ADA: Demographic Methods

Monica Alexander

Week 2: Mortality

Overview

- Life tables
- Cause-deleted life tables
- Lifespan disparity
- Decomposition

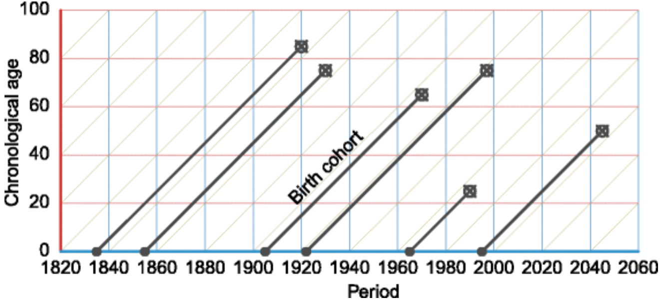
Life tables

What is a life table?

A life table describes the **survivorship by age** for a certain population.

There are different ways of describing survivorship (e.g. probability still alive, probability of dying, years of life left to live...), so a life table has many different columns.

Summarizing survivorship by age



(Riffe et al. 2017)

The life table

x = age

n = length of age interval

x	n	l_x	${}_nd_x$	${}_nq_x$	${}_np_x$	${}_na_x$	${}_nL_x$	T_x	e_x
0	1								
1	4								
5	5								
10	5								
...									
100	∞								

The life table

- how we calculate life expectancy
- used a lot by actuaries
- summary measures to compare populations
- tell us something about the implied stationary population (more later)

History

17. In the next place, whereas many persons live in great fear and apprehension of some of the more formidable and notorious diseases following; I shall only set down how many died of each: that the respective numbers, being compared with the total 229,250, those persons may the better understand the hazard they are in.

Table of notorious diseases		Table of casualties	
Apoplexy	1,306	*Bleeding*	69
Cut of the Stone	38	*Burnt*, and *Scalded*	125
Falling Sickness	74	*Drowned*	829
Dead in the streets	243	*Excessive drinking*	2
Gowt	134	*Frighted*	22
Head-Ache	51	*Grief*	279
Jaundice	998	*Hanged themselves*	222
Lethargy	67	*Killed by several accidents*	1,021
Leprosy	6	*Murdered*	86
Lunatick	158	*Poisoned*	14
Overlaid, and *Starved*	529	*Smothered*	26
Palsy	423	*Shot*	7
Rupture	201	*Starved*	51
Stone and *Strangury*,	863	*Vomiting*	136
Sciatica	5		
Sodainly	454		

Graunt, 1662 (from Smith and Keyfitz)

History

9. Whereas we have found that of 100 quick conceptions about 36 of them die before they be six years old, and that perhaps but one surviveth 76, we, having seven decades between six and 76, we sought six mean proportional numbers between 64, the remainder living at six years, and the one which survives 76, and find that the numbers following are practically near enough to the truth; for men do not die in exact proportions, nor in fractions: from whence arises this Table following:

Viz. of 100 there dies within the first six years	36	The fourth	6
The next ten years, or decade	24	The next	4
The second decade	15	The next	3
The third decade	9	The next	2
		The next	1

Graunt, 1662 (from Smith and Keyfitz)

History

10. From whence it follows, that of the said 100 conceived there remains alive at six years end 64.

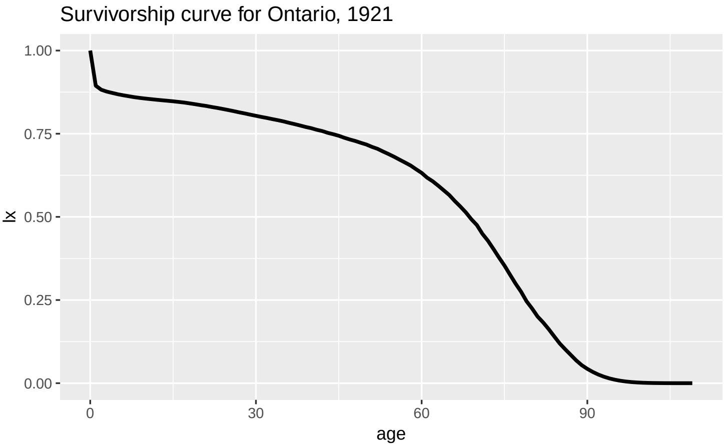
At sixteen years end	40	At fifty-six	6
At twenty-six	25	At sixty-six	3
At thirty-six	16	At seventy-six	1
At forty-six	10	At eighty	0

Graunt, 1662 (from Smith and Keyfitz)

The survivorship function l_x

$l_x =$ the number of survivors at age x

l_0 is the 'radix'. If $l_0 = 1$, then l_x is a probability of survival.



x	n	$\mathbf{l_x}$	${}_nd_x$	${}_nq_x$	${}_np_x$	${}_na_x$	${}_nL_x$	T_x	e_x
0	1	1							
1	4	0.89							
5	5	0.86							
10	5	0.85							
...	...	...							
100	∞	0.0014							

Start by thinking of survival of a cohort of people moving through time.

Deaths ${}_nd_x$

${}_nd_x$ is the number of deaths between ages x and $x + n$.

$$ {}_nd_x = l_x - l_{x+n}. $$

x	n	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	${}_nq_x$	${}_np_x$	${}_na_x$	${}_nL_x$	T_x	e_x
0	1	1	0.11						
1	4	0.89	0.03						
5	5	0.86	0.01						
10	5	0.85	...						
...	...	...	...						
100	∞	0.0014	0.0014						

Note that everyone who survived to the last age group must die (*memento mori*)

Probability of death ${}_nq_x$

${}_nq_x$ is the probability of dying between ages x and $x + n$.

$$ {}_nq_x = \frac{{}_nd_x}{l_x} $$

Note that this is conditional on having survived to age x.

x	n	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	$\mathbf{{}_nq_x}$	${}_np_x$	${}_na_x$	${}_nL_x$	T_x	e_x
0	1	1	0.11	0.11					
1	4	0.89	0.03	0.034					
5	5	0.86	0.01	0.011					
10	5	0.85	...	...					
...	...	...	...	...					
100	∞	0.0014	0.0014	1					

Again, note last interval!

Probability of surviving ${}_np_x$

${}_nq_x$ is the probability of surviving between ages x and $x + n$.

$$ {}_np_x = 1 -_n q_x. $$

- Note again: conditional probability.
- How else could we calculate the probability of surviving?

x	n	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	$\mathbf{{}_nq_x}$	$\mathbf{{}_np_x}$	${}_na_x$	${}_nL_x$	T_x	e_x
0	1	1	0.11	0.11	0.89				
1	4	0.89	0.03	0.034	0.97				
5	5	0.86	0.01	0.011	0.989				
10	5	0.85	...	...	...				
...	...	...	...	...	...				
100	∞	0.0014	0.0014	1	0				

Average years lived ${}_na_x$

${}_na_x$ is the number of years lived by those who died between ages x and $x + n$.

(pretend for now we observe the lifelines of all people so would have this info)

x	n	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	$\mathbf{{}_nq_x}$	$\mathbf{{}_np_x}$	$\mathbf{{}_na_x}$	${}_nL_x$	T_x	e_x
0	1	1	0.11	0.11	0.89	0.3			
1	4	0.89	0.03	0.034	0.97	1.5			
5	5	0.86	0.01	0.011	0.989	2.5			
10	5	0.85	...	...	...	...			
...	...	...	...	...	...	...			
100	∞	0.0014	0.0014	1	0	1.5			

Person-years lived ${}_nL_x$

${}_nL_x$ is the number of person-years lived between ages x and $x + n$.

Total PYL = PYL by those who survived + PYL by those who died.

$${}_nL_x = n \cdot l_{x+n} +_n a_x \cdot_n d_x.$$

Note: last age interval, no survivors, so we just have ${}_na_x \cdot {}_nd_x$

x	n	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	$\mathbf{{}_nq_x}$	$\mathbf{{}_np_x}$	$\mathbf{{}_na_x}$	$\mathbf{{}_nL_x}$	T_x	e_x
0	1	1	0.11	0.11	0.89	0.3	0.92		
1	4	0.89	0.03	0.034	0.97	1.5	3.485		
5	5	0.86	0.01	0.011	0.989	2.5	4.275		
10	5	0.85	...	...	...	...	...		
...	...	...	...	...	...	...	...		
100	∞	0.0014	0.0014	1	0	1.5	0.0021		

- What's the max that ${}_nL_x$ can be?
- How does ${}_nL_x$ relate to the survival curve?
- So what's the continuous version (L_x)?

Person-years lived above age x T_x

T_x is the number of person-years lived above age x.

It is the sum of all the ${}_nL_x$'s at above age x, i.e.

$$T_x = \sum_{x}^{\infty} {}_nL_x$$

x	n	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	$\mathbf{{}_nq_x}$	$\mathbf{{}_np_x}$	$\mathbf{{}_na_x}$	$\mathbf{{}_nL_x}$	$\mathbf{T_x}$	e_x
0	1	1	0.11	0.11	0.89	0.3	0.92	58.01	
1	4	0.89	0.03	0.034	0.97	1.5	3.485	57.09	
5	5	0.86	0.01	0.011	0.989	2.5	4.275	53.61	
10	5	0.85	...	...	...	...	...	...	
...	...	...	...	...	...	...	...	...	
100	∞	0.0014	0.0014	1	0	1.5	0.0021	0.0021	

- How does T_x relate to the survival curve?
- So what's the continuous version of T_x?

Life expectancy e_x

e_x is the average number of remaining years of life for those who reach age x.

$$e_x = \frac{T_x}{l_x}$$

- you may be familiar with e_0, life expectancy at birth

x	n	$\mathbf{l_x}$	$\mathbf{{}_nd_x}$	$\mathbf{{}_nq_x}$	$\mathbf{{}_np_x}$	$\mathbf{{}_na_x}$	$\mathbf{{}_nL_x}$	$\mathbf{T_x}$	$\mathbf{e_x}$
0	1	1	0.11	0.11	0.89	0.3	0.92	58.01	58.01
1	4	0.89	0.03	0.034	0.97	1.5	3.485	57.09	64.1
5	5	0.86	0.01	0.011	0.989	2.5	4.275	53.61	62.4
10	5	0.85	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
100	∞	0.0014	0.0014	1	0	1.5	0.0021	0.0021	1.5

- Does life expectancy have to monotonically decrease over age?

More on survival and probabilities

Example:

- For Kenyan children born in 2005, ${}_1q_0 = 0.053$ and ${}_4q_1 = 0.26$. What is ${}_5q_0$?

Survival probabilities multiply

Converting between 1- and 5-year

- Life table above was (mostly 5-year age groups)
- What if we wanted to convert these to 1-year age groups?
- E.g. For Taiwan in 1978, ${}_5q_{25} = 0.0247$. What is ${}_1q_{27}$?

Assume probability of surviving is constant in the absence of any other information.

Period life tables

Creating a synthetic cohort

- So far, we have assumed we are dealing with cohort data
- But cohort data are not available in a timely manner
- Create a synthetic cohort using mortality rates observed in a period
- Construct life table the same way
- Interpretation is different: e.g. life expectancy at birth is the number of years a newborn could expect to life if all age-specific mortality rates stayed the same in future

${}_nq_x$ conversion

- We observe period mortality rates ${}_nM_x$
- Recall from last week, these are number of deaths / person years lived. So in life table notation:

$$ {}_nM_x = \frac{{}_ndx}{{}_nLx} $$

- We want ${}_nq_x$ (then calculate a whole life table)
- Use conversion formula:

$$ {}_nq_x = \frac{n \cdot {}_nM_x}{1 + (n - {}_na_x) \cdot {}_nM_x} $$

then all other columns can be derived as before, expect. . . .

Average years lived ${}_na_x$

${}_na_x$ is the number of years lived by those who died between ages x and $x + n$.

- If we have individual lifelines, we can work this out
- We don't for period data. **For most age groups** using

$$ {}_na_x = n/2 $$

is a fine approximation.

For which age groups would it not be fine?

Average years lived ${}_na_x$

Somewhat rough approximations:

- First age group:

$$ {}_1a_0 = 0.07 + 1.7\,{}_1M_0 $$

- Second age group:

$$ {}_4a_1 = 1.5 $$

- Last age group:

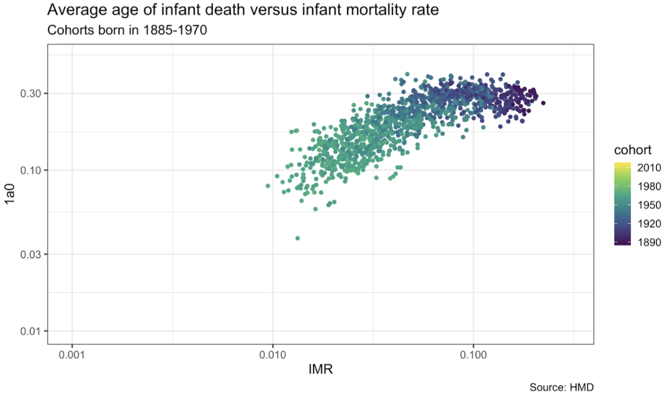
$$ {}_\infty a_\omega = 1/{}_\infty M_\omega $$

where ${}_\infty M_\omega$ is the age-specific mortality rate for the last age interval (check units).

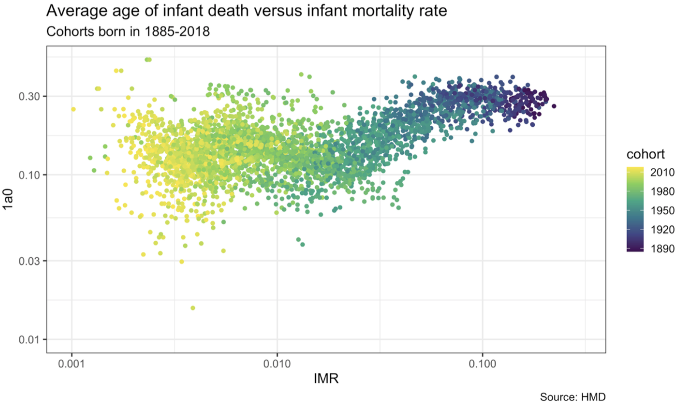
A note on ${}_1a_0$

$$ {}_1a_0 = 0.07 + 1.7\,{}_1M_0 $$

Where does this come from?



. . . oh no

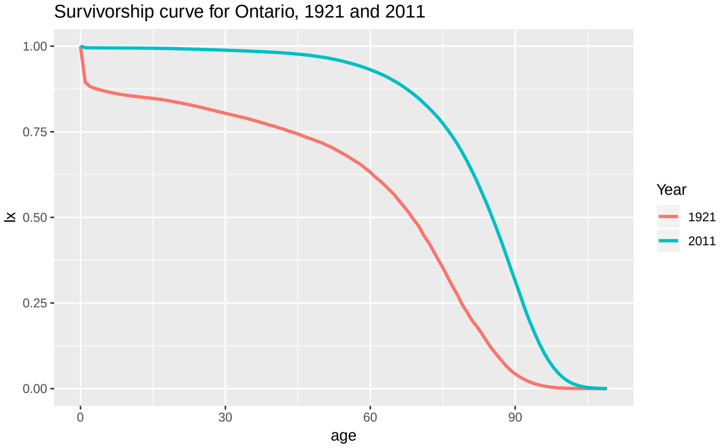


Average age of infant death versus infant mortality rate
Cohorts born in 1885-2018
0.30
0.10
0.03
0.01
1a0
0.001
0.010
0.100
IMR
cohort
2010
1980
1950
1920
1890
Source: HMD

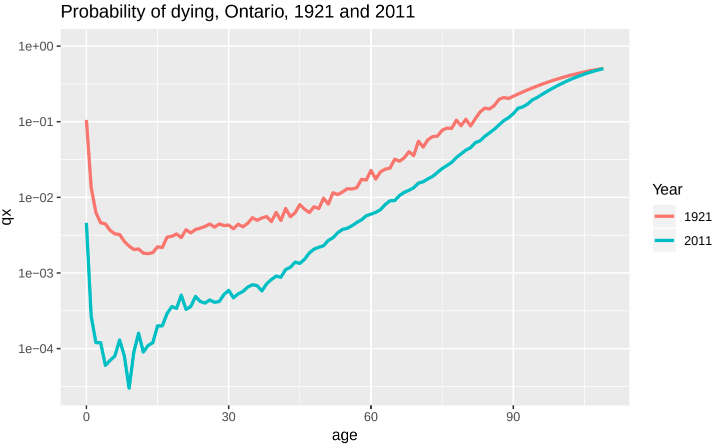
Competing effects

- When mortality is low, the relationship between average age of infant death and infant mortality becomes less clear
- Importance of level of premature births (which often have very young ages of death)
- Read more: Alexander and Root 2022

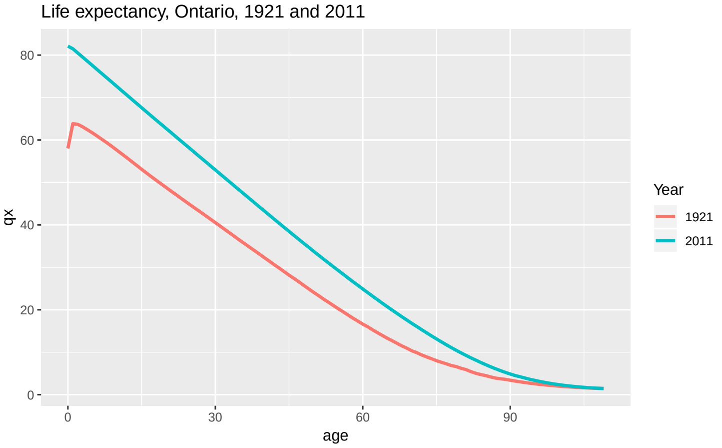
Characteristic shapes



Characteristic shapes



Characteristic shapes



Period life expectancy and interpretation

THE HILL

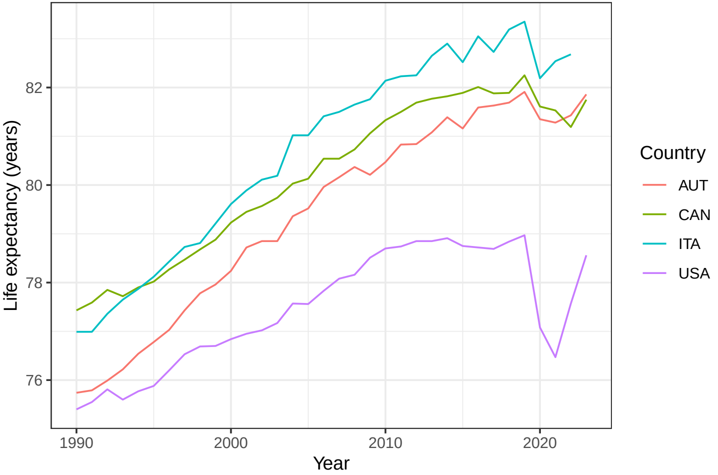
(NEXSTAR) – As Ben Franklin once famously said, "In this world nothing can be said to be certain, except death and taxes." Eventually, we all die, but that may appear sooner than others for a variety of reasons, and where you live in the country may play a role in that as well.

A new study from MoneyGeek researched healthcare across the U.S., with one of its variables being the average life expectancy for every state in the country.

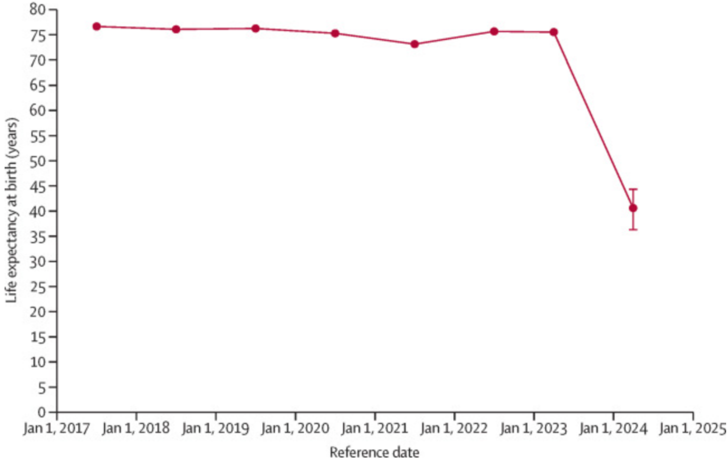
What they found was that the average life expectancy for a U.S. resident was 76.4 years old. And that Americans in general pass away sometime in their 70s on average.

The state where residents tend to live the longest is Hawaii, with an average life expectancy of 79.9.

Period life expectancy and interpretation



Period life expectancy and interpretation



From Guilot et al (2025)

Multiple decrements and cause-deleted life tables

Single versus multi-decrements

- So far we have only considered decrements over age due to all deaths
- But people die of different things, and age patterns of causes of death are important
- Can extend the single-decrement life table into a series of **multiple decrement life tables**

Not just mortality! Life table approach could be used to marriage (divorce, widowhood), contraception use, etc.

Multiple decrements

Say we observe mortality rates by cause.

${}_nM_x^i$ is the mortality due to cause i.

${}_nM_x^{-i}$ is the mortality due to all causes apart from i.

Note the conversion formula, thus the dependence of probabilities of dying:

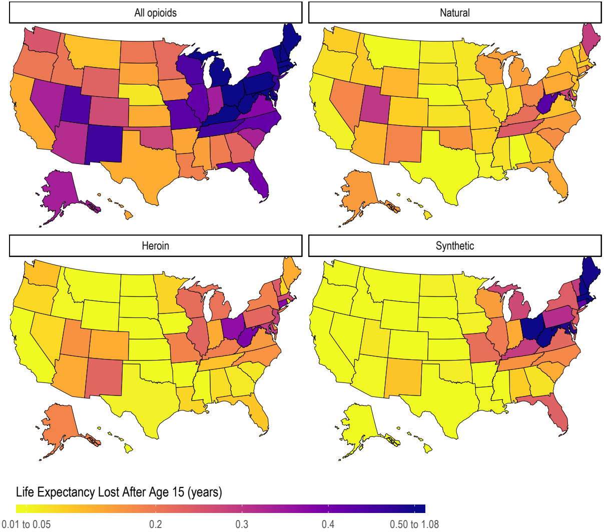
$$ {}_nq_x^i = \frac{n \cdot {}_nM_x^i}{1 + (n - {}_na_x) \cdot {}_nM_x} $$

Also note we don't have to use this:

$$ {}_nq_x^i = {}_nq_x \cdot \frac{{}_nMx^i}{{}_nMx} = {}_nq_x \cdot \frac{{}_nDx^i}{{}_nDx} $$

Use this technique to create **cause-deleted life tables**, looking at the hypothetical situation where we omit all deaths due to one cause but everything else stays the same.

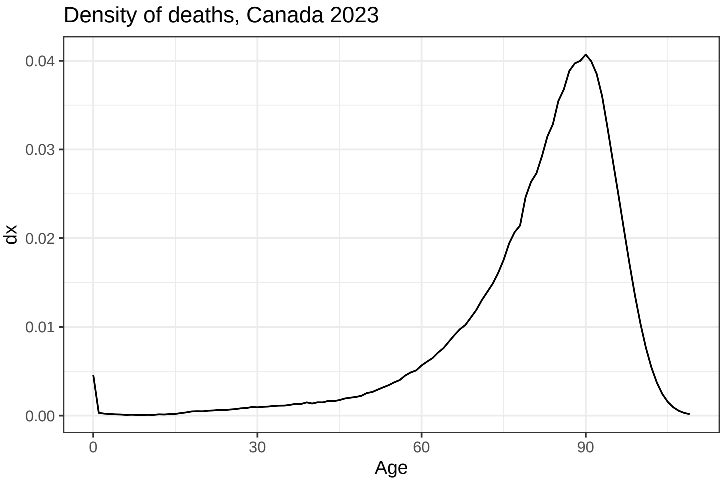
Life expectancy lost due to opioids



More here: https://sanjaybasu.shinyapps.io/opioid_geographic/

Lifespan disparity

Death density dunction d_x



- This is a pdf, $\int d_x = 1$
- What is e_0 in relation to this?
- What else might be care about?

Lifespan disparity

- "Years of life lost to death"

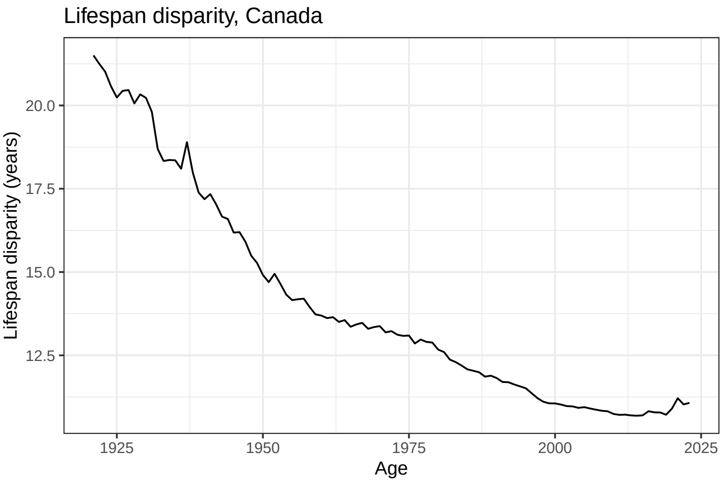
$$e^{\dagger} = \sum_{x} {}_{n}d_{x} e_{x}$$

or in continuous form

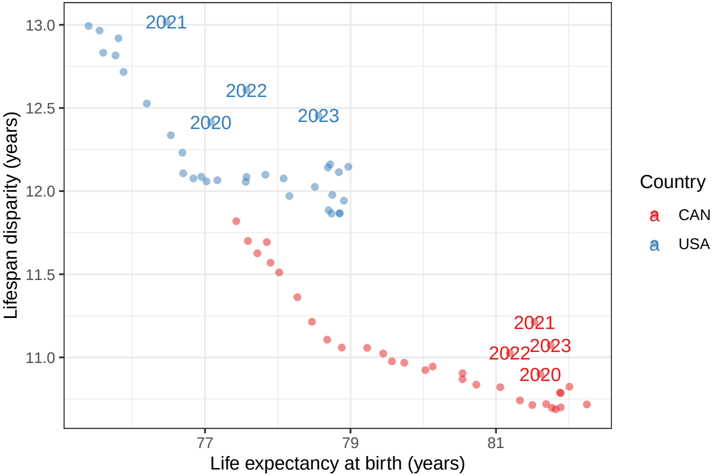
$$e^{\dagger} = \int_{x} d(x) e(x) \mathrm{d}x$$

- Tells you something about the spread of ages of death in a population

Lifespan disparity in Canada



Relationship with life expectancy



Decomposition

Standardization, revisited

- CDR in Canada: 7.9 per 1,000 people
- CDR in Kenya: 7.2 per 1,000 people

Recall: standardization was controlling for differences in age structure

Using Canada's population:

- Age-standardized mortality rate in Canada: 7.9 per 1,000 people
- Age-standardized mortality rate in Kenya: 22 per 1,000 people

Decomposition

- Decomposition is a technique that allows for quantifying the impact of age structure (and other characteristics)
- Asking the question: how much is a difference between death rates (or life expectancy, etc) due to age structure versus differences in mortality risk?

Kitagawa decomposition

Write the difference between two CDR rates as

$$\Delta = CDR^B - CDR^A = \sum_i C_i^B \cdot M_i^B - \sum_i C_i^A \cdot M_i^A$$

- C_i is age structure (proportion of population in age group i)
- M_i is age-specific mortality rate

Kitagawa decomposition

Can rewrite this as

$$\Delta = \sum_i \left(C_i^B - C_i^A\right) \cdot \left[\frac{M_i^B + M_i^A}{2}\right] + \sum_i \left(M_i^B - M_i^A\right) \cdot \left[\frac{C_i^A + C_i^B}{2}\right]$$

$$\begin{aligned}
&= \begin{array}{l}\text{difference in age}\\ \text{composition}\end{array} \cdot \left[\begin{array}{l}\text{weighted by average}\\ \text{age-specific mortality}\end{array}\right] \\
&+ \begin{array}{l}\text{difference in rate}\\ \text{schedules}\end{array} \cdot \left[\begin{array}{l}\text{weighted by}\\ \text{average age}\\ \text{composition}\end{array}\right] \\
&= \begin{array}{c}\text{contribution of age compositional}\\ \text{differences to } \Delta\end{array} + \begin{array}{c}\text{contribution of rate schedule}\\ \text{differences to } \Delta\end{array}
\end{aligned}$$

Kitagawa decomposition

$$\Delta = \sum_i \left(C_i^B - C_i^A \right) \cdot \left[\frac{M_i^B + M_i^A}{2} \right] + \sum_i \left(M_i^B - M_i^A \right) \cdot \left[\frac{C_i^A + C_i^B}{2} \right]$$

- Decompose into age differences and mortality differences
- Evelyn Kitagawa (1955)

Decomposition of Canada versus Kenya

- CDR in Canada: 7.9 per 1,000 people
- CDR in Kenya: 7.2 per 1,000 people

$$\Delta = CDR_K - CDR_C = -0.000724$$

- Age contribution: -0.0107
- Rate contribution: 0.0099

Decomposition in the wild

Home | JAMA Network Open | Vol. 6, No. 5

Original Investigation | Infectious Diseases

COVID-19 Mortality by Race and Ethnicity in US Metropolitan and Nonmetropolitan Areas, March 2020 to February 2022

Dielle J. Lundberg, MPH[1,2]; Elizabeth Wrigley-Field, PhD[3,4]; Ahyoung Cho, MPP[5,6] ; Rafeya Raquib, MS[1]; Elaine O. Nsoesie, PhD, MS[1,5]; Eugenio Paglino, MSc[7,8]; Ruijia Chen, ScD[9]; Mathew V. Kiang, ScD[10]; Alicia R. Riley, PhD[11]; Yea-Hung Chen, PhD[9]; Marie-Laure Charpignon, MSc[12]; Katherine Hempstead, PhD[13]; Samuel H. Preston, PhD[7,8]; Irma T. Elo, PhD[7,8]; M. Maria Glymour, ScD, MS[9]; Andrew C. Stokes, PhD[1]

» Author Affiliations | Article Information

Key Points

Question Why did racial and ethnic disparities in COVID-19 mortality in the US decrease in the Omicron wave compared with the initial wave of the pandemic?

Findings In this cross-sectional study of 977 018 adults who died from COVID-19, 60.3% of the national decrease in disparities in COVID-19 mortality for non-Hispanic Black compared with non-Hispanic White adults between the initial and Omicron waves could be explained by increases in mortality among non-Hispanic White adults and shifts in mortality to nonmetropolitan areas, where more non-Hispanic White adults reside.

Meaning This study found that racial and ethnic disparities in COVID-19 mortality decreased nationally for some groups during the first 2 years of the pandemic, but this decrease was mostly explained by increases in mortality for non-Hispanic White adults and changes in pandemic geography.

Statistical Analysis

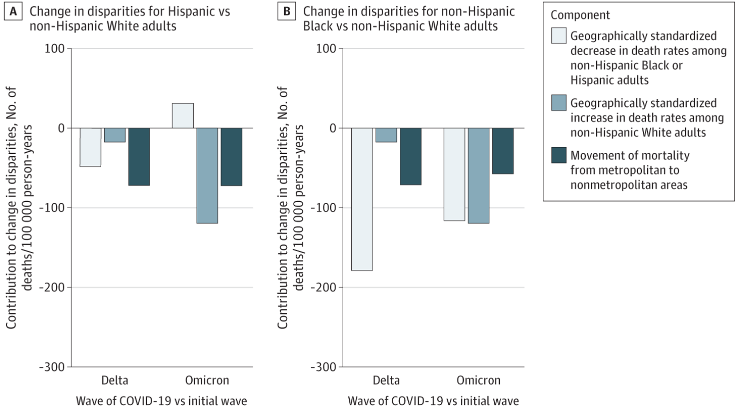
We decomposed the national change between the initial wave of the pandemic and the Omicron wave in the absolute disparity in age-standardized COVID-19 death rates among non-Hispanic Black compared with non-Hispanic White adults. We also decomposed the national change in disparities for Hispanic compared with non-Hispanic White adults and compared the initial wave with the second, Alpha, and Delta waves. We sought to understand the contribution of the following 4 components to national changes in disparities:

1. The geographically standardized decrease in death rates among non-Hispanic Black or Hispanic adults in a hypothetical population in which the non-Hispanic Black or Hispanic population had the geographic distribution of a standard population.

2. The geographically standardized increase in death rates among non-Hispanic White adults in a hypothetical population in which the non-Hispanic White population had the geographic distribution of a standard population.

3. The change in mortality outcomes associated with shifts in where deaths occurred from metropolitan to nonmetropolitan areas, where more non-Hispanic White adults reside relative to the national geographic distribution (ie, the differential outcomes associated with changes in racial or ethnic–specific and geography-specific mortality rates because of the actual geographic distribution of the US population, with non-Hispanic White adults overrepresented in nonmetropolitan areas).

4. The change in mortality outcomes associated with shifts in the racial and ethnic population composition in metropolitan and nonmetropolitan areas, which we expected to be minor.



A Change in disparities for Hispanic vs non-Hispanic White adults
B Change in disparities for non-Hispanic Black vs non-Hispanic White adults
Contribution to change in disparities, No. of deaths/100 000 person-years
100
0
-100
-200
-300
Delta
Omicron
Wave of COVID-19 vs initial wave
Component
Geographically standardized decrease in death rates among non-Hispanic Black or Hispanic adults
Geographically standardized increase in death rates among non-Hispanic White adults
Movement of mortality from metropolitan to nonmetropolitan areas

The change between periods in racial and ethnic disparities in mortality is equivalent to the racial and ethnic disparity in the change in mortality:

$$(\overline{m}_{b,2} - \overline{m}_{w,2}) - (\overline{m}_{b,1} - \overline{m}_{w,1}) = (\overline{m}_{b,2} - \overline{m}_{b,1}) - (\overline{m}_{w,1} - \overline{m}_{w,1}) \tag{1}$$

Due to that equivalence, we proceed by geographically decomposing the changes in race and ethnicity-specific aggregate mortality $(\overline{m}_{i,2} - \overline{m}_{i,1})$ before plugging those components into Eq. (1) to produce our final decomposition equation.

The change in race and ethnicity-specific aggregate mortality can be geographically decomposed as:

$$\begin{aligned}
\overline{m}_{i,2} - \overline{m}_{i,1} &= \sum_k \left[c_{i,1,k}(m_{i,2,k} - m_{i,1,k}) + (c_{i,2,k} - c_{i,1,k})m_{i,2,k}\right] \\
&= \sum_k \left[(c_k^{st} + \tilde{c}_{i,1,k})(m_{i,2,k} - m_{i,1,k}) + \left(c_k^{st} + \tilde{c}_{i,2,k} - (c_k^{st} + \tilde{c}_{i,1,k})\right) m_{i,2,k}\right] \\
&= \sum_k \left[c_k^{st}(m_{i,2,k} - m_{i,1,k}) + \tilde{c}_{i,1,k}(m_{i,2,k} - m_{i,1,k}) + (\tilde{c}_{i,2,k} - \tilde{c}_{i,1,k})m_{i,2,k}\right]
\end{aligned}$$

More on decomposition next week

- Life expectancy
- Extension to cause of death

But now: Lab