Decomposition techniques in population health research

José Manuel Aburto

EUROPEAN DOCTORAL SCHOOL OF DEMOGRAPHY

Recap

To avoid the interaction term, Kitagawa suggests

$$\Delta CDR = \underbrace{\sum_x \left(\frac{M_x(t_2) + M_x(t_1)}{2} \right) \left(\frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)} \right)}_{\text{Changes in x-composition}} + \underbrace{\sum_x \left(\frac{\frac{N_x(t_2)}{N(t_2)} + \frac{N_x(t_1)}{N(t_1)}}{2} \right) (M_x(t_2) - M_x(t_1))}_{\text{Changes in rates}} \tag{1}$$

Recap

$$\dot{\bar{v}} = \underbrace{\bar{\dot{v}}}_{\text{Direct component}} + \underbrace{Cov(v, \acute{w})}_{\text{Structural or compositional component}} \quad (2)$$

Main result in Vaupel & Canudas-Romo (2002)

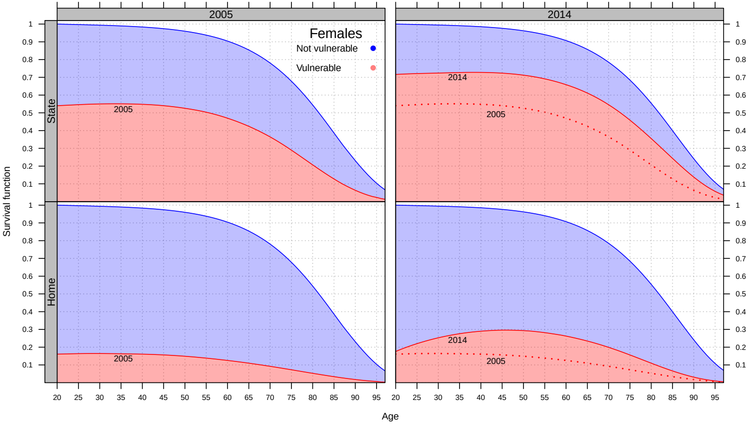
Recap

Life expectancy at age x is defined as

$$e(x) = \frac{\int_x^\infty \ell(a)da}{\ell(x)}$$

Then we can define disability-free life expectancy as

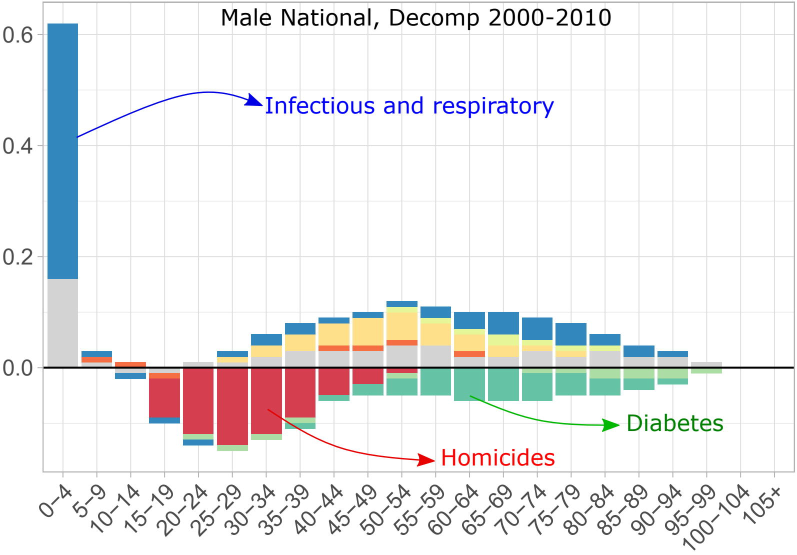
$$e^{DF}(x) = \frac{\int_x^\infty [1 - \pi(a)]\ell(a)da}{\ell(x)}$$


2005
2014
Females
Not vulnerable
Vulnerable
State
Home
Survival function
Age
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95

Recap

Arriaga (1984) Effects of mortality change by age groups on life expectancies ($\sum_n \Delta_x =$ Total change):

$$
{}_n\Delta_x = \underbrace{\frac{\ell_x^1}{\ell_0^1}\left(\frac{{}_nL_x^2}{\ell_x^2} - \frac{{}_nL_x^1}{\ell_x^1}\right)}_{\text{Direct effect}} + \underbrace{\frac{T_{x+n}^2}{\ell_0^1}\left(\frac{\ell_x^1}{\ell_x^2} - \frac{\ell_{x+n}^1}{\ell_{x+n}^2}\right)}_{\text{Indirect and interaction effects}}
$$


Male National, Decomp 2000-2010
0.6
0.4
0.2
0.0
Infectious and respiratory
Diabetes
Homicides
0–4
5–9
10–14
15–19
20–24
25–29
30–34
35–39
40–44
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80–84
85–89
90–94
95–99
100–104
105+

Outline

- Cause-deleted life tables
- Measures of variation in ages at death
- Linear integral decomposition

Cause-deleted life tables (based on Chapter 4, Preston et al 2001)

Often we are interested in knowing what a life table would look if only one cause of death were operating to diminish a cohort.

Cause-deleted life tables (based on Chapter 4, Preston et al 2001)

Often we are interested in knowing what a life table would look if only one cause of death were operating to diminish a cohort.
The activity of a particular decrement i will be observed as it works itself out within a multiple decrement process.

Cause-deleted life tables (based on Chapter 4, Preston et al 2001)

Often we are interested in knowing what a life table would look if only one cause of death were operating to diminish a cohort.
The activity of a particular decrement i will be observed as it works itself out within a multiple decrement process.

'What would happen if...'

- The rate of decrement from $\mu^i(x)$ if i were the only decrement $({}^*m^i(x))$ differs from what it would be if i were working in the presence of other decrements $(m^i(x))$.

- Assume the force of decrement of cause i is proportional all-cause force of decrement:

$$\mu^i(a) = R_i \cdot \mu(a) \quad \text{for} x \leq a \leq x + n$$

- Assume the force of decrement of cause i is proportional all-cause force of decrement:

$$\mu^i(a) = R_i \cdot \mu(a) \quad \text{for} x \leq a \leq x + n \tag{3}$$

- Since, by assupmtion,

$$_n^*p_x^i = e^{-\int_x^{x+n} \mu^i(a)da} = e^{-\int_x^{x+n} R_i \cdot \mu(a)da}$$

- Assume the force of decrement of cause i is proportional all-cause force of decrement:

$$\mu^i(a) = R_i \cdot \mu(a) \quad \text{for} x \leq a \leq x + n \tag{3}$$

- Since, by assupmtion,

$${}_n^*p_x^i = e^{-\int_x^{x+n} \mu^i(a)da} = e^{-\int_x^{x+n} R_i \cdot \mu(a)da}$$

$${}_n^*p_x^i = e^{-R_i \int_x^{x+n} \mu(a)da} = [e^{-\int_x^{x+n} \mu(a)da}]^{R_i}$$

- Assume the force of decrement of cause i is proportional all-cause force of decrement:

$$\mu^i(a) = R_i \cdot \mu(a) \quad \text{for} x \leq a \leq x + n \tag{3}$$

- Since, by assupmtion,

$$ {}_n^*p_x^i = e^{-\int_x^{x+n} \mu^i(a)da} = e^{-\int_x^{x+n} R_i \cdot \mu(a)da}$$

$$ {}_n^*p_x^i = e^{-R_i \int_x^{x+n} \mu(a)da} = [e^{-\int_x^{x+n} \mu(a)da}]^{R_i}$$

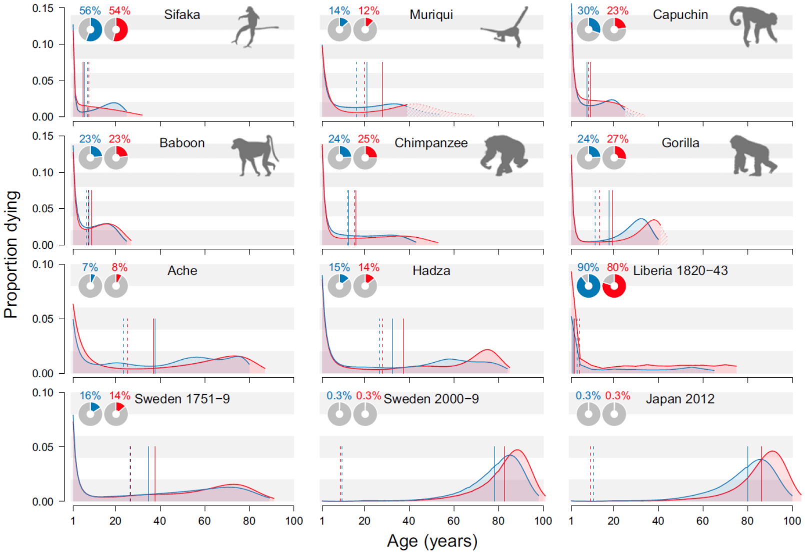
$$ {}_n^*p_x^i = [{}_np_x]^{R_i}$$

$$ {}_n^{*}p_x^{i} = \left[{}_n p_x\right]^{R_i} $$

$$ = \left[{}_n p_x\right]^{\frac{{}_n D_x^i}{{}_n D_x}} $$

Now with this simple relation we can create hypothetical scenarios.

`https://population-health.shinyapps.io/saudi-arabia-health-profile/`


Sifaka
56% 54%
Muriqui
14% 12%
Capuchin
30% 23%
Baboon
23% 23%
Chimpanzee
24% 25%
Gorilla
24% 27%
Ache
7% 8%
Hadza
15% 14%
Liberia 1820−43
90% 80%
Sweden 1751−9
16% 14%
Sweden 2000−9
0.3% 0.3%
Japan 2012
0.3% 0.3%
Proportion dying
0.15
0.10
0.05
0.00
1 20 40 60 80 100
Age (years)

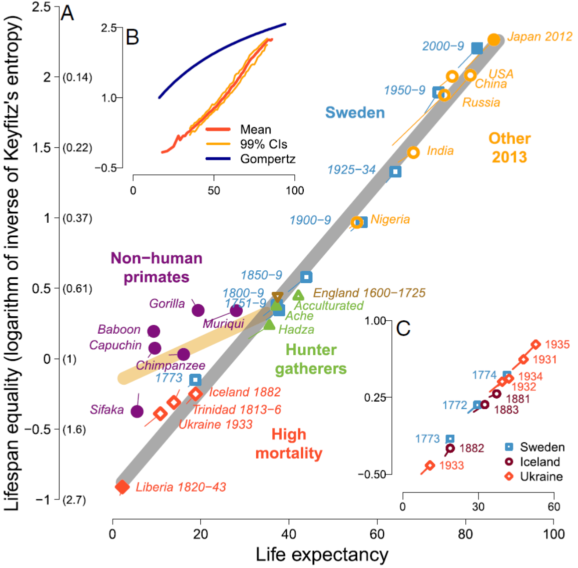
Lifespan variation indicators

Indicators	•	Conventional life table notation
Life disparity	$e^\dagger$	$\sum_{y=0}^{\omega} d_y e(y)$
Gini coefficient	G	$1 - \frac{1}{e_0} \sum_{y=0}^{\omega} \ell_{y+1}^2$
Theil's index	T	$\sum_{y=0}^{\omega} d_y \left[\frac{\bar{x}_y}{e_0} \ln \frac{\bar{x}_y}{e_0}\right]$
Mean logarithmic deviation	MLD	$\sum_{y=0}^{\omega} d_y (\ln(e_0/\bar{x}_y)$
Variance	V	$\sum_{y=0}^{\omega} d_y (\bar{x}_y - e_0)^2$
Standard deviation	SD	$\sqrt{V}$
Interquartile range	IQR	$\hat{x}_3 - \hat{x}_1$

Table 1 Pearson correlation coefficients between pairs of indices, calculated from birth (ages 0–110+) in the top panel and calculated conditional on survival to age 10 (ages 10–110+) in the bottom panel, for all female and male life tables in the Human Mortality Database (7,516 in total)

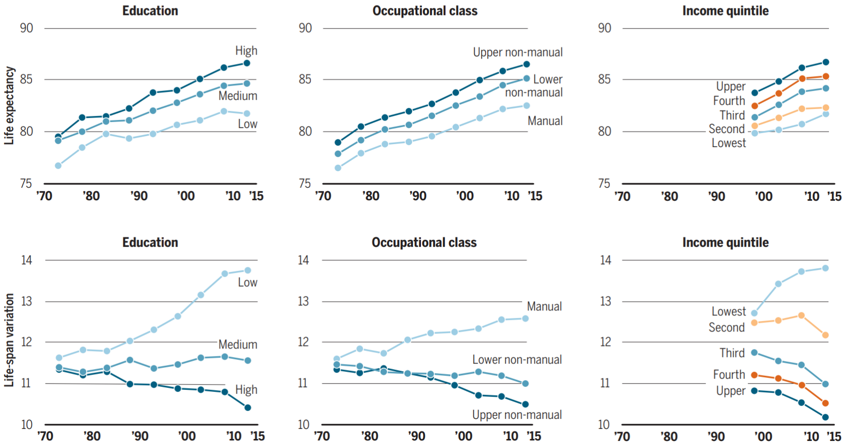
	$e^\dagger$	G	T	MLD	S	V	IQR
$e^\dagger$	1.000						
G	.978	1.000					
T	.947	.991	1.000				
MLD	.965	.991	.992	1.000			
S	.981	.933	.893	.930	1.000		
V	.987	.945	.911	.944	.996	1.000	
IQR	.967	.966	.948	.956	.920	.944	1.000
	$e^\dagger_{10}$	G_{10}	T_{10}	MLD_{10}	S_{10}	V_{10}	IQR_{10}
$e^\dagger_{10}$	1.000						
G_{10}	.986	1.000					
T_{10}	.978	.995	1.000				
MLD_{10}	.979	.990	.995	1.000			
S_{10}	.986	.962	.961	.973	1.000		
V_{10}	.984	.964	.971	.980	.998	1.000	
IQR_{10}	.981	.978	.977	.976	.958	.966	1.000

Source: van Raalte & Caswell (2013)


A
B
C
Lifespan equality (logarithm of inverse of Keyfitz's entropy)
Life expectancy
2.5
2 (0.14)
1.5 (0.22)
1 (0.37)
0.5 (0.61)
0 (1)
−0.5 (1.6)
−1 (2.7)
0
20
40
60
80
100
Mean
99% CIs
Gompertz
Sweden
Other 2013
Non−human primates
Hunter gatherers
High mortality
Japan 2012
2000−9
USA
China
Russia
1950−9
India
1925−34
1900−9
Nigeria
1850−9
1800−9
1751−9
England 1600−1725
Acculturated
Ache
Hadza
Gorilla
Muriqui
Baboon
Capuchin
Chimpanzee
1773
Iceland 1882
Trinidad 1813−6
Ukraine 1933
Sifaka
Liberia 1820−43
1935
1931
1774
1934
1932
1772
1881
1883
1773
1882
1933
Sweden
Iceland
Ukraine

Trends in life expectancy and life-span variation for Finnish females, 1971–1975 to 2011–2014

Life expectancy is the average age at death, and life-span variation is the standard deviation, conditional upon survival to age 30, with age-specific death rates frozen at those observed in the given year. See supplementary materials for data and methods, including trends for males (which are qualitatively similar), and robustness checks using alternative measures of life-span variation.



How can we decompose by age and cause of death these (any) function?

Linear integral decomposition (Horiuchi et al 2008)

- Relies on the assumption that values of the covariates change **continuously**, or gradually, along an actual or hypothetical dimension

Linear integral decomposition (Horiuchi et al 2008)

- Relies on the assumption that values of the covariates change **continuously**, or gradually, along an actual or hypothetical dimension
- Applied when difference in a dependent variable is expressed as a sum of the effects of differences in its covariates. (TFR,mean completed parity)

- Let y be a demographic function (e.g. $e^{\dagger}$), which is differentiable, of n covariates (e.g. rates) denoted by $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$.

- Let y be a demographic function (e.g. $e^{\dagger}$), which is differentiable, of n covariates (e.g. rates) denoted by $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$.
- Assume they both depend on an underlying dimension t (e.g. time).

- Let y be a demographic function (e.g. $e^\dagger$), which is differentiable, of n covariates (e.g. rates) denoted by $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$.
- Assume they both depend on an underlying dimension t (e.g. time).
- We have observations at two time points and $\boldsymbol{x}$ is a differentiable vector function of t between t_1 and t_2.

- Let y be a demographic function (e.g. $e^{\dagger}$), which is differentiable, of n covariates (e.g. rates) denoted by $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$.
- Assume they both depend on an underlying dimension t (e.g. time).
- We have observations at two time points and $\boldsymbol{x}$ is a differentiable vector function of t between t_1 and t_2.

$$y(t) = f(\boldsymbol{x}(t)) = f(x_1(t), x_2(t), \ldots, x_n(t)) \quad (4)$$

By the fundamental theorem of calculus

$$y(t_2) - y(t_1) = \int_{t_1}^{t_2} \frac{d}{dt} y(t) dt$$

By the fundamental theorem of calculus

$$y(t_2) - y(t_1) = \int_{t_1}^{t_2} \frac{d}{dt} y(t) dt \tag{5}$$

Applying the chain rule for partial derivatives of a composite function

$$y(t_2) - y(t_1) = \int_{t_1}^{t_2} \left[\sum_{i=1}^{n} \frac{\partial}{\partial x_i(t)} y(t) \cdot \frac{d}{dt} x_i(t) \right] dt \tag{6}$$

Exchange of integration and summation, and applying the substitution rule for definite integrals

$$y(t_2) - y(t_1) = \sum_{i=1}^{n} \int_{x_i(t_1)}^{x_i(t_2)} \frac{\partial}{\partial x_i(t)} y(t) dx_i(t)$$

Exchange of integration and summation, and applying the substitution rule for definite integrals

$$y(t_2) - y(t_1) = \sum_{i=1}^{n} \int_{x_i(t_1)}^{x_i(t_2)} \frac{\partial}{\partial x_i(t)} y(t) dx_i(t) \quad (7)$$

(Simplifying notation)

$$y_2 - y_1 = \sum_{i=1}^{n} \int_{x_{i1}}^{x_{i2}} \frac{\partial y}{\partial x_{i1}} dx_i = \sum_{i=1}^{n} c_i \quad (8)$$

$$y_2 - y_1 = \sum_{i=1}^{n} c_i \tag{9}$$

c_i is the total change in y produced by changes in the i-th covariate, x_i.

Important:Theoretical foundation for decomp analysis: implies that even if a dependent variable is not an additive function of its covariates, a change in the dependent variable can be expressed as a sum of effects of the covariates.

Exercise 1: by age

International Journal of
Epidemiology

Issues | Advance articles | Submit ▾ | Purchase | Alerts | About ▾

Article Contents

Abstract
Introduction
Data
Analytical methods
Results
Discussion
Funding
Acknowledgements
References
Supplementary data

The impact of violence on Venezuelan life expectancy and lifespan inequality

Jenny García, José Manuel Aburto

International Journal of Epidemiology, dyz072, https://doi.org/10.1093/ije/dyz072

Published: 21 April 2019 **Article history** ▾

Split View | PDF | Cite | Permissions | Share ▾

Abstract

Background

Venezuela is one of the most violent countries in the world. According to the United Nations, homicide rates in the country increased from 32.9 to 61.9 per 100 000 people between 2000 and 2014. This upsurge coincided with a slowdown in life expectancy improvements. We estimate mortality trends and quantify the impact of violence-related deaths and other causes of death on life expectancy and lifespan inequality in Venezuela.

Extending to causes of death

Lifespan inequality in Denmark, Sweden and Norway

Country
Denmark

Initial year
1975

Final year
1995

Comparison year
2014

Females

Year	L. Exp	L. Inequality
1975	77	0.20613
1995	77.83	0.18797
Difference	0.83	-0.01816

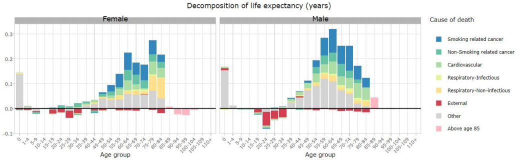
Males

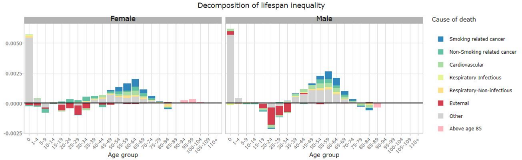
Year	L. Exp	L. Inequality
1975	71.28	0.23749
1995	72.72	0.21034
Difference	1.44	-0.02715

Difference with Sweden

Indicator	Sex	
	Females	Males
Life expectancy	1.38	1.79
cv	-0.0117	-0.0121

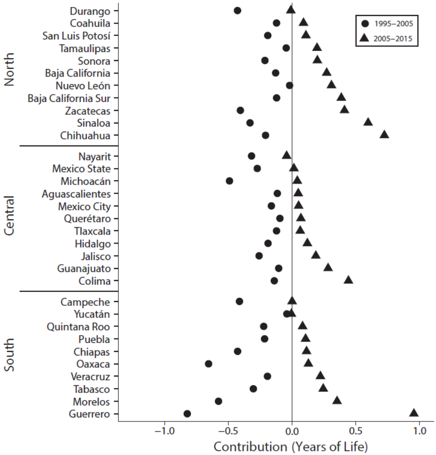
Life expectancy and lifespan inequality trends | Decomposition results | Cause-specific summary | Comparison with Sweden | Early life ASMR





Positive (negative) values increase (decrease) the gap in life expectancy and lifespan inequality with Sweden

`https://jmaburto.shinyapps.io/DK_App/`



Note. This figure shows how homicides contributed to changes in life span inequality (i.e., panel b of Figure 1) in 2 periods: positive values suggest increases in years of life lost because of homicides and negative values correspond to reductions in life years lost because of death. This figure shows each of the 32 Mexican states grouped in broad regions: north, central, and south. Within each region, states are ordered according to the magnitude of the impact of homicides to life span inequality at age 15 years in the period 2005 to 2015.

FIGURE 2—Contribution of Homicide Mortality to Changes in Male Life Span Inequality by State and Period: Mexico, 1995–2005 and 2005–2015

Exercise 2: by cause of death

Challenge 1 Use the linear integral model to decompose the change in the standard deviation of the age-at-death distribution and life expectancy by age and cause of death for 3 countries you might be interested in (over time or between them). Interpret the results of life expectancy alongside standard deviation. Make it interesting. You can use data from HCoD, HMD, WHO, GBD.