Lexis diagram and system of demographic indicators

16. 11. 2022, 23. 11. 2022

Bety Ukolova – ukolovae@natur.cuni.cz

Lexis diagram: exercise

a. Deaths at completed age 2 years, from birth cohort 1989

b. Deaths at reached age 2 years from birth cohort 1989

c. Death at completed age 2 years in 1991.

d. Deaths at completed age 11 years during 2002–2003

e. Deaths at reached age 11 years during 2002 and 2003

f. Deaths at completed age 2 years in 1991 from birth cohort 1989

g. Deaths at completed age 11 and 12 years in 2002–2003

h. Persons living at exact age 36 in years 2002–2003

i. Persons living at completed age 15–17 at 1. 7. 1998.

Demographic indicators (I)

Absolute measures

- Components of population change ($P_{t+1} = P_t + N_t - D_t + I_t - E_t$)

Relative measures

- Relative frequencies, proportions (extensive)
- Rates and probabilities (intensive)

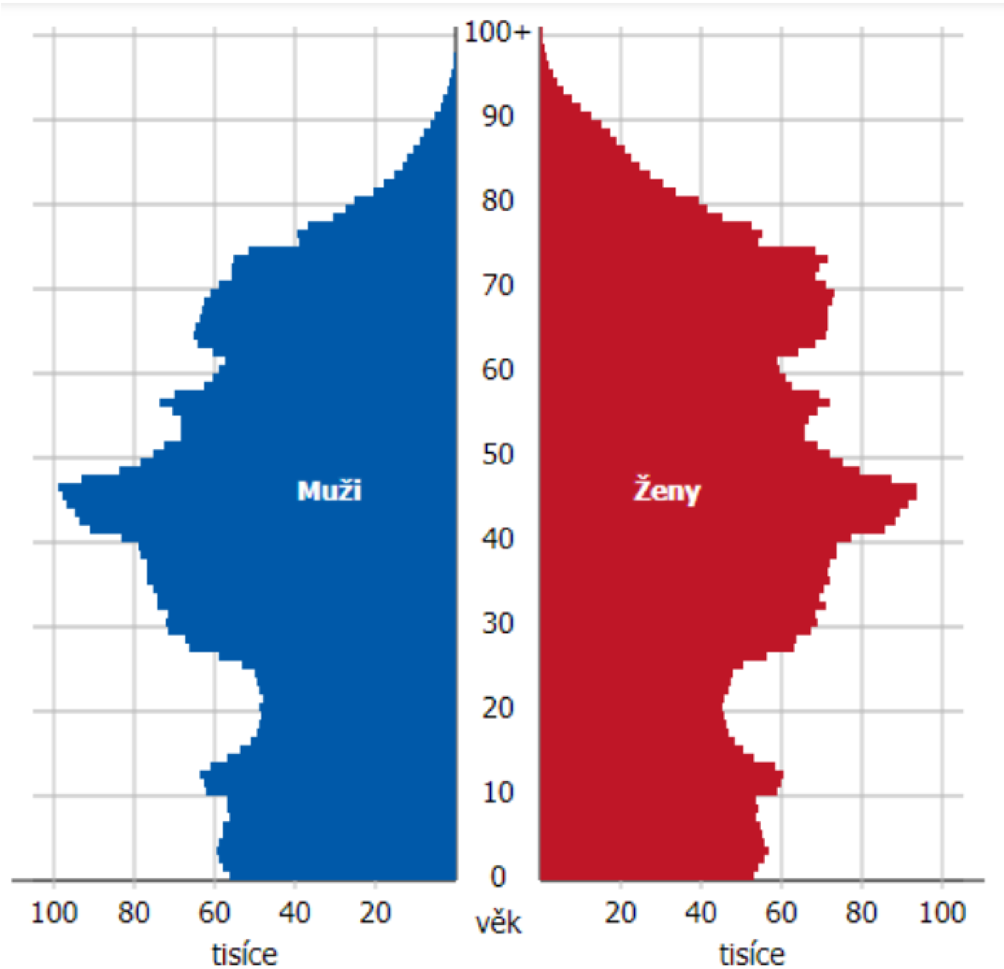
Ratios

- Indices

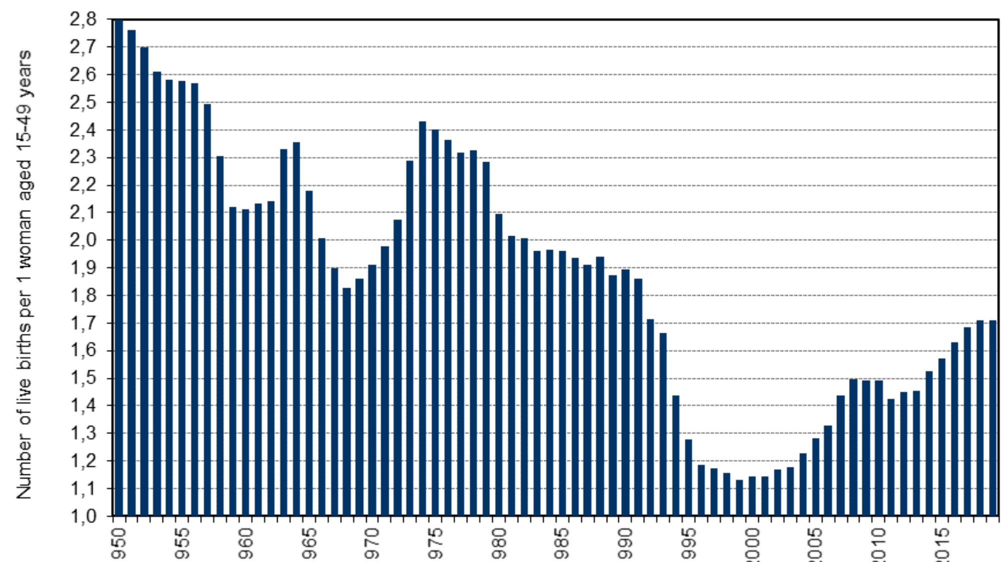
Demographic indicators (I) – drawbacks

Drawbacks examples: Analysing fertility with absolute numbers, sever covid and vaccination thru relative measures, age specific index of abortion

Age structure of Czechia at 1. 1. 2021



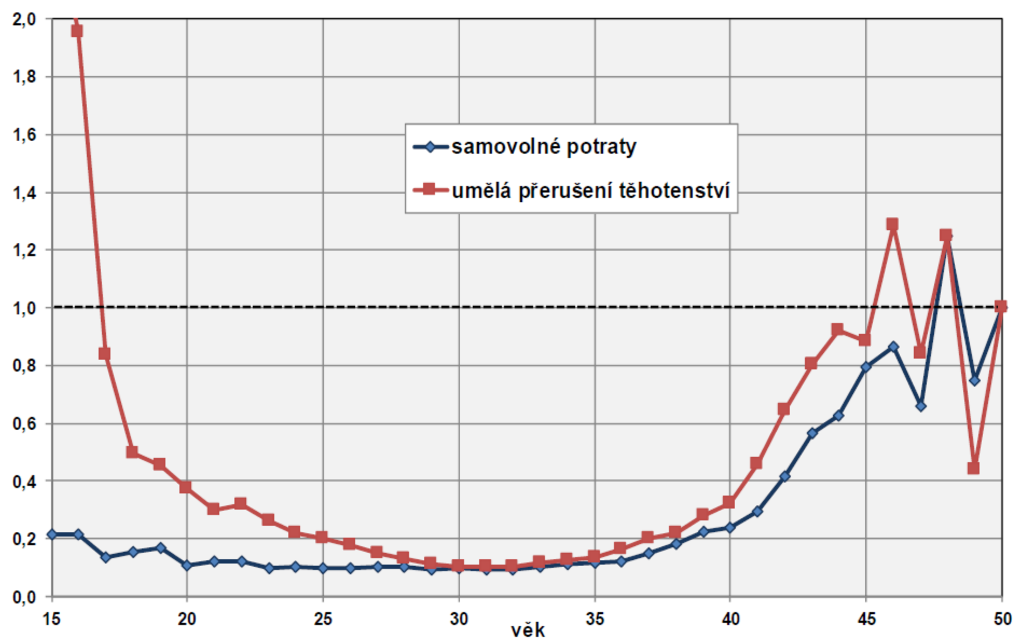
Total fertility rate (average number of children per woman), Czechia, 1950–2019



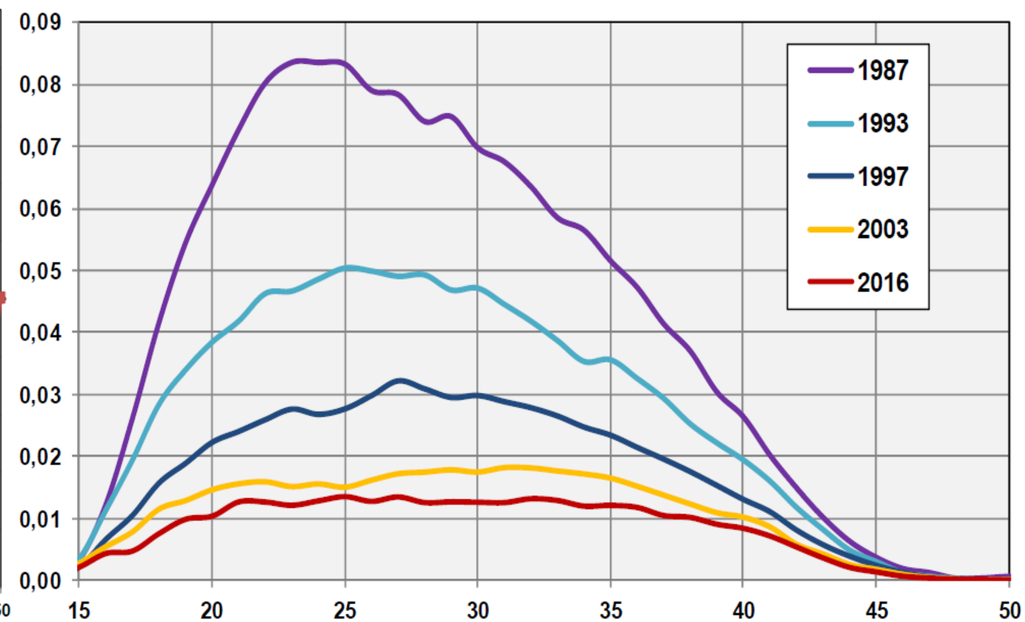
Demographic indicators (I) – drawbacks

Drawbacks examples: Analysing fertility with absolute numbers, sever covid and vaccination thru relative measures, age specific index of abortion

Age specific index of abortion, Czechia, 2016



Age specific abortion rates, Czechia, 2016



Demographic indicators (II)

Crude

- Crude death rate, crude birth rate…

General

- General fertility rate…

Specific

- Age specific death rate…

Differential

- Age specific death rates by educational attainment, by parity…

Analytical (synthetic)

- Total fertility rate, life expectancy…

Standardized, adjusted, corrected

- Standardized death rate…

Demographic indicators (III)

Provisional
Final
Revised
Corrected

$$\frac{numerator}{denominator}$$

Rates and probabilities

= Two major categories of measures used by demographers

$$rates = \frac{Number\ of\ occurences\ of\ events}{Exposure}$$

Exposure ~ Average size of a population during a given period ~ Number of person years exposure to risk ~ Mid-year population at risk

$$probabilities = \frac{Number\ of\ occurences\ of\ events}{Population\ at\ risk\ at\ the\ beginning\ of\ the\ interval}$$

$Population\ at\ risk\ at\ the\ beginning\ of\ the\ interval$ *~ Population on 1st of January*

Age specific mortality rate is denoted as u_x or m_x. Age specific probability is denoted as q_x. From that, these measures are sometimes called *m-rates* and *q-rates*, even tough **both of them are not rates**.

Probabilities are sometimes called „quotients“ **[2]**.

Rates are central death rates.

Age specific mortality rates

- Events are either evitable (**avoidable**; birth of a child) or inevitable (**unavoidable**; death). Events can also be **repeatable** (birth of a child) or **unrepeatable** (death). These differences have implication for computations of rates and probabilities.
- If the events are avoidable, we distinguish between

  Rates of the first type (conditional rates, occurrence-exposure rates)

  Rates of the second type (unconditional rates, reduced rates, frequencies)

Age specific unconditional fertility rate: $$f_x^i = \frac{N_x^i}{P_x^F}$$

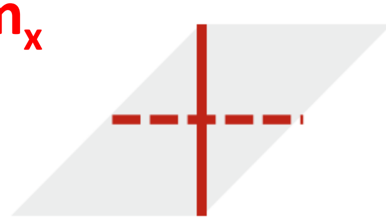
Age specific conditional fertility rate: $$f_x^i = \frac{N_x^i}{P_x^{F,i-1}}$$

i	birth order of a child
F	female
x	age
N	number of birth
P	population exposure

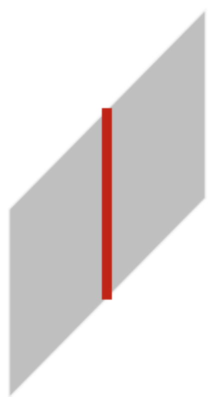
Age specific mortality rates

- Mortality rates can also be distinguished according to the sets of events they belong to.

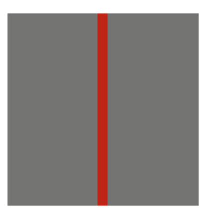
1st primary sets ${}_{t-t+n}m_x$
- Rates by horizontal parallelograms
- Rates by age completed and cohort



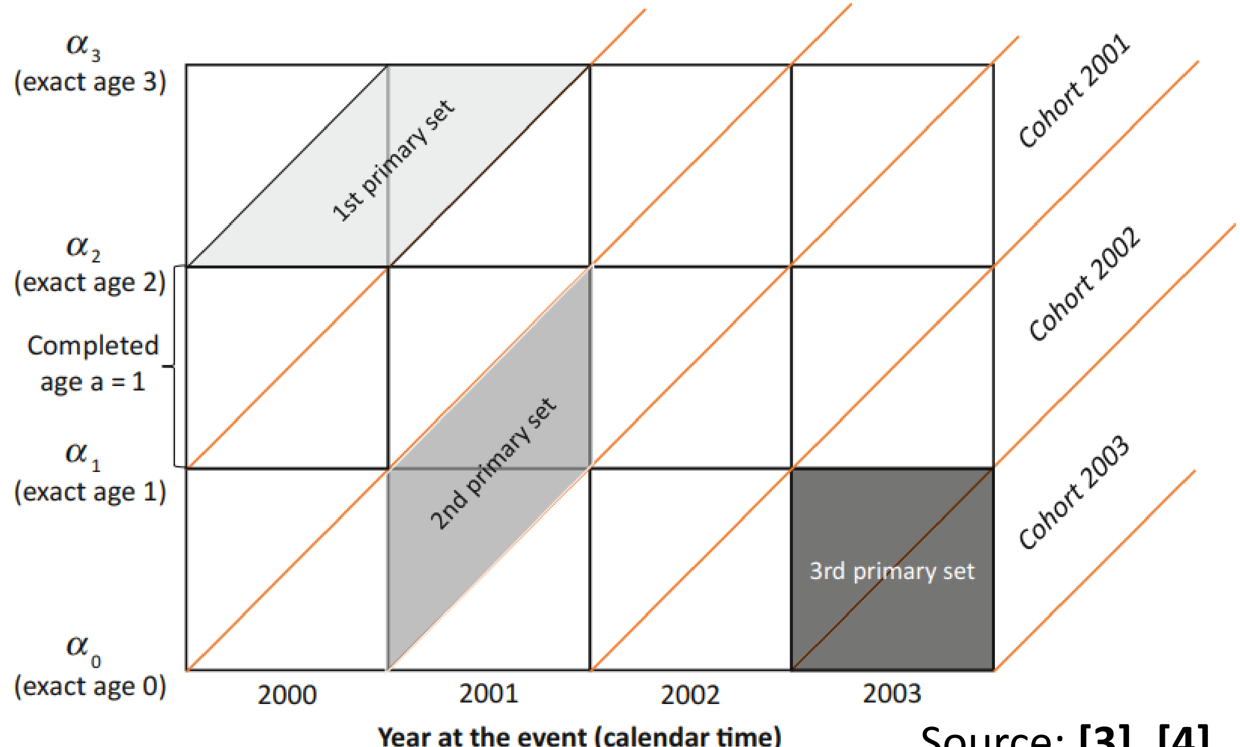
2nd primary sets ${}_{t}m_{x-x+n}$
- Rates by vertical parallelograms
- Rates by age reached and a cohort



3rd primary sets ${}_{t}m_x$
- Rates by squares
- Rates by age completed and a year (period)



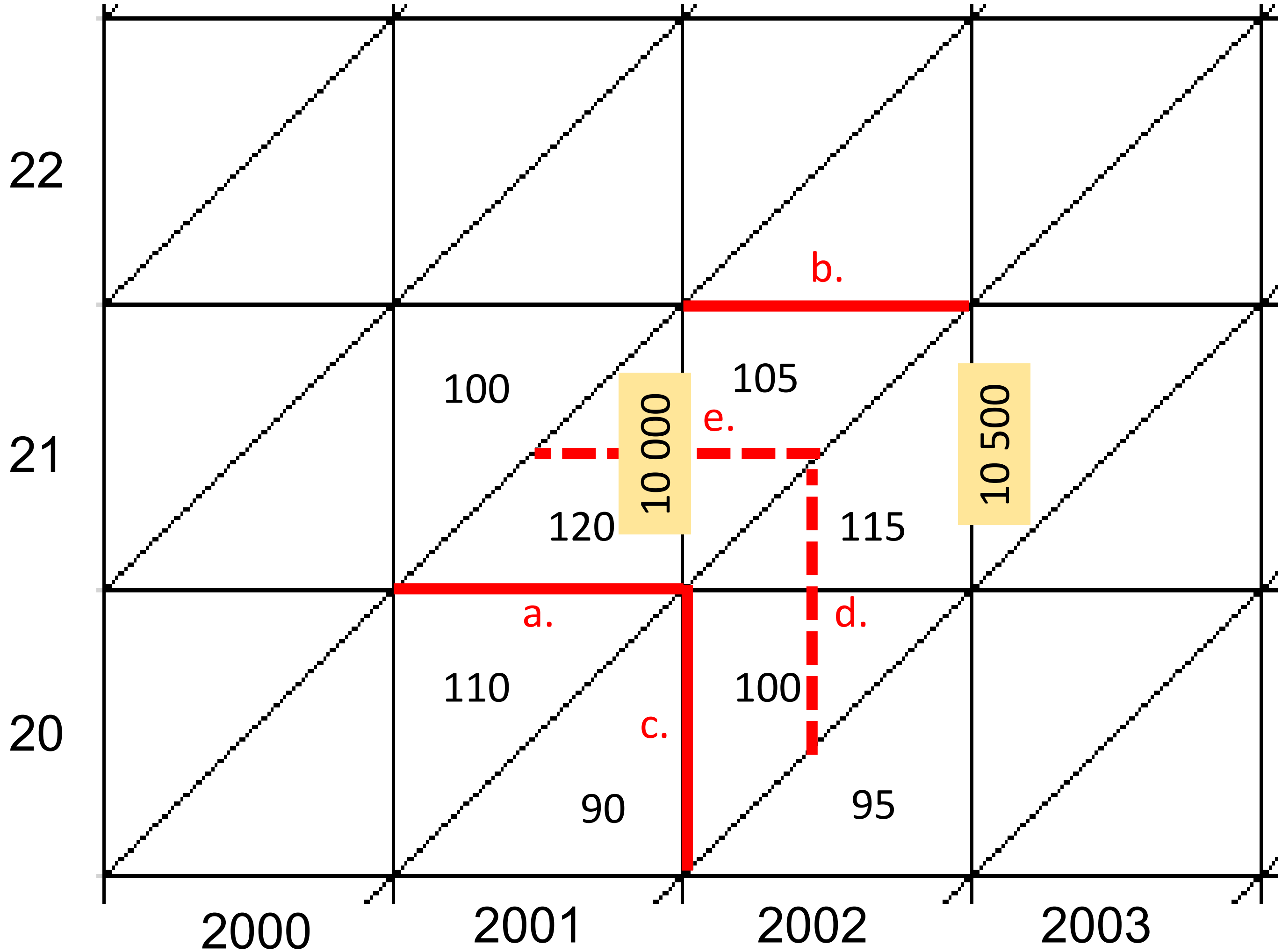
To calculate the rates, we need to sum events located in right elementary sets of events and consequently, estimate the population exposure.



Source: **[3]**, **[4]**

Exercise 2

Estimate a., b., c., d. and e.



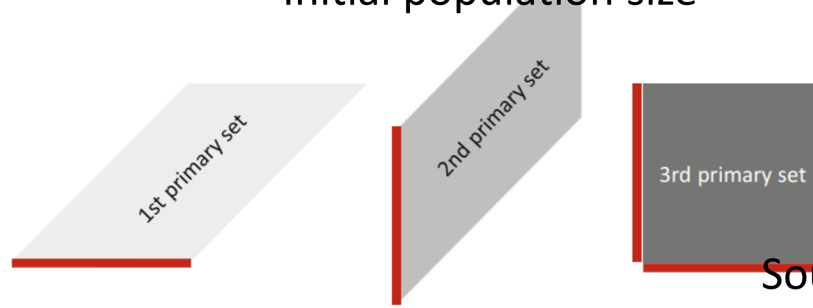
Age specific probabilities of death

- Age specific probabilities of death can be computed either directly or indirectly.
- The **indirect estimation** relies on derivation of probabilities from mortality rates.
- The **direct estimation** relies in direct computations from data (dividing events by population at the beginning of the interval, in other words by the initial population).

q_x
- Indirect method
  - Linear method
  - Exponential method
- Direct method

Today we will only estimate the probabilities with direct method, however, this does not mean, that this approach is more or less important than the other one.

Initial population size



Source: **[3]**

Direct calculation of rates and probabilities: example

Draw following data to Lexis diagram:

Age	Birth cohort	$_{1996}D_x$	Birth cohort	$_{1997}D_x$	$_{1.1.1996}P_x$	$_{1.1.1997}P_x$
50	1946	**351**	1947	360	**67 125**	**82 567**
	1945	**285**	1946	**349**		
51	1945	**307**	1946	374	**70 224**	**66 609**
	1944	342	1945	319		

another
cohort D *attribute*
period *age*

Calculate age specific mortality rates a probabilities of dying:

$_{1996-1997}m_{50}$

$_{1996}m_{50-51}$

$_{1996}m_{51}$

$_{1996-1997}q_{50}$

$_{1996}q_{50-51}$

Can be directly estimated from Lexis diagram

$_{1996}q_{51}$

Can be directly estimated from Lexis diagram and a formulae has to be applied

Remember (in 3rd sets of events)

$$ {}^{c-(c+1)}_{t}q_x = (q' + q'') - (q' \times q'') $$

$$ q' = \frac{Deaths\ in\ upper\ elementary\ set\ of\ events}{Population\ on\ 1.1.\ at\ given\ period} $$

$$ q'' = \frac{Deaths\ in\ lower\ elementary\ set\ of\ events}{Population\ at\ given\ exact\ age} $$

Direct calculation of rates and probabilities: example

Draw following data to Lexis diagram:

Age	Birth cohort	$_{1996}D_x$	Birth cohort	$_{1997}D_x$	$_{1.1.1996}P_x$	$_{1.1.1997}P_x$
50	1946	**351**	1947	360	**67 125**	**82 567**
	1945	**285**	1946	**349**		
51	1945	**307**	1946	374	**70 224**	**66 609**
	1944	342	1945	319		

$$ {}^{\text{another}}_{\text{cohort}}{}_{\text{period}} D^{\text{attribute}}_{\text{age}} $$

Calculate age specific mortality rates a probabilities of dying:

$_{1996\text{–}1997}m_{50}$

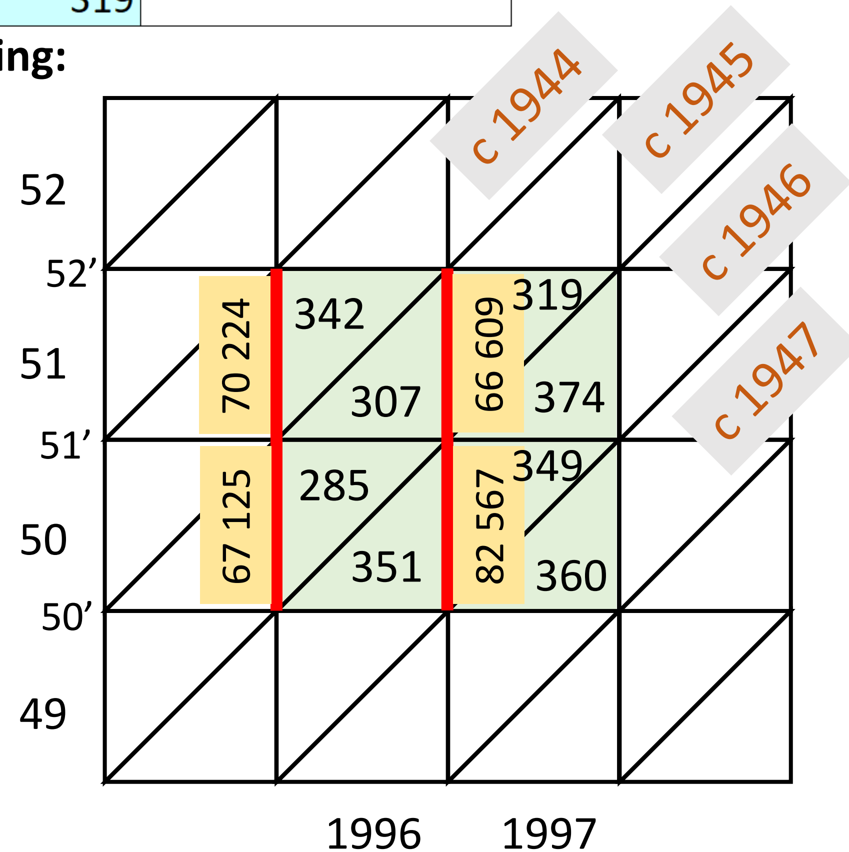
$_{1996}m_{50\text{–}51}$

$_{1996}m_{51}$

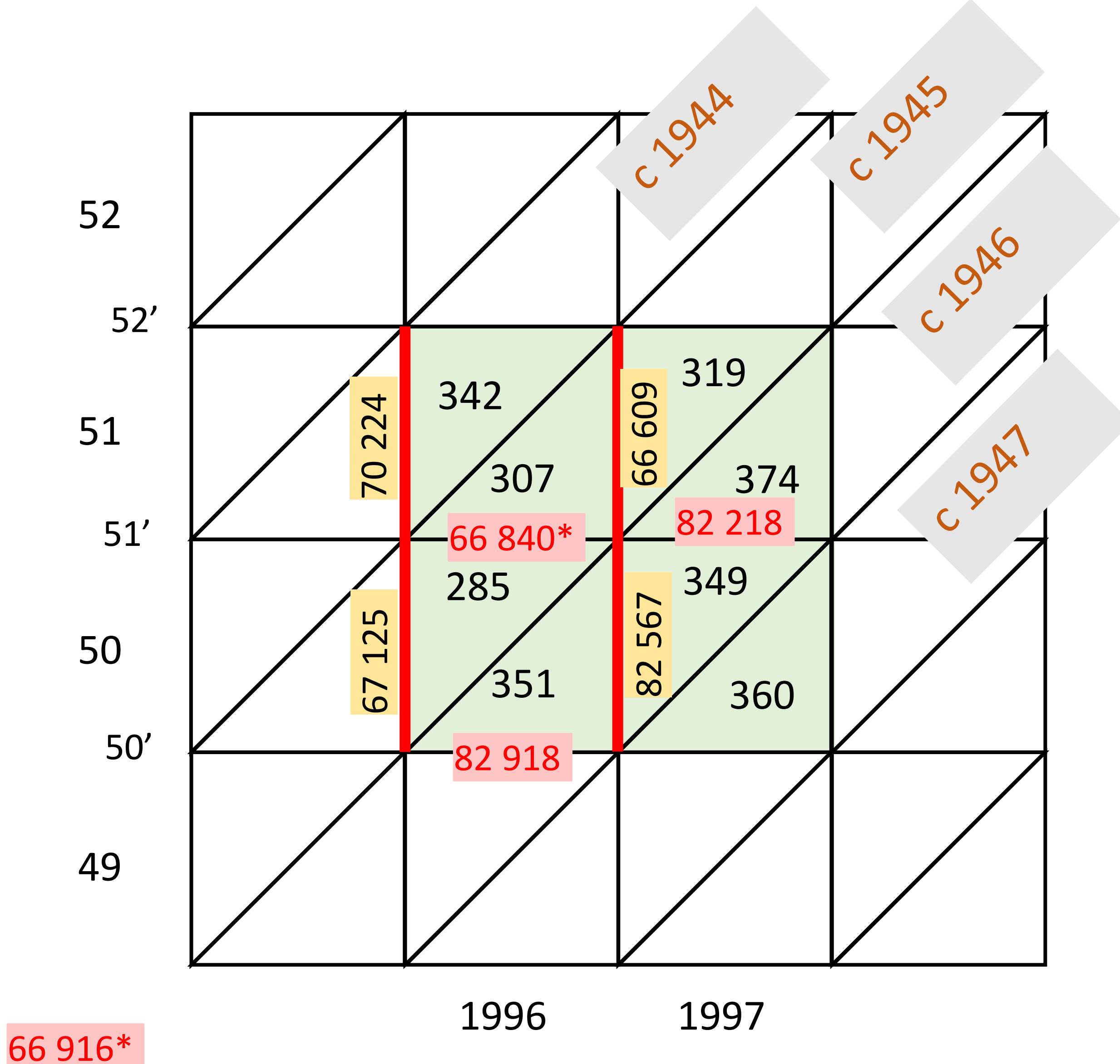
$_{1996\text{–}1997}q_{50}$

$_{1996}q_{50\text{–}51}$

$_{1996}q_{51}$



Direct calculation of rates and probabilities: example



another
cohort D attribute
period age

Direct calculation of rates and probabilities: example

$_{1996-1997}m_{50}$

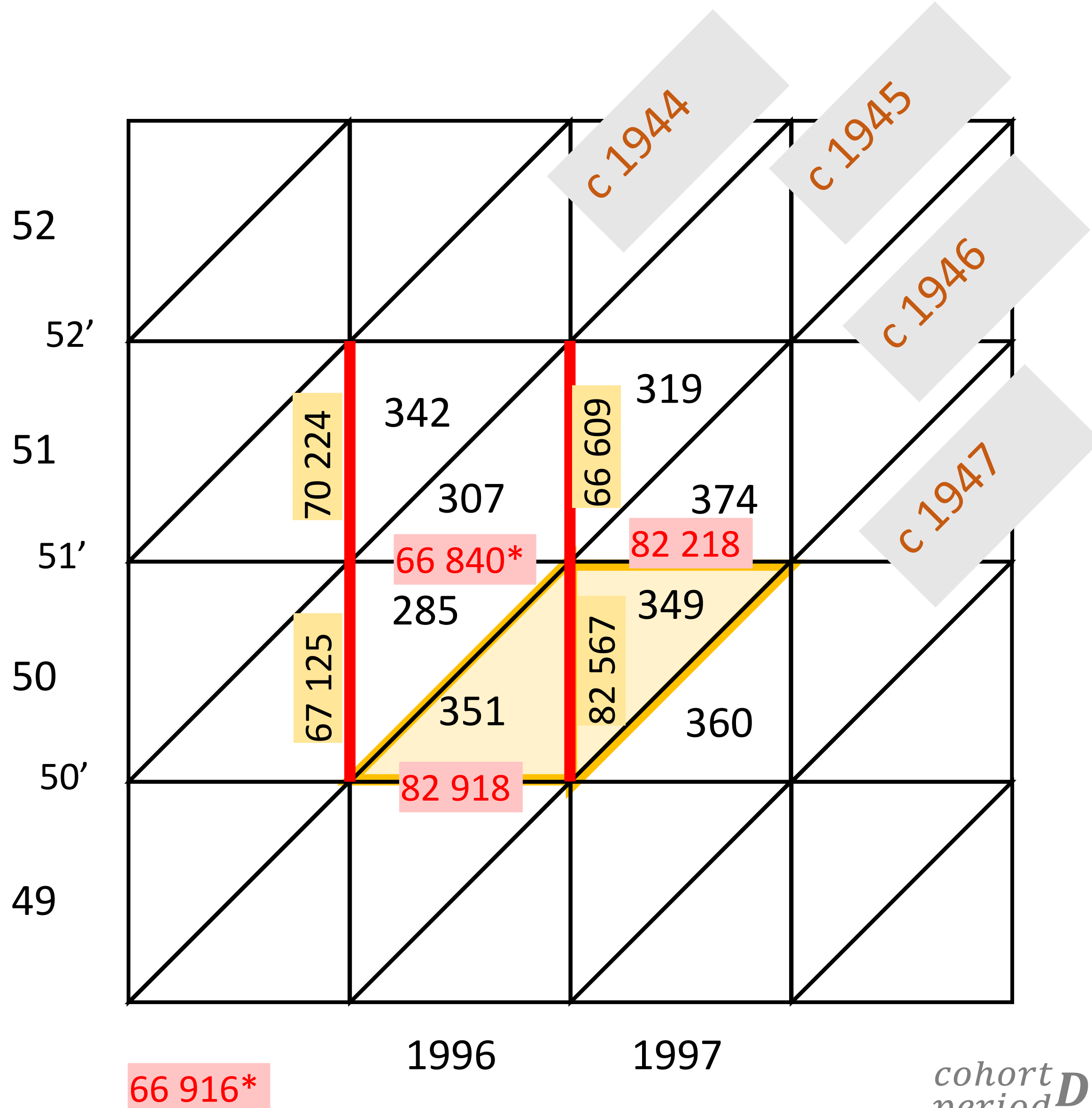
$_{1996}m_{50-51}$

$_{1996}m_{51}$

$_{1996-1997}q_{50}$

$_{1996}q_{50-51}$

$_{1996}q_{51}$



Direct calculation of rates and probabilities: example

${}_{1996\text{–}1997}m_{50}$

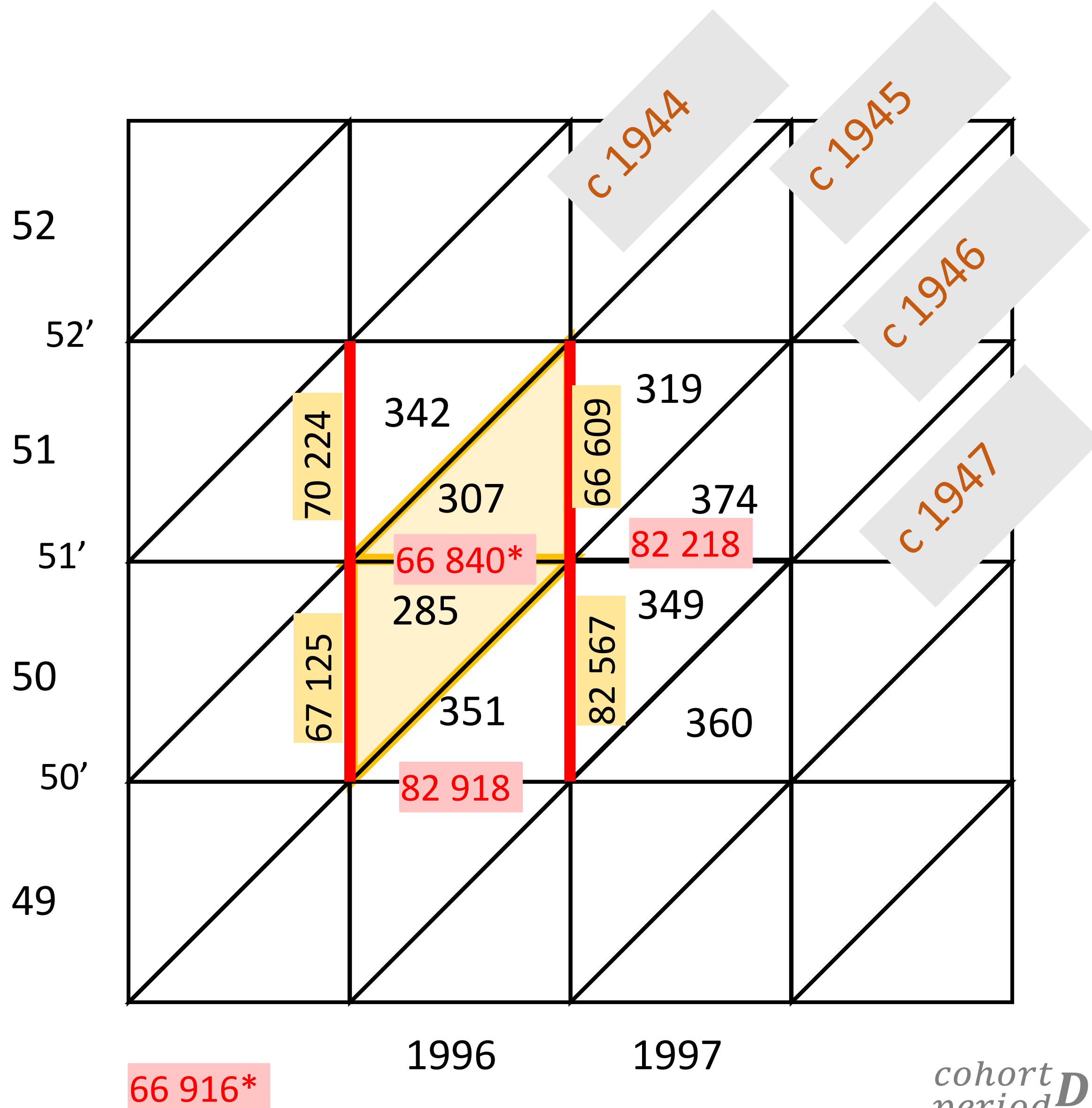
$\boxed{{}_{1996}m_{50\text{–}51}}$

${}_{1996}m_{51}$

${}_{1996\text{–}1997}q_{50}$

$\boxed{{}_{1996}q_{50\text{–}51}}$

${}_{1996}q_{51}$



another
cohort D attribute
period age

Direct calculation of rates and probabilities: example

${}_{1996–1997}m_{50}$

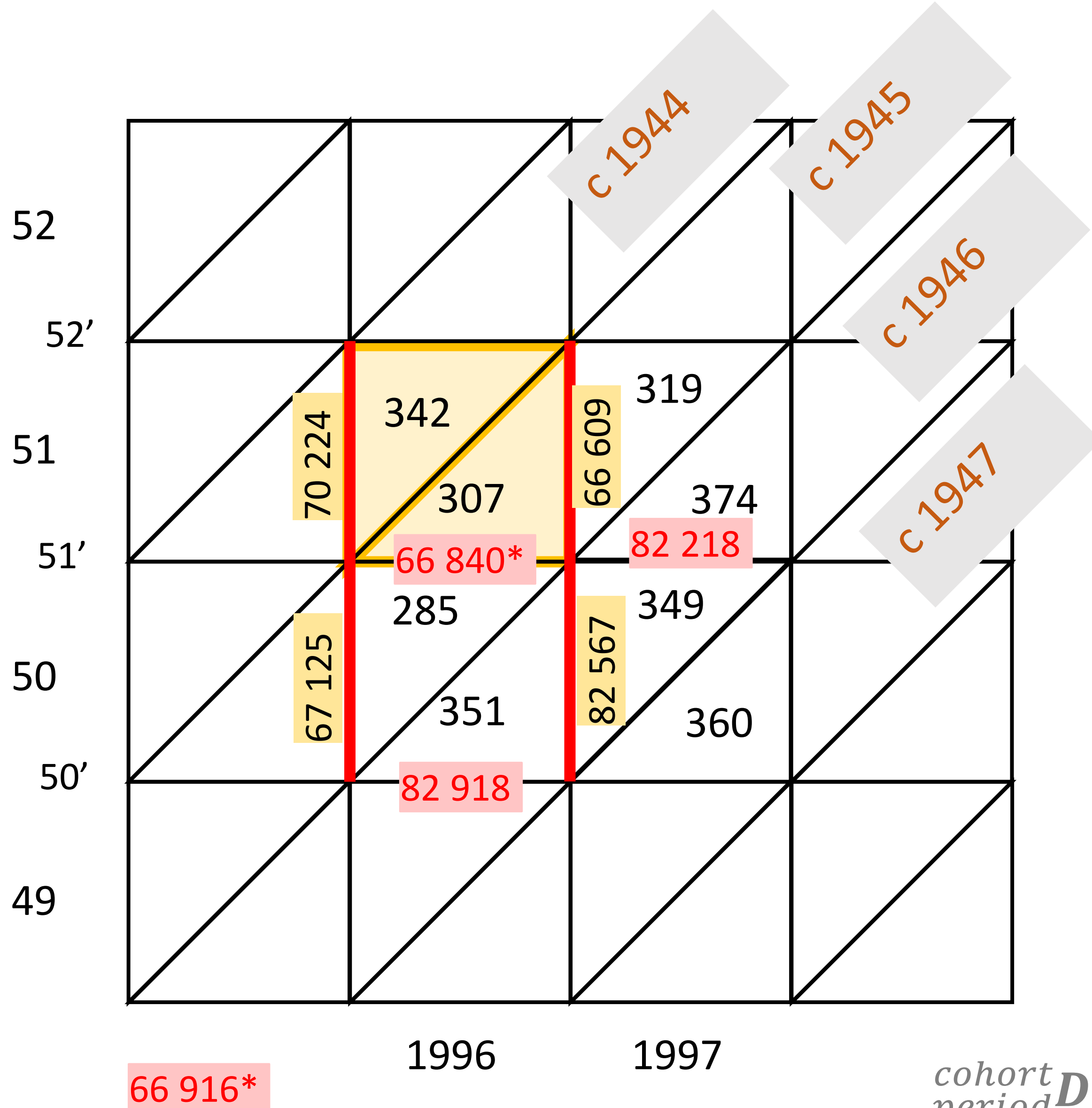
${}_{1996}m_{50–51}$

$\boxed{{}_{1996}m_{51}}$

${}_{1996–1997}q_{50}$

${}_{1996}q_{50–51}$

$\boxed{{}_{1996}q_{51}}$



Rates and probabilities: remarks

- Rates are not dependant upon the length of intervals, whereas probabilities are.

		Age			
		50		50–54	
		m_x	q_x	m_x	q_x
Period	2015	0,002140	0,002140	0,002484	0,012360
	2015–2019	0,002032	0,002030	0,002390	0,011880

- We assume that those, who undergo competing events, have same mortality as those, who do not undergo competing events (f. e. migration)
- We assume uniform distribution of deaths over the time and age.

Source: [5]

Rates and probabilities: remarks

- We assume uniform distribution of deaths over the time and age.

$$P_{x+1} = P_x - D_x - E_x + I_x$$

$$0{,}5(P_x + P_{x+1}) = 0{,}5(P_x + P_x - D_x - E_x + I_x) = P_x - 0{,}5D_x - 0{,}5E_x + 0{,}5I_x$$

$$\boldsymbol{m_x = \frac{D_x}{P_x - 0,5D_x - 0,5E_x + 0,5I_x}}$$

$-0{,}5D_x$ → average contribution to exposure of those who died in given one year age interval was 0,5, which is the implication of uniformity assumption

- In certain age groups, the assumption of uniform distribution od deaths over an age intrval does not hold (f. e. age 0, very high ages). Therefore, at these ages we consider non-uniformity by using adjusted formulas or different methods to estimate the intensity of mortality.

Source: [2]

Indirect estimation of age specific probabilities of death

- q_x
  - Indirect method
    - Linear method
    - Exponential method
  - Direct method

Linear method: $$q_x = \frac{2 \times n \ \times m_x}{2 + n \times m_x}$$

Exponential method: $$q_x = 1 - e^{(-n \times m_x)}$$

These two methods are based on different assumptions about the shape of survival curve over an age interval (x, x+n]. The expression for **linear method** assumes that l_x (survivors at age x) is a straight line over a single age interval (x, x+n]. The expression for **exponential method** assumes, that l_x is an exponential curve over an age interval (x, x+n).

Source: [6]

Indirect estimation of age specific probabilities of death

What is behind the formula $$q_x = \frac{2 \times n \times m_x}{2 + n \times m_x} \;?$$

For n = 1:

$$q_x = \frac{2m_x}{2+m_x} \quad \text{where} \quad \boxed{m_x = \frac{D_x}{P_x - 0.5D_x - 0.5E_x + 0.5I_x}}$$

$$q_x = 2m_x \cdot \frac{1}{2+m_x} = \frac{2D_x}{P_x-0.5D_x-0.5E_x+0.5I_x} \cdot \frac{1}{2+\dfrac{D_x}{P_x-0.5D_x-0.5E_x+0.5I_x}} =$$

$$= \frac{2D_x}{P_x-0.5D_x-0.5E_x+0.5I_x} \cdot \frac{1}{\dfrac{2(P_x-0.5D_x-0.5E_x+0.5I_x)+D_x}{P_x-0.5D_x-0.5E_x+0.5I_x}} =$$

$$= \frac{2D_x}{\cancel{P_x-0.5D_x-0.5E_x+0.5I_x}} \cdot \frac{\cancel{P_x-0.5D_x-0.5E_x+0.5I_x}}{2(P_x-0.5D_x-0.5E_x+0.5I_x)+D_x} =$$

$$= \frac{2D_x}{2(P_x-0.5D_x-0.5E_x+0.5I_x)+D_x} = \frac{D_x}{P_x\cancel{-0.5D_x}-0.5E_x+0.5I_x\cancel{+0.5D_x}} =$$

$$= \boxed{\frac{D_x}{P_x-0.5E_x+0.5I_x}}$$ ← Net probability of dying (Berkson)

The difference between m_x and q_x lies in denominator ($-0,5D_x$). The presence of *"$-0,5Dx$"* in denominator of the fraction implies, that for age group (x, x+n] mortality rates shall be higher than probabilities of death (exception: age 0, extremely high ages).

Migration ~ Competing event, assuming no migration, the net probability equals to ***Deaths/Population at the beginning of the interval***

Indirect estimation of age specific probabilities of death

What is behind the formula $$q_x = 1 - e^{(-n \times m_x)} \;?$$

Relationship between q_x and m_x can be expressed thru other life table functions.
Most directly thru l_x (number of persons left alive at exact age x). It applies:

$${}_n m_x = \frac{l_x - l_{x+n}}{\int_x^{x+n} l_x dx}$$

The difference between persons at exact ages x and x+n equals to number of deaths (in model, closed population, like LT)

We consider l_x is a discrete function. Integrating discrete functions can be understood as ∑ the elements within bounds, in this case (x, x+n]. We divide deaths by sum of persons in exacts ages in interval (x, x+n], which is our exposure.

$${}_n q_x = \frac{l_x - l_{x+n}}{l_x}$$

Numerator is same as above. The deaths are divided by persons in exact age x (beginning of the interval).

So far, everything comes back to the very basic definition of rates and probabilities.

Source: [7]

Indirect estimation of age specific probabilities of death

What is behind the formula $\boxed{q_x = 1 - e^{(-n \times m_x)} \text{ ?}}$

By the Gompertz law of mortality, it applies: $\boldsymbol{l_x = e^{a+bx}}$

Substitute formulas below with $\boldsymbol{l_x = e^{a+bx}}$:

$$_n m_x = \frac{l_x - l_{x+n}}{\int_x^{x+n} l_x dx}$$

$$_n q_x = \frac{l_x - l_{x+n}}{l_x}$$

Get this:

$$_n m_x = \frac{e^{a+bx} - e^{a+b(x+n)}}{\int_x^{x+n} e^{a+bx} dx} = -b$$

$$_n q_x = \frac{e^{a+bx}(1 - e^{bn})}{e^{a+bx}} = 1 - e^{bn}.$$

In formula for q_x put m_x instead of b:

$$\boxed{_n q_x = 1 - e^{-n\ _n m_x}}$$

Source: [7]

Indirect estimation of age specific probabilities of death

$$ {}_n m_x = \frac{l_x - l_{x+n}}{\int_x^{x+n} l_x \, dx} \quad \text{where } l_x = e^{a+bx} $$

$$ \frac{e^{a+bx} - e^{a+b(x+n)}}{\underbrace{\int_x^{x+n} e^{a+bx} \, dx}_{\text{solve integral}}} = \frac{e^{a+bx} - e^{a+b(x+n)}}{1} : \frac{e^{a+bx+bn} - e^{a+bx}}{b} = \frac{-(\cancel{e^{a+bx+bn}} - \cancel{e^{a+bx}})}{1} \cdot \frac{b}{\cancel{e^{a+bx+bn}} - \cancel{e^{a+bx}}} = -b $$

solve integral: $\int e^a e^{bx} \, dx = e^a \int e^{bx} = e^a \int e^u \, du = e^a \frac{1}{b} \int e^u = \frac{e^a e^u}{b} = \boxed{\frac{e^a e^{bx}}{b}}$

substitute $bx = u$

$dx = du$

definite solution: $\int_x^{x+n} e^a e^{bx} \, dx = \boxed{\frac{e^a e^{b(x+n)}}{b}} - \boxed{\frac{e^a e^{bx}}{b}} = \frac{e^{a+bx+bn} - e^{a+bx}}{b}$

$$ q_x = \frac{l_x - l_{x+n}}{l_x} \quad \text{where } l_x = e^{a+bx} $$

$$ \frac{e^{a+bx} - e^{a+b(x+m)}}{e^{a+bx}} = \frac{\cancel{e^{a+bx}}}{\cancel{e^{a+bx}}} + \frac{-e^{a+bx+bn}}{e^{a+bx}} = 1 - \frac{\cancel{e^a} \cancel{e^{bx}} e^{bn}}{\cancel{e^a} \cancel{e^{bx}}} = 1 - e^{bn} $$

Indirect calculation of probabilities: example

Calculate age specific probabilities of death for **French females in 2015, single ages**. Apply linear as well as exponential method. Plot the results. Discuss the differences in estimated rates and probabilities and general shape of „mortality curves".

Data source: Human Mortality Database (**Registration is needed**)

Sources

[0] Slides provided by dr. Burcin

[2] CASELLI, Graziella; VALLIN, Jacques; WUNSCH, Guillaume. *Demography: Analysis and Synthesis, Four Volume Set: A Treatise in Population*. Elsevier, 2005. Part II, Chapter 8. Pages 79–86.

[3] TESÁRKOVÁ, Klára Hulíková; KURTINOVÁ, Olga. *Lexis in Demography*. Springer International Publishing, 2018. Pages 11–14 & 46–48. Available online per CKIS.

[4] Methods protocol of *Human Mortality Database*. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available online: https://www.mortality.org/File/GetDocument/Public/Docs/MethodsProtocolV6.pdf.

[5] SWANSON D. A., SIEGEL J. S. The methods and materials of demography. Elsevier. San Diego. 2004. ISBN: 0-12-641955-8. Available online per CKIS or at https://demographybook.weebly.com/uploads/2/7/2/5/27251849/david_a._swanson_jacob_s._siegel_the_methods_and_materials_of_demography_second_edition__2004.pdf. Pages 265–340.

[6] Fergany, N. (1971). On the Human Survivorship Function and Life Table Construction. *Demography*, *8*(3), 331–334. https://doi.org/10.2307/2060621. Available online: https://www.jstor.org/stable/pdf/2060621.pdf?refreqid=excelsior%3Ada90c51056d89db94ef8d79f5eaf46fd&ab_segments=&origin=.

[7] Reed LJ, Merrell M. A short method for constructing an abridged life table. 1939. Am J Epidemiol. 1995 Jun 1;141(11):993-1022; discussion 991-2. doi: 10.1007/978-3-642-81046-6_7. PMID: 7771448.