Measures of disease frequency and effects
Analysis of epidemiological data

Esa Läärä
University of Oulu, Finland
esa.laara@oulu.fi http://stat.oulu.fi/laara

Bendix Carstensen
Steno Diabetes Center, Denmark
& Department of Biostatistics, University of Copenhagen
bxc@steno.dk www.biostat.ku.dk/~bxc

Nordic Summer School in Cancer Epidemiology
August 2011, Danish Cancer Society, Copenhagen
www.biostat.ku.dk/~bxc/NSCE

Outline

Introduction

Basic concepts

Frequency

Comparison

Age, period, etc.

Standardization

Survival

Conclusion

Introduction
Measures of Disease Occurrence

Bendix Carstensen & Esa Läärä

Nordic Summerschool of Cancer Epidemiology
15-26 August 2011
Copenhagen
www.biostat.ku.dk/~bxc/NSCE

Key references

IS: dos Santos Silva, I. (1999).
Cancer Epidemiology: Principles and Methods.
International Agency for Research on Cancer (IARC), Lyon.

B&D: Breslow, N.E., Day, N.E. (1987).
Statistical Methods in Cancer Research Volume II – The Design and Analysis of Cohort Studies.
IARC Scientific Publications No. 82, IARC, Lyon.

C&H: Clayton, D., Hills, M. (1993).
Statistical Models in Epidemiology. OUP, Oxford.

Internet resources on cancer statistics

NORDCAN : Cancer Incidence and Mortality in the Nordic Countries, Version 4.0. Association of Nordic Cancer Registries, Danish Cancer Society, 2002.
http://www-dep.iarc.fr/nordcan.htm
NORDCAN is a graphical package providing data on the incidence of, and mortality from 40 major cancers for 80 regions of the Nordic countries (Denmark, Finland, Iceland, Norway and Sweden). Using NORDCAN, these data can be presented as a variety of tables and graphs that can be easily exported or printed. NORDCAN allows countries and cancer sites to be grouped and compared as desired.

GLOBOCAN 2008 : Cancer Incidence and Mortality Worldwide in 2008 http://globocan.iarc.fr/

Basic Concepts
Measures of Disease Occurrence

Bendix Carstensen & Esa Läärä

Nordic Summerschool of Cancer Epidemiology
15-26 August 2011
Copenhagen
www.biostat.ku.dk/~bxc/NSCE

What is Epidemiology?

Some textbook definitions of epidemiology:
Greek: *epi* = upon, *demos* = people

- “study of the **distribution** and **determinants** of disease **frequency** in man” (MacMahon and Pugh, 1970)
- “study of the distribution and determinants of health related **states** and **events** in specified populations,. . . ” (Last (ed.) Dictionary of Epidemiology, 2000)
- “discipline on principles of **occurrence** research in medicine” (Miettinen, 1985)

Different epidemiologies

- **descriptive** epidemiology
— monitoring & surveillance of diseases for planning of health services
— a major activity of cancer registries.
- **etiologic** or “analytic” epidemiology
— study of cause-effect relationships
- **disease** epidemiologies — *e.g.* of cancer, cardiovascular diseases, infectious diseases, musculoskeletal disorders, mental health, . . .
- **determinant-based** epidemiologies — *e.g.* occupational epidemiology, nutritional epidemiology, . . .
- **clinical** epidemiology — study of diagnosis, prognosis and effectiveness of therapies in patient populations
— basis of evidence-based medicine

Frequency (from Webster's Dictionary)

Etymology: *frequentia* = assembly, multitude, crowd.

1. Also, **frequency**. the state or fact of being frequent; frequent occurrence. We are alarmed by the frequency of fires in the neighborhood.
2. Rate of occurrence:
   The doctor has increased the frequency of his visits.
3. *Physics*: number of periods or . . . regularly occurring events . . . of any given kind in unit of time, usually in one second.
4. *Math*: the number of times a value recurs in a unit change of the independent variable of a given function.
5. *Statistics*: the number of items occurring in a given category. Cf. **relative frequency**.

Meanings 2 and 5 are both relevant in epidemiology.

But what is "rate" and "occurrence"?

Cancer i Norden 1997 (NORDCAN)

Frequency of cancer (all sites excl. non-melanoma skin) in Nordic male populations expressed by different measures:

	New	Crude	ASR	Cumul.	
	cases	rate	(World)	risk	SIR
Denmark	11,787	452	281	27.8	104
Finland	10,058	401	269	26.5	101
Iceland	633	464	**347**	**32.6**	**132**
Norway	10,246	**469**	294	29.4	109
Sweden	**19 908**	455	249	25.4	93

- Where is the frequency truly **highest**, where **lowest**?
- What do these measures mean?

Questions on frequency & occurrence

How many women in Denmark:

- are carriers of breast cancer today? — **prevalence**
- will contract a new breast ca. during 2007? — **incidence**
- die from breast ca. in 2007? — **mortality**
- will be alive after 5 years since diagnosis among those getting breast ca. in 2007? — **survival**
- are cured from breast cancer during 2007? — **cure**

Questions on frequency & occurrence

- What is the relative frequency or/and rate of occurrence of these states and events?
- How great are the **risks** of these events?
- Is the frequency/occurrence/risk of breast cancer greater among nulliparous than parous women?
- What are the **excess** and **relative risks** for nulliparous compared to parous women?
- What is the **dose-response relationship** between occupational exposure to crystalline silica and the risk of getting lung cancer in terms of level and length of exposure?

What is risk?

What do we mean by "risk of disease S"?

(a) **probability** of *getting* S during a given **risk period**
→ **incidence** probability,

(b) **rate** of change of that probability
→ **hazard** or intensity,

(c) **probability** of *carrying* S at a given *time point*
→ **prevalence** probability.

Most commonly meaning (a) is attached with risk.
NB: "Risk" should not be used in the meaning of **risk factor**
However, in **risk assessment** literature: "hazard" is often used in that meaning. In statistics, though, hazard refers to notion (b): change of probability per unit time.

Risks are conditional probabilities

- There are no "absolute risks".
- All risks are conditional on a multitude of factors, like
  - length of risk period (*e.g.* next week or lifetime),
  - age and gender,
  - genetic constitution,
  - health behaviour & environmental exposures.
- In principle each individual has a "personal" value for the risk of given disease in any defined risk period, depending on his/her own risk factor profile.
- Yet, these individual risks are latent and unmeasurable.
- **Average risks** of disease in large groups sharing common characteristics (like gender, age, smoking status) are estimable from appropriate epidemiologic studies by pertinent **measures of occurrence**.

Types of epidemiologic studies

Can crudely be classified along the following axes:

- *study question*: descriptive ↔ causal
- *study unit*: individual ↔ aggregate (ecological study)
- *allocation of exposure*: experimental ↔ observational
- *population*: closed (cohort) ↔ open (dynamic)
- *dimensionality*: cross-sectional ↔ longitudinal
- *timing of observations*: concurrent ↔ historical ("pro-" vs. "retrospective")
- *sampling of exposure data*: cohort ↔ case-control

Focus in this course: *observational*, and *longitudinal cohort and case-control* studies.

Descriptive and causal questions

Descriptive: What is the occurrence of outcome C in different population groups.
— Medical demography

Descriptive (II) — groups defined *e.g.* by exposure to a determinant or risk factor X?

Causal (also **etiological** or "analytical"): What is the occurrence of outcome C in a population exposed to risk factor X as compared to . . . what the occurrence in the same population *would have been, if not* exposed?

N.B.: Causal question — *counterfactual conditional*!

Challenge: How to find a *comparable* group of unexposed?

Experimental and observational studies

Allocation of exposure in etiologic studies?

• **Experimental**: Exposure controlled by investigators, its levels being **randomized** among the study subjects.

+ **Comparability** of exposure groups.
+ Feasible in clinical and preventive trials.
– Ethically impossible for hazardous exposures.

• **Observational**: Exposure imposed by the own behaviour of the subjects themselves & and by their environment.

– Possibility of **confounding**: due to other determinants of the outcome, correlated with exposure.
* Challenges: **Valid**: and **efficient** non-randomized design and statistical analysis.

Experimental and observational studies

Allocation of exposure or risk factor in causal studies?

Experimental (Intervention trial): Exposure is controlled by investigators; its levels are allocated among recruited subjects by **randomization**,

⇒ **comparability** of exposure groups.

Observational: Exposure imposed by own behaviour of study subjects and/or by their environment,

⇒ possibility of **confounding** due to other determinants.

Time dimensionality of a study

Cross-sectional: Outcome *status* and its *prevalence* in population at given *time point* are studied, *e.g.*

- number of Danish citizens living with existing cancer on 13 August 2007.

Longitudinal: *Change* in health status, like the *incidence* of new cases over a *time period* is of interest, *e.g.*

- number of Danish citizens getting a new cancer diagnosed during year 2007.

Causal question ⟶ longitudinal study preferred.

Study population & study base

Types of **study population** & its membership defined

- **closed** – **cohort**: members taken by certain event, *e.g.*
  1. birth cohort, people born during same year,
  2. workers employed by Carlsberg brewery during 1970's, followed up since then, even after retirement
- **open** – **dynamic**: defined by changeable status, *e.g.*
  1. citizens of Copenhagen, currently resident;
  2. *catchment population* of the Oncological Clinic at Rigshospitalet (CPH),

Study base = study population × its experience in time.

Study base (SB): population experience

Cross-sectional: SB = study population at a *time point*,

Longitudinal: SB comprises **follow-up times** of individuals in the study population over a given *period*.

Cohort: Follow-up time = period from **entry** until a single **exit** at which **outcome** or **censoring** occurs.

Dynamic: Follow-up time consists of possibly several periods of membership since the first entry until the final exit.

- Follow-up calculation complicated.
- Approximation by *mid-population*.

R and how we use it

Measures of Disease Occurrence

Bendix Carstensen & Esa Läärä

Nordic Summerschool of Cancer Epidemiology
15-26 August 2011
Copenhagen
`www.biostat.ku.dk/~bxc/NSCE`

What is R?

- A practical calculator:
  - You can see what you compute
  - ...and change easily to do similar calculations.
- A statistical program.
- An environment for data analysis and graphics.
- Free.
- Runs on any computer.
- Updated every 6 months.

A simple calculator

R lets you enter simple arithmetic and giver you back the ansver straightaway:

```
> 5+8
[1] 13
> sqrt( 1/12 + 1/17 )
[1] 0.3770370
> exp( 1.96 * sqrt( 1/12 + 1/17 ) )
[1] 2.093825
> D0 <- 12
> D1 <- 17
> exp( 1.96 * sqrt( 1/D0 + 1/D1 ) )
[1] 2.093825
```

Handy in daily life too.

A smart calculator

Case-control study of MI:

	Men		Women	
PA index	Case	Cont	Case	Cont
2500+ kcals	141	208	49	58
$<$ 2500 kcals	144	112	32	45
Total	285	320	81	103

```
> (141/208)/(144/112)
[1] 0.5272436
> (49/58)/(32/45)
[1] 1.188039
```

A smart calculator

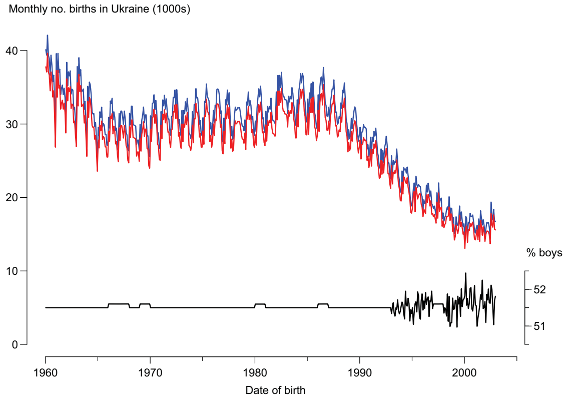
```
> D1 <- c(141, 49)
> D0 <- c(144, 32)
> H1 <- c(208, 58)
> H0 <- c(112, 45)
> OR <- (D1/D0)/(H1/H0)
> OR
[1] 0.5272436 1.1880388
```

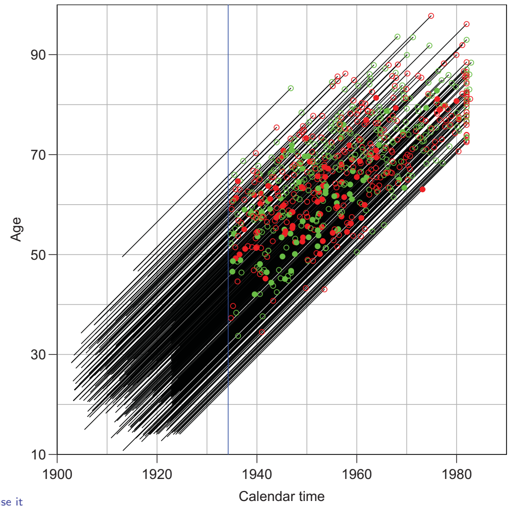
Things done in parallel for the two exposure groups.

R for epidemiology

Versatile graphics:

- Simple graphs easy
- Complicated graphs possible
- You can add things to a graph
- Interactive graphs:
  - Put things on with the mouse
  - Identify points with the mouse





Getting your graphs out

You can save graphs to disk and later fetch them into your documents in almost any format you like: (`.eps`, `.pdf`, `.emf`, `.bmp`, `.png`).

You can choose to save graphs from the screen or to write directly to a file.

Tools for anything!

- More than 1500 add-on packages.
- Several packages for epidemiology:
  - `Epi`: Mostly chronic disease epidemiology:
    - Cohort studies, split follow-up time
    - Lexis diagram, sevral timescales
    - Multistate model support
    - Advanced tabulation
    - Parameter reporting
  - `epicalc`: For a book by Virasakdi Chongsuvivatwong.
  - `epitools`: Mostly infectious diseases.
  - `epiR`: Leaning towards veterinay epidemiology.
- Install and update packages from within R.

Versatility is paid by steep learning curve

Command line interface:

- You must write commands
- You must know what they are called
- Easy to repeat analyses, because you always have a script of what you did.
- There is a simple editor built into R.
- A good workbook introduction is: `www.mhills.pwp.blueyonder.co.uk/Rwork_book.html`
- Many other introductions to R on the R homepage.

R in this course

- Only use R as a simple calculator.
- No need for for a lot of fancy stuff.
- The script editor (we will show you what that is) will help you keep your solutions for future reference.
- A short recap of exercises tomorrow morning, and tomorrow afternoon.
- After the course, solutions to all exercises will be provided.

Frequency measures

Measures of Disease Occurrence

Bendix Carstensen & Esa Läärä

Nordic Summerschool of Cancer Epidemiology
15-26 August 2011
Copenhagen
`www.biostat.ku.dk/~bxc/NSCE`

Measuring frequency: Cases, population, time

Quantification of the occurence of disease (or any other health-related state or event) requires specification of:

1. what is meant by a **case**, i.e., an individual in a population who has or gets the disease (more generally: possesses the state or undergoes the event of interest).
   $\Rightarrow$ challenge to accurate diagnosis and classification!
2. the **population** from which the cases originate.
3. the **time point** or **period** of observation.

Types of occurrence measures

- Longitudinal – **incidence** measures.
- Cross-sectional – **prevalence** measures.

General form of frequency or occurrence measures

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

Numerator: number of cases observed in the population — at a certain time point or during a specified period.

Denominator: generally proportional to the size of the population from which the cases emerge.

Numerator and denominator must cover the *same population*.

Prevalence

Prevalence:
Point prevalence, is the proportion of existing cases (old and new) in a population at a single point of time.

$$P = \frac{\text{No. of existing cases in a population at one point of time}}{\text{No. of people in the population at the same point of time}}$$

This measure is called point prevalence, because it refers to a single point in time. It is often referred to simply as prevalence.

Incidence measures

Incidence proportion (Q) over a fixed *risk period*:

$$Q = \frac{\text{number of incident (new) cases during period}}{\text{size of pop'n at risk at start of the period}}$$

Also called **cumulative incidence** or **cumulative risk** (*e.g.* by **IS**).

Indidence rate (I) over a defined observation period:

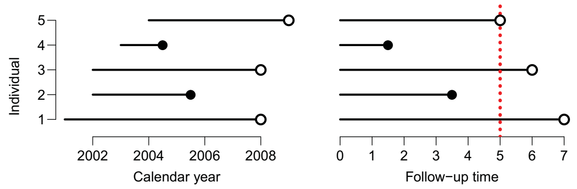
$$I = \frac{\text{number of incident (new) cases during period}}{\text{sum of follow-up times of pop'n at risk}}$$

Also called **incidence density** or **hazard**.

Later we will provide a more precise mathematical definition of the concepts.

Example: Follow-up of a small cohort

○ = exit with censoring; outcome not observed,
● = exit with outcome event (disease onset) observed



$$\text{Inc. rate} = \frac{2 \text{ cases}}{5 + 3.5 + 5 + 1.5 + 5 \text{ years}} = 10 \text{ per } 100 \text{ years}$$

No censoring in the 5-year risk period $\Rightarrow$ can calculate:

$$\text{Inc. prop.} = 2/5 = 0.4 \ (40\ \%)$$

Properties of incidence proportion

- Dimensionless quantity ranging from 0 to 1 (0% to 100%) = *relative frequency*,
- Estimates the average theoretical **risk** or probability of the outcome occurring during the risk period, in the **population at risk** — *i.e.* among those who are still free from the outcome at the start of the period,
- Simple formula valid when the follow-up time is fixed & equals the risk period, and when there are no **competing events** or **censoring** (see below),
- Competing events & censoring $\Rightarrow$ Calculations need to be corrected using special methods of survival analysis.

Properties of incidence rate

- Like *a frequency* quantity in physics; it is a scaled quantity; it is measured in time^{-1}: cases/1000 Y, say.
- Estimates the average underlying **intensity** or **hazard rate** of the outcome in a population,
- Estimation accurate in the **constant hazard model**,
- Calculation straightforward also with competing events and censored observations.
- Hazard depends on age (& other time variables) ⇒ rates *specific to age group etc.* needed,
- Incidence proportions can be estimated from rates. In the constant hazard model with no competing risks:

$$Q = 1 - \exp(-I \times \Delta) \approx I \times \Delta$$

(we shall return to the derivation of this).

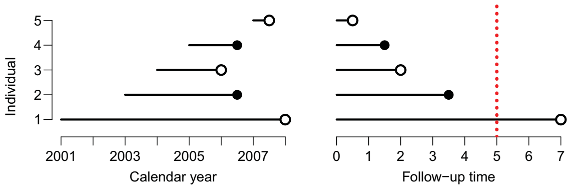
Competing events and censoring

The outcome event of interest (*e.g.* onset of disease) is not always observed for all subjects during the chosen risk period.

- Some subjects die (from other causes) before the event. ⇒ Death is a **competing event** after which the outcome cannot occur any more.
- Others emigrate and escape national disease registration, or the whole study is closed "now", which prematurely interrupts the follow-up of some individuals, ⇒ **censoring**, **withdrawal**, or **loss to follow-up**

Competing events and censorings require special statistical treatment in incidence and risk calculations.

Follow-up of another small cohort



Two censored observations ⇒ can calculate the rate:

$$I = 2/12.5 \text{ y} = 16 \text{ per } 100 \text{ years}$$

but the 5-year Q **is no more** $2/5$!
However, under constant rate model

$$Q = 1 - \exp(-5 \times 2/12.5) = 0.55$$

Person-years in dynamic populations

With dynamic study population individual follow-up times are always variable and impossible to measure accurately.

Common approximation – **mid-population** principle:

- Let the population size be N_{t-1} at start and N_t at the end of the observation period t with length L_t years,
- Mid-population for the period: $\bar{N}_t = \frac{1}{2} \times (N_{t-1} + N_t)$.
- Approximate person-years: $Y_t \approx \bar{N}_t \times L_t$.

NB. The actual study population often contains also some already affected, who thus do not belong to the population at risk. With rare outcomes the influence of this is small.

Male person-years in Finland 1991-95

Total male population (1000s) on 31 December by year:

1990	1991	1992	1993	1994	1995
2431	2443	2457	2470	2482	2492

Approximate person-years (1000s):

$$\begin{array}{rlr} \text{1992:} & \frac{1}{2} \times (2443 + 2457) \times 1 = & 2450 \\ \text{1993-94:} & \frac{1}{2} \times (2457 + 2482) \times 2 = & 4937 \\ \text{1991-95:} & \frac{1}{2} \times (2431 + 2492) \times 5 = & 12307.5 \end{array}$$

Relationships between incidence measures

With constant incidence rate over risk period (length $= \Delta$), incidence proportion Q and rate I are related:

$$\begin{aligned} Q &= 1 - \exp(-I \times \Delta) \approx I \times \Delta \\ I &= -\log(1 - Q)/\Delta \approx Q/\Delta, \end{aligned}$$

The approximations are good when

- the incidence proportion is "small" (under 10 %).
  - incidence rate (I) is small
  - the risk period (Δ) is small

Mortality

Cause-specific mortality from disease C is described by **mortality rate** (and proportion), defined like I (and Q), but

- cases are only *deaths* from cause C, and
- follow-up is extended until death (from *any* acuse) or censoring

The cumulative risk of death from a given cause (cause-specific mortality proportion/risk) requires correction for *competing events*. **Total mortality**: cases are deaths from any cause. Mortality depends on the incidence and the **prognosis** or fatality of the disease, *i.e.* the **survival** of those affected.

Theoretical concepts behind incidences

Analysis of incidences
= analysis of **time to event** or **failure time** or **survival** data.

Mathematical concepts:

$$\begin{aligned} T &= \text{ time to outcome event – random variable,} \\ S(t) &= P(T > t) = \textbf{ survival} \text{ function of } T, \\ &= \text{ probability of avoiding the event up to given time } t, \\ \lambda(t) &= -S'(t)/S(t) = \textbf{ intensity} \text{ or } \textbf{hazard} \text{ function,} \\ \Lambda(t) &= \int_0^t \lambda(u)du = -\log S(t) = \textbf{cumulative hazard}, \\ F(t) &= 1 - S(t) = 1 - \exp\{-\Lambda(t)\} = \textbf{risk} \text{ function} \\ &= \text{probability of the outcome to occur before } t \end{aligned}$$

Intensity or hazard function

Can be viewed as *theoretical incidence rate*. Formally:

$$\lambda(t) = \lim_{\Delta \to 0} \frac{P(t < T \leq t + \Delta \mid T > t)}{\Delta}$$

≈ Probability of outcome event occurring in a short risk period $]t, t + \Delta]$, given "survival" or avoidance of the event up to the start t, divided by the period length — "risk per time".

This is equivalent to saying that over a short interval

risk ≈ intensity × length of interval

or $P(t < T \leq t + \Delta \mid T > t) \approx \lambda(t) \times \Delta$.

Exponential survival times (constant hazard)

Simplest probability model for time to event:

Exponential distribution, Exp(λ), in which

rate $\lambda(t) = \lambda$ (constant) ⇒ risk over $]0, t] = 1 - \exp(-\lambda t)$

Analysis of event data of n individuals. For subject i let

y_i = time to event or censoring, total: $Y = \sum y_i$

d_i = 1/0-indicator for observing event, total: $D = \sum d_i$

Exp(λ) model ⇒ **Likelihood function** of λ is equivalent to that when number of cases D is *Poisson*-distributed

(Analysis part of the course)

Basic statistical analysis of empirical rates

Asymptotic statistical inference based on likelihood:

- **Maximum likelihood estimator** (MLE) of λ is

$$\widehat{\lambda} = \frac{D}{Y} = \frac{\text{number of cases}}{\text{total person-time}} = I, \text{ empirical incidence rate!}$$

- **Standard error** of the empirical rate is $I/\sqrt{D}$

⇒ The more cases, the greater is **precision** in rate!

- Approximate **confidence interval** for "true" rate λ:

estimator ± $1.96 \times$ standard error

More about these issues in the analysis lectures.

Prevalence measures

Point prevalence or simply **prevalence** P of a health state C in a population at a given time point t is defined

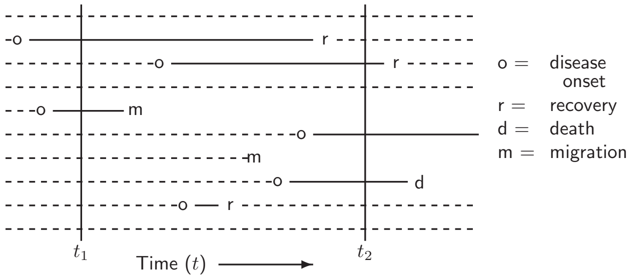
$$P = \frac{\text{number of existing or prevalent cases of } C}{\text{size of the whole population}}$$

This is calculable from a cross-sectional study base.

Period prevalence for period from t_1 to t_2 is like P but

- numerator refers to all cases prevalent already at t_1 plus new cases occurring during the period, and
- denominator is the population size at t_2.

Example 4.1 (IS: p. 59)



Prevalence at time t_1 : $2/10 = 0.2 = 20\%$
Prevalence at time t_2 : $3/8 = 0.38 = 38\%$
Period prevalence: $5/8 = 0.62 = 62\%$

Relationships between measures

Point prevalence of C at given time point t depends on

- *incidence* of new cases of C before t
- *duration* of C, depending in turn on the probability of **cure** or recovery from C or **survival** of those affected.

Stationary ("stable") population: prevalence (P), incidence (I), and average duration ($\bar{d}$) of C are related:

$$P = \frac{I \times \bar{d}}{I \times \bar{d} + 1} \approx I \times \bar{d}$$

prevalence = incidence × duration

The approximation works well, when $P < 0.1$ (10%).

Prevalence of cancer?

Difficult to ascertain, whether and when a cancer is cured.

⇒ Existing or prevalent cancer case problematic to define.

Cancer registry practice: Prevalence of cancer C at time point t in the target population refers to the

number & proportion of population members who

- are alive and resident in the population at t, and
- have a record of incident cancer C diagnosed before t.

Often further classified by years since diagnosis.

Example: Liver and testis cancer

Crude comparison of incidence, mortality and prevalence in the male population of Finland 1999

	Liver	Testis
No. of new cases during 1999	119	103
No. of deaths during 1999	123	8
No. of prevalent cases 1.1.2000	120	1337
– " – diagnosed < 1 y ago	*36*	*97*
– " – diagnosed 1-< 5 y ago	*53*	*291*
– " – diagnosed 5-< 10 y ago	*17*	*304*
– " – diagnosed > 10 y ago	*14*	*642*

Comparative measures

Measures of Disease Occurrence

Bendix Carstensen & Esa Läärä

Nordic Summerschool of Cancer Epidemiology
15-26 August 2011
Copenhagen
`www.biostat.ku.dk/~bxc/NSCE`

Relative and absolute comparisons (IS: Ch 5.2)

Quantification of the **association** between a determinant (risk factor or exposure) and an outcome (disease) is based on

comparison of occurrence between the *index* ("exposed") and the *reference* ("unexposed") groups or populations by

- relative measures (ratio)
- absolute measures (difference)

In causal studies these are used to estimate the **causal effect** of the exposure factor on the disease risk.

⇒ **comparative measures** ≈ **effect measures**

Relative comparative measures

Generic name **"relative risk"** RR comparing occurrences between exposed (1) and unexposed (0) groups can be

- incidence rate ratio I_1/I_0,
- incidence proportion ratio Q_1/Q_0,
- incidence odds ratio $[Q_1/(1-Q_1)]/[Q_0/(1-Q_0)]$,
- prevalence ratio P_1/P_0, or
- prevalence odds ratio $[P_1/(1-P_1)]/[P_0/(1-P_0)]$,

depending on study base and details of its design.

Absolute comparative measures

Generic **"excess risk"** btw exposed and unexposed can be

- incidence rate difference $I_1 - I_0$,
- incidence proportion difference $Q_1 - Q_0$,
- prevalence difference $P_1 - P_0$.

Use of relative and absolute comparisons

Ratio – describes the *biological strength* of the exposure

Difference – informs about its *public health importance*.

Example: (IS, Table 5.2, p.97)

Relative and absolute comparisons between the exposed and the unexposed to risk factor X in two diseases.

	Disease A	Disease B
Incidence rate among exposed[a]	20	80
Incidence rate among unexposed[a]	5	40
Rate ratio	4.0	2.0
Rate difference[a]	15	40

[a] Rates per 100 000 pyrs.

Factor X has a stronger biological potency for disease A, but it has a greater public health importance for disease B.

Ratio measures in "rare diseases" (IS: Ex 5.13)

	Exposure	
	Yes	No
No. initially at risk	4 000	16 000
Deaths	30	60
Person-years at risk	7 970	31 940

$$\text{Inc. prop'n ratio} = \frac{30/4\,000}{60/16\,000} = \frac{7.5 \text{ per } 1\,000}{3.75 \text{ per } 1\,000} = 2.0000$$

$$\text{Inc. rate ratio} = \frac{30/7\,970 \text{ y}}{60/31\,940 \text{ y}} = \frac{3.76 \text{ per } 1\,000 \text{ y}}{1.88 \text{ per } 1\,000 \text{ y}} = 2.0038$$

$$= \frac{0.00756}{0.00376} = 2.0076$$

Attributable fraction

Combine absolute and relative comparisons.

When incidence is higher for the exposed, we can calculate

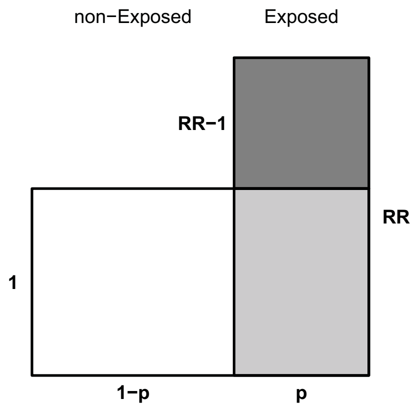
$$\textbf{Excess fraction, EF} = \frac{Q_1 - Q_0}{Q_1} = \frac{\text{RR} - 1}{\text{RR}}$$

also called **attributable fraction**, AF or **attributable risk**.

EF Estimates the fraction out of all new cases among those exposed, which are "caused" by the exposure itself, and which thus could be "avoided" if the exposure were absent

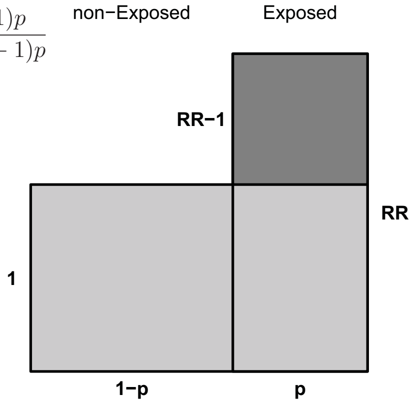
Attributable fraction, AF

$$\text{AF} = \frac{RR - 1}{RR}$$



Population attributable fraction, PAF

$$\mathrm{PAF} = \frac{(RR-1)p}{1+(RR-1)p}$$



Population attributable fraction

If we instead ask:
"How large a fraction of **all** cases would be prevented if exposure was abolished?".

Depends on the fraction of the population which is exposed

$$\mathrm{PAF} = \frac{(\mathrm{RR}-1)p}{1+(\mathrm{RR}-1)p}$$

PAF Estimates the fraction out of all new cases, which are "caused" by the exposure itself, and which thus could be "avoided" if the exposure were absent.

AF is a "biological" measure.
PAF is a "population level" measure.

Measures of potential impact (cont'd)

When the exposed have a lower incidence, we can calculate

$$\textbf{Preventive fraction, PF} = \frac{Q_0 - Q_1}{Q_0} = 1 - \mathrm{RR}$$

also called **relative risk reduction** = percentage of cases prevented among the exposed due to the exposure.

Used to evaluate the relative effect of a preventive intervention (exposed) vs. no intervention (unexposed).

Effect of smoking on mortality by cause

(**IS**: Example 5.14, p. 98)

Underlying cause of death	Never smoked regularly Rate[b]	Current cigarette smoker Rate[b]	Rate ratio	Rate difference[b]	Excess fraction (%)
	(1)	(2)	(2)/(1)	(2) − (1)	$\frac{(2)-(1)}{(2)} \times 100$
Cancer					
All sites	305	656	2.2	351	54
Lung	14	209	14.9	195	93
Oesophagus	4	30	7.5	26	87
Bladder	13	30	2.3	17	57
Respiratory diseases					
(except cancer)	107	313	2.9	206	66
Vascular diseases	1037	1643	1.6	606	37
All causes	1706	3038	1.8	1332	44

[a] Data from Doll *et al.*, 1994a.
[b] Age-adjusted rates per 100 000 pyrs.

Time scales

Measures of Disease Occurrence

Bendix Carstensen & Esa Läärä

Nordic Summerschool of Cancer Epidemiology
15-26 August 2011
Copenhagen
`www.biostat.ku.dk/~bxc/NSCE`

Incidence by age, calendar year, and other time variables

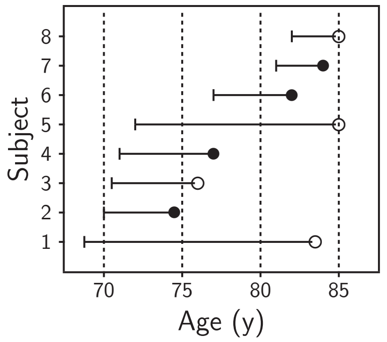
Incidence can be studied on various **time scales**, *e.g.*:

Time scale	Origin (date of:)
age	birth
exposure time	first exposure
follow-up time	entry to study
duration of disease	diagnosis

Age is usully the strongest time-dependent determinant of health outcomes.

Age is also often correlated with duration of "chronic" exposure (e.g. years of smoking).

Follow-up of a geriatric cohort



Overall rate: 4 cases/53.5 person-years = 7.5 per 100 y
Hides the fact that the "true" rate probably varies by age, being higher among the old.

Person-years and cases in agebands: age-specific rates

	Ageband			
Subject	70-74	75-79	80-84	Total
1	5.0	5.0	3.5	13.5
2	4.5	-	-	4.5
3	4.5	1.0	-	5.5
4	4.0	2.0	-	6.0
5	3.0	5.0	5.0	13.0
6	-	3.0	2.0	5.0
7	-	-	3.0	3.0
8	-	-	3.0	3.0
Sum of person-years	21.0	16.0	16.5	53.5
Cases	1	1	2	4
Rate (/100 y)	4.8	6.2	12.1	7.5
	Age-specific rates			overall

Lung cancer incidence rates in Finland by age, period and cohort

Calendar	Age group (y)									
period	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85+
1953-57	21	61	119	209	276	340	295	279	193	93
1958-62	22	65	135	243	360	405	429	368	265	224
1963-67	24	61	143	258	395	487	509	479	430	280
1968-72	21	61	134	278	424	529	614	563	471	358
1973-77	16	50	134	251	413	541	629	580	490	392
1978-82	13	36	115	234	369	514	621	653	593	442
1983-87	11	31	74	186	347	450	566	635	592	447
1988-92	9	25	57	128	262	411	506	507	471	441
1993-97	7	22	48	106	188	329	467	533	487	367
1998-02	5	14	46	77	150	239	358	445	396	346

- Rows: age-incidence pattern in different calendar periods.
- Columns: Trends of age-specific rates over calendar time.
- Diagonals: age-incidence pattern in birth cohorts.

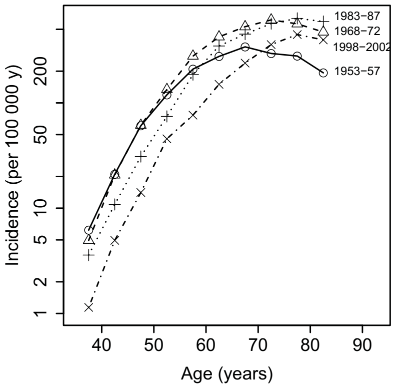
Incidence by age, calendar time & birth cohort

- **Secular trends** of specific and adjusted rates show, how the "cancer burden" has developed over periods of calendar time.

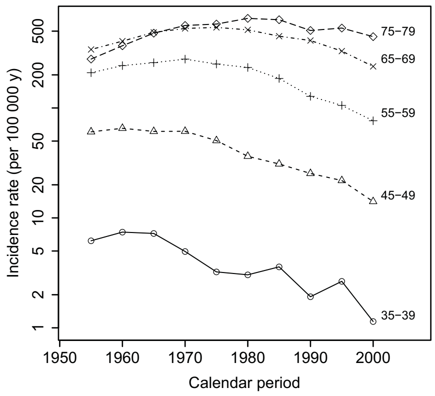
Birth cohort = people born during the same limited time interval, *e.g.* single calendar year, or 5 years period.

- Analysis of rates by birth cohort reveals, how the level of incidence (or mortality) differs between successive generations.
- Often more informative about "true" age-incidence pattern than age-specific incidences of single calendar period.

Age-incidence curves by period (rows)



Time trends by age (columns)

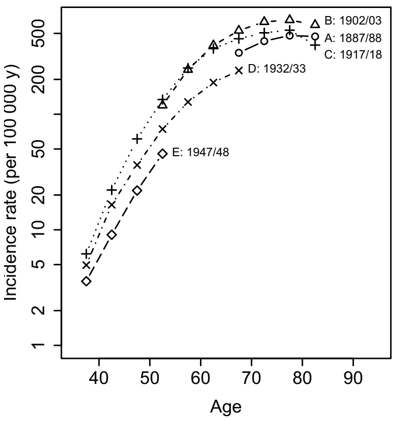


Age-specific rates by birth cohort

Calendar	Age group (y)								
period	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	
1953-57	21	61	$\boxed{119}$	209	276	$\boxed{340}$	295	279	
1958-62	$\boxed{22}$	65	135	$\boxed{243}$	360	405	$\boxed{429}$	368	
1963-67	24	$\boxed{61}$	143	258	$\boxed{395}$	487	509	$\boxed{479}$	A
1968-72	21	61	$\boxed{134}$	278	424	$\boxed{529}$	614	563	
1973-77	$\boxed{16}$	50	134	$\boxed{251}$	413	541	$\boxed{629}$	580	
1978-82	13	$\boxed{36}$	115	234	$\boxed{369}$	514	621	$\boxed{653}$	B
1983-87	11	31	$\boxed{74}$	186	347	$\boxed{450}$	566	635	
1988-92	$\boxed{9}$	25	57	$\boxed{128}$	262	411	$\boxed{506}$	507	
1993-97	7	$\boxed{22}$	48	106	$\boxed{188}$	329	467	$\boxed{533}$	C
1998-02	5	14	$\boxed{46}$	77	150	$\boxed{239}$	358	445	
			E: 1947/48			D: 1932/33			

A = synthetic cohort born around 1887/88, B: 1902/03, C: 1917/18

Age-incidence curves in 5 birth cohorts



Split of follow-up by age and period

Incidence of (or mortality from) disease C in special study cohort (e.g. occupational group, users of certain medicine)

→ often compared to incidence in a *reference* or "general" population

Appropriate adjustment for age and calendar time needed in this, *e.g.* by comparing *observed* to *expected* cases with SIR (see p. 70-71).

⇒ Cases and person-years in the study cohort must be split by more than one time scale (age).

Example of follow-up

Entry and exit dates for a small cohort of four subjects

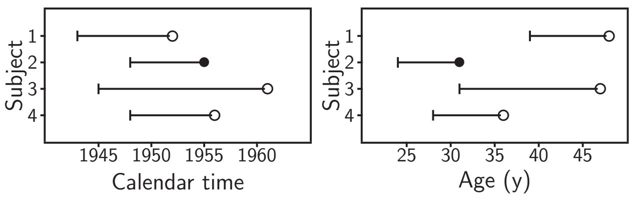
Subject	Born	Entry	Exit	Age at entry	Outcome
1	1904	1943	1952	39	Migrated
2	1924	1948	1955	24	Disease C
3	1914	1945	1961	31	Study ends
4	1920	1948	1956	28	Unrelated death

Subject 1: Follow-up time spent in each ageband

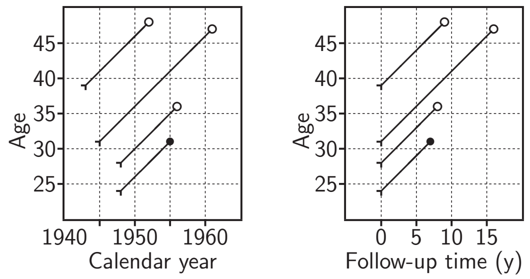
Age band	Date in	Date out	Time (years)
35–39	1943	1944	1
40–44	1944	1949	5
45–49	1949	1952	3

Follow-up of cohort members by calendar time and age

| entry
- ● exit because of disease onset (outcome of interest)
- ○ exit due to other reason (censoring)



Follow-up in Lexis-diagrams — by age and period



Follow-up lines run diagonally through different ages and calendar periods.

Standardization

Measures of Disease Occurrence

Bendix Carstensen & Esa Läärä

Nordic Summerschool of Cancer Epidemiology
15-26 August 2011
Copenhagen
www.biostat.ku.dk/~bxc/NSCE

Crude & adjusted rates

- Incidence of most cancers (and many other diseases) increases strongly by age in all populations.
  ⇒ Most of the caseload comes from older age groups.
- **Crude incidence rate** is a rate in which:
  - numerator = sum of age-specific numbers of cases,
  - denominator = sum of age-specific person-years.
- This is generally a poor **summary measure**.
- Comparisons of crude incidences between populations can be very misleading, when the age structures differ.
- Solution: Standardization.

Stomach cancer in Cali and Birmingham (IS, Table 4.2, p. 71)

	Cali			Birmingham			
Age (y)	No. of Male cases 1982 -86	Male Population 1984 (10^3s)	Inci- Rate ($/10^5$ y) 1982 -86	No. of Male cases 1983 -86	Male Population 1985 (10^3s)	Inci- Rate ($/10^5$ y) 1983 -86	Rate ratio
0–44	39	524.2	**1.5**	79	1 683.6	**1.2**	*1.25*
45-64	266	76.3	**69.7**	1037	581.5	**44.6**	*1.56*
65+	315	22.4	**281.3**	2352	291.1	**202.0**	*1.39*
Total	620	622.9	**19.9**	3468	2 556.2	**33.9**	*0.59*

In each age group Cali has a higher incidence but the crude incidence is higher in Birmingham. *Is there a paradox*?

Comparison of age structures (IS, Tables 4.3,4.4)

	% of male population			
Age (years)	Cali 1984	B'ham 1985	Finland 1999	World Stand.
0–44	84	66	61	74
45–64	12	23	27	19
65+	4	11	12	7
All ages	100	100	100	100

- The fraction of old men greater in Birmingham than in Cali.

⇒ The crude rates are **confounded** by age.
⇒ Any summary rate must be **adjusted for age**.

Age-adjustment by standardisation

Age-standardised incidence rate (ASR):

$$\text{ASR} = \sum_{k=1}^{K} \text{weight}_k \times \text{rate}_k \;/\; \text{sum of weights}$$

= **Weighted average** of age-specific rates over the age-groups $k = 1, \ldots, K$.

- Weights describe age distribution of some **standard population**.
- Standard population can be real (*e.g.* one of the populations under comparison, or their average) or fictitious (*e.g.* World Standard Population, WSP)

Some standard populations:

Age group (years)	African	World	European	Truncated
0	2 000	2 400	1 600	–
1–4	8 000	9 600	6 400	–
5–9	10 000	10 000	7 000	–
10–14	10 000	9 000	7 000	–
15–19	10 000	9 000	7 000	–
20–24	10 000	8 000	7 000	–
25–29	10 000	8 000	7 000	–
30–34	10 000	6 000	7 000	–
35–39	10 000	6 000	7 000	6 000
40–44	5 000	6 000	7 000	6 000
45–49	5 000	6 000	7 000	6 000
50–54	3 000	5 000	7 000	5 000
55–59	2 000	4 000	6 000	4 000
60–64	2 000	4 000	5 000	4 000
65–69	1 000	3 000	4 000	–
70–74	1 000	2 000	3 000	–
75–79	500	1 000	2 000	–
80–84	300	500	1 000	–
85+	200	500	1 000	–
Total	100 000	100 000	100 000	31 000

Stomach cancer in Cali & B'ham

Age-standardized rates by the World Standard Population:

	Cali		Birmingham	
Age	Rate[a]	Weight	Rate[a]	Weight
0–44	1.5 ×	0.74 = 1.11	1.2 ×	0.74 = 0.89
45–64	69.7 ×	0.19 = 13.24	44.6 ×	0.19 = 8.47
65+	281.3 ×	0.07 = 19.69	202.0 ×	0.07 = 14.14
Age-standardised rate		**34.04**		**23.50**

ASR in Cali higher – coherent with the age-specific rates.
Summary rate ratio estimate: **standardized rate ratio**

$$\text{SRR} = 34.0/23.5 = 1.44$$

Known as **comparative mortality figure (CMF)** when the outcome is death (from specific cause C or all causes).

Cumulative rate and cumulative risk

- Choice of standard population weights somewhat arbitrary.
- Alternative and perhaps more "natural" method for age-adjustment is provided by:

$$\textbf{Cumulative rate} = \sum_{k=1}^{K} \text{width}_k \times \text{rate}_k$$

- Weigths are widths of the agebands to be included:

$$\textbf{Cumulative risk} = 1 - \exp(-\text{cumul. rate}) \approx \text{cumul. rate}$$

- Usually calculated up to 65 or 75 years with 5-year agebands.
- These estimate the average risk in the population to get the disease by 65 or 75 years given survival until then.
- The competing acuses of exit (death) is **not** taken into account.

Stomach cancer in Cali & B'ham

From age-specific rates of Table 4.2. the cumulative rates up to 65 years and their ratio are

$$\begin{aligned} \text{Cali:} \quad & 45\ y \times \tfrac{1.5}{10^5 y} + 20\ y \times \tfrac{69.7}{10^5 y} = 0.0146 = \mathbf{1.46} \text{ per } 100 \\ \text{B'ham:} \quad & 45\ y \times \tfrac{1.2}{10^5 y} + 20\ y \times \tfrac{44.6}{10^5 y} = 0.0095 = \mathbf{0.95} \text{ per } 100 \\ \text{ratio:} \quad & 1.46/0.95 = \mathbf{1.54} \end{aligned}$$

Cumulative risks (inc. proportions) & their ratio up to 65 y:

$$\begin{aligned} \text{Cali:} \quad & 1 - \exp(-0.0146) = 0.0145 = \mathbf{1.45}\% \\ \text{B'ham:} \quad & 1 - \exp(-0.0095) = 0.0094 = \mathbf{0.94}\% \\ \text{ratio:} \quad & 1.45/0.94 = \mathbf{1.54} \end{aligned}$$

Cumulative measures in 5-y groups

Age-group (years)	Incidence rate (per 100 000 pyrs)
0–4, . . . , 15–19	0.0
20–24, 25–29	0.1
30–34	0.9
35–39	3.5
40–44	6.7
45–49	14.5
50–54	26.8
55–59	52.6
60–64	87.2
65–69	141.7
70–74	190.8
Sum	524.9

$$\begin{aligned} \text{Cum. rate 0-75 y} &= 5\text{ y} \times \frac{524.9}{10^5\text{ y}} = 0.0262 = \mathbf{2.6}\% \\ \text{Cum. risk 0-75 y} &= 1 - \exp(-0.0262) = 0.0259 = \mathbf{2.6}\%. \end{aligned}$$

Observed and expected cases

- Suppose O cases are **observed** in an **index** population of interest (*e.g.* an occupational cohort) during its follow-up over a lengthy calendar period.
- *Question*: What would be the **expected number of cases** E, if the age- and period-specific rates of a **reference** population for comparison were valid for the index population?
- The ratio "observed/expected" estimates of the "true" rate ratio between the index and the reference populations jointly adjusted for age and period.

Standardized incidence ratio, SIR

Let λ_{kl} = incidence rate in a Lexis-diagram cell defined by ageband k and period l in the reference population. Hence,

$$\text{expected number } (E) = \sum_{k=1}^{K} \sum_{l=1}^{L} \lambda_{kl} \times Y_{kl},$$

where Y_{kl} is the person-years in cell kl of the index population.

The **standardised incidence ratio** (SIR) is defined

$$\text{SIR} = \frac{O}{E}$$

When the outcome is death, this measure is called **standardized mortality ratio**, SMR.

SIR for Cali with Birmingham as reference

Total person-years at risk and expected number of cases in Cali 1982-86 based on age-specific rates in Birmingham (**IS**: Fig. 4.9, p. 74)

Age	Person-years	Expected cases in Cali
0–44	524 220×5= 2 621 100	0.000012×2 621 100= 31.45
45–64	76 304×5= 381 520	0.000446× 381 520=170.15
65+	22 398×5= 111 990	0.002020× 111 990=226.00
All ages	=**3 114 610**	**Total expected (E) 427.82**

Total observed number $O = 620$. Standardised incidence ratio:

$$\text{SIR} = \frac{O}{E} = \frac{620}{427.8} = 1.45 \quad (\text{or } 145 \text{ per } 100)$$

Crude and adjusted measures

(IS: Table 4.6, p. 78, extended)

	Cali, 1982-86	B'ham, 1983-86	Rate ratio
Crude rates ($/10^5$ y)	19.9	33.9	*0.59*
ASR ($/10^5$ y)^B with 3 broad age groups	48.0	33.9	*1.42*
ASR ($/10^5$ y)^C –"–	19.9	14.4	*1.38*
ASR ($/10^5$ y)^W –"–	34.0	23.5	*1.44*
Cum. rate $<$ 65 y (per 1000) –"–	14.6	9.5	*1.54*
ASR ($/10^5$ y)^W with 18 5-year age groups	36.3	21.2	*1.71*
Cum. rate $<$ 75 y (per 1000) –"–	46.0	26.0	*1.77*

Standard population: ^{B} Birmingham 1985, ^{C} Cali 1985, ^{W} World SP

NB: The ratios of age-adjusted rates appear less dependent on the choice of standard weights than on the coarseness of age grouping. 5-year age groups are preferred.

Survival

Measures of Disease Occurrence

Bendix Carstensen & Esa Läärä

Nordic Summerschool of Cancer Epidemiology
15-26 August 2011
Copenhagen
www.biostat.ku.dk/~bxc/NSCE

Survival analysis

The **prognosis** of cancer patients:
what is their chance to **survive** 1 year, 5 years etc. after diagnosis?

Survival analysis: In principle like incidence analysis but

- population at risk = patients with cancer,
- basic time variable = time since the date of diagnosis, at which the follow-up starts,
- outcome event of interest = death,
- measures and methods used somewhat different from those used in incidence analysis.

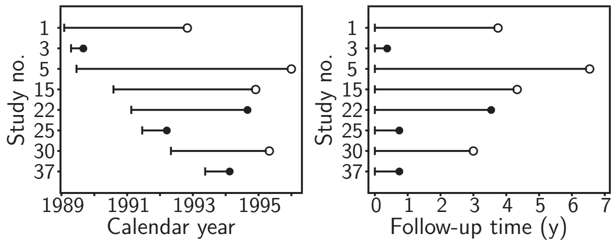
Follow-up of 8 out of 40 breast cancer patients (from IS, table 12.1., p. 264)

No.	Age (y)	Sta-ge[a]	Date of diag-nosis	Date at end of follow -up	Vital status at end of follow -up	Cause of death[c]	Full years from diagn's up to end of follow -up	Days from diagn's up to end of follow -up
1	39	1	01/02/89	23/10/92	A	–	3	1360
3	56	2	16/04/89	05/09/89	D	BC	0	142
5	62	2	12/06/89	28/12/95	A	–	6	2390
15	60	2	03/08/90	27/11/94	A	–	4	1577
22	64	2	17/02/91	06/09/94	D	O	3	1297
25	42	2	20/06/91	15/03/92	D	BC	0	269
30	77	1	05/05/92	10/05/95	A	–	3	1100
37	45	1	11/05/93	07/02/94	D	BC	0	272

[a] 1 = absence of regional lymph node involment and metastases
2 = involvment of regional lymph node and/or presence of metastases
[b] A = alive; D = dead; [c] BC = breast cancer; O = other causes

Follow-up of breast cancer patients (cont'd)

| entry = diagnosis; ● exit = death; ○ exit = censoring



(**IS**: Figure 12.1, p. 265)

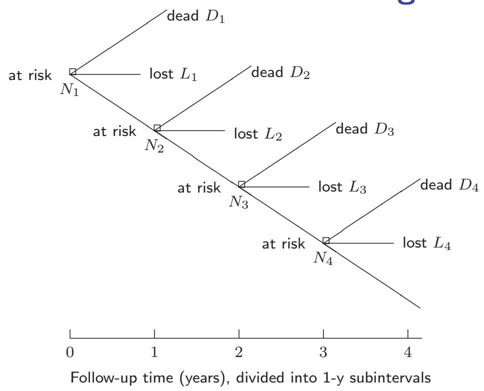
Life table or "actuarial" method

(1) Divide the follow-up time into subintervals $k = 1, \ldots K$; usually each with 1 year width.

(2) Tabulate from original data for each interval

$N_k =$ size of the **risk set**, *i.e.* the no. of subjects still alive and under follow-up at the start of interval,
$D_k =$ no. of **cases**, *i.e.* deaths observed in the interval,
$L_k =$ no. of **losses**, *i.e.* individuals **censored** during the interval before being observed to die.

Life table items in a tree diagram



N_k = population at risk at the start of the kth subinterval
D_k = no. of deaths, L_k = no. of losses or censorings in interval k

Life table items for breast ca. patients

(**IS**: Table 12.2., p. 273, first 4 columns)

Inter-val (k)	Years since diagnosis	No. at start of interval (N_k)	No. of deaths (D_k)	No. of losses (L_k)
1	$0-<1$	40	7	0
2	$1-<2$	33	3	6
3	$2-<3$	24	4	3
4	$3-<4$	17	4	4
5	$4-<5$	9	2	3
6	$5-<6$	4	1	2
7	$6-<7$	1	0	1
Total			21	19

Life table calculations (cont'd)

(3) Calculate and tabulate for each interval

$N'_k = N_k - L_k/2 =$ corrected size of the risk set, or "effective denominator" at start of the interval,

$q_k = D_k/N'_k =$ estimated conditional probability of dying during the interval given survival up to its start,

$p_k = 1 - q_k =$ conditional survival proportion over the int'l,

$S_k = p_1 \times \cdots \times p_k =$ **cumulative survival proportion** from date of diagnosis until the end of the kth interval

$=$ estimate of **survival probability** up to this time point.

Follow-up of breast ca. patients (cont'd)

Actuarial life table completed (IS, table 12.2, p. 273)

Inter-val	Years since dia-gnosis	No. at start of in-terval	No. of deaths	No. of losses	Effec-tive deno-minator	Cond'l prop'n of deaths during int'l	Survival prop'n over int'l	Cumul. survival; est'd survival prob'ty
(k)		(N_k)	(D_k)	(L_k)	(N'_k)	(q_k)	(p_k)	(S_k)
1	$0\text{–} < 1$	40	7	0	40.0	0.175	0.825	0.825
2	$1\text{–} < 2$	33	3	6	30.0	0.100	0.900	0.743
3	$2\text{–} < 3$	24	4	3	22.5	0.178	0.822	0.610
4	$3\text{–} < 4$	17	4	4	15.0	0.267	0.733	0.447
5	$4\text{–} < 5$	9	2	3	7.5	0.267	0.733	0.328
6	$5\text{–} < 6$	4	1	2	3.0	0.333	0.667	0.219
7	$6\text{–} < 7$	1	0	1	0.5	0.0	1.0	0.219

1-year survival probability is thus estimated 82.5% and 5-year probability 32.8%.

Comparison to previous measures and methods

Complement of survival proportion $Q_k = 1 - S_k$ is actually incidence proportion of deaths. It estimates cumulative risk of death from start of follow-up till end of kth interval.

"Actuarial" indidence rate in the kth interval:

$$I_k = \frac{\text{number of cases } (D_k)}{\text{approximate person-time}}$$

where the person-time is approximated by

$$\left[N_k - \frac{1}{2}(D_k + L_k)\right] \times \text{length of interval}$$

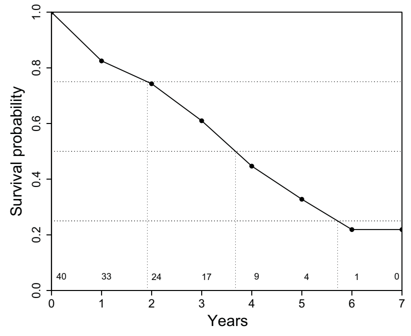
Survival curve and other measures

Line diagram of survival proportions through interval endpoints provides graphical estimates of interesting parameters of the survival time distribution, *e.g.*:

- **median** and **quartiles**: time points at which the curve crosses the 50%, 75%, and 25% levels
- **mean residual lifetime**: area under the curve, given that it decreases all the way down to the 0% level.

NB. Often the curve ends at higher level than 0%, in which case some measures cannot be calculated.

Survical curve of breast ca. patients (IS: Fig 12.8)



Numbers above x-axis show the size of population at risk.

Cause-specific and relative survival

Cause-specific survival analysis:

- outcome event: death from the disease C itself that defines study population
- deaths from other causes → losses.
- problem: ambiguity in cause of death.

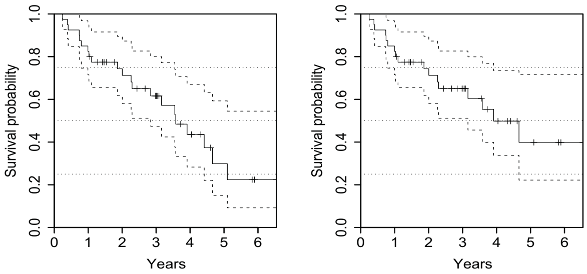
Relative survival: $S_k^{\text{rel}} = S_k^{\text{obs}}/S_k^{\text{exp}}$, i.e. ratio of

- **observed** survival proportion S_k^{obs} in the study population, and
- **expected** survival proportion S_k^{exp} based on age-specific mortalities in the reference (national) population. (See SIR!)

Breast Cancer patients (cont'd)

Overall and cause-specific (death from breast ca.) survival (**IS**: Fig 12.9 & 12.12, p. 271-3)

Kaplan-Meier curves – alternative to "actuarial":



Conclusion

Measures of Disease Occurrence

Bendix Carstensen & Esa Läärä

Nordic Summerschool of Cancer Epidemiology
15-26 August 2011
Copenhagen
www.biostat.ku.dk/~bxc/NSCE

Conclusion

Measuring and comparing disease frequencies

- not a trivial task but
- demands expert skills in epidemiologic methods.

Major challenges:

- obtain the right denominator for each numerator,
- valid calculation of person-years,
- appropriate treatment of time and its various aspects,
- removal of confounding from comparisons.