Decomposition- Class 1

José Manuel Aburto

Preliminaries

- Introduction
- Course materials
  `https://github.com/jmaburto/Anahuac-Decomposition-Workshop`
- Teams
- Challenges in class. I expect an Rmarkdown file and PDF or HTML.

Outline

- The first decomposition method: Kitagawa (1955)
- Parts of the survival function

Origins of decomposition

- We love rates.

Origins of decomposition

- We love rates.
- But crude rates are summary measures of population change

Origins of decomposition

- We love rates.
- But crude rates are summary measures of population change
- Due to births, deaths or migration.
- These can differ due to underlying rates and population structure

Origins of decomposition

- Methods of standardization
  Aim: Eliminate compositional effect from overall rates of some phenomenon.

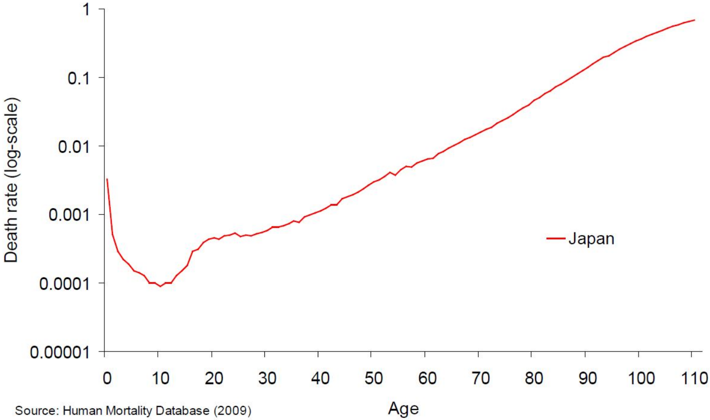
Origins of decomposition

- Methods of standardization
  Aim: Eliminate compositional effect from overall rates of some phenomenon.
  - Indirect standardization $\longrightarrow$ 1876
  - Direct standardization $\longrightarrow$ 1844

Origins of decomposition

- Methods of standardization
  Aim: Eliminate compositional effect from overall rates of some phenomenon.
  - Indirect standardization $\longrightarrow$ 1876
  - Direct standardization $\longrightarrow$ 1844
- Unreliable due to their dependence on an arbitrary standard

Figure 1. Age-specific death rates for the total population of Japan in 2000.



Source: Human Mortality Database (2009)

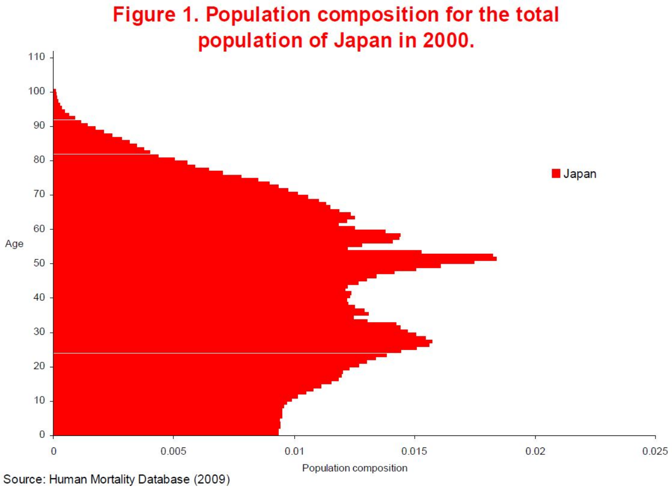


Figure 1. Population composition for the total population of Japan in 2000.
Age
110
100
90
80
70
60
50
40
30
20
10
0
0
0.005
0.01
0.015
0.02
0.025
Population composition
Japan
Source: Human Mortality Database (2009)

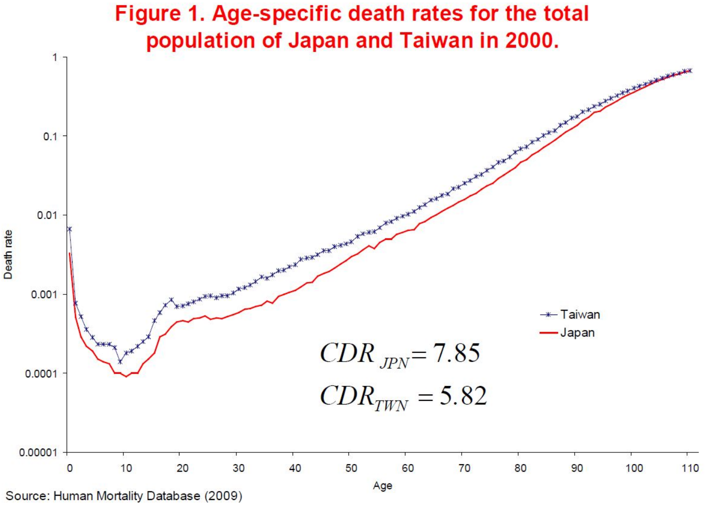
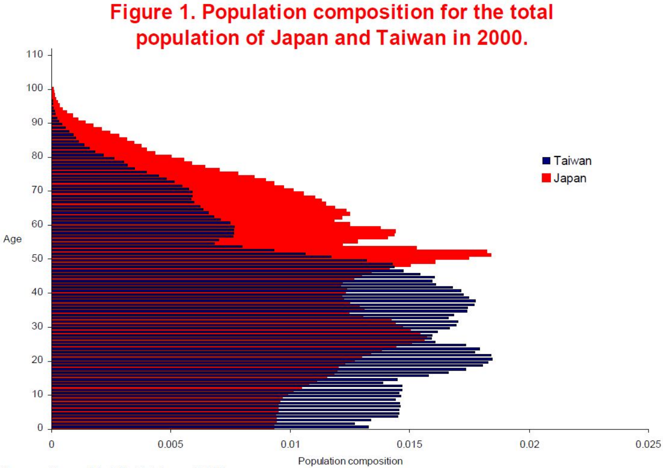


Figure 1. Age-specific death rates for the total population of Japan and Taiwan in 2000.
Death rate
1
0.1
0.01
0.001
0.0001
0.00001
0
10
20
30
40
50
60
70
80
90
100
110
Age
Taiwan
Japan
CDR JPN = 7.85
CDR TWN = 5.82
Source: Human Mortality Database (2009)



Source: Human Mortality Database (2009)

Crude Death rate (CDR)

	JAPAN	TAIWAN
CDR	7.85	5.82
SCDR Direct Stand.	5.79	8.72
SCDR Indirect Stand.	9.73	4.78

Motivation to develop further methods of comparison: Decomposition

Kitagawa (1955)

t_1 is the initial period and t_2 is the final period
D_x = number of deaths at age x
M_x = death rate
N_x = is the mid-year population
N = total population over ages
Note that $D_x = M_x * N_x$

Kitagawa (1955)

$$\Delta CDR = \sum_x M_x(t_2)\frac{N_x(t_2)}{N(t_2)} - \sum_x M_x(t_1)\frac{N_x(t_1)}{N(t_1)} \tag{1}$$

Kitagawa (1955)

The difference between the crude rates can be expressed as

$$\Delta CDR = \underbrace{\sum_x M_x(t_1)\left[\frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)}\right]}_{\text{Changes in x-composition}} + \underbrace{\sum_x \frac{N_x(t_1)}{N(t_1)}\left[M_x(t_2) - M_x(t_1)\right]}_{\text{Change in rates with pop 1 as standard}} + \underbrace{\sum_x (M_x(t_2) - M_x(t_1))\left[\frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)}\right]}_{\text{Interaction of rates and compositions}} \tag{2}$$

Kitagawa (1955)

To avoid the interaction term, Kitagawa suggests

$$\Delta CDR = \underbrace{\sum_x \left(\frac{M_x(t_2) + M_x(t_1)}{2} \right) \left(\frac{N_x(t_2)}{N(t_2)} - \frac{N_x(t_1)}{N(t_1)} \right)}_{\text{Changes in x-composition}} + \underbrace{\sum_x \left(\frac{\frac{N_x(t_2)}{N(t_2)} + \frac{N_x(t_1)}{N(t_1)}}{2} \right) (M_x(t_2) - M_x(t_1))}_{\text{Changes in rates}} \quad (3)$$

Challenge 1: show that (1) can be expressed as (3) (step by step)

Example: Nishikido, Cui and Esteve 2022

Aim: To examine the role of partnership formation in explaining the gap between Sweden and Spain regarding transitions to first birth.

Data: 2012/2013 Swedish Generations and Gender Survey (GGS) and the 2018 Spanish Fertility Survey (SFS)

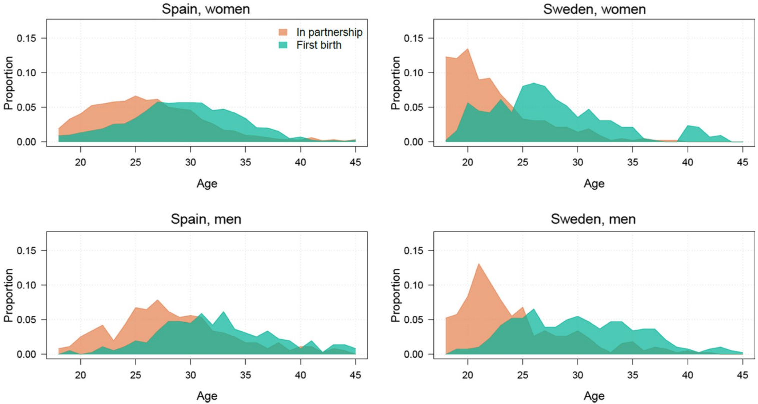


Fig. 2 Timing and intensity of first event occurrences by age, 1965–1969. Source: Calculated by authors based on the Spanish Fertility Survey (2018) and Swedish Generations and Gender Survey (2013) from the Harmonized Histories dataset

1) Transition probabilities $\longrightarrow$ focus on behavior across the reproductive period, i.e. whether fertility schedule differences exist after controlling for partnership status.

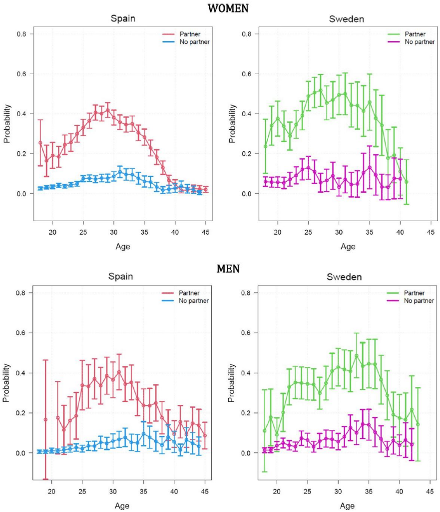


Fig. 3 First-birth probability within the next 3 years at a given age (with 95% confidence intervals) based on partnership status, 1965–1969. Source: Calculated by authors based on the Spanish Fertility Survey (2018) and Swedish Generations and Gender Survey (2013) from the Harmonized Histories dataset

2) The decomposition analysis $\longrightarrow$ allows us to explore how much of the gap in first-order TFRs between Sweden and Spain is attributable to age-specific (1) differences in fertility behavior (e.g., the partnered first-order TFRs) or (2) differences in composition (e.g., the proportion of those in a stable partnership).

We compute the first-order TFR based on partnership status, which is expressed as:

$$\mathrm{TFR}(t) = \sum_{x=\alpha}^{\beta} \mathrm{ASFR}(x,t) = \sum_{x=\alpha}^{\beta} \sum_{p=0}^{1} \mathrm{ASFR}(x,t,p) C(x,t,p), \tag{2}$$

where α and β represent the minimum and maximum reproductive ages. $\mathrm{ASFR}(x,t,p)$ and $C(x,t,p)$ are the age-specific fertility rate and the proportion of women, respectively, given age, cohort, and partnership status. We follow the Kitagawa (1955) approach and decompose the difference in first-order TFR as the following:

$$\Delta\mathrm{TFR}(c) = \sum_{x=\alpha}^{\beta} \sum_{p=0}^{1} \Delta\mathrm{ASFR}(x,t,p)\overline{C}(x,t,p) + \overline{\mathrm{ASFR}}(x,t,p)\Delta C(x,t,p), \tag{3}$$

where Δ and overbar mean the difference and average between two populations, separately. For example, $\Delta\mathrm{ASFR}(x,t,p) = \mathrm{ASFR}(x,t,p,\mathrm{SWE}) - \mathrm{ASFR}(x,t,p,\mathrm{ESP})$, and $\overline{C}(x,t,p) = \frac{C(x,t,p,\mathrm{SWE})+C(x,t,p,\mathrm{ESP})}{2}$. Based on Eq. (3), the two effects can be defined as:

$$\text{Rate effect} = \sum_{x=\alpha}^{\beta} \sum_{p=0}^{1} \Delta\mathrm{ASFR}(x,t,p)\overline{C}(x,t,p), \text{and} \tag{4}$$

$$\text{Composition effect} = \sum_{x=\alpha}^{\beta} \sum_{p=0}^{1} \overline{\mathrm{ASFR}}(x,t,p)\Delta C(x,t,p). \tag{5}$$

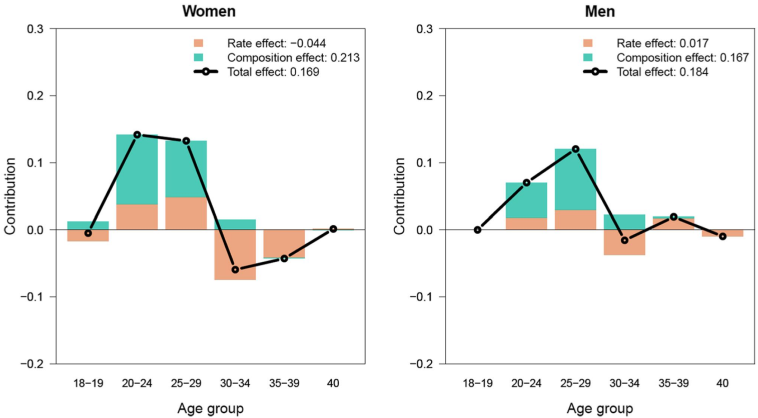


Fig. 4 Contribution of first-order rate and partnership composition to the first-order TFR differential by age group, 1965–1969. Source: Calculated by authors based on the Spanish Fertility Survey (2018) and Swedish Generations and Gender Survey (2013) from the Harmonized Histories dataset. $\Delta TFR = TFR_{SWEDEN} - TFR_{SPAIN}$

3) Standardization $\longrightarrow$ Read the paper

Kitagawa (1955)

Challenge 2: With data on fertility (e.g. HFD or WPP) select 5 countries and analyze the change in their crude fertility rate (CFR) in a recent period (10 years) and decompose these changes following Kitagawa's decomposition and describe your results.

Then for the most recent period select the two countries (among the 5) with the highest and lowest CFR and decompose their difference and describe your results.

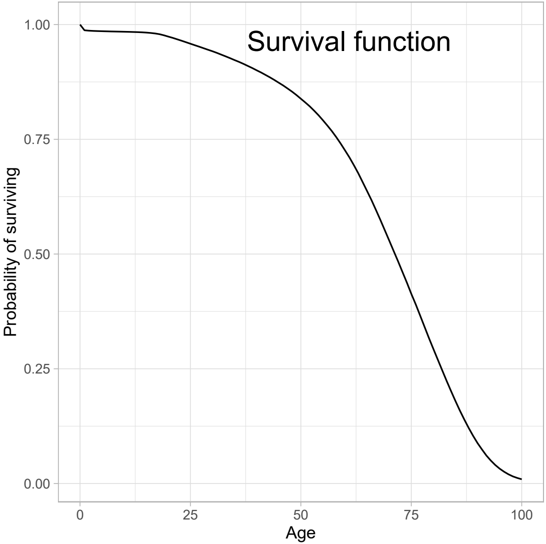
Further reading

- Gupta, Prithwis Das. "A general method of decomposing a difference between two rates into several components." Demography 15.1 (1978): 99-112.
- Cho, L. J., & Retherford, R. D. (1973). Comparative analysis of recent fertility trends in East Asia.
- Gonalons-Pons, P., & Schwartz, C. R. (2017). "Trends in Economic Homogamy: Changes in Assortative Mating or the Division of Labor in Marriage?." Demography, 54(3), 985-1005.

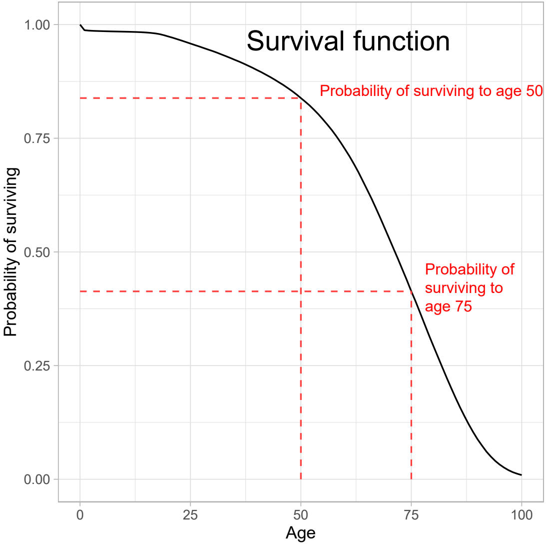
Kitagawa (1955)

Exercise 1

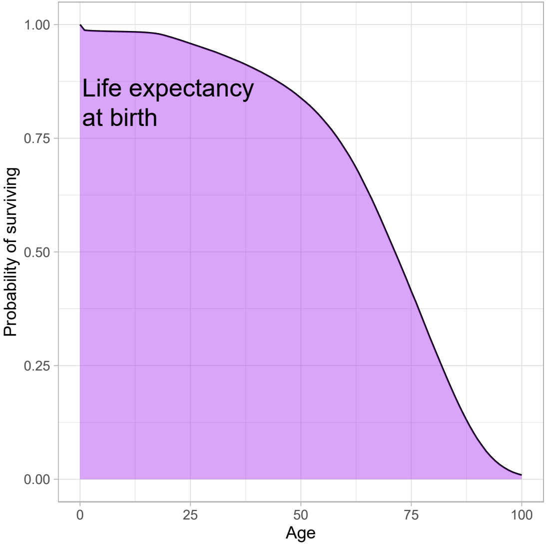
Black men USA 2020



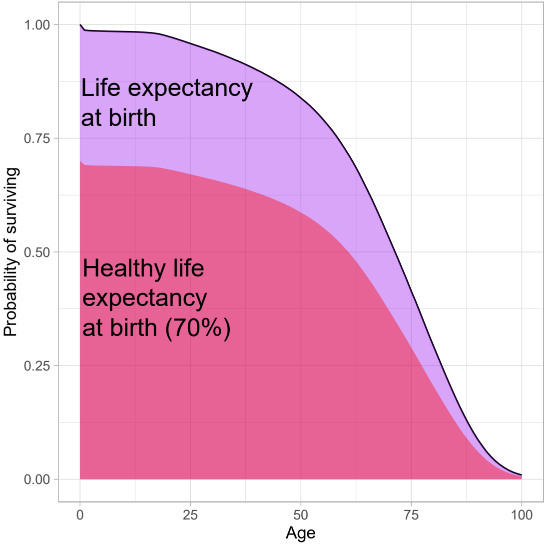
Black men USA 2020



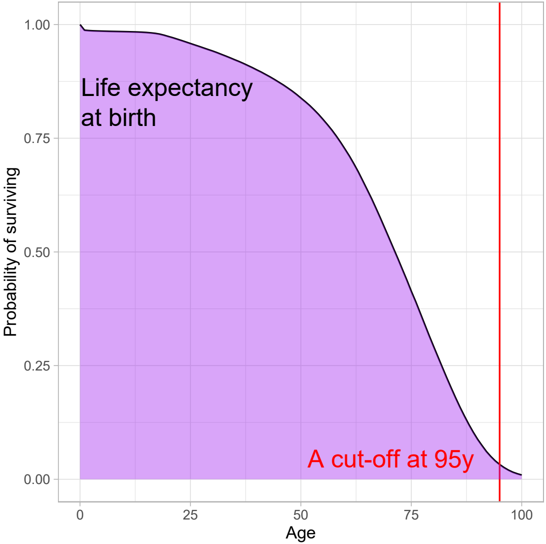
Black men USA 2020
Life expectancy = 67.8y



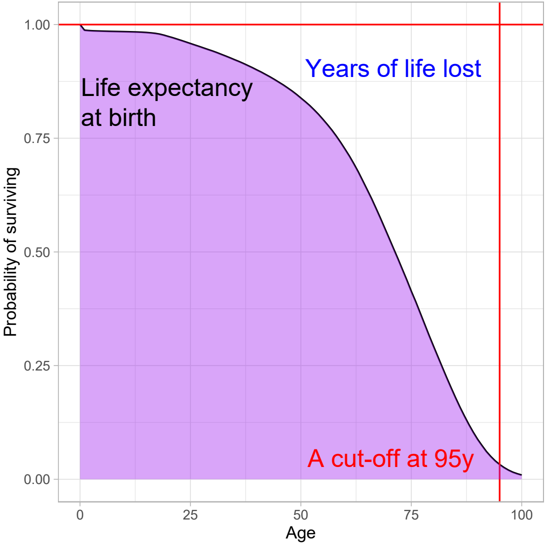
Black men USA 2020
Life expectancy = 67.8y



Black men USA 2020
Life expectancy = 67.8y

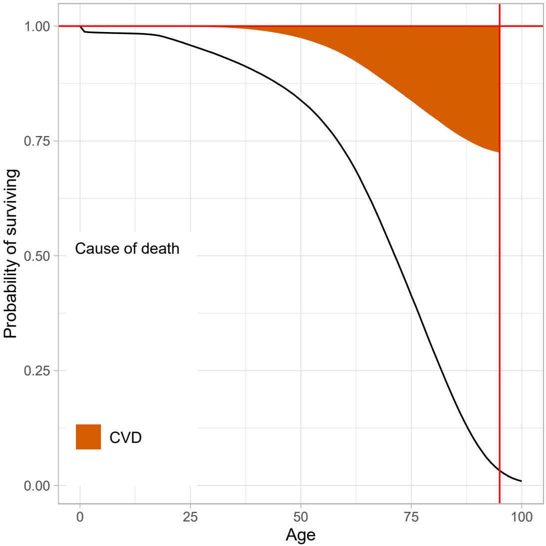


Black men USA 2020
Life expectancy = 67.8y
Years lost = 28.3y



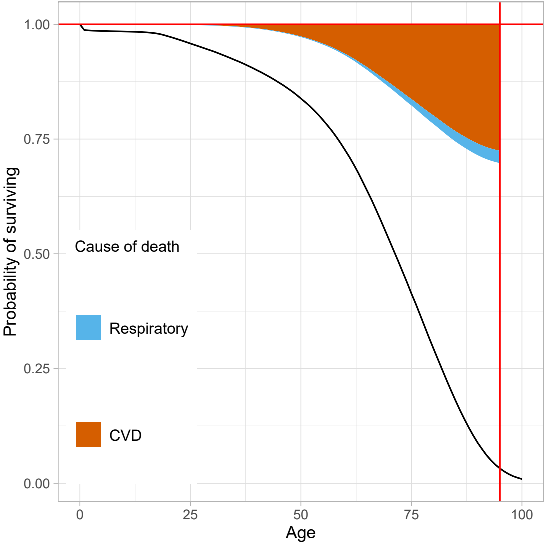
Black men USA 2020
Life expectancy = 67.8y
Years lost = 28.3y

CVD 7y ⟶ 24.4%



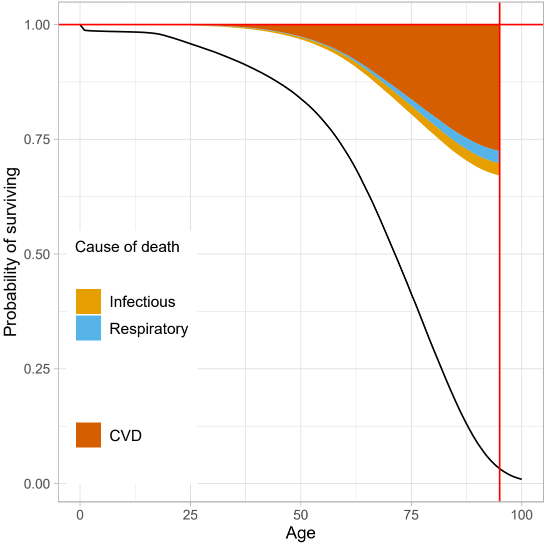
Black men USA 2020
Life expectancy = 67.8y
Years lost = 28.3y

CVD 7y ⟶ 24.4%
Respiratory .6y ⟶ 2.2%



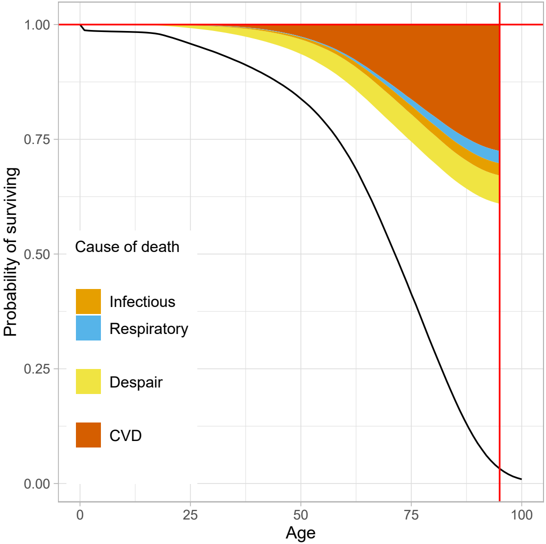
Black men USA 2020
Life expectancy = 67.8y
Years lost = 28.3y

CVD 7y ⟶ 24.4%
Respiratory .6y ⟶ 2.2%
Infectious .8y ⟶ 2.8%



Black men USA 2020
Life expectancy = 67.8y
Years lost = 28.3y

CVD 7y ⟶ 24.4%
Respiratory .6y ⟶ 2.2%
Infectious .8y ⟶ 2.8%
Despair 2.9y ⟶ 10.2%



Black men USA 2020

Life expectancy = 67.8y

Years lost = 28.3y

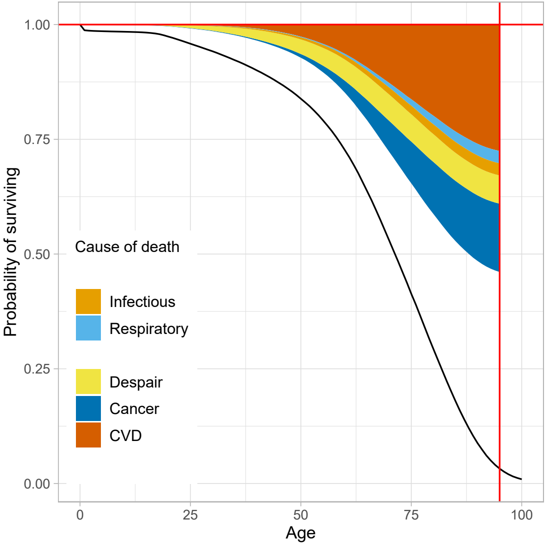
CVD 7y ⟶ 24.4%

Respiratory .6y ⟶ 2.2%

Infectious .8y ⟶ 2.8%

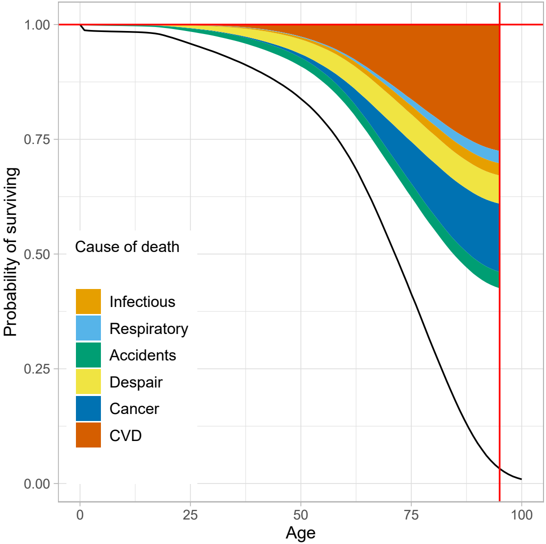
Despair 2.9y ⟶ 10.2%

Cancer 3.7y ⟶ 12.9%



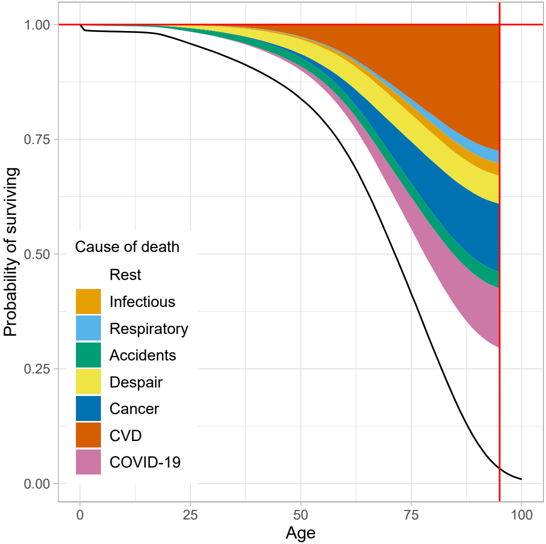
Black men USA 2020
Life expectancy = 67.8y
Years lost = 28.3y

CVD 7y ⟶ 24.4%
Respiratory .6y ⟶ 2.2%
Infectious .8y ⟶ 2.8%
Despair 2.9y ⟶ 10.2%
Cancer 3.7y ⟶ 12.9%
Accidents 1.7y ⟶ 5.9%

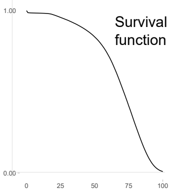


Black men USA 2020
Life expectancy = 67.8y
Years lost = 28.3y

CVD 7y ⟶ 24.4%
Respiratory .6y ⟶ 2.2%
Infectious .8y ⟶ 2.8%
Despair 2.9y ⟶ 10.2%
Cancer 3.7y ⟶ 12.9%
Accidents 1.7y ⟶ 5.9%
COVID-19 3y ⟶ 10.6%
Rest 8.7y ⟶ 31%



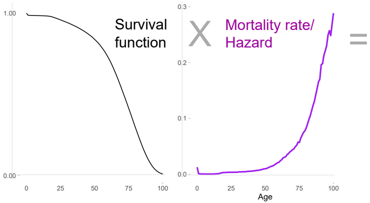
Black men USA 2020
Life expectancy = 67.8y
Years lost = 28.3y



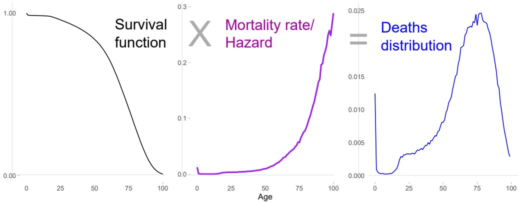
Black men USA 2020

Life expectancy = 67.8y

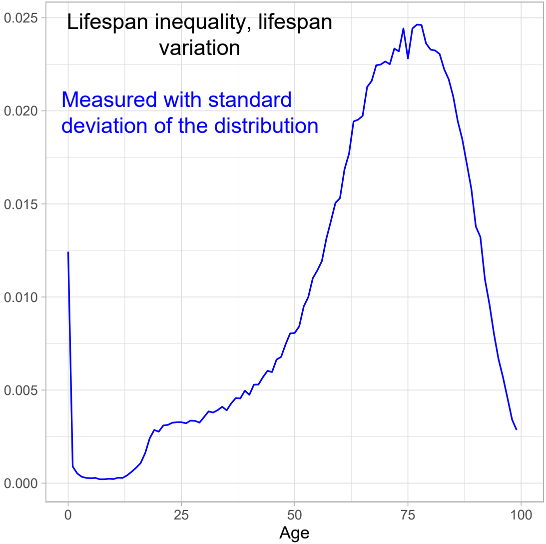
Years lost = 28.3y



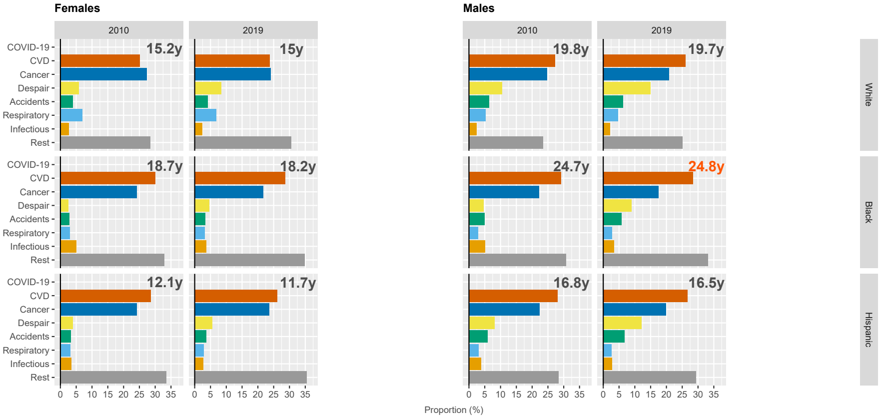
Black men USA 2020
Life expectancy = 67.8y
Years lost = 28.3y



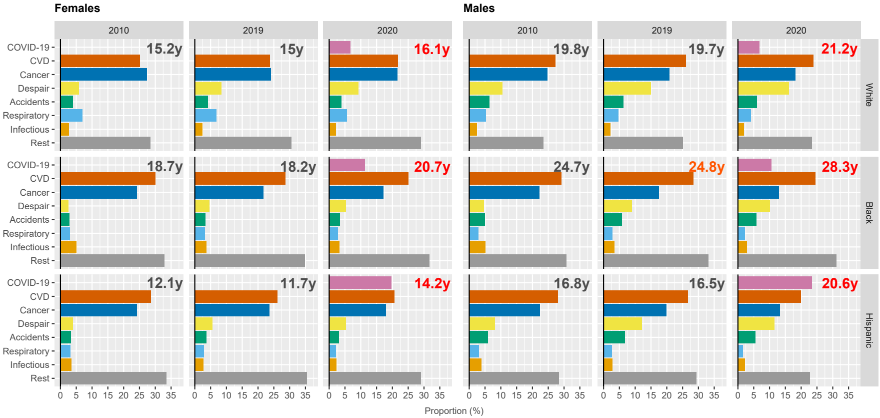
Black men USA 2020
Life expectancy $= 67.8$y
Years lost $= 28.3$y
Lifespan inequality $=$ 19.7y



Average years of life lost



Average years of life lost



Further reading

- Aburto, J. M., Tilstra, A. M., Floridi, G., and Dowd, J. B. (2022). Significant impacts of the COVID-19 pandemic on race/ethnic differences in US mortality. Proceedings of the National Academy of Sciences, 119(35), e2205813119.
- Andersen, Per Kragh, Vladimir Canudas-Romo, and Niels Keiding. Cause-specific measures of life years lost. Demographic Research 29 (2013): 1127-1152.