

Demographic Research Methods and the PyCCM library: Lecture Two

Charles Rahal and Jiani Yan

Leverhulme Centre for Demographic Science

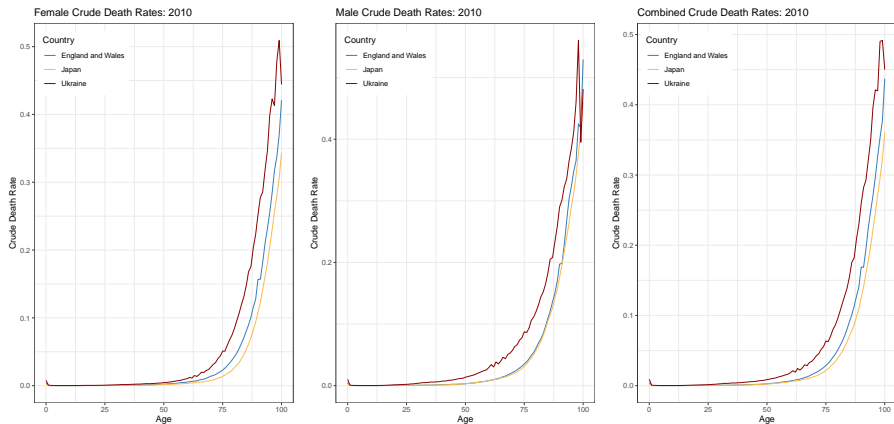
A lecture delivered at the Banco de la Republica, October 2025



Why are Crude Rates Crude?

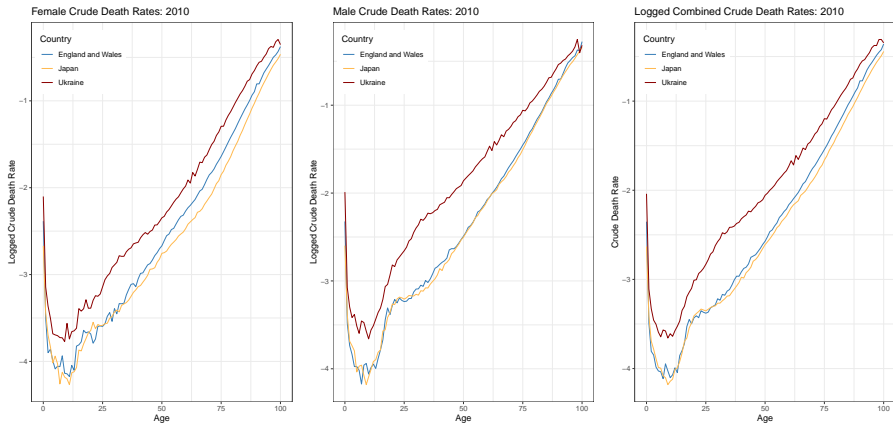
- Crude rates don't account for variation in the age of occurrence of demographic events (they're 'age blind').
- Rates of occurrence of typical demographic events vary significantly by age (i.e. marriage, having children, dying).
- This largely reflects variations in physiological capacity.
- For migration, it reflects economic and social gains from movement, such as contributions which can be made to/wages to be earned from labour markets.
- Lets look at some 'Age specific death rates' from the **Human Mortality Database**.
- Lets then follow this with some fertility data from the **Human Fertility Database**

Age Specific Death Rates



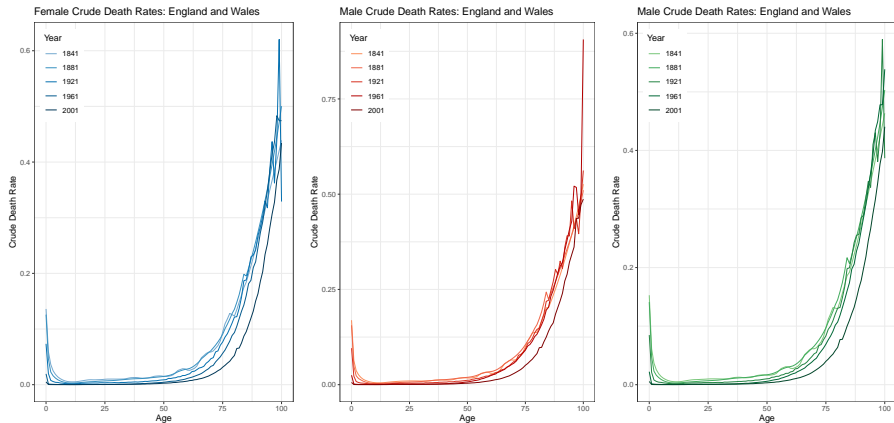
Source: Human Mortality Database

Logged Age-Specific Death Rates



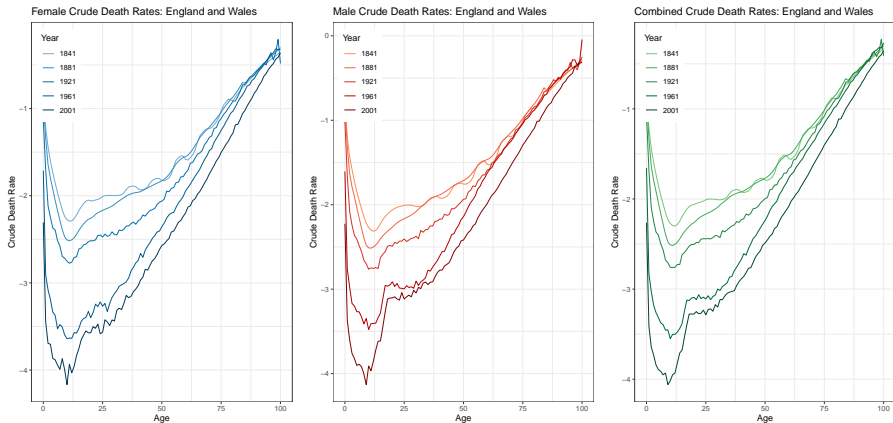
Source: Human Mortality Database

Age Specific Death Rates Over Time



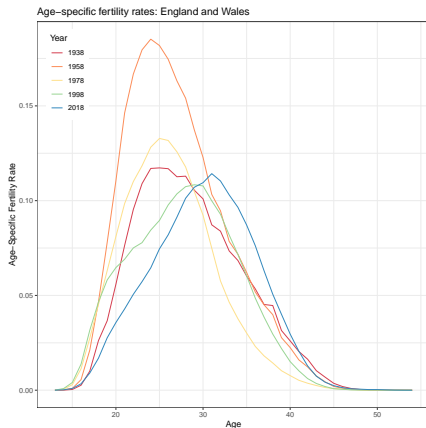
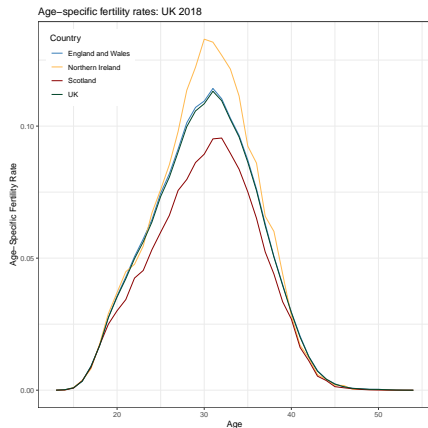
Source: Human Mortality Database

Logged Age Specific Death Rates Over Time



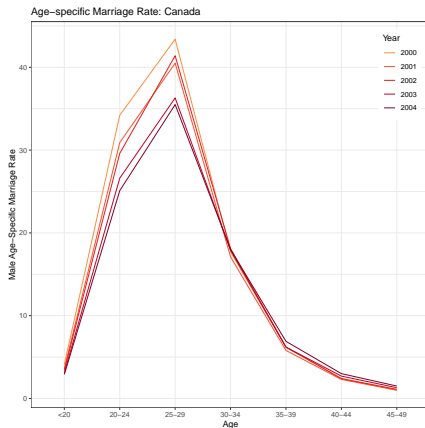
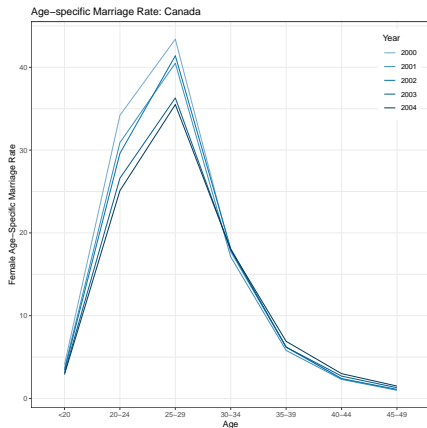
Source: Human Mortality Database

Age Specific Fertility Rates



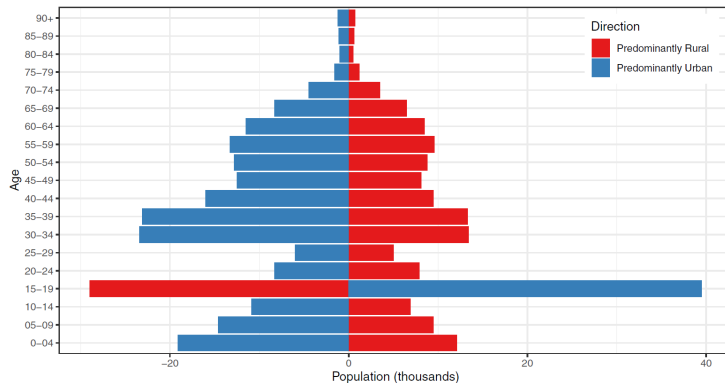
Source: Human Fertility Database

Age Specific Marriage Rates



Source: Statistics Canada

Age Specific Internal Migration



Net Internal (within the UK) migration. Source: Office for National Statistics

Why are Crude Rates Crude? (Cont.)

- So far we've defined a 'Crude' rate which focuses on a **periodic** dimension.
 - For example, this relates to the interval $[0, T]$ in the CDR calculation:

$$\text{Crude Death Rate } [0, T] = \frac{\# \text{ deaths in } [0, T]}{\# \text{ PY lived in } [0, T]} \quad (1)$$

- Lets now define a **period age-specific rate**:

$${}_nM_x[0, T] = \frac{\# \text{ deaths in ages } x \text{ to } x+n \text{ between } 0 \text{ and } T}{\# \text{ PY lived in ages } x \text{ to } x+n \text{ between } 0 \text{ and } T} \quad (2)$$

- This is necessary due to age variation at which specific events (birth, death, migration) happen.

More on General Age-Specific Rates

- Lets unpack this notation a little and give an example of an age-specific fertility rate.
- The general formula is:

$${}_nF_x[0, T] = \frac{\# \text{ births in } 0 \text{ to } T \text{ to women aged } x \text{ to } x+n}{\# \text{ PY between } 0 \text{ to } T \text{ by women aged } x \text{ to } x+n} \quad (3)$$

Where:

- n is the age interval.
- x is the starting age.
- 0 is the starting period.
- T is the period when the accounting stops.

An example of Age-specific Rates

- If we consider just women giving birth between 2015-2019 between the ages of 25-29, this can be written as:

$${}_5F_{25}[2015, 2020) = \frac{\# \text{ births in 2015 to 2020 to women aged 25 to 30}}{\# \text{ PY 2015 to 2020 by women aged 25 to 30}} \quad (4)$$

- These are 'left inclusive': **from** (and including) 25 **until** (excluding) 30.
- These are 'left inclusive': **from** (and including) 2015 **until** (excluding) 2020.
- Note change in notation in the final equation (i.e. [2015, 2020)).
- Note the ${}_5$ year interval. These intervals are going to become increasingly important.

Bringing Age into Crude Rates

- Crude rates can be expressed as a weighted average of age-specific rates.
- The weights are the population's proportionate age distribution.
- Lets look at the CDR as an example:

$$\begin{aligned} CDR &= \frac{D}{N} \\ &= \frac{\sum_n D_x}{N} \\ &= \frac{\sum \frac{{}_n D_x}{{}_n N_x}}{{}_n N_x} N \\ &= \sum \frac{{}_n D_x}{{}_n N_x} \times \frac{{}_n N_x}{N} \\ &= \sum {}_n M_x \times {}_n C_x \end{aligned} \tag{5}$$

Where ${}_n M_x$ is an age-specific mortality rate, and ${}_n C_x$ is the fraction of the population at interval x to $x + n$

Standardisation Necessary for Comparison

- Human populations differ considerably in age composition.
- A lot of demographic analysis is comparative/relative, rather than absolute.
- This means we need to standardise our rate calculations to be able to ask a question like:

‘What would be the mortality rate if population A had the same age structure of population B?’

- For example, the ONS uses standardised rates based on the European Standard Population (ESP) 2013.
- The ESP is an artificial population structure which is used in the weighting of mortality or incidence data to produce ASRs.

Age Standardisation Example

Lets consider the ASCDR for the United Kingdom constituent countries:

$$C_i^{UK} = \frac{C_i^E + C_i^S + C_i^W + C_i^{NI}}{4} = \text{Average age distribution} \quad (6)$$

$$ASCDR^E = \sum_{i=1}^{\infty} M_i^E \times C_i^{UK} = \text{Age standardised CDR: England} \quad (7)$$

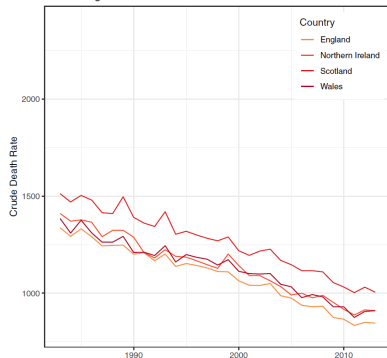
$$ASCDR^S = \sum_{i=1}^{\infty} M_i^S \times C_i^{UK} = \text{Age standardised CDR: Scotland} \quad (8)$$

$$ASCDR^W = \sum_{i=1}^{\infty} M_i^W \times C_i^{UK} = \text{Age standardised CDR: Wales} \quad (9)$$

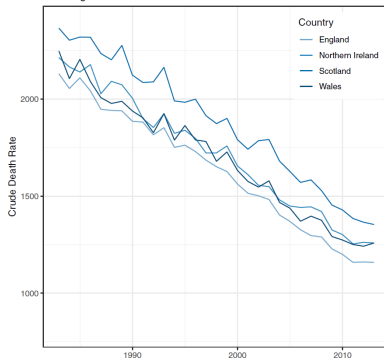
$$ASCDR^{NI} = \sum_{i=1}^{\infty} M_i^{NI} \times C_i^{UK} = \text{Age standardised CDR: Northern Ireland} \quad (10)$$

Age Standardisation Example (Cont.)

Female Age-Standardised Death Rates: 1983–2013



Male Age-Standardised Death Rates: 1983–2013



Source: Office for National Statistics

The ASMRs for the period 1983 to 2013 show a decline in mortality of 45% for males in the UK and 36% for females.

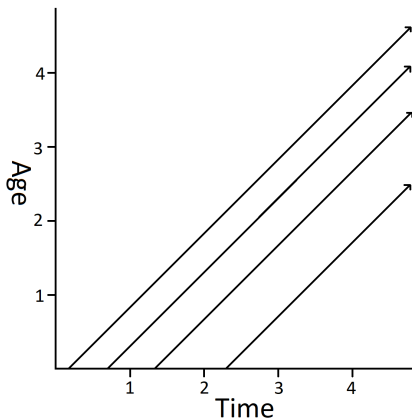
Demographic Time

- We've talked about chronological time ('calendar time', or 'periods').
- A different way is through 'personal time', measured by age, unique across people who share the same birthdate.
- We can visualise individual trajectories through chronological (period) and personal time (age) using the Lexis diagram, as well as also considering the periods overlaid on top of this.
- The Lexis diagram helps clarify the relationship between exposure segments for cohorts and exposure segments for periods.
- The Lexis diagram is to a demographer as a wrench is to a mechanic.
- This diagram is named after economist and social scientist Wilhelm Lexis.
 - However, this is likely an example of **Stigler's Law**!

(Stigler's law of eponymy states no scientific discovery is named after its original discoverer)

Lifelines beginning at birth in the Lexis Diagram

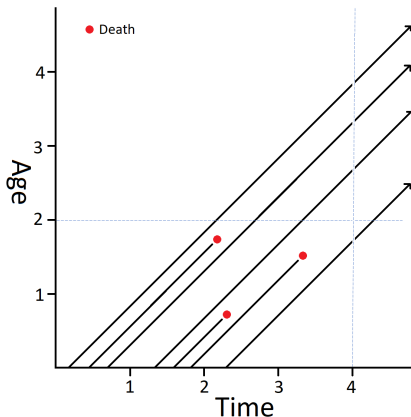
- Lets first see a very basic Lexis Diagram with only births.



- Calendar time is usually represented on the x-axis.
- Age is represented on the vertical y-axis.
- In this example, the line begins on the time axis at the time of the person's birth.
- Each line is diagonal and represents one person because people age linearly over time.

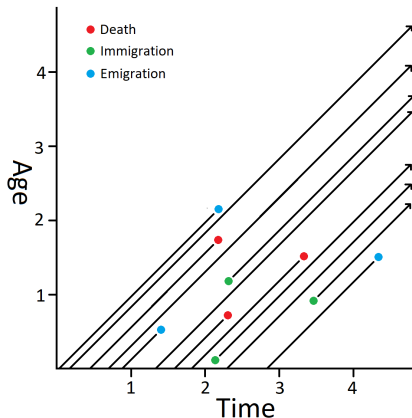
Lifelines in the Lexis Diagram

- However, we can represent more than just births!



- In this example, the red circles represent deaths.
- The sum of the lengths of all lifelines in a particular portion of the diagram represents person-years lived or exposure.
- Critically, that can be considered for just a specific area.
 - For example, what would PY between Time 2-3 look like?

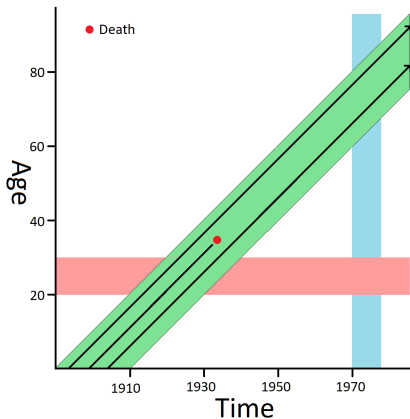
Lifelines in the Lexis Diagram



- We can also represent other events, such as immigration and emigration.
- In this version of the diagram:
 - Green circles represent immigration.
 - Blue circles represent emigration.
- Lifelines usually consist of people's life timelines, but it may also consist of entities, such as marriages, formed by a demographic event.

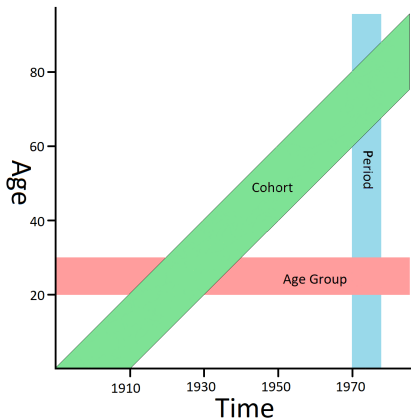
Lifelines in the Lexis Diagram

- Lets introduce the essential concept of a **cohort**.



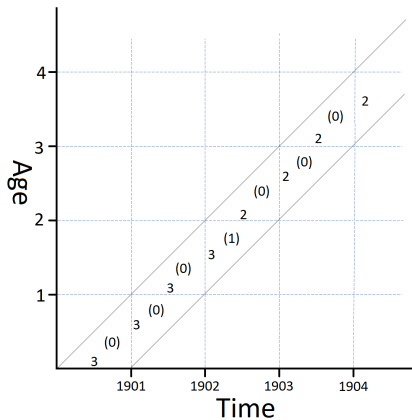
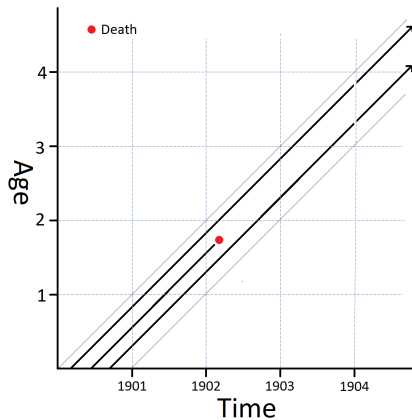
- A cohort advances through life at 45°
 - People age one year in one year!
- A cohort is a group of individuals followed simultaneously through time and age: 'shared experiences'.
- A cohort could also refer to the year an individual enters university, or signs a contract.
- Aggregate of all units that experience a particular demographic event during an interval.

Lifelines in the Lexis Diagram

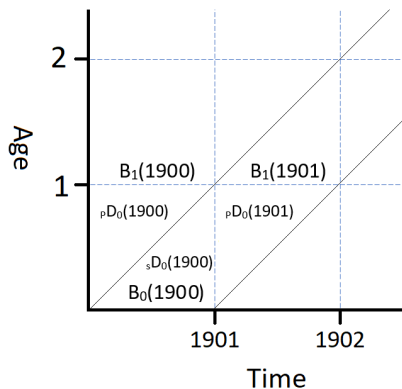


- Tools like this help us better understand and conduct 'age-period-cohort' analysis.
- The figure on this slide is a common representation of the Lexis diagram.
- For example, we see (highlighted):
 1. The birth cohort of 1890-1910.
 2. The period of all people alive in 1970-1980.
 3. All people between age 20-30.
- The most frequently encountered type of cohort is the birth cohort.

Lexis Parallelograms



Lexis Parallelograms



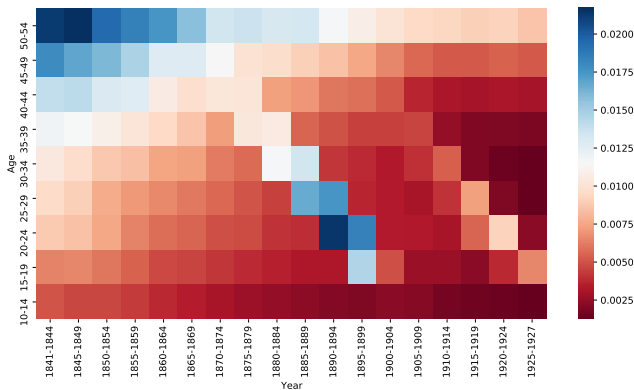
- The triangles in the Lexis diagram separate events occurring in a particular period into two cohorts.
- The probability that a member of the cohort will die after reaching 1 before reaching 2 is:

$${}_1q_1^{1900_c} = \frac{\#D \text{ to } 1900_c, \text{ ages } 1-2}{\# \text{persons reaching } 1} \quad (11)$$

- Can you figure out exactly what this number is, using the right hand side figure on the slide above?

- That's right! its $\frac{1}{3}$! What about ${}_1q_2^{1900_c}$?

Mortality in Specific Cohorts



- We can also see mortality of cohorts in relation to specific events
- The figure plots historic data for England and Wales.
- The colormap plots **cohort** 'mx' (age specific death rates)!

Importance of Cohorts

Why is the concept of a cohort so crucial?

- Ryder (1965) founding paper, presenting a demographic approach to the study of social change.
- 'Demographic metabolism' (i.e. the succession of cohorts entering and leaving the society) as the motor of social change.
- Continual emergence of new participants into the social process and the continual withdrawal of their predecessors compensate the society for limited individual flexibility..
- Society whose members were immortal would resemble a “stagnant pond”.

Importance of Cohorts (Cont)

- Inter-cohort comparison becomes a way to study change.
- This is the idea behind a 'life course approach' to study population change.
- Cohorts are affected by change in a different way, as they have different ages in a given moment in time.
- Each cohort is distinct by features that are constant throughout time.
- Richard Easterlin ('the Easterlin hypothesis'): Large cohorts thought to have more problems (but more political power), smaller cohorts fewer problems (but less political power).

The APC Identification problem

- APC models are used when individuals/populations followed over time.
- When passage of time is explanatory, there is risk of confused interpretation.
- The models involve three time scales for age, period and cohort.
 - Interlinked since the calendar period is the sum of the cohort and the age
- At the core of the models is a linear predictor of the form:

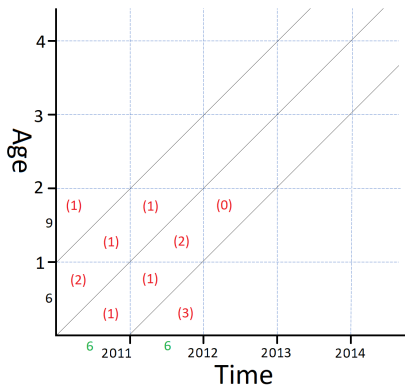
$$\mu_{\text{age,coh}} = \alpha_{\text{age}} + \beta_{\text{per}} + \gamma_{\text{coh}} + \delta \quad (12)$$

- Different values of the RHS result in the same LHS.
- Linear parts from the three APC effects cannot be disentangled

Solutions fall into three broad categories:

1. Identify time effects through non-testable constraints
2. Use graphs to get an impression of time effects or replace the time effects
3. Reparametrize in terms of invariant, non-linear parts of the time effects

Recap: Moving from Lexis to Probabilities (Cont.)



What is the probability that a baby born in 2010 dies between exact age 1 and 2?.

$$\begin{aligned}
 {}_1q_1^{2010c} &= \frac{\# \text{ 2010}_c \text{ reach age 2}}{\# \text{ reach age } 1_c} \\
 &= \frac{2}{4} \\
 &= .5
 \end{aligned} \tag{13}$$

- What is the compliment to dying? That's right! Surviving!

Recap: Moving from Lexis to Probabilities (Cont.)

- Lets now slightly formalise our notation.
- What is the probability that a baby born in 2010 **survives** until age 2?

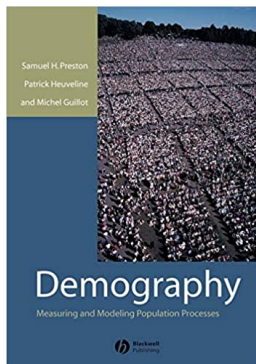
$$\begin{aligned} {}_2p_0^{2010c} &= (1 - {}_2q_0^{2010c}) \\ &= (1 - \frac{4}{6}) \\ &= (1 - .6666) \\ &= 0.3333 \end{aligned} \tag{14}$$

Introducing Lifetables

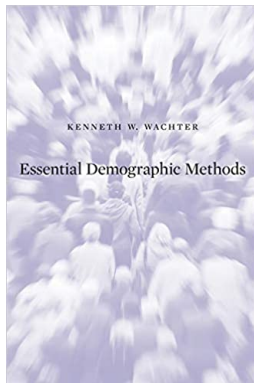
Lets quickly note the difference between **cohort** and **period** based lifetables.

- Waiting for all of the **cohort** to die would allow us to say something about the average duration of life for the lines on our Lexis diagram.
- Waiting for entire cohorts to die would mean waiting for a *really* along time.
- Remember this in terms of the Lexis Diagram: we'd have to wait the entire duration of the x-axis!
- To generate estimates more regularly and frequently, we derive information from a 'synthetic' cohort of individuals at different ages in a given **period**.

Key Reading



(a) Preston et al: Chapter 3



(b) Wachter: Chapters 3 and 7

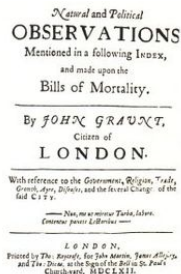
What can a lifetable be used for?

- A life table is the study of the distribution of the time it takes a cohort ('real' or 'synthetic') of individuals (couples, institutions...) to exit a state (to experience an 'event').
- Note the important terminology here: 'synthetic' → period life table.
- Examples include:
 - From alive to dead (event is death)
 - From single to married (event is marriage)
 - From with 0 children to with 1 child (event is birth of the first child)

The lifetable as an accounting device

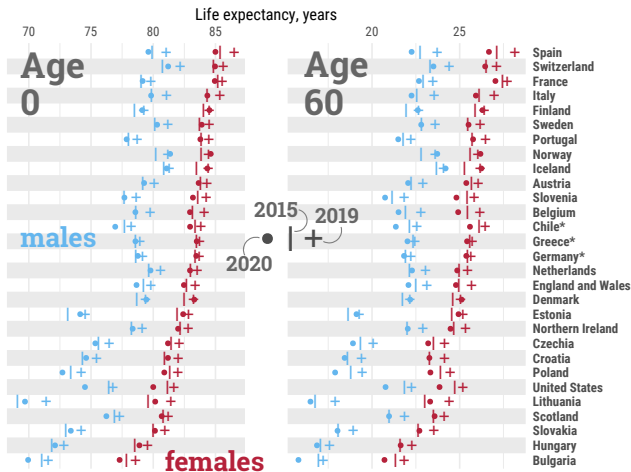
- A life table is an accounting device with different columns.
- Columns present different measures and different ways of presenting information about the mortality/survival of cohort (synthetic or otherwise).
- The life table can be applied to any (single or multiple) decrement process.
 - A decrement is like an attrition factor; a possible route of termination out of the lifetable.
 - But again, it is most commonly applied to mortality.
 - **Class quiz:** Is mortality a multiple or single decrement process?
- Amongst other things, one of our goals will be to describe the duration of life, synthesised in the measure called 'life expectancy'.

The History of the Lifetable



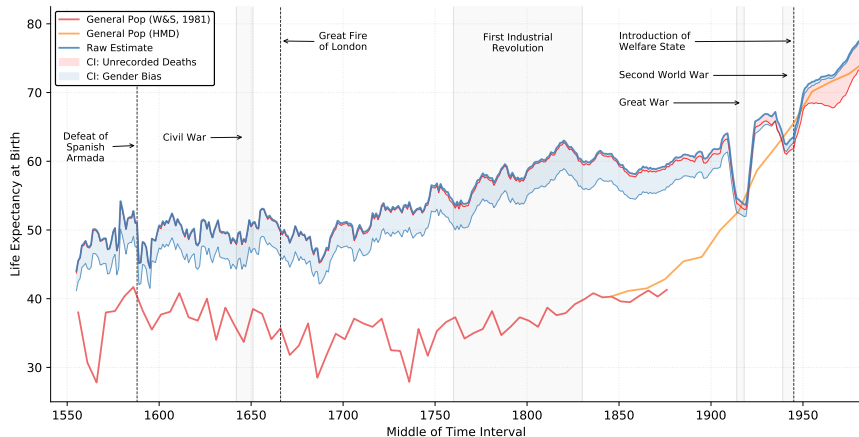
- The idea was originally devised by John Graunt (1662) using 'bulletins' containing deaths by age in London.
 - Graunt is regarded as the founder of demography.
 - Graunt 're-purposed' data collected for London merchants wanting early warning signs for outbreak of plagues.
 - When burials went up, they shut down shops.
-
- The merchants paid a subscription to have numbers of burials collected from parish registers and published each week.
 - This database formed the Observations from which Graunt derived insights from bits of empirical information about survivor-ship to different ages.

Motivation: Recent Applications



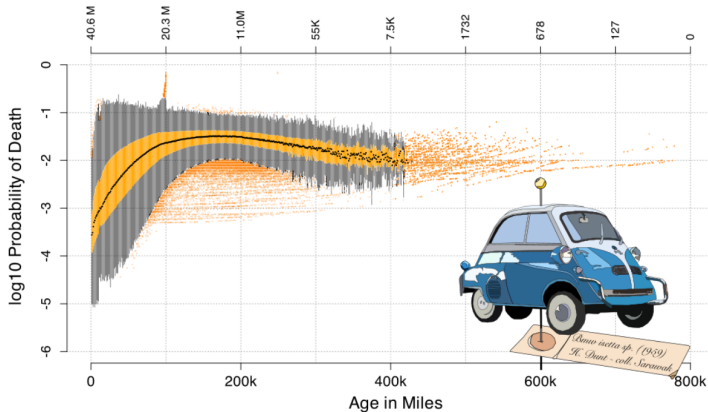
- Perhaps you saw this recent paper in the news?

Motivation: Recent Applications



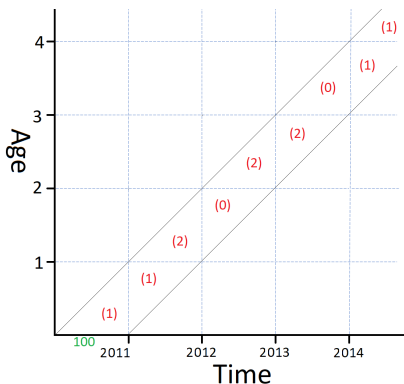
- A 5-year ongoing project!

Very flexible...



- They can even be measured to calculate the life expectancy of cars!

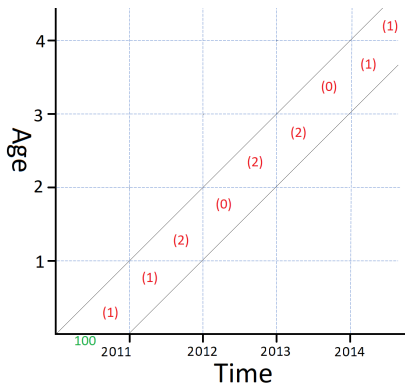
More on Probabilities



- Assume we have data on the initial number of members in a cohort (e.g. 100 here).
- Based on this Lexis diagram, what can we say about the numbers of survivors to age 1, 2, ...?
- What can we say about probabilities of death between ages 0–1, 1–2, 2–3?

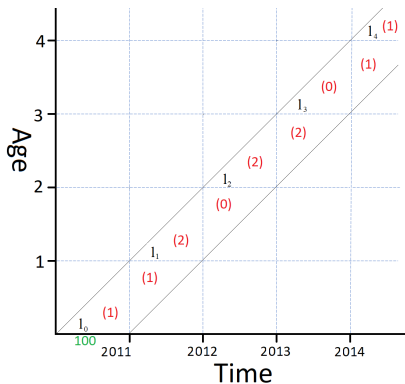
- We're now going to show how we can calculate these two basic inputs:
number born and **deaths** at each period, into full life tables.

Calculating Lifetable Columns



- In lifetable notation, l_x refers to the number of survivors to age x .
- ${}_n d_x$ refers to the number of deaths between age x (start of the age interval) to $x + n$ (width of the age interval).
- **Class Quiz!:** what would ${}_1 d_3$ be?
- ${}_n q_x$ refers to the probability of death between x and $x + n$.
- Note: ${}_n q_x = \frac{{}_n d_x}{l_x}$.
- So here, ${}_1 q_0 = \frac{2}{100} = 0.02$

Calculating Lifetable Columns



- For those who have reached (exact) age 1 we can compute the probability that they reach age 2 (conditional probability)

- This would be:

$$\begin{aligned} {}_1p_1 &= 1 - {}_1q_1 \\ &= 1 - \frac{2}{98} \\ &= 0.9796 \end{aligned} \quad (15)$$

- Note! This equals l_2/l_1 :

$${}_1p_1 = \frac{l_2}{l_1} \quad (16)$$

Generalising Lifetable Notation

It's important to realise the generalisable nature of our notation:

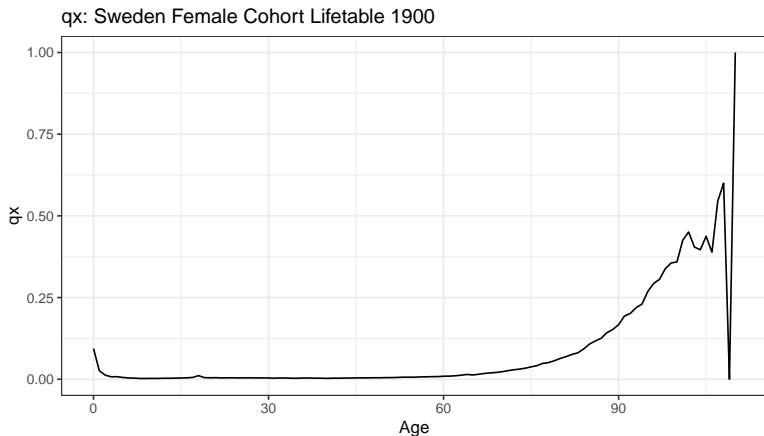
- ${}_n d_x$ refers to the number of deaths to members of the cohort between ages x and $x + n$.
- ${}_n q_x$ is the probability of dying between ages x and $x + n$ among cohort members alive at age x .
- ${}_n q_x = \frac{{}_n d_x}{l_x}$
- l_x is the number of survivors at exact age x for the cohort.
- Survival probabilities multiply: $l_{x+n} = l_x \times (1 - {}_n q_x)$
- Complement of ${}_n q_x$ is probability of surviving from age x to age $x + n$:

$${}_n p_x = 1 - {}_n q_x \quad (17)$$

Generalising Lifetable Notation

- Life tables rescale the initial size of the cohort to a round number (comparison between cohorts is then by default feasible) called radix of the table, l_0 .
- The radix represents – in the life table – the number of members of the cohort alive at (exact) age 0.
 - Usually, it is a power of 10 (such as 100 or 1000).
- The radix (latin for root) is completely arbitrary.
- Lets look at some lifetable column construction, and simultaneously on visualising those columns for a cohort of Swedish females in 1900.
- Lets try and reconstruct a lifetable from *just* from ${}_nq_x$.
- Check out the `lifetable_construction_cohort_sweden_f_1900.xlsx` file on Canvas to follow along here.
- Remember that here we're talking about **cohorts** rather than **periods**.
- The data comes from the HMD. What does ${}_nq_x$ look like, though?

The Cohort Life Table



- If we have data on ages at death (and on the number of population entering into an age interval at a point in time), we can compute ${}_nq_x$ directly.

The Cohort Life Table

Age	q_x	l_x	d_x	p_x
0	0.09399	100000	9399	0.90601
1	0.02587	90601	2343.848	0.97413
2	0.01264	88257.15	1115.57	0.98736
3	0.00758	87141.58	660.5332	0.99242
4	0.00775	86481.05	670.2281	0.99225
5	0.00547	85810.82	469.3852	0.99453
6	0.00399	85341.44	340.5123	0.99601
7	0.00343	85000.92	291.5532	0.99657
8	0.00239	84709.37	202.4554	0.99761
9	0.00227	84506.91	191.8307	0.99773

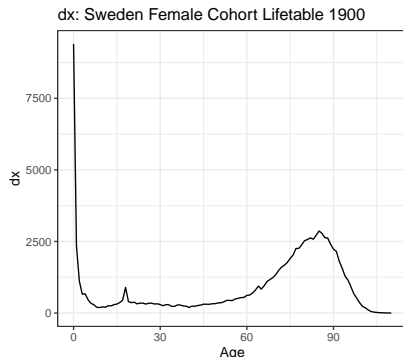
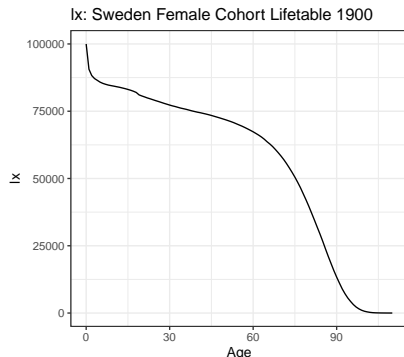
- From ${}_1q_x$ we can derive other columns of the life table: l_x , ${}_1d_x$, ${}_1p_x$.

The Cohort Life Table

Just to be as verbose as possible, we'll reinforce the excel calculations here:

- **Age** acts as our 'index' in these calculations.
- ${}_nq_x$ as our **probability of death** can either be calculated endogenously or be provided externally.
- l_x is an arbitrarily defined **radix** for $x = 0$, and then $l_x = l_{x-1} \times (1 - q_{x-1})$.
- d_x is simply the $l_x \times {}_nq_x$.

The Cohort Life Table



- Just to again reinforce:

- From ${}_1q_x$ we can derive other columns of the life table: l_x , ${}_1d_x$, ${}_1p_x$.

The Cohort Life Table

- In order to obtain life expectancy, we need another set of quantities.
 - These come from either data or additional hypotheses.
- How many years do people dying between x and $x + 1$ live?
- A simple hypothesis: they live on average 0.5 years (i.e. they all die at the middle of the year, or the distribution of deaths is uniform over the interval).
- **But** this is not necessarily true, since we know that:
 - During the first year of life deaths are very concentrated towards the very beginning of the year.
 - Mortality may be high and thus on average individuals will live less than half the interval.

The Cohort Life Table

- For that reason, we introduce some other columns in the life table.
- ${}_na_x$ is the average number of persons-years lived by individuals who die between (exact) age x and (exact) age $x + n$.
- The (total) persons-years lived by individuals between (exact) age x and (exact) age $x + n$ is:

$${}_nL_x = (n \times l_{x+n}) + ({}_na_x \times {}_nd_x) \quad (18)$$

- $n \times l_{x+n}$: The number of 'full' years lived by people who don't die.
- ${}_na_x \times {}_nd_x$: The years lived by people dying in the interval.

The Cohort Life Table

- If we sum up all the person-years lived beyond x , we obtain T_x .
- The main use of T_x is for computing the expectation (or average) of further life beyond age x .
- This key final column in the lifetable is called e_x .
- More clearly, T_x is calculated as:

$$T_x = {}_nL_x + {}_nL_{x+n} + {}_nL_{x+2n} + \dots \quad (19)$$

- We can think of this as all of the years remaining left to be lived, reverse summed to period x

The Cohort Life Table

- For this it is easiest to first fill in the ${}_nL_x$ columns, and then cumulate sums from the bottom up.
- Once we have T_x , we can get life expectancy at age x :

$$e_x = \frac{T_x}{l_x} \quad (20)$$

- Life expectancy at age x is the average number of years lived by individuals who are alive at (exact) age x .
- It is one of the most important quantities in science, and usually (but not always) the main source of interest in a lifetable.

The Cohort Life Table

Age	q_x	l_x	d_x	p_x	L_x	T_x	e_x
0	0.094	100000.000	9399.000	0.906	95300.500	6132074.221	61.321
1	0.026	90601.000	2343.848	0.974	89429.076	6036773.721	66.630
2	0.013	88257.152	1115.570	0.987	87699.367	5947344.645	67.387
3	0.008	87141.582	660.533	0.992	86811.315	5859645.278	67.243
4	0.008	86481.049	670.228	0.992	86145.934	5772833.963	66.753
5	0.005	85810.820	469.385	0.995	85576.128	5686688.029	66.270
6	0.004	85341.435	340.512	0.996	85171.179	5601111.901	65.632
7	0.003	85000.923	291.553	0.997	84855.146	5515940.722	64.893
8	0.002	84709.370	202.455	0.998	84608.142	5431085.575	64.114
9	0.002	84506.914	191.831	0.998	84410.999	5346477.433	63.267

- **Class Quiz:** Why does life expectancy rise after age 1?
- Note: we've only talked about the first 10 years of life here.
- Where is ${}_na_x$ here?

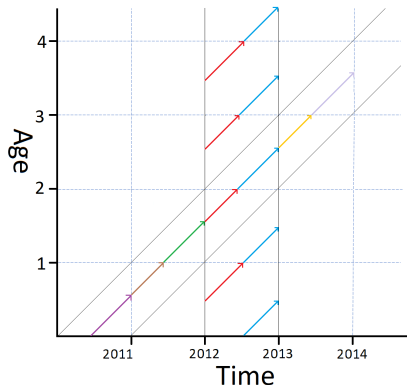
Abridged and Complete Life Tables

- Usually there is a distinction between:
 1. Complete life tables: life tables built using 1-year age intervals up to a certain age, then an open age interval.
 2. Abridged life tables: life tables built using 5-year age intervals up to a certain age then an open age interval.
- When using abridged lifetables, there is usually an exception for the first 5-year interval (divided in 1 and 4 years respectively).
- **Class quiz:** is ${}_4d_1$ abridged? how about ${}_1d_3$?
- Note also that some functions (e.g. T_x) refer to a specific age, and others to intervals (i.e. ${}_nq_x$).
- **Class quiz:** Is accuracy increasing or decreasing in the size of n ? Why ever use $n > 1$?

Introducing Period Lifetables

- So far we have been working through an example of a cohort life table where we had information on probabilities of dying at different ages for the members of the cohort.
- However, if we waited for all members of a cohort to die to compute measures of life expectancy, we would have to wait a long time.
- As a result, instead of following a real cohort, we compute measures of life expectancy using age-specific mortality information from the current period. This is a 'synthetic' cohort.
- We commonly hear about life expectancy in 2020 or life expectancy in 2015-2020.
- This information comes from the period life table.
- Main differences from cohort life tables are:
 - Input data: **cohort** probabilities versus **period** age-specific death rates
 - Interpretation: real cohorts versus synthetic cohorts

Introducing Synthetic Cohorts



- Assumption: when today's babies are 40, their death rates would be same as today's 40-year olds.
- What would happen to a cohort if it were subjected to the same mortality conditions of that period?
- Contrast with cohort tables which simply record.
- From period age-specific mortality rates (faster, easier), we can build period life tables.

- This gives us the most commonly used measures of period life expectancy.

Introducing the Period Life Table

- For the cohort life table, we started with probabilities.
- Remember:

$${}_nq_x = \frac{{}_nd_x}{l_x} \quad (21)$$

- In the period life table, we start with the more readily available age-specific death rates.
- Remember that age-specific rate is the number of deaths in an age interval/person-years in an age interval.
- In life table terms, we introduce a new function for age-specific rates: ${}_nm_x$
- From ${}_nm_x$ values we need to derive ${}_nq_x$ values to proceed with the usual sequence of steps.

Greville-Chiang Conversion

- The Greville-Chiang conversion equation helps us do this:

$${}_nq_x = \frac{{}_n \times_n m_x}{1 + ({}_n - {}_n a_x) \times_n m_x} \quad (22)$$

- If we assume ${}_n a_x = \frac{n}{2}$:

$${}_nq_x = \frac{{}_n \times_n m_x}{1 + \frac{n}{2} \times_n m_x} = \frac{2{}_n \times_n m_x}{2 + n \times_n m_x} \quad (23)$$

- See PHG Section 3.1 for further derivation of this important equation.

Choosing ${}_na_x$

- The ${}_na_x$ is important in making the ${}_nm_x \rightarrow {}_nq_x$ conversion.
- Strategies for choosing ${}_na_x$:
 1. If we have data on exact age at death, we can compute it exactly.
 2. Infer it from the graduation of the ${}_nm_x$ function.
 3. Borrow ${}_na_x$ from another population with similar ${}_nm_x$ curve.
 4. Use a rule of thumb such as ${}_na_x = \frac{n}{2}$.

Choosing ${}_na_x$

- Life expectancy estimates are particularly sensitive to very young and high mortality ages.
- Coale and Demeny fit a line to intertemporal and international data to give slightly better estimates than $\frac{n}{2}$:

	Males	Females
Value of ${}_1a_0$		
If ${}_1m_0 \geq .107$	0.33	0.35
If ${}_1m_0 \leq 0.107$	$0.045 + 2.684 \times {}_1m_0$	$0.053 + 2.800 \times {}_1m_0$
Value of ${}_4a_1$		
If ${}_1m_0 \geq .107$	1.352	1.361
If ${}_1m_0 \leq .107$	$1.651 - 2.816 \times {}_1m_0$	$1.522 - 1.518 \times {}_1m_0$

Open ended age interval

- Procedures for dealing with the open-ended interval are now important.
- In this interval, n is in effect infinity.
- The number of persons dying in the cohort above x must equal the number of persons surviving to x :

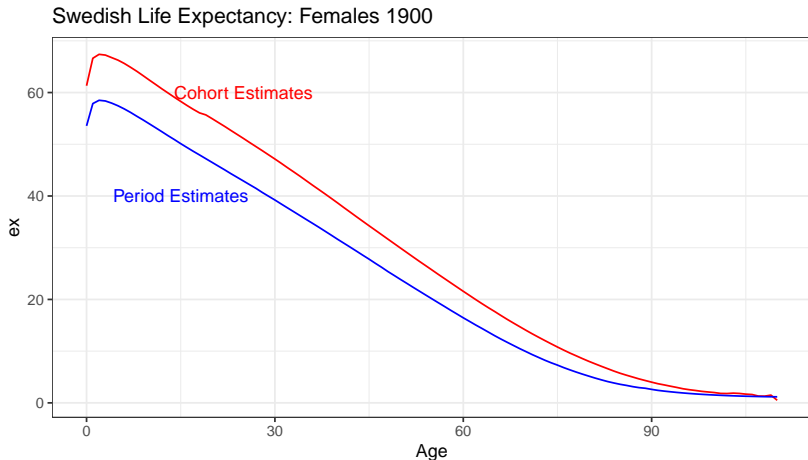
$${}_{\infty}L_x = \frac{l_x}{{}_{\infty}m_x} \quad (24)$$

- Naturally, we also have ${}_{\infty}q_x = 1.00$ and ${}_{\infty}p_x = 0$.
- **Class Quiz:** Why is the open interval not important for cohort life tables?

Our Period Life Table for Swedish Females 1900

Age	m_x	q_x	l_x	d_x	p_x	L_x	T_x	e_x
0	0.095	0.091	100000	9091.083	0.909	95454.46	5355760	53.558
1	0.028	0.028	90908.92	2522.683	0.972	89647.58	5260306	57.863
2	0.015	0.015	88386.23	1303.733	0.985	87734.37	5170658	58.501
3	0.010	0.010	87082.5	872.5276	0.990	86646.24	5082924	58.369
4	0.008	0.008	86209.97	726.2643	0.992	85846.84	4996278	57.955
5	0.007	0.007	85483.71	589.5076	0.993	85188.96	4910431	57.443
6	0.006	0.006	84894.2	490.1185	0.994	84649.14	4825242	56.838
7	0.005	0.005	84404.08	388.208	0.995	84209.98	4740593	56.165
8	0.005	0.005	84015.88	429.0629	0.995	83801.34	4656383	55.423
9	0.005	0.005	83586.81	376.9605	0.995	83398.33	4572581	54.705

Cohort vs. Period Life Expectancies



- Which is greater? Why? What is the 'gap'? Can we calculate the 'lag'?

Abridged Example: Our estimates vs HMD



- Where is the biggest divergence? Why?

Age–profile smoothing (rolling windows): overview I

- Input abridged exposures E_x and deaths D_x are aligned and crude age–specific death rates are computed as:

$$m_x = \frac{D_x}{E_x}.$$

- A *centered moving average on the log–scale* is applied to m_x across age:

$$\tilde{m}_x = \exp \left(\frac{1}{W} \sum_{k \in \mathcal{N}_w(x)} \log m_k \right), \quad W = |\mathcal{N}_w(x)|.$$

- Window size is a user option (`ma_window`, default = 5), with edge handling via reduced windows.
- Implemented in `make_lifetable()` with `pandas.rolling()` on $\log(m_x)$ then exponentiation back.

Age-profile smoothing (rolling windows): overview II

- The formula computes the **geometric mean** of the m_k values in the age window $\mathcal{N}_w(x)$ (of width W):

$$\tilde{m}_x = \left(\prod_{k \in \mathcal{N}_w(x)} m_k \right)^{1/W}.$$

- Averaging on the log scale:
 - Smooths the mortality curve by reducing random fluctuations.
 - Downweights extreme outliers compared to arithmetic averaging.
 - Preserves the multiplicative nature of mortality rates.
 - Guarantees $\tilde{m}_x > 0$ and approximately unbiasedness on the original scale.
- This produces a regularized age profile of mortality that is then used to derive smoothed q_x , p_x , and life table functions.

From smoothed m_x to q_x (life table pipeline)

- After smoothing $m_x \mapsto \tilde{m}_x$, the standard Greville–Chiang conversion is applied:

$$q_x = \frac{n \tilde{m}_x}{1 + (n - a_x) \tilde{m}_x}, \quad p_x = 1 - q_x.$$

- Default $a_x = n/2$ except for infancy, where special $a_0(m_0)$ rules are used.
- Survivorship and person–years then follow:

$$\ell_{x+n} = \ell_x p_x, \quad L_x = n \ell_{x+n} + a_x d_x, \quad T_x = \sum_{j \geq x} L_j, \quad e_x = \frac{T_x}{\ell_x}.$$

- Nonterminal $q_x < 1$, open interval $q_\omega = 1$, and checks for L_ω monotonicity are included.

What the “rolling window” actually does

- **Scope:** smoothing is *across ages* within a period (no temporal smoothing).
- **Scale:** operates on $\log m_x$ for multiplicative stability and positivity.
- **Centering:** the window is centered in age; at edges, it shrinks (`min_periods=1`).
- **Auditability:** the chosen window size is stored in output metadata.

Mortality improvements: parametric reduction factor

- Improvements are represented by a multiplicative factor $F(t) \in (0, 1]$ reducing baseline mortality after t years:

$$F(t) = \exp(-G S(t)), \quad G = -\log(1 - \text{improvement_total}).$$

- $S(t) \in [0, 1]$ is a smooth convergence profile reaching ≈ 1 at $t = \text{convergence_years}$.
- $F(t)$ is applied to projected mortality surfaces by geography and sex.

Two smoother choices for $S(t)$

Exponential:

$$S_{\text{exp}}(t) = 1 - \exp(-\kappa t), \quad \kappa = \frac{-\log(1 - \text{converge_frac})}{\text{convergence_years}}.$$

Logistic:

$$S_{\text{log}}(t) = \frac{1}{1 + \exp\left(-s\left(\frac{t}{\text{conv}} - \text{mid_frac}\right)\right)}.$$

Parameters: midpoint fraction `mid_frac` and steepness (derived if absent).

Putting it together in the PyCCM pipeline

1. Build life tables per period using rolling smoothing across age on $\log m_x$.
 2. For projection year t , compute $F(t)$ via the selected smoother and parameters.
 3. Apply $F(t)$ as a scalar mortality reduction by year, geography, and sex.
- **Note:** the code does not perform rolling smoothing across time; age-smoothing only, with temporal change handled by $F(t)$.

Quick Review

Lets quickly review all of the columns in our lifetables before we move on:

- x : exact age
- l_x : number of people alive at x
- ${}_n d_x$: Number dying between x and $x + n$
- ${}_n q_x$: Probability dying between x and $x + n$
- ${}_n p_x$: Probability surviving between x and $x + n$
- ${}_n L_x$: PY lived between ages x and $x + n$.
- T_x : PY lived above x
- e_x : Life expectancy at age x
- ${}_n m_x$: ASDR between x and $x + n$
- ${}_n a_x$: Average PY lived between x and $x + n$ for those dying.

Quick Review (Cont.)

Observed Data:

- ${}_nN_x$: mid year population in age interval x to $x + n$.
- ${}_nD_x$: deaths between ages x and $x + n$ during the year.

Steps for period life table construction:

1. ${}_nM_x$: $\frac{{}_nD_x}{{}_nN_x}$
2. ${}_na_x$: calculated or pre-specified.
3. ${}_nq_x = \frac{n \times {}_n m_x}{1 + (n - {}_n a_x) \times {}_n m_x}$
 ${}_nq_{85}$: 1.00
4. ${}_np_x = 1 - {}_nq_x$
5. l_0 : 100,000
 l_{x+n} : $l_x \times {}_np_x$
6. ${}_nL_x$: $(n \times l_{x+n}) + ({}_na_x \times {}_nd_x)$
 ${}_nL_x$: $\frac{l_x}{{}_nm_x}$
7. ${}_nd_x$: $l_x - l_{x+n}$
8. T_x : $\sum_n {}_nL_a$
9. e_x : $\frac{T_x}{l_x}$

Stationary Populations

- A stationary population is one that meets the following conditions:
 1. Age-specific death rates are unchanging over time.
 2. A flow of births into the population is unchanging over time.
 - The same number of newborns are added into the population.
 - Constant death rates implies number of deaths are the same.
 - This implies that $R = 0$.
 3. The population has no net migration.

The Life Table as a Stationary Population

- The life table can be conceived of as a stationary population.
- If the conditions of a stationary population hold, then we can relate different demographic parameters surrounding mortality and fertility to one another.
- Crude Death Rate = Crude Birth Rate = $\frac{1}{e_x^0}$
- It connects a period fertility measure to a cohort mortality measure.
- In a stationary population, the mean age of people dying in a particular year would equal the life expectancy at birth.

Modelling Duration Data

- The life table presents one formulation with dealing with duration data where we are interested in modelling the time to a specific event.
- Wider branch of statistics that deals with duration data: Survival Analysis.
- In sociological applications, researchers are interested in modelling sequences of events as a part of a life course.
- Another way to express the pace of mortality in a cohort, whether actual or synthetic, is through the hazard rate, closely related to the, ${}_n m_x$ function.
- If we let the age interval n grow smaller and smaller, we obtain the hazard, $h(x)$ or sometimes called $\mu(x)$: ‘the force of mortality’.

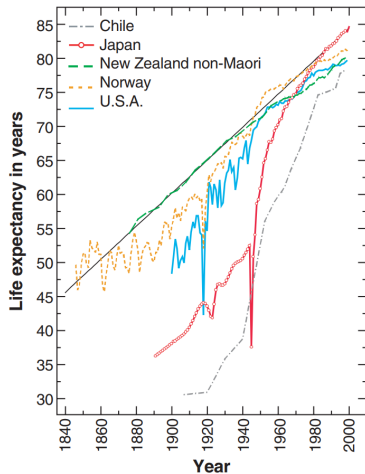
Mortality Transition and the Epidemiological Transition

- So far we have been talking about mortality transitions purely in terms of age, and falling death rates and rising life expectancy.
- Omran (1971) argued for the importance of including morbidity in the story of the demographic transition: calling it the 'epidemiological transition'.
- Divided into three stages:
 1. **Stage 1** (pre-demographic transition): age of pestilence and famine → high mortality rates with oscillations, very low life expectancy
 2. **Stage 2** (beginning of the demographic transition): age of preceding pandemics → epidemics and infectious diseases become less frequent, degenerative diseases start rising.
 3. **Stage 3** (end of the demographic transition): age of degenerative and man-made diseases at very low levels of mortality.

Epidemiological Transition: The Fourth Stage

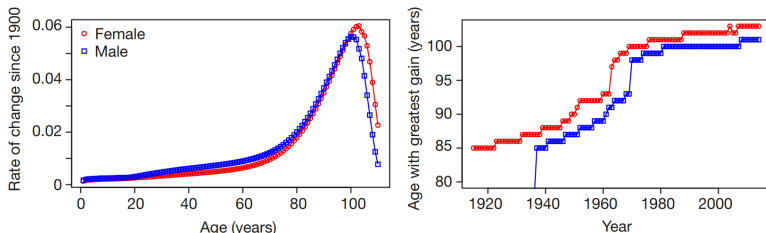
- Olshansky and Ault (1986) speak of the emergence of a **fourth stage**: the age of delayed degenerative diseases. There are (at least) two conflicting views on the fourth stage.
- They have to do with the possible limits to human longevity. To what extent do they exist?
- Fries (1980): compression of morbidity/rectangularization. Life expectancy free of disease expands and diseases are compressed. There is a rectangularization of survival curves, and there is a limit (e.g. because of genetic endowment)
- Variability in age at death becomes lower.
- Following from the idea there has been no increase in the maximum age at death in human populations

Optimists (Oeppen and James W.Vaupel)



- “The notion of a fixed life-span evolved into a belief in a looming limit to life expectancy”
- Limits have been ‘broken’, on average 5 years after publication
- Best-practice life expectancy has increased by 2.5 years per decade for a century and a half
- If life expectancy were close to a maximum, then the increase in the record expectation of life should be slowing.
- The apparent leveling off of life expectancy in various countries is an artifact of laggards catching up and leaders falling behind

Pessimists (Dong et al. 2016)



- However, note that there are people more, and less optimistic about 'plateauing' in life expectancy!
- This *might* be one of the topics which we offer in Assignment One, or one you might like to propose.

Conclusion

- In this lesson we mostly focused on lifetables and their construction.
 - Note: we have only been dealing with a *single decrement* process.
 - Individuals (alive) have only one recognized mode of exit (death).
- Now, on to the computer lab:
 1. Take the mortality rates we computed yesterday.
 2. Build them into full lifetables.
 3. Plot this for multiple DPTOs (in a loop or otherwise).
 4. Adjust your code to smooth (using PyCCM functionality or otherwise).
 5. Adjust your code to improve (using PyCCM functionality or otherwise).