MPIDR-NES Training Programme

Moscow, New Economic School, 14th January - 1st February 2013

Population and Health

Лекция 1: События, длительности, сетка Лексиса, коэффициенты
Lecture 1: Events, durations, Lexis diagram, rates

MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH
MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG

РЭШ
Российская экономическая школа

States and events

States

At every time people constituting a population can be described by certain characteristics (being in certain *states*).

- **Demographic states:** being alive, living alone or together with partner, being married or divorced, being a parent, living in country of origin or in a host country.

- **Socio-economic states:** having certain level of education, social position, profession, level of wealth, job etc.

- **Health states:** being in good (bad) health, being free of disease or of a specific disease, being able (unable) to perform normal daily activities, being able to see (hear, feel), having good (bad) memory.

Events

Events designate transitions from one state to another.

<u>Examples:</u> birth, death, marriage, divorce, conception, childbearing, getting (losing) a job, contracting a disease, being diagnosed with a disease, recovering from disease, getting handicapped, losing acuity,...

Concept of „event“ is based on assumption that there is a specific moment in time at which this event occurs. However, transitions between some states are fuzzy in respect to time (developing of a chronic disease, gaining or losing health) and also difficult to register.

In these cases, one is restricted to analysis of population *prevalence* of respective states.

Classification of events-1

<u>Repeatable and non-repeatable events</u>.

Death can be experienced only once.

Motherhood and *migration* are repeatable. They can be considered as non-repeatable for a fixed rank of event. It is possible to give birth to a first child and to enter parity one only once. Only women of parity 0 can pass to parity 1, women of parity 1 to parity 2 etc.

Some events are hardly classifiable on a one-dimensional ordinal scale. Work careers: waiter-student-engineer-businessman. Some rankings are still possible: jobs by level of salary or by level of qualification, grades of civil service etc.

Classification of events-2

<u>Excluding and non-excluding events</u>.

Some events like *out-migrations or deaths* **withdraw subjects from the population.**
Births and marriages do not have this property. Therefore, they can be studied retrospectively.

This difference implies an important distinction for types of event-rates: **event frequencies for non-excluding events** and **occurrence-exposure rates for excluding events**.

Concerns for *retrospective studying of non-excluding events* (births, marriages, illnesses):
- misreporting (or “forgetting”);
- selectivity of death or migration in respect to rank of the event in interest (number of past births, migrations, diseases).

Absorbing and transient events.

Absorbing state (AS) is a state such that once entered, it can not be left (backward transition is impossible). *Absorbing event* is a transition to AS. Death. Entering second parity (parity can only become higher), marriage (once married one can not return to being “never married”), getting a non-curable disease etc.

Transient events correspond to two-way transitions. From good health to illness and back. Out- and in-migrations.

Demographic models often assume absorbing events even if in the reality back transitions are possible. Back transitions are often hardly observable in demographic surveillance.

Classification of events-4

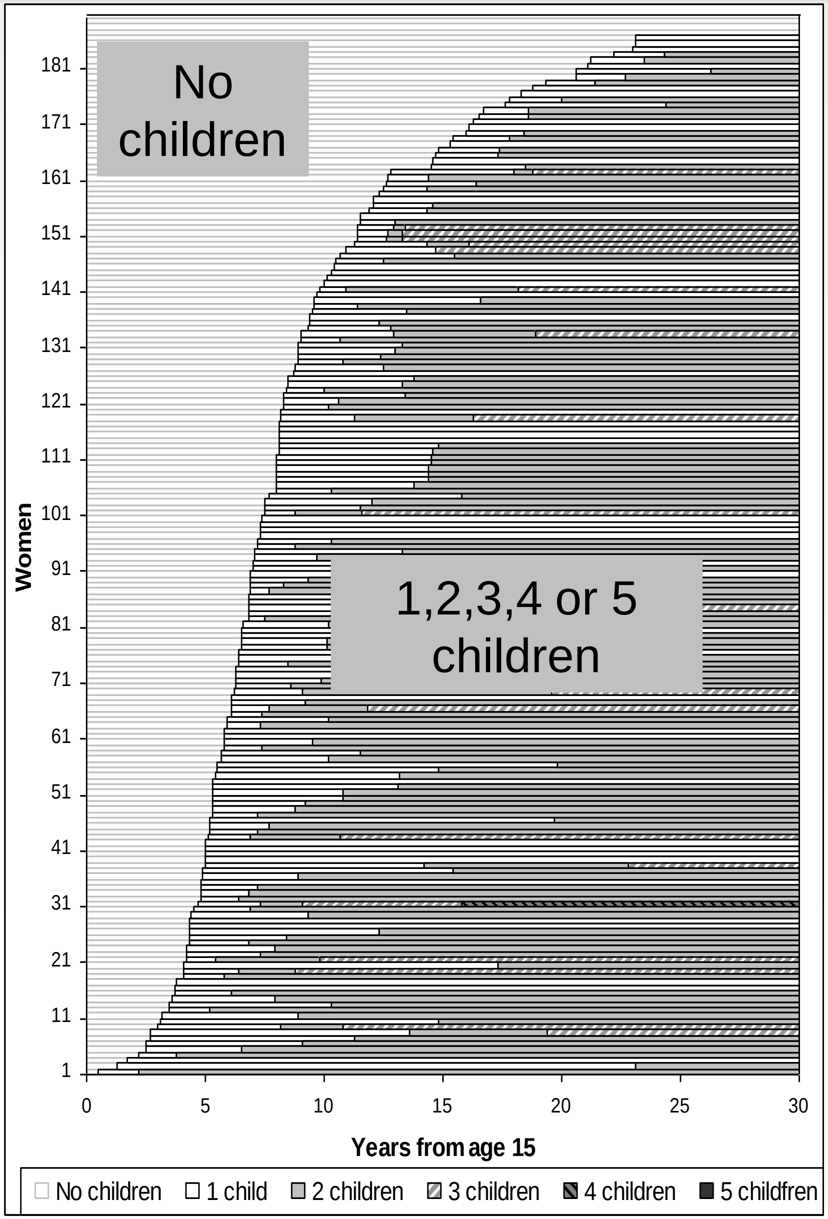
<u>Possible and fatal events</u>.

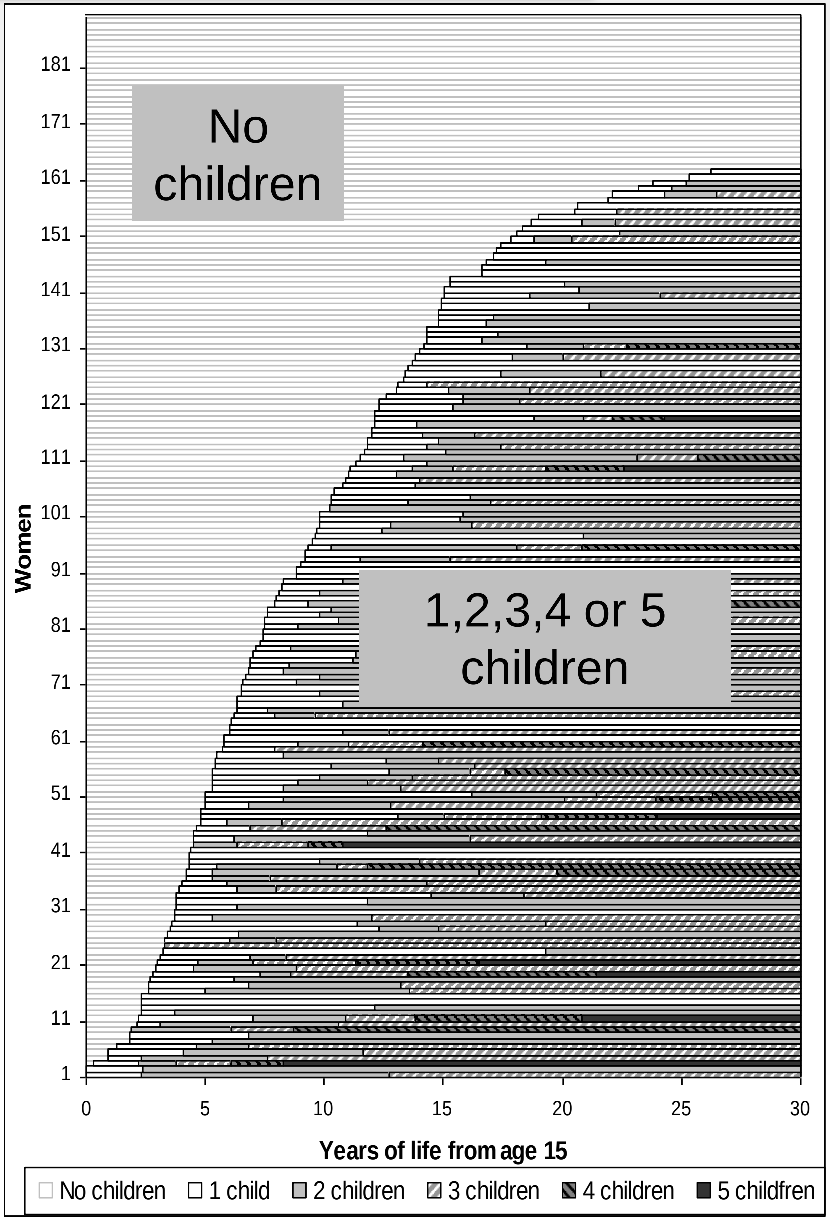
Death is not only absorbing, but also fatal. All people will eventually die. During every year of life death is *possible*. At the scale of one‘s lifetime, the odds of death is 100%. From this viewpoint, death is *fatal*.

Many other events are inevitable conditioned on other events. All women become sterile after reaching certain age (menopause). No one can avoid biological degeneration due to aging ...

Life trajectories, durations, and events

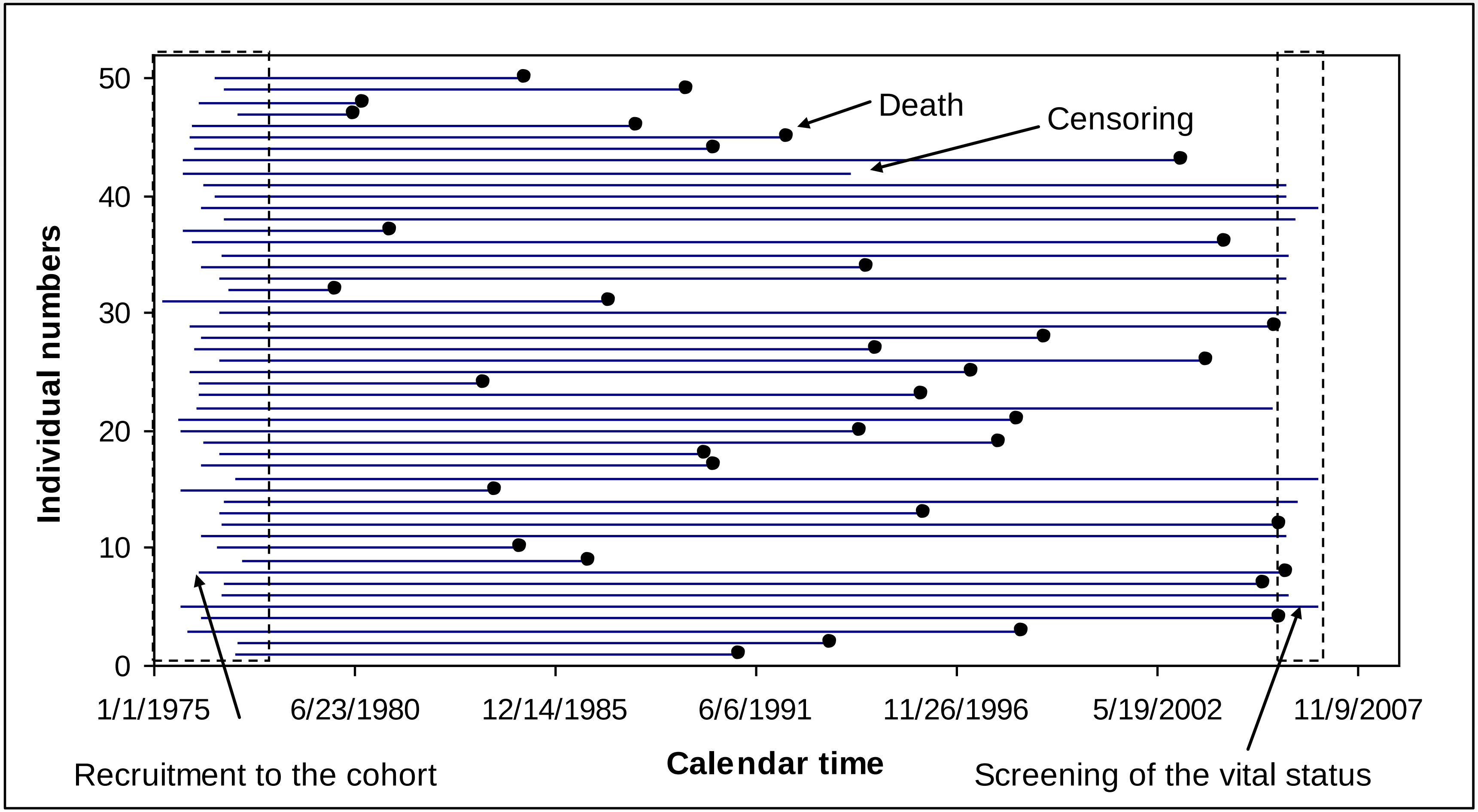
States and events. Maternity trajectories of 200 Bulgarian (left panel) and 200 US women (right panel)





Source: Houle, Shkolnikov 2002

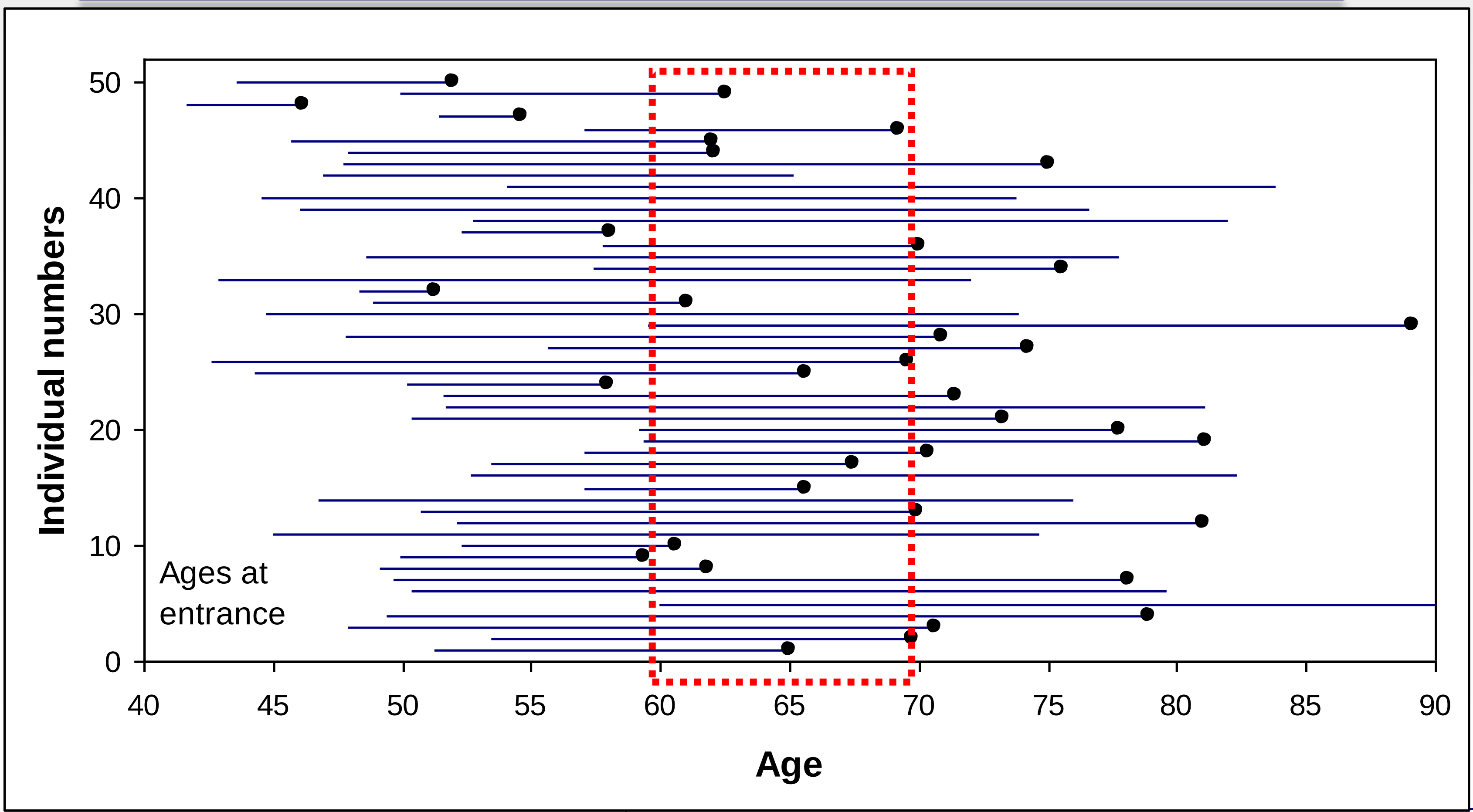
Example: observation on survival of 50 individuals over time, 1975-77 to 2006



Duration: time

Initial population: N=50
Deaths: D=38

Example: survival of the same 50 individuals over age: from age 40 to 59



Duration: age

Age=(CurDate-BirthDate)/365.25

For the whole region of observations:
Deaths=38
Population=50
Population exposure=987 person-years.
Mortality rate: M=38/987=0.036 or 36 per 1000

Occurrence-exposure rates

Population attrition and survival functions

Let *f(t)* be a probability density function expressing the probability that event (say death) happens to an individual in a very short interval

$$f(t) = \lim_{\Delta t \to 0} \frac{\Pr(t < T < t + \Delta t)}{\Delta t}$$

Then the probability distribution function is

$$F(t) = \Pr(T < t) = \int_0^t f(\tau) d\tau$$

The hazard function is defined as a probability of death of an individual in a very short time interval given survival to the beginning of this interval.

The survival function: $S(t) = \Pr(T \geq t) = 1 - F(t)$

The hazard function: $h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t < T < t + \Delta t)}{S(t)\Delta t}$

In mortality studies, this measure is also called force of mortality or instant death rate.

Hazard, probability and occurrence-exposure rate (mortality rate)

From the previous formula it follows that

$$h(t) = -\frac{1}{S(t)} \lim_{\Delta t \to 0} \frac{S(t+\Delta t) - S(t)}{\Delta t} = -\frac{dS(t)}{S(t)dt} = -\frac{d(\ln S(t))}{dt}$$

Consequently: $$S(t) = S(0) \cdot e^{-\int_0^t h(\tau)d\tau}$$

For two moments of time *t* and *t+a*: $$S(t+a) = S(t) \cdot e^{-\int_t^{t+a} h(\tau)d\tau}$$

Probabilities of survival and death between *t=0* and *t* $$p_t = \frac{S(t)}{S(0)} = e^{-\int_0^t h(\tau)d\tau}, q_t = 1 - p_t$$

For a time interval, mortality rate *M(t,t+a)* is defined as an average hazard within the interval.

$$S(t+a) = S(t) \cdot e^{-M(t,t+a)\cdot a}$$

Mathematics: mortality rate as a mean hazard

M(t,t+a) (or ${}_aM_t$) is mortality rate. It is defined as a mean hazard (force of mortality) over a time interval *(t, t+a)*.

By definition, death hazard corresponds to a very short time interval.

Mortality rate can be computed for any interval of any length.

$$\int_t^{t+a} h(\tau)d\tau = {}_aM_t \cdot a$$

${}_aM_t$ = Number of observed events (deaths) relative to person-years of exposure

$$S(t+a) = S(t) \cdot e^{-M(t,t+a)\cdot a}$$

Life trajectories and occurrence-exposure tables

For certain reasons, one might want to transform the original individual-level data containing individuals‘ dates of birth, entrance, out-migration, death, and censoring into occurrence-exposure table that has year-age cells with numbers of death-events and amounts of exposure-time in them.

These two ways of presenting the data are in many ways equivalent. While the individual trajectories can be analyzed by means of proportional hazard or event-history regression models, the event-exposure data can be analyzed by means of Poisson regression models.

Why one might want to do it ?

Individual life trajectories and occurrence-exposure table

Imagine a list based on observation of many individuals during a long time. For each individual dates of birth, beginning of follow-up (entrance), death, and end of follow-up (censoring) are known. One has to produce two matrices corresponding to experience of this population during the observation period: matrix of events (death counts) and matrix of population-exposure (person-years).

Royal Society (Academy of Sciences in England)

INPUT

ID	MonthBirth	YearBirth	MonthEntrance	YearEntrance	MonthDeath	YearDeath	DayCensoring	MonthCensoring	YearCensoring
5	10	1757	2	1793	5	1829	1	7	2007
7	7	1756	6	1793	9	1794	1	7	2007
....	....	....	....	....	....	....	....	....	....
21260	1	1917	3	1955			1	7	2007

OUTPUT1:

	Year			
Age	1750	1751	…..	2006
20				
21		E(Age, Year) - population exposure		
22				
….				
100				

OUTPUT2:

	Year			
Age	1750	1751	…..	2006
20				
21		D(Age, Year) - death events		
22				
….				
100				

Automating the transformation

The algorithm for transformation of individual life histories into event-exposure tables is simple. The program will have to run across individuals. For each individual, it runs across calendar years of his life and adds exposure-time contributions and 0/1 death counts to appropriate age-year cells of the *E* and the *D* matrices.

It should be possible also to organize an opposite process. One could run across age-year cells, find individuals who *contribute* to these cells, and calculate their contributions in terms of person-years and deaths.

Lexis diagram, routine tables, and death rates

Registration of population and demographic events in statistical tables

Routine population statistics is always produced by summation of individual records. However, it would be impossible to publish tabulations across detailed ages and calendar time (say on daily or monthly basis).

Individual census records $\Rightarrow$ Age-specific population counts at the census date

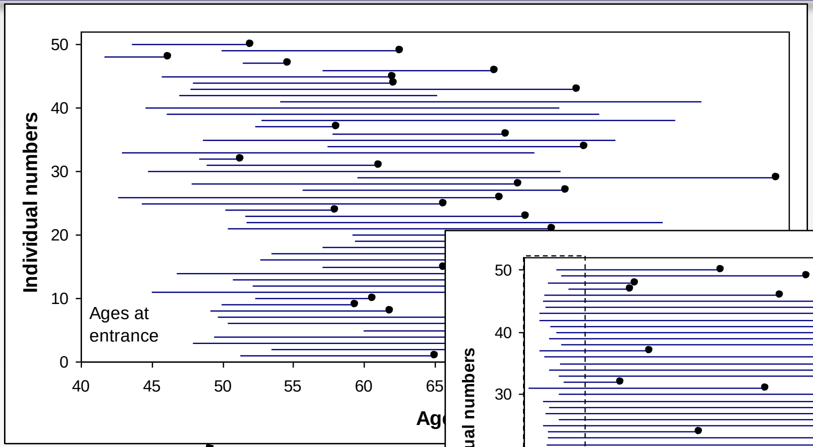
Exact date of death and exact date of birth $\Rightarrow$ Approximate age at death in completed years

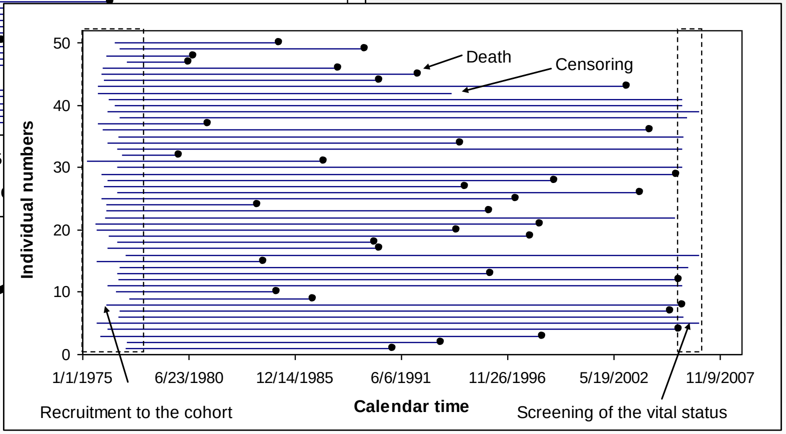
Exact date of birth and exact date of mother's birth $\Rightarrow$ Approximate mother's age at birth in completed years

Lexis diagram is a useful tool for looking at event counts, population exposures and population sizes in terms of integer ages and calendar years.

Lexis diagram

Lexis diagram invented (promoted?) by Wilhelm Lexis (1875).
It shows demographic events and sets of people classified by time, age, and year of birth (cohort).





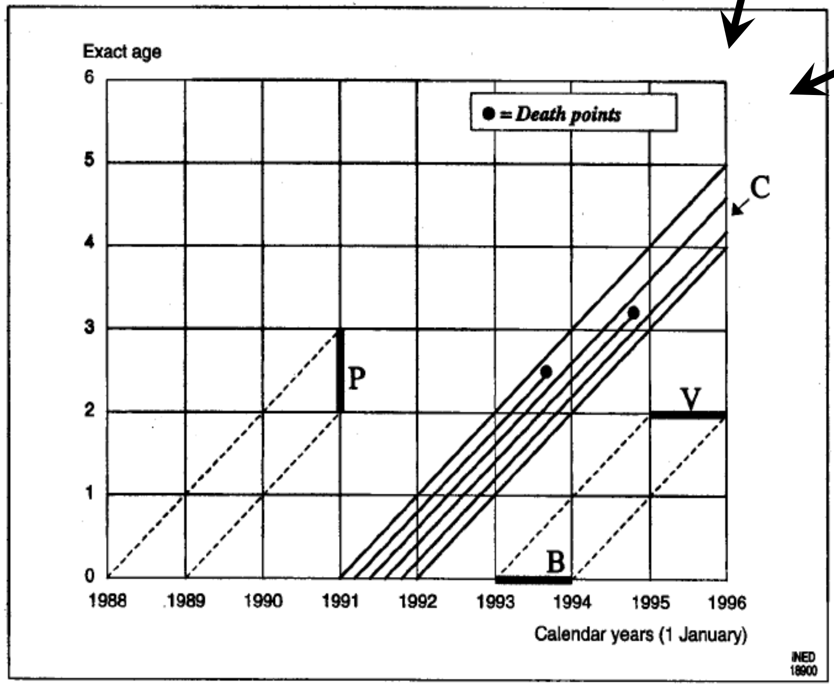


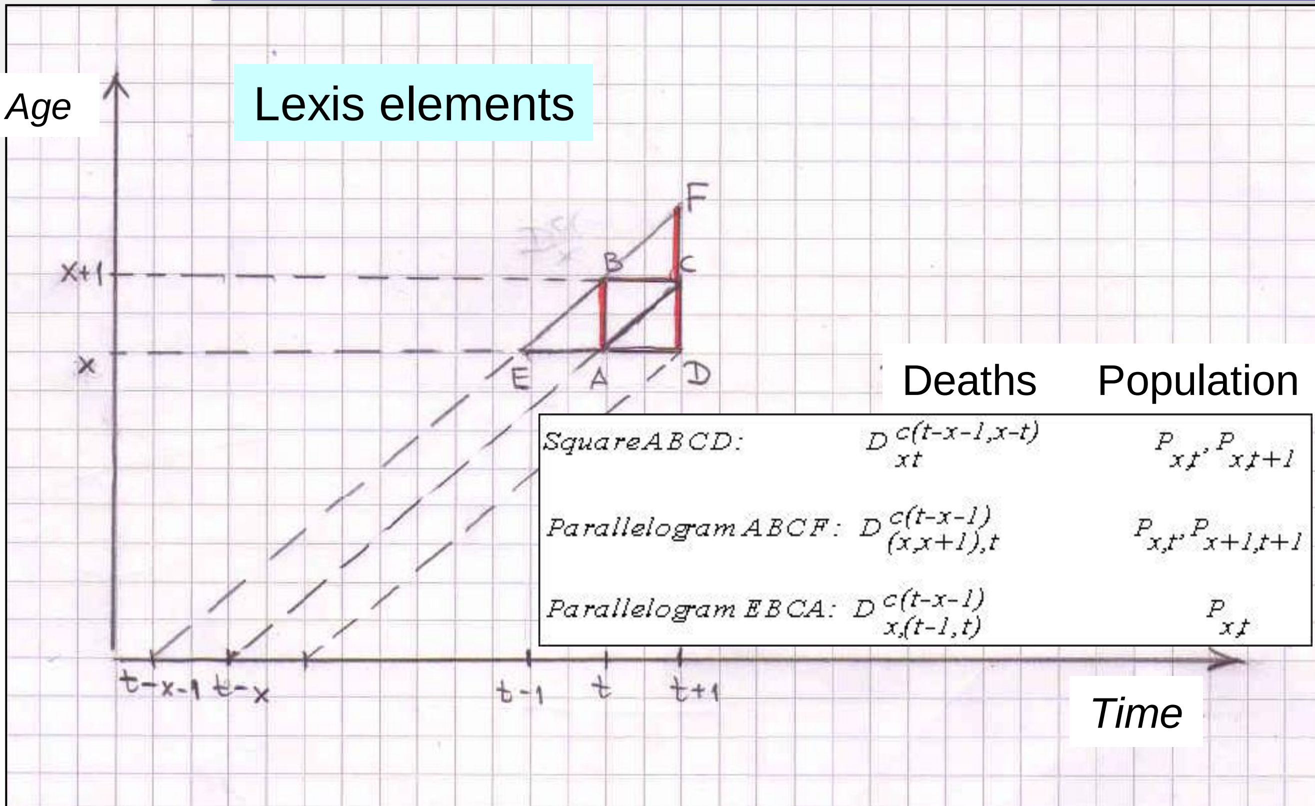
FIGURE 6–4 Lexis grid with the births of 1 year, the population by age, and the column of lifelines of a birth cohort.

Life trajectories of a birth cohort.

P - population being at age 2 on Jan 1, 1991.

B - people born in 1993, V are people of the same cohort at exact age two.

Classification of events and populations according to age, calendar year, and year of birth



	Deaths	Population
Square ABCD:	$D^{c(t-x-1,x-t)}_{xt}$	$P_{x,t}, P_{x,t+1}$
Parallelogram ABCF:	$D^{c(t-x-1)}_{(x,x+1),t}$	$P_{x,t}, P_{x+1,t+1}$
Parallelogram EBCA:	$D^{c(t-x-1)}_{x,(t-1,t)}$	$P_{x,t}$

Three types of Lexis elements as a basis for calculation of rates

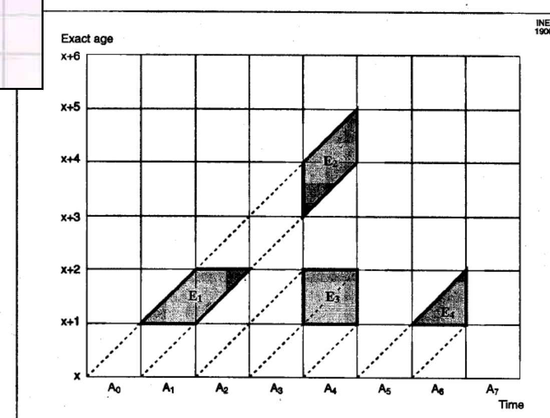
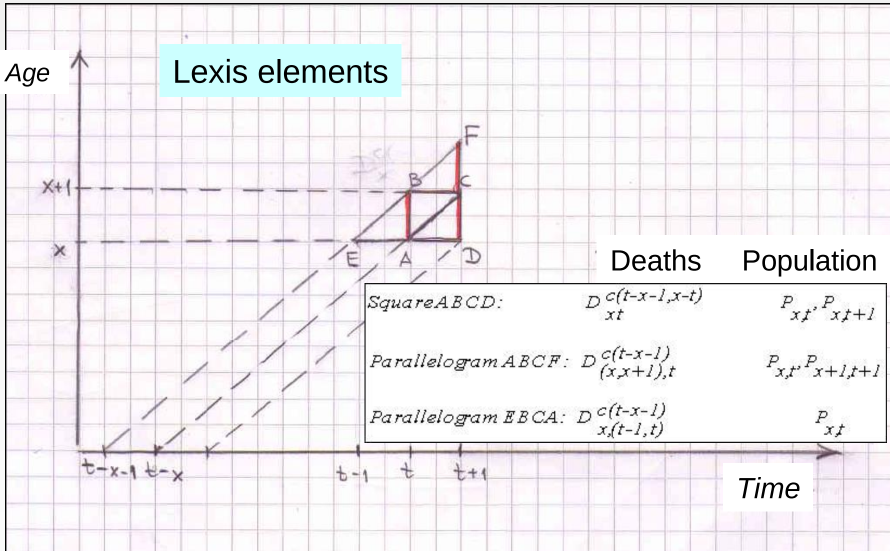


FIGURE 6–5 Representation of events on the Lexis diagram according to different possible annual classification modes.

Source: Caselli, Vallin, 2002

Rates corresponding to one-year Lexis elements



Type 1: $$M_{x,t} = \frac{D_{x,t}}{0.5(P_{x,t} + P_{x,t+1})}$$

Type 2: $$M_{(x,x+1),t}^{C(t-x-1)} = \frac{D_{(x,x+1),t}^{C(t-x-1)}}{0.5(P_{x,t} + P_{x+1,t+1})}$$

Type 3: $$M_{x,(t-1,t)}^{C(t-x-1)} = \frac{D_{x,(t-1,t)}^{C(t-x-1)}}{P_{x,t}}$$

Infant mortality „rate“ (death probability):

$$IMR_{0,t} = \frac{D_{0,t}}{B_t}$$

B_t – births in year t

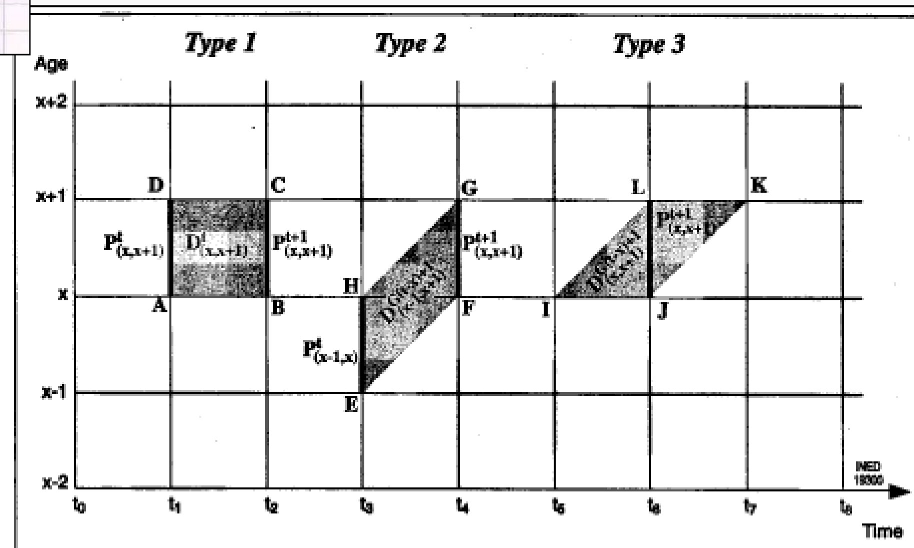
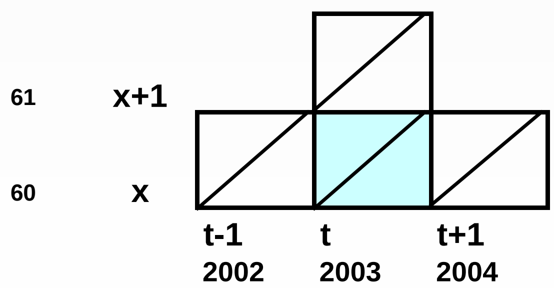


FIGURE 6–8 Representation of the elements required for computing a death rate by age according to the classification mode of the deaths.

Mortality rates by one-year age group for the three types of Lexis elements

West Germany, Males

Age	Death						Population		Rates (for 1000) in 2003					
	2002		2003		2004		2002	2003	Type 1	Type 2		Type 3		
x	C	C-1	C	C-1	C	C-1			Lexis square	x	(x+x+1)/2	2002-03	2003-4	Average
60	2173	2685	2120	2119	2138	2129	463989	385332	9.98			9.25	11.03	10.14
61	2735	2876	2270	2824	2221	2172	489987	458191	10.74	9.52	10.69	11.34	9.69	10.52
62	3360	3277	2949	3172	2421	2915	478015	483349	12.73	11.86	12.88	13.66	12.13	12.90
63	3524	3314	3418	3393	2996	3269	445364	470878	14.87	13.89	14.96	15.53	14.20	14.87
64	3466	3220	3694	3511	3600	3519	413812	438317	16.91	16.04	16.71	16.86	16.46	16.66
65	3605	3535	3622	3648	3866	3605	397057	406642	18.09	17.39	18.28	18.27	17.77	18.02
66	3645	3688	3895	3781	3800	3600	377339	389616	20.02	19.18	19.95	19.68	19.24	19.46
67	4118	4071	3957	4072	3797	3829	346253	369300	22.44	20.73	22.77	23.65	21.08	22.37
68	3955	3229	4417	4231	4131	3909	276267	337993	28.16	24.81	27.65	29.63	24.63	27.13
69	3582	3484	4085	3552	4276	4183	269775	269206	28.34	30.49	28.71	26.44	30.71	28.58
70	3941	3881	3610	3766	4319	3619	269242	262190	27.76	26.93	28.23	28.62	27.57	28.10
71	4356	4559	4064	4139	3645	3786	278063	260934	30.44	29.53	30.64	30.55	30.08	30.32
72	4907	4802	4537	4956	4040	4140	263652	268567	35.68	31.74	35.30	37.41	32.31	34.86
73	5212	5135	5093	5008	4620	4719	257257	253596	39.54	38.86	39.85	39.73	38.69	39.21
74	5463	4781	5282	5492	5118	5027	224980	246693	45.68	40.83	45.53	48.69	41.79	45.24
75	5073	4717	5560	5144	5408	5362	203364	215177	51.15	50.22		50.24	50.76	50.50



Lengths of intervals over age and time

Rates over five-year age groups

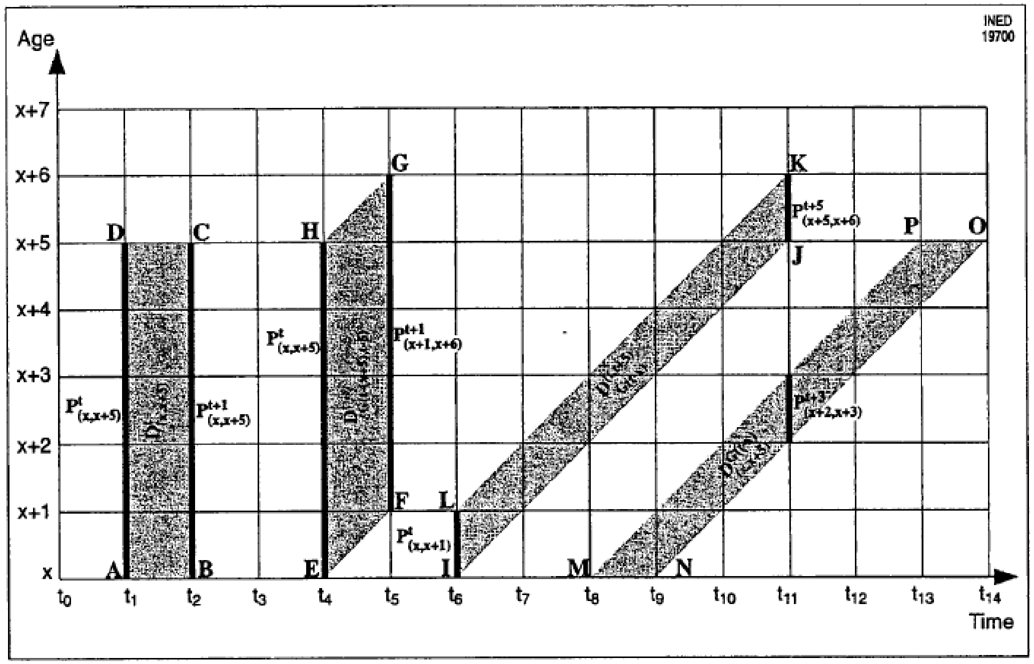


FIGURE 6–12 Representation of the elements needed for the calculation of death rates by age according to how deaths are classified.

Source: Caselli, Vallin, 2002

Age is split into 5-year groups.

During the first year of life the risk of death rapidly decreases and after age 1 pace of the decrease greatly slows down. To capture these quick changes the first 5-year group 0-4 is usually split into two age groups, 0 and 1-4.

Five-year (abridged) age scales are often used for compact presentation of the data especially in analyses of mortality by cause.

Type 1 (ABCD)

$${}_5M_{x,t} = \frac{{}_5D_{x,t}}{0.5(P_{(x,x+5),t} + P_{(x,x+5),t+1})}$$

Type 2 (EHGF)

$${}_5M^{C(t-x-5,t-x)}_{x,t} = \frac{{}_5D^{C(t-x-5,t-x)}_{x,t}}{0.5(P_{(x,x+5),t} + P_{(x+1,x+6),t+1})}$$

The end