \int_0^\infty p_2(a)\,da - \int_0^\infty p_1(a)\,da = \int_0^\infty p_2^1 p_2^2 p_2^3\,da - \int_0^\infty p_1^1 p_1^2 p_1^3\,da . \tag{1}$$

From Das Gupta (1993:8), the difference between two rates, where each rate is written as a product of three factors, can be expressed as:

$p_2 - p_1$

$$= \left(p_2^1 - p_1^1\right)\left[\frac{p_1^2 p_1^3 + p_2^2 p_2^3}{3} + \frac{p_1^2 p_2^3 + p_2^2 p_1^3}{6}\right] + \left(p_2^2 - p_1^2\right)\left[\frac{p_1^1 p_1^3 + p_2^1 p_2^3}{3} + \frac{p_1^1 p_2^3 + p_2^1 p_1^3}{6}\right] +$$

$$\left(p_2^3 - p_1^3\right)\left[\frac{p_1^1 p_1^2 + p_2^1 p_2^2}{3} + \frac{p_1^1 p_2^2 + p_2^1 p_1^2}{6}\right].$$

The above equation can be applied to equation (1), thus:

$$e_2 - e_1 = \int_0^\infty \left(p_2^1 - p_1^1\right)\left[\frac{p_1^{-1} + p_2^{-1}}{3} + \frac{p_1^2 p_2^3 + p_2^2 p_1^3}{6}\right]da + \int_0^\infty \left(p_2^2 - p_1^2\right)\left[\frac{p_1^{-2} + p_2^{-2}}{3} + \frac{p_1^1 p_2^3 + p_2^1 p_1^3}{6}\right]da$$

$$+ \int_0^\infty \left(p_2^3 - p_1^3\right)\left[\frac{p_1^{-1} + p_2^{-1}}{3} + \frac{p_1^1 p_2^2 + p_2^1 p_1^2}{6}\right]da \, . \tag{2}$$

Where p_1^{-i} and p_2^{-i} are the product of survival probabilities for all causes but cause i in the population at times 1 and 2, respectively (e.g. $p_1^{-1} = p_1^2 p_1^3$).

Let $\frac{p_1^{-i} + p_2^{-i}}{3} = \frac{p_1^{-i} + p_2^{-i}}{2} - \frac{p_1^{-i} + p_2^{-i}}{6}$, substituting this expression back into equation (2) we obtain that the difference in life expectancy at birth for the population at times 1 and 2 is given by:

$$e_2 - e_1 = \int_0^\infty \left(p_2^1 - p_1^1\right)\left[\frac{p_1^{-1} + p_2^{-1}}{2} + \frac{\left(p_1^2 - p_2^2\right)\left(p_2^3 - p_1^3\right)}{6}\right]da +$$

$$\int_0^\infty \left(p_2^2 - p_1^2\right)\left[\frac{p_1^{-2} + p_2^{-2}}{2} + \frac{\left(p_1^1 - p_2^1\right)\left(p_2^3 - p_1^3\right)}{6}\right]da + \int_0^\infty \left(p_2^3 - p_1^3\right)\left[\frac{p_1^{-3} + p_2^{-3}}{2} + \frac{\left(p_1^2 - p_2^2\right)\left(p_2^1 - p_1^1\right)}{6}\right]da \, .$$

The above equation is equivalent to

$$e_2 - e_1 = \int_0^\infty \left(p_2^1 - p_1^1\right)\left[\frac{p_1^{-1} + p_2^{-1}}{2}\right]da + \int_0^\infty \left(p_2^2 - p_1^2\right)\left[\frac{p_1^{-2} + p_2^{-2}}{2}\right]da + \int_0^\infty \left(p_2^3 - p_1^3\right)\left[\frac{p_1^{-3} + p_2^{-3}}{2}\right]da - \int_0^\infty \frac{\left(p_2^1 - p_1^1\right)\left(p_2^2 - p_1^2\right)\left(p_2^3 - p_1^3\right)}{2}da$$

Notice that the fractions in square brackets do not involve the cause of interest, e.g. when we take the difference in survival probabilities due to cause 1, the square bracket does not involve cause 1.

Four causes of death

Let $p_1(a) = p_1^1 p_1^2 p_1^3 p_1^4$ and $p_2(a) = p_2^1 p_2^2 p_2^3 p_2^4$ (where $p_t^i(a) = p_t^i$, for $i=1,2,3,4$ and $t=1,2$). Then, the difference in life expectancy at birth for a population at times 1 and 2 is given by:

$$e_2 - e_1 = \int_0^\infty p_2(a)da - \int_0^\infty p_1(a)da = \int_0^\infty p_2^1 p_2^2 p_2^3 p_2^4 da - \int_0^\infty p_1^1 p_1^2 p_1^3 p_1^4 da \, . \tag{3}$$

From Das Gupta (1993:10) the difference in two rates, where each rate is written as a product of four factors, can be expressed as:

$$p_2 - p_1 = \left(p_2^1 - p_1^1\right)\left[\frac{p_1^2 p_1^3 p_1^4 + p_2^2 p_2^3 p_2^4}{4} + \frac{p_1^2 p_1^3 p_2^4 + p_1^2 p_2^3 p_1^4 + p_2^2 p_1^3 p_1^4 + p_2^2 p_2^3 p_1^4 + p_2^2 p_1^3 p_2^4 + p_1^2 p_2^3 p_2^4}{12}\right] +$$

$$\left(p_2^2 - p_1^2\right)\left[\frac{p_1^1 p_1^3 p_1^4 + p_2^1 p_2^3 p_2^4}{4} + \frac{p_1^1 p_1^3 p_2^4 + p_1^1 p_2^3 p_1^4 + p_2^1 p_1^3 p_1^4 + p_2^1 p_2^3 p_1^4 + p_2^1 p_1^3 p_2^4 + p_1^1 p_2^3 p_2^4}{12}\right] +$$

$$\left(p_2^3 - p_1^3\right)\left[\frac{p_1^1 p_1^2 p_1^4 + p_2^1 p_2^2 p_2^4}{4} + \frac{p_1^1 p_1^2 p_2^4 + p_1^1 p_2^2 p_1^4 + p_2^1 p_1^2 p_1^4 + p_2^1 p_2^2 p_1^4 + p_2^1 p_1^2 p_2^4 + p_1^1 p_2^2 p_2^4}{12}\right] +$$

$$\left(p_2^4 - p_1^4\right)\left[\frac{p_1^1 p_1^2 p_1^3 + p_2^1 p_2^2 p_2^3}{4} + \frac{p_1^1 p_1^2 p_2^3 + p_1^1 p_2^2 p_1^3 + p_2^1 p_1^2 p_1^3 + p_2^1 p_2^2 p_1^3 + p_2^1 p_1^2 p_2^3 + p_1^1 p_2^2 p_2^3}{12}\right].$$

The above equation can be applied to equation (3), thus:

$$e_2 - e_1 =$$

$$\int_0^\infty \left(p_2^1 - p_1^1\right)\left[\frac{p_1^{-1} + p_2^{-1}}{4} + \frac{p_1^2 p_1^3 p_2^4 + p_1^2 p_2^3 p_1^4 + p_2^2 p_1^3 p_1^4 + p_2^2 p_2^3 p_1^4 + p_2^2 p_1^3 p_2^4 + p_1^2 p_2^3 p_2^4}{12}\right]da +$$

$$\int_0^\infty \left(p_2^2 - p_1^2\right)\left[\frac{p_1^{-1} + p_2^{-1}}{4} + \frac{p_1^1 p_1^3 p_2^4 + p_1^1 p_2^3 p_1^4 + p_2^1 p_1^3 p_1^4 + p_2^1 p_2^3 p_1^4 + p_2^1 p_1^3 p_2^4 + p_1^1 p_2^3 p_2^4}{12}\right]da +$$

$$\int_0^\infty \left(p_2^3 - p_1^3\right)\left[\frac{p_1^{-1} + p_2^{-1}}{4} + \frac{p_1^1 p_1^2 p_2^4 + p_1^1 p_2^2 p_1^4 + p_2^1 p_1^2 p_1^4 + p_2^1 p_2^2 p_1^4 + p_2^1 p_1^2 p_2^4 + p_1^1 p_2^2 p_2^4}{12}\right]da +$$

$$\int_0^\infty \left(p_2^4 - p_1^4\right)\left[\frac{p_1^{-1} + p_2^{-1}}{4} + \frac{p_1^1 p_1^2 p_2^3 + p_1^1 p_2^2 p_1^3 + p_2^1 p_1^2 p_1^3 + p_2^1 p_2^2 p_1^3 + p_2^1 p_1^2 p_2^3 + p_1^1 p_2^2 p_2^3}{12}\right]da \ .$$

Let $\dfrac{p_1^{-i} + p_2^{-i}}{4} = \dfrac{p_1^{-i} + p_2^{-i}}{2} - \dfrac{3\left(p_1^{-i} + p_2^{-i}\right)}{12}$, substituting this expression back into the above equation we obtain that the difference in life expectancy at birth for the population at times 1 and 2 is given by:

$$e_2 - e_1 =$$

$$\int_0^\infty \left(p_2^1 - p_1^1\right)\left[\frac{p_1^{-1} + p_2^{-1}}{2} - \frac{\left(p_2^2 - p_1^2\right)\left(p_2^3 p_2^4 - p_1^3 p_1^4\right) + \left(p_2^3 - p_1^3\right)\left(p_2^2 p_2^4 - p_1^2 p_1^4\right) + \left(p_2^4 - p_1^4\right)\left(p_2^2 p_2^3 - p_1^2 p_1^3\right)}{12}\right]da +$$

$$\int_0^\infty \left(p_2^2 - p_1^2\right)\left[\frac{p_1^{-2} + p_2^{-2}}{2} - \frac{\left(p_2^1 - p_1^1\right)\left(p_2^3 p_2^4 - p_1^3 p_1^4\right) + \left(p_2^3 - p_1^3\right)\left(p_2^1 p_2^4 - p_1^1 p_1^4\right) + \left(p_2^4 - p_1^4\right)\left(p_2^1 p_2^3 - p_1^1 p_1^3\right)}{12}\right]da +$$

$$\int_0^\infty \left(p_2^3 - p_1^3\right)\left[\frac{p_1^{-3} + p_2^{-3}}{2} - \frac{\left(p_2^1 - p_1^1\right)\left(p_2^2 p_2^4 - p_1^2 p_1^4\right) + \left(p_2^2 - p_1^2\right)\left(p_2^1 p_2^4 - p_1^1 p_1^4\right) + \left(p_2^4 - p_1^4\right)\left(p_2^1 p_2^2 - p_1^1 p_1^2\right)}{12}\right]da +$$

$$\int_0^\infty \left(p_2^4 - p_1^4\right)\left[\frac{p_1^{-4} + p_2^{-4}}{2} - \frac{\left(p_2^1 - p_1^1\right)\left(p_2^2 p_2^3 - p_1^2 p_1^3\right) + \left(p_2^2 - p_1^2\right)\left(p_2^1 p_2^3 - p_1^1 p_1^3\right) + \left(p_2^3 - p_1^3\right)\left(p_2^1 p_2^2 - p_1^1 p_1^2\right)}{12}\right]da \ .$$

Similarly, the above can be simplified as:

$$e_2 - e_1 = \sum_{i=1}^4 \int_0^\infty \left(p_2^i - p_1^i\right)\left[\frac{p_1^{-i} + p_2^{-i}}{2}\right]da - \sum_{i=1}^4 \int_0^\infty \left(p_2^i - p_1^i\right)\left(\frac{\displaystyle\sum_{\substack{j=1 \\ j\neq i}}^4 \left(p_2^j - p_1^j\right)\left(\prod_{\substack{k=1 \\ k\neq i,j}}^4 p_2^k - \prod_{\substack{k=1 \\ k\neq i,j}}^4 p_1^k\right)}{12}\right)da \ .$$

Five causes of death

Let $p_1(a) = p_1^1 p_1^2 p_1^3 p_1^4 p_1^5$ and $p_2(a) = p_2^1 p_2^2 p_2^3 p_2^4 p_2^5$ (where $p_t^i(a) = p_t^i$, for i=1,2,3,4,5 and t=1,2). Then the difference in life expectancy at birth for the population at times 1 and 2 is given by:

$$e_2 - e_1 = \int_0^\infty p_2(a)\,da - \int_0^\infty p_1(a)\,da = \int_0^\infty p_2^1 p_2^2 p_2^3 p_2^4 p_2^5\,da - \int_0^\infty p_1^1 p_1^2 p_1^3 p_1^4 p_1^5\,da . \tag{4}$$

From Das Gupta (1993:13) the difference in two rates, where each rate is written as a product of five factors, can be expressed as:

$$p_2 - p_1 =$$

$$\left(p_2^1 - p_1^1\right)\left\{ \frac{p_1^2 p_1^3 p_1^4 p_1^5 + p_2^2 p_2^3 p_2^4 p_2^5}{5} + \right.$$

$$\frac{p_1^2 p_1^3 p_1^4 p_2^5 + p_1^2 p_1^3 p_2^4 p_1^5 + p_1^2 p_2^3 p_1^4 p_1^5 + p_2^2 p_1^3 p_1^4 p_1^5 + p_2^2 p_2^3 p_2^4 p_1^5 + p_2^2 p_2^3 p_1^4 p_2^5 + p_2^2 p_1^3 p_2^4 p_2^5 + p_1^2 p_2^3 p_2^4 p_2^5}{20} +$$

$$\left. \frac{p_1^2 p_1^3 p_2^4 p_2^5 + p_1^2 p_2^3 p_1^4 p_2^5 + p_1^2 p_2^3 p_2^4 p_1^5 + p_2^2 p_2^3 p_1^4 p_1^5 + p_2^2 p_1^3 p_2^4 p_1^5 + p_2^2 p_1^3 p_1^4 p_2^5}{30} \right\}$$

$$+$$
$$\vdots$$
$$+$$

$$\left(p_2^5 - p_1^5\right)\left\{ \frac{p_1^1 p_1^2 p_1^3 p_1^4 + p_2^1 p_2^2 p_2^3 p_2^4}{5} + \right.$$

$$\frac{p_1^1 p_1^2 p_1^3 p_2^4 + p_1^1 p_1^2 p_2^3 p_1^4 + p_2^1 p_2^2 p_1^3 p_1^4 + p_1^1 p_2^2 p_1^3 p_1^4 + p_2^1 p_2^2 p_2^3 p_1^4 + p_1^1 p_2^2 p_2^3 p_1^4 + p_2^1 p_1^2 p_1^3 p_2^4 + p_1^1 p_2^2 p_2^3 p_2^4}{20} +$$

$$\left. \frac{p_1^1 p_1^2 p_2^3 p_2^4 + p_1^1 p_2^2 p_1^3 p_2^4 + p_1^1 p_2^2 p_2^3 p_1^4 + p_2^1 p_2^2 p_1^3 p_1^4 + p_2^1 p_1^2 p_2^3 p_1^4 + p_2^1 p_1^2 p_1^3 p_2^4}{30} \right\} da .$$

The above equation can be applied to equation (4), thus:

$$\int_0^\infty \left(p_2^1 - p_1^1\right)\left\{ \frac{p_1^{-1} + p_2^{-1}}{5} + \right.$$

$$\frac{p_1^2 p_1^3 p_1^4 p_2^5 + p_1^2 p_1^3 p_2^4 p_1^5 + p_1^2 p_2^3 p_1^4 p_1^5 + p_2^2 p_1^3 p_1^4 p_1^5 + p_2^2 p_2^3 p_2^4 p_1^5 + p_2^2 p_2^3 p_1^4 p_2^5 + p_2^2 p_1^3 p_2^4 p_2^5 + p_1^2 p_2^3 p_2^4 p_2^5}{20} +$$

$$\left. \frac{p_1^2 p_1^3 p_2^4 p_2^5 + p_1^2 p_2^3 p_1^4 p_2^5 + p_1^2 p_2^3 p_2^4 p_1^5 + p_2^2 p_2^3 p_1^4 p_1^5 + p_2^2 p_1^3 p_2^4 p_1^5 + p_2^2 p_1^3 p_1^4 p_2^5}{30} \right\} da$$

$$+$$
$$\vdots$$
$$+$$

$$\int_0^\infty \left(p_2^5 - p_1^5\right)\left\{ \frac{p_1^{-5} + p_2^{-5}}{5} + \right.$$

$$\frac{p_1^1 p_1^2 p_1^3 p_2^4 + p_1^1 p_1^2 p_2^3 p_1^4 + p_2^1 p_2^2 p_1^3 p_1^4 + p_2^1 p_1^2 p_1^3 p_1^4 + p_2^1 p_2^2 p_2^3 p_1^4 + p_1^1 p_2^2 p_2^3 p_1^4 + p_2^1 p_1^2 p_2^3 p_2^4 + p_1^1 p_2^2 p_2^3 p_2^4}{20} +$$

$$\left.\frac{p_1^1 p_1^2 p_2^3 p_2^4 + p_1^1 p_2^2 p_1^3 p_2^4 + p_1^1 p_2^2 p_2^3 p_1^4 + p_2^1 p_2^2 p_1^3 p_1^4 + p_2^1 p_1^2 p_2^3 p_1^4 + p_2^1 p_1^2 p_1^3 p_2^4}{30}\right\} da .$$

Let $\dfrac{p_1^{-i} + p_2^{-i}}{5} = \dfrac{p_1^{-i} + p_2^{-i}}{2} - \dfrac{4\left(p_1^{-i} + p_2^{-i}\right)}{20} - \dfrac{3\left(p_1^{-i} + p_2^{-i}\right)}{30}$, substituting this expression back into the above equation we obtain that the difference in life expectancy at birth for the population at times 1 and 2 is given by:

$$e_2 - e_1 =$$

$$\int_0^\infty \left(p_2^1 - p_1^1\right)\left\{ \frac{p_1^{-1} + p_2^{-1}}{2} - \right.$$

$$\frac{\left(p_2^2 - p_1^2\right)\left(p_2^3 p_2^4 p_2^5 - p_1^3 p_1^4 p_1^5\right) + \left(p_2^3 - p_1^3\right)\left(p_2^2 p_2^4 p_2^5 - p_1^2 p_1^4 p_1^5\right) + \left(p_2^4 - p_1^4\right)\left(p_2^2 p_2^3 p_2^5 - p_1^2 p_1^3 p_1^5\right) + \left(p_2^5 - p_1^5\right)\left(p_2^2 p_2^3 p_2^4 - p_1^2 p_1^3 p_1^4\right)}{20} -$$

$$\left.\frac{\left(p_2^2 p_2^3 - p_1^2 p_1^3\right)\left(p_2^4 p_2^5 - p_1^4 p_1^5\right) + \left(p_2^2 p_2^4 - p_1^2 p_1^4\right)\left(p_2^3 p_2^5 - p_1^3 p_1^5\right) + \left(p_2^2 p_2^5 - p_1^2 p_1^5\right)\left(p_2^3 p_2^4 - p_1^3 p_1^4\right)}{30}\right\} da$$

$$+$$
$$\vdots$$
$$+$$

$$\int_0^\infty \left(p_2^5 - p_1^5\right)\left\{ \frac{p_1^{-5} + p_2^{-5}}{2} - \right.$$

$$\frac{\left(p_2^1 - p_1^1\right)\left(p_2^2 p_2^3 p_2^4 - p_1^1 p_1^3 p_1^4\right) + \left(p_2^2 - p_1^2\right)\left(p_2^1 p_2^3 p_2^4 - p_1^1 p_1^3 p_1^4\right) + \left(p_2^3 - p_1^3\right)\left(p_2^1 p_2^2 p_2^4 - p_1^1 p_1^2 p_1^4\right) + \left(p_2^4 - p_1^4\right)\left(p_2^1 p_2^2 p_2^3 - p_1^1 p_1^2 p_1^3\right)}{20} -$$

$$\left.\frac{\left(p_2^1 p_2^2 - p_1^1 p_1^2\right)\left(p_2^3 p_2^4 - p_1^3 p_1^4\right) + \left(p_2^1 p_2^3 - p_1^1 p_1^3\right)\left(p_2^2 p_2^4 - p_1^2 p_1^4\right) + \left(p_2^1 p_2^4 - p_1^1 p_1^4\right)\left(p_2^2 p_2^3 - p_1^2 p_1^3\right)}{30}\right\} da .$$

Similarly, the above equation can be rewritten as:

$$e_2 - e_1 = \sum_{i=1}^5 \int_0^\infty \left(p_2^i - p_1^i\right)\left[\frac{p_1^{-i} + p_2^{-i}}{2}\right] da - \sum_{i=1}^5 \int_0^\infty \left(p_2^i - p_1^i\right)\left(\frac{\sum_{\substack{j=1 \\ j \neq i}}^5 \left(p_2^j - p_1^j\right)\left(\prod_{\substack{k=1 \\ k \neq i,j}}^5 p_2^k - \prod_{\substack{k=1 \\ k \neq i,j}}^5 p_1^k\right)}{20}\right) da -$$

$$\sum_{i=1}^{5}\int_{0}^{\infty}\left(p_2^i - p_1^i\right)\left(\dfrac{\displaystyle\sum_{\substack{k=1\\k\neq i,\beta}}^{5}\left\{\left(p_2^{\beta}p_2^k - p_1^{\beta}p_1^k\right)\left(\prod_{\substack{m=1\\m\neq i,j,k}}^{5}p_2^m - \prod_{\substack{m=1\\m\neq i,j,m}}^{5}p_1^m\right)\right\}}{30}\right)da, \text{ where } \beta = \begin{cases}2 \text{ if } i=1\\1 \text{ if } i>1\end{cases}.$$

Six causes of death

$$e_2 - e_1 = \sum_{i=1}^{6}\int_{0}^{\infty}\left(p_2^i - p_1^i\right)\left[\dfrac{p_1^{-i} + p_2^{-i}}{2}\right]da - \sum_{i=1}^{6}\int_{0}^{\infty}\left(p_2^i - p_1^i\right)\left(\dfrac{\displaystyle\sum_{\substack{j=1\\j\neq i}}^{6}\left(p_2^j - p_1^j\right)\left(\prod_{\substack{k=1\\k\neq i,j}}^{6}p_2^k - \prod_{\substack{k=1\\k\neq i,j}}^{6}p_1^k\right)}{30}\right)da -$$

$$\sum_{i=1}^{6}\int_{0}^{\infty}\left(p_2^i - p_1^i\right)\left(\dfrac{\displaystyle\sum_{\substack{k=1\\k\neq i}}^{5}\sum_{\substack{j=k+1\\j\neq k,i}}^{6}\left\{\left(p_2^k p_2^j - p_1^k p_1^j\right)\left(\prod_{\substack{m=1\\m\neq i,j,k}}^{6}p_2^m - \prod_{\substack{m=1\\m\neq i,j,m}}^{6}p_1^m\right)\right\}}{60}\right)da.$$

General form, *n* causes of death

Let $p_1(a) = p_1^1 p_1^2 \ldots p_1^n$ and $p_2(a) = p_2^1 p_2^2 \ldots p_2^n$ (where $p_t^i(a) = p_t^i$, for $i=1,2,\ldots,n$ and $t=1,2$). Then the difference in life expectancy at birth for a population at times 1 and 2 is given by:

$$e_2 - e_1 = \int_0^\infty p_2(a)da - \int_0^\infty p_1(a)da = \int_0^\infty p_2^1 p_2^2 \ldots p_2^n da - \int_0^\infty p_1^1 p_1^2 \ldots p_1^n da.$$

Then, the decomposition of life expectancy at birth for a population at times 1 and 2 into *n* causes of death is given by ($n>5$):

$$e_2 - e_1 = \sum_{i=1}^{n} \int_0^\infty \left(p_2^i - p_1^i \right) \left[\frac{p_2^{-i} + p_1^{-i}}{2} \right] da -$$

$$\sum_{i=1}^{n} \int_0^\infty \left(p_2^i - p_1^i \right) \left\{ \sum_{r=1}^{\delta} \frac{\displaystyle\sum_{\substack{j=1 \\ j\neq i}}^{n-(r-1)} \sum_{\substack{k=j+1 \\ k\neq i,j}}^{n-(r-2)} \sum_{\substack{l=k+1 \\ l\neq i,j,k}}^{n-(r-3)} \cdots \sum_{\substack{z=y+1 \\ z\neq i,j,\ldots,y}}^{n} \left(p_2^j p_2^k p_2^l \ldots p_2^z - p_1^j p_1^k p_1^l \ldots p_1^z \right) \left(\displaystyle\prod_{\substack{\theta=1 \\ \theta\neq i,j,\ldots,z}}^{n} p_2^\theta - \prod_{\substack{\theta=1 \\ \theta\neq i,j,\ldots,z}}^{n} p_1^\theta \right)}{n \binom{n-1}{r}} \right\} da$$

where $\delta = \begin{cases} \dfrac{n-1}{2} & \text{if } n \text{ is odd} \\[2mm] \dfrac{n-2}{2} & \text{if } n \text{ is even} \end{cases}$.

Appendix 2. Implementing the decompositional system

We approximate the decomposition by computing the first part of equation (3), which is

in fact a generalization of equation (2), that is, $e_2 - e_1 = \sum_{i=1}^{n} \int_{0}^{\infty} \left(p_2^i - p_1^i \right) \left(\frac{p_1^{-i} + p_2^{-i}}{2} \right) da$.

In the discrete case, the approximation formula is equivalent to:

$e_2 - e_1 = \sum_{i=1}^{n} \left({}_nL_{x_1}^i - {}_nL_{x_2}^i \right) \left(\frac{{}_nL_{x_1}^{-i} + {}_nL_{x_2}^{-i}}{2 \cdot n} \right)$ for $l_0 = 1$, where ${}_nL_{x_t}^i$ and ${}_nL_{x_t}^{-i}$ represent the person-

years lived between ages x and $x+n$ at time t (t=1,2) in the life tables for cause i and cause $-i$, respectively.

In order to estimate ${}_nL_{x_t}^i$, we assume that the force of decrement function from cause i is proportional to the force of decrement function from all causes combined within the interval x to $x+n$ (Preston et al. 2001:82). The ${}_na_x^i$ values are obtained through graduation using equations 4.8 for ages 0, 1, 5 and 95 and equation 4.6 for ages 10 to 90 from Preston et al. (2001:82-84).

Having estimated ${}_nL_{x_t}^i$, we then estimate the person-years lived in the cause-deleted

life tables for each cause as: ${}_nL_{x_t}^{-i} = \left(\frac{{}_nL_{x_t}}{{}_nL_{x_t}^i} \right) \cdot n$, for i=1,2,3,…,k and t=1,2. (2.1)

where ${}_nL_{x_t}$ are the person-years lived between ages x and $x+n$ in the master life table for all causes of death combined at time t (t=1,2). The construction of the master life tables for 1970 and 2000 follows the methodology of Preston et al. (2001: chap. 3), including the use of graduation to determine the values of ${}_na_x$ as described above.

Equation 2.1 preserves the multiplicative property whereby the product of the probabilities of survival to a particular age in the associated single decrement table for cause i and in the cause-deleted table for cause i equals the probability of survival to that age for all causes combined.

We also wish to preserve that multiplicative property when the associated single decrement life tables for each individual cause are combined. We ensure this property through the formula by which we estimate person-years lived in the residual category of cause, the

category that includes all causes of death that are not individually modeled. In particular, the person-years lived for the remaining causes (cause k, the *residual*) are computed as[5]:

$$_nL_{x_t}^k = \left(\frac{_nL_{x_t}}{\prod_{m=1}^{k-1} {_nL_{x_t}^m}}\right) \cdot n^{(k-1)} \quad \text{for } t=1,2.$$ (2.2)

For the open ended interval, which in our applications begins at age 100, we assume mortality rates to be constant and that no person-years are lived above age 110. In this case, the person-years lived for the master and the associated single decrement life tables are computed as:

$$_+L_{100} = l_{100} \cdot \int_{100}^{110} e^{-M(x-100)} dx = l_{100} \cdot \left[-\frac{1}{M}\left(e^{-10M} - 1\right)\right] = l_{100} \cdot \frac{1-e^{-10M}}{M}, \text{ and } _+L_{100}^i = l_{100}^i \cdot \frac{1-e^{-10M^i}}{M^i},$$

respectively, where M and M^i represent the death rate above age 100 from all causes and from cause of death i.

---

[5] There are two possibilities when computing the person-years lived (PY) in the associated (ALT) and cause-deleted (CDLT) life tables. On the one hand, we can model each cause of death (including the remaining causes) as an ALT and compute the PY of the CDLT as a ratio using equation 2.1. On the other hand, we can model the first n-1 causes of death as an ALT, compute the PY for the remaining causes of death (residual) using equation 2.2, and then estimate the PY for the CDLT using equation 2.1. There is little difference in the results between the two options, but we prefer to use the second possibility given that it provides better estimates of the interaction terms.

<div align="center">References</div>

Andreev, E.M., Vladimir M. Shkolnikov, and Alexander Z. Begun. 2002. "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.

Arias, E. 2002. "United States Life Tables, 2000." Hyattsville, MD: National Center for Health Statistics.

Arriaga, E. 1982. "A Note on the Use of Temporary Life Expectancies for Analyzing Changes and Differentials of Mortality." Pp. 559-62 in *Mortality in South and East Asia: A Review of Changing Trends and Patterns, Manila 1980*, WHO/ESCAP. Geneva: World Health Organization.

―――. 1984. "Measuring and Explaining the Change in Life Expectancies." *Demography* 21(1):83-96.

―――. 1989. "Changing Trends in Mortality Decline During the Last Decade." Pp. 105-29 in *Differential Mortality: Methodological Issues and Biosocial Factors*, Eds. L. Ruzicka, W. Guillaume, and P. Kane. Oxford, England: Clarendon, Press. International Studies in Demography.

Centers for Disease Control. 2001. "Comparability Across ICD Revisions for Selected Causes." http://www.cdc.gov/nchs/data/dvs/comp2.pdf

―――. 2005. "10 Leading Causes of Death, United States 2000, All Races, Both Sexes." http://webapp.cdc.gov/sasweb/ncipc/leadcaus10.html

Cutler, D. 2004. *Your Money or Your Life*. Oxford, England: Oxford University Press.

Das Gupta, P. 1978. "General Method of Decomposing a Difference Between 2 Rates Into Several Components." *Demography* 15(1):99-112.

―――. 1989. "Methods of Decomposing the Difference Between 2 Rates With Applications to Race-Sex Inequalities in Earnings." *Mathematical Population Studies* 2(1):15-36.

―――. 1991. "Decomposition of The Difference Betwen Two Rates and Its Consistency When More Than Two Populations Are Involved." *Mathematical Population Studies* 3(2):105-25.

―――. 1993. "Standardization and Decomposition of Rates: a User's Manual." *Current Population Reports, Series P-23, 186*. Washington, D.C.: U.S. Bureau of the Census.

Ergin, A., P. Muntner, R. Sherwin, and J. He. 2004. "Secular Trends in Cardiovascular Disease Mortality, Incidence, and Case Fatality Rates in Adults in the United States." *American Journal of Medicine* 117:219-27.

Kitagawa, E. 1955. "Components of a Difference Between Two Rates." *Journal of American Statistical Association* 50(272):1168-94.

National Cancer Institute. 2007. "U.S. Standard Population 2000." Surveillance Epidemiology and End Results. http://seer.cancer.gov/stdpopultions/stdpopdic.html

National Center for Health Statistics. 1974. "Vital Statistics of the United States, 1970: Life Tables. Vol. 2, Sec.5." Washington, DC: U.S. Government Printing Office.

———. 2002. *Multiple Cause of Death Public Use Files for 2000, 2001, and 2002* [computer file]. NCHS CD-ROM; Series 20. Hyattsville, Md.: U.S. Dept. of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention.

———. 2001. Multiple Cause Of Death, 1968-1973 [Computer File]. Inter-university Consortium for Political and Social Research (ICPSR) release. Hyattsville, MD: U.S. Dept. of Health and Human Services, National Center for Health Statistics [producer]. Ann Arbor, MI: ICPSR [distributor], 2004.

Pollard, J.H. 1982. "The Expectation of Life and Its Relationship to Mortality." *Journal of Institute of Actuaries* 109(Part II, No 442):225-40.

———. 1988. "On the Decomposition of Changes in Expectation of Life and Differentials in Life Expectancy." *Demography* 25(2):265-76.

Pressat, R. 1985. "Contribution Des Écarts De Mortalité Par Âge ŕ La Différence Des Vies Moyennes." *Population* 4-5:765-70.

Preston, S.H., I.T. Elo, A. Foster, and H.S. Fu. 1998. "Reconstructing the Size of the African American Population by Age and Sex, 1930-1990." *Demography* 35(1):1-21.

Preston, S.H., P. Heuveline, and M. Guillot. 2001. *Demography. Measuring and Modeling Population Processes*. 2002 ed. Malden, Massachusetts: Blackwell Publishers Inc.

Preston, S.H., N. Keyfitz, and R. Schoen. 1972. *Causes of Death: Life Tables for National Populations*. New York and London : Seminar Press.

Siegel, J.S. 1974. "Estimates of Coverage of Population by Sex, Race, and Age in 1970 Census." *Demography* 11(1):1-23.

Siegel, J.S. and J.S. Passel. 1976. "New Estimates of Number of Centenarians in United-States." *Journal of the American Statistical Association* 71(355):559-66.

U.S. Bureau of the Census. 1971. *U.S. Census of Population and Housing: 1970. "General Population Characteristics."* Washington, DC.:U.S. Government Printing Office.

———. 2005. "Monthly Postcensal Resident Population, by Single Year of Age, Sex, Race, and Hispanic Origin, 2000-2005."
http://www.census.gov/popest/national/asrh/2005_nat_res.html.

Vaupel, J.W. and V.C. Romo. 2002. "Decomposing Demographic Change into Direct Vs. Compositional Components." *Demographic Research* 7(1):1-14.

———. 2003. "Decomposing Change in Life Expectancy: a Bouquet of Formulas in Honor of Nathan Keyfitz's 90th Birthday." *Demography* 40(2):201-16.