Nordic Summer School in Cancer Epidemiology

17 to 18 August, 2009

Danish Cancer Society, Copenhagen

Measures of disease frequency and effects

Esa Läärä

Department of Mathematical Sciences, University of Oulu, Finland

esa.laara@oulu.fi http://math.oulu.fi/

with contributions from **Bendix Carstensen**

Steno Diabetes Center, Denmark

& Department of Biostatistics, University of Copenhagen

bxc@steno.dk www.biostat.ku.dk/~bxc

Outline

Basic concepts

Frequency

Comparison

Age, period, etc.

Standardisation

Survival

Conclusion

Key references

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

Key references

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

B&D: Breslow, N.E., Day, N.E. (1987).
Statistical Methods in Cancer Research Vol. II – The Design and Analysis of Cohort Studies.
IARC, Lyon.

Key references

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

B&D: Breslow, N.E., Day, N.E. (1987).
Statistical Methods in Cancer Research Vol. II – The Design and Analysis of Cohort Studies.
IARC, Lyon.

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

BASIC CONCEPTS

BASIC CONCEPTS

What is epidemiology?

BASIC CONCEPTS

What is epidemiology?

Some textbook definitions:

- "study of the **distribution** and **determinants** of disease **frequency** in man" (MacMahon and Pugh 1970)

BASIC CONCEPTS

What is epidemiology?

Some textbook definitions:

- "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, . . ." (Porta (ed.) Dictionary of Epidemiology, 2008)

BASIC CONCEPTS

What is epidemiology?

Some textbook definitions:

- "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, . . ." (Porta (ed.) Dictionary of Epidemiology, 2008)
- "discipline on principles of **occurrence** research in medicine" (Miettinen 1985)

Different epidemiologies

Different epidemiologies

- **descriptive** epidemiology – monitoring & surveillance of diseases for planning of health services

Different epidemiologies

- **descriptive** epidemiology – monitoring & surveillance of diseases for planning of health services
  – a major activity of cancer registries.

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

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, . . .

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, . . .

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: $<$ L *frequentia* $=$ assembly, multitude, crowd.

Frequency (from Webster's Dictionary)

Etymology: $<$ L *frequentia* $=$ assembly, multitude, crowd.

2. rate of occurrence

Frequency (from Webster's Dictionary)

Etymology: $<$ L *frequentia* $=$ assembly, multitude, crowd.

2. rate of occurrence
3. *Physics*. number of ... regularly occurring events ... in unit of time,

Frequency (from Webster's Dictionary)

Etymology: $<$ L *frequentia* $=$ assembly, multitude, crowd.

2. rate of occurrence
3. *Physics*. number of ... regularly occurring events ... in unit of time,
5. *Statistics*. the number of items occurring in a given category. Cf. **relative frequency**.

Frequency (from Webster's Dictionary)

Etymology: $<$ L *frequentia* $=$ assembly, multitude, crowd.

2. rate of occurrence

3. *Physics*. number of ... regularly occurring events ... in unit of time,

5. *Statistics*. the number of items occurring in a given category. Cf. **relative frequency**.

Meanings 3. and 5. are both relevant in epidemiology.

Frequency (from Webster's Dictionary)

Etymology: $<$ L *frequentia* $=$ assembly, multitude, crowd.

2. rate of occurrence
3. *Physics*. number of ... regularly occurring events ... in unit of time,
5. *Statistics*. the number of items occurring in a given category. Cf. **relative frequency**.

Meanings 3. and 5. are both relevant in epidemiology.

But what are **rate** and **occurrence**?

Cancer in Norden 1997 (NORDCAN)

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

	New cases	Crude rate	ASR (World)	Cumul. 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

Cancer in Norden 1997 (NORDCAN)

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

	New cases	Crude rate	ASR (World)	Cumul. 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?

Cancer in Norden 1997 (NORDCAN)

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

	New cases	Crude rate	ASR (World)	Cumul. 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

Questions on frequency & occurrence

How many women in Denmark

- are carriers of breast cancer today at 12? – **prevalence**

Questions on frequency & occurrence

How many women in Denmark

- are carriers of breast cancer today at 12? – **prevalence**
- will contract a new breast ca. during 2009? – **incidence**

Questions on frequency & occurrence

How many women in Denmark

- are carriers of breast cancer today at 12? – **prevalence**
- will contract a new breast ca. during 2009? – **incidence**
- die from breast ca. in 2009? – **mortality**

Questions on frequency & occurrence

How many women in Denmark

- are carriers of breast cancer today at 12? – **prevalence**
- will contract a new breast ca. during 2009? – **incidence**
- die from breast ca. in 2009? – **mortality**
- will be alive after 5 years since diagnosis among those getting breast ca. in 2009? – **survival**

Questions on frequency & occurrence

How many women in Denmark

- are carriers of breast cancer today at 12? – **prevalence**
- will contract a new breast ca. during 2009? – **incidence**
- die from breast ca. in 2009? – **mortality**
- will be alive after 5 years since diagnosis among those getting breast ca. in 2009? – **survival**
- are cured of breast cancer during 2009? – **cure**

Questions on frequency & occurrence

How many women in Denmark

- are carriers of breast cancer today at 12? – **prevalence**
- will contract a new breast ca. during 2009? – **incidence**
- die from breast ca. in 2009? – **mortality**
- will be alive after 5 years since diagnosis among those getting breast ca. in 2009? – **survival**
- are cured of breast cancer during 2009? – **cure**

What are the **proportions** or/and **rates** of occurrence of these states and events?

Questions on frequency & occurrence

Questions on frequency & occurrence

- How great are the **risks** of these events?

Questions on frequency & occurrence

- How great are the **risks** of these events?
- Is the frequency/occurrence/risk of breast ca. greater among nulliparous than parous women?

Questions on frequency & occurrence

- How great are the **risks** of these events?
- Is the frequency/occurrence/risk of breast ca. greater among nulliparous than parous women?
- What are the **excess** and **relative risks** for nulliparous compared to parous women?

Questions on frequency & occurrence

- How great are the **risks** of these events?
- Is the frequency/occurrence/risk of breast ca. 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 is risk?

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

What is risk?

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

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

What is risk?

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

(a) **probability** of *getting* S during a given **risk period**
$\rightarrow$ **incidence** probability,

What is risk?

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

(a) **probability** of *getting* S during a given **risk period**
$\rightarrow$ **incidence** probability,

(b) **rate** of change of that probability

What is risk?

What do we mean by “risk of disease S”?

(a) **probability** of *getting* S during a given **risk period**
$\rightarrow$ **incidence** probability,

(b) **rate** of change of that probability
$\rightarrow$ **hazard** or intensity,

What is risk?

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

(a) **probability** of *getting* S during a given **risk period**
$\rightarrow$ **incidence** probability,

(b) **rate** of change of that probability
$\rightarrow$ **hazard** or intensity,

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

What is risk?

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

(a) **probability** of *getting* S during a given **risk period**
$\rightarrow$ **incidence** probability,

(b) **rate** of change of that probability
$\rightarrow$ **hazard** or intensity,

(c) **probability** of *carrying* S at a given *time point*
$\rightarrow$ **prevalence** probability.

What is risk?

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

(a) **probability** of *getting* S during a given **risk period**
$\rightarrow$ **incidence** probability,

(b) **rate** of change of that probability
$\rightarrow$ **hazard** or intensity,

(c) **probability** of *carrying* S at a given *time point*
$\rightarrow$ **prevalence** probability.

Most commonly meaning (a) is attached with risk.

What is risk?

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

(a) **probability** of *getting* S during a given **risk period**
$\rightarrow$ **incidence** probability,

(b) **rate** of change of that probability
$\rightarrow$ **hazard** or intensity,

(c) **probability** of *carrying* S at a given *time point*
$\rightarrow$ **prevalence** probability.

Most commonly meaning (a) is attached with risk.

NB. "Risk" should not be used in the meaning of **risk factor**

What is risk?

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

(a) **probability** of *getting* S during a given **risk period**
$\rightarrow$ **incidence** probability,

(b) **rate** of change of that probability
$\rightarrow$ **hazard** or intensity,

(c) **probability** of *carrying* S at a given *time point*
$\rightarrow$ **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.

What is risk?

What do we mean by “risk of disease S”?

(a) **probability** of *getting* S during a given **risk period**
$\rightarrow$ **incidence** probability,

(b) **rate** of change of that probability
$\rightarrow$ **hazard** or intensity,

(c) **probability** of *carrying* S at a given *time point*
$\rightarrow$ **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

Risks are conditional probabilities

- There are no “absolute risks”.

Risks are conditional probabilities

- There are no "absolute risks".
- All risks are conditional on a multitude of factors, like

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),

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,

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,

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.

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 an own quantitative value for the risk of given disease in any defined risk period, depending on his/her own risk factor profile.

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 an own quantitative 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.

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 an own quantitative 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)

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 an own quantitative 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**.

Descriptive and causal studies

Descriptive and causal studies

- **Descriptive:** What is the occurrence of lung cancer workers exposed to silica dust as compared to that in subjects of other occupations?

Descriptive and causal studies

- **Descriptive:** What is the occurrence of lung cancer workers exposed to silica dust as compared to that in subjects of other occupations?
- **Causal:** What is the risk of lung cancer among silica dust workers *as compared to* . . .

Descriptive and causal studies

- **Descriptive:** What is the occurrence of lung cancer workers exposed to silica dust as compared to that in subjects of other occupations?
- **Causal:** What is the risk of lung cancer among silica dust workers *as compared to* ... what the risk in these same men *would be, had they not been* exposed to silica?

Descriptive and causal studies

- **Descriptive:** What is the occurrence of lung cancer workers exposed to silica dust as compared to that in subjects of other occupations?
- **Causal:** What is the risk of lung cancer among silica dust workers *as compared to* ... what the risk in these same men *would be, had they not been* exposed to silica?

NB. Causal question – **counterfactual conditional**!

Descriptive and causal studies

- **Descriptive:** What is the occurrence of lung cancer workers exposed to silica dust as compared to that in subjects of other occupations?
- **Causal:** What is the risk of lung cancer among silica dust workers *as compared to* . . . what the risk in these same men *would be, had they not been* exposed to silica?

NB. Causal question – **counterfactual conditional**!

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

Types of epidemiologic studies

Types of epidemiologic studies

Can crudely be classified in following axes:

Types of epidemiologic studies

Can crudely be classified in following axes:

- *study unit*: individual – aggregate (ecological study)

Types of epidemiologic studies

Can crudely be classified in following axes:

- *study unit*: individual – aggregate (ecological study)
- *allocation of exposure*: experimental – observational

Types of epidemiologic studies

Can crudely be classified in following axes:

- *study unit*: individual – aggregate (ecological study)
- *allocation of exposure*: experimental – observational
- *population*: closed (cohort) – open (dynamic)

Types of epidemiologic studies

Can crudely be classified in following axes:

- *study unit*: individual – aggregate (ecological study)
- *allocation of exposure*: experimental – observational
- *population*: closed (cohort) – open (dynamic)
- *dimensionality*: cross-sectional – longitudinal

Types of epidemiologic studies

Can crudely be classified in following axes:

- *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")

Types of epidemiologic studies

Can crudely be classified in following axes:

- *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

Types of epidemiologic studies

Can crudely be classified in following axes:

- *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 & case-control* studies.

Experimental and observational studies

Experimental and observational studies

Allocation of exposure in etiologic studies?

Experimental and observational studies

Allocation of exposure in etiologic studies?

- **Experimental**:

Experimental and observational studies

Allocation of exposure in etiologic studies?

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

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.

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.

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.

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.

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.

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.

Study population

Study population

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

Study population

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

- **closed – cohort**: members taken by certain event, *e.g.*

Study population

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

- **closed – cohort**: members taken by certain event, *e.g.*

    (i) birth cohort, people born during same year,

Study population

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

- **closed – cohort**: members taken by certain event, *e.g.*

  (i) birth cohort, people born during same year,
  (ii) workers employed by Carlsberg brewery during 1970's,

Study population

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

- **closed – cohort**: members taken by certain event, *e.g.*

  (i) birth cohort, people born during same year,
  (ii) workers employed by Carlsberg brewery during 1970's, followed up since then, even after retirement

Study population

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

- **closed – cohort**: members taken by certain event, *e.g.*

  (i) birth cohort, people born during same year,
  (ii) workers employed by Carlsberg brewery during 1970's, followed up since then, even after retirement

  Once taken in, you can't escape from a cohort.

Study population

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

- **closed – cohort**: members taken by certain event, *e.g.*

  (i) birth cohort, people born during same year,
  (ii) workers employed by Carlsberg brewery during 1970's, followed up since then, even after retirement

  Once taken in, you can't escape from a cohort.

- **open – dynamic**: defined by changeable status, *e.g.*

Study population

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

- **closed – cohort**: members taken by certain event, *e.g.*

  (i) birth cohort, people born during same year,
  (ii) workers employed by Carlsberg brewery during 1970's, followed up since then, even after retirement

  Once taken in, you can't escape from a cohort.

- **open – dynamic**: defined by changeable status, *e.g.*

  (i) citizens of Copenhagen, currently resident;

  One may leave an open pop'n and come back to it.

Study population

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

- **closed – cohort**: members taken by certain event, *e.g.*

  (i) birth cohort, people born during same year,
  (ii) workers employed by Carlsberg brewery during 1970's, followed up since then, even after retirement

  Once taken in, you can't escape from a cohort.

- **open – dynamic**: defined by changeable status, *e.g.*

  (i) citizens of Copenhagen, currently resident;
  (ii) *catchment population* of the Oncological Clinic at Rigshospitalet (CPH),

  One may leave an open pop'n and come back to it.

Study base and its dimensionality

Study base and its dimensionality

Study base

$=$ Study population $\times$ its experience in time.

Study base and its dimensionality

Study base
= Study population $\times$ its experience in time.

Dimensionality of study base

- **Cross-sectional**:
  Study base = study population at a defined time point.

Study base and its dimensionality

Study base
= Study population × its experience in time.

Dimensionality of study base

- **Cross-sectional**:

  Study base = study population at a defined time point.

  - *e.g.* all newborn in Denmark 2009 at their dates of birth.

Study base and its dimensionality

Study base
= Study population $\times$ its experience in time.

Dimensionality of study base

- **Cross-sectional**:

  Study base = study population at a defined time point.

  - *e.g.* all newborn in Denmark 2009 at their dates of birth.

- **Longitudinal**:

  Study base comprises **follow-up times** of individuals in the study population over a given period.

Study base and its dimensionality

Study base
= Study population $\times$ its experience in time.

Dimensionality of study base

- **Cross-sectional**:
  Study base = study population at a defined time point.
  - *e.g.* all newborn in Denmark 2009 at their dates of birth.
- **Longitudinal**:
  Study base comprises **follow-up times** of individuals in the study population over a given period.

Causal research $\rightarrow$ longitudinal base preferred.

Study base (cont'd)

Longitudinal base:

Study base (cont'd)

Longitudinal base:

(a) *Cohort:* Individual time intervals from **entry** until **exit**, at which the *outcome* or *censoring* occurs.

Study base (cont'd)

Longitudinal base:

(a) *Cohort:* Individual time intervals from **entry** until **exit**, at which the *outcome* or *censoring* occurs.

(b) *Dynamic population:* Each subject contributes possibly several time intervals of membership since the 1st entry until the 'final' exit.

Study base (cont'd)

Longitudinal base:

(a) *Cohort:* Individual time intervals from
entry until **exit**, at which
the *outcome* or *censoring* occurs.

(b) *Dynamic population:* Each subject contributes possibly several time intervals of membership since the 1st entry until the 'final' exit.

- Person-time calculation complicated.

Study base (cont'd)

Longitudinal base:

(a) *Cohort:* Individual time intervals from
entry until **exit**, at which
the *outcome* or *censoring* occurs.

(b) *Dynamic population:* Each subject contributes possibly several time intervals of membership since the 1st entry until the 'final' exit.

- Person-time calculation complicated.
- Population-based annual (or 5-year period) incidence and mortality statistics:
  $Y \approx$ *mid-population* $\times$ length of period.

Measurement of exposures and outcomes (IS, ch.2)

In epidemiological studies, it is neccessary to measure

Measurement of exposures and outcomes (IS, ch.2)

In epidemiological studies, it is neccessary to measure

(1) the primary **exposure**(s) of interest,

Measurement of exposures and outcomes (IS, ch.2)

In epidemiological studies, it is neccessary to measure

(1) the primary **exposure**(s) of interest,

(2) other exposure(s), potential **confounders** and **modifiers**,

Measurement of exposures and outcomes (IS, ch.2)

In epidemiological studies, it is neccessary to measure

(1) the primary **exposure**(s) of interest,

(2) other exposure(s), potential **confounders** and **modifiers**,

(3) the **outcome**(s) of interest.

Measurement of exposures and outcomes (IS, ch.2)

In epidemiological studies, it is neccessary to measure

(1) the primary **exposure**(s) of interest,

(2) other exposure(s), potential **confounders** and **modifiers**,

(3) the **outcome**(s) of interest.

Many approaches, *e.g.*

Measurement of exposures and outcomes (IS, ch.2)

In epidemiological studies, it is neccessary to measure

(1) the primary **exposure**(s) of interest,
(2) other exposure(s), potential **confounders** and **modifiers**,
(3) the **outcome**(s) of interest.

Many approaches, *e.g.*

- personal interviews & questionnaires, diaries,

Measurement of exposures and outcomes (IS, ch.2)

In epidemiological studies, it is neccessary to measure

(1) the primary **exposure**(s) of interest,
(2) other exposure(s), potential **confounders** and **modifiers**,
(3) the **outcome**(s) of interest.

Many approaches, *e.g.*

- personal interviews & questionnaires, diaries,
- hospital records, other routine data,

Measurement of exposures and outcomes (IS, ch.2)

In epidemiological studies, it is neccessary to measure

(1) the primary **exposure**(s) of interest,
(2) other exposure(s), potential **confounders** and **modifiers**,
(3) the **outcome**(s) of interest.

Many approaches, *e.g.*

- personal interviews & questionnaires, diaries,
- hospital records, other routine data,
- biological and environmental measurements.

MEASURING FREQUENCY

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

MEASURING FREQUENCY

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

MEASURING FREQUENCY

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).

MEASURING FREQUENCY

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).

⇒ challenge to accurate diagnosis and classification!

MEASURING FREQUENCY

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).

⇒ challenge to accurate diagnosis and classification!

(2) the **population** from which the cases originate.

MEASURING FREQUENCY

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

Types of occurrence measures

- Longitudinal – **incidence** measures.

General form of frequency or occurrence measures

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

Types of occurrence measures

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

General form of frequency or occurrence measures

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

Types of occurrence measures

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

General form of frequency or occurrence measures

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

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.

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.

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

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.

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

Numerator and denominator must cover the *same population*.

Incidence measures

Incidence measures

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

Incidence measures

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

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

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 **"risk"** (*e.g.* **IS**).

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 **"risk"** (*e.g.* **IS**).

NB. "Cumulative incidence" has other meanings, too.

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 **"risk"** (*e.g.* **IS**).

  NB. "Cumulative incidence" has other meanings, too.

- **Indidence rate** (I) over a defined observation period:

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 **"risk"** (*e.g.* **IS**).

  NB. "Cumulative incidence" has other meanings, too.

- **Indidence rate** (I) over a defined observation period:

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

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 **“risk”** (*e.g.* **IS**).

  NB. “Cumulative incidence” has other meanings, too.

- **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**.

Example: Follow-up of a small cohort

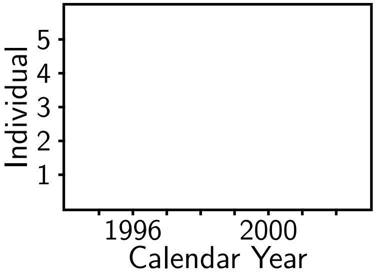
$|$ = entry, $\circ$ = exit with censoring; outcome not observed,

$\bullet$ = exit with outcome event (disease onset) observed

Example: Follow-up of a small cohort

$|$ = entry, $\circ$ = exit with censoring; outcome not observed,

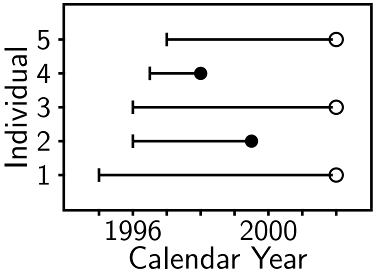
$\bullet$ = exit with outcome event (disease onset) observed



Example: Follow-up of a small cohort

$|$ = entry, $\circ$ = exit with censoring; outcome not observed,

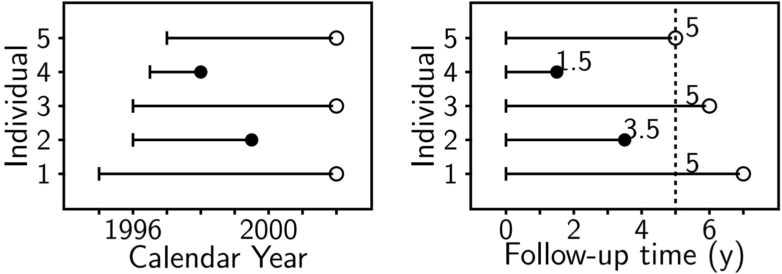
$\bullet$ = exit with outcome event (disease onset) observed



Example: Follow-up of a small cohort

$|$ = entry, $\circ$ = exit with censoring; outcome not observed,

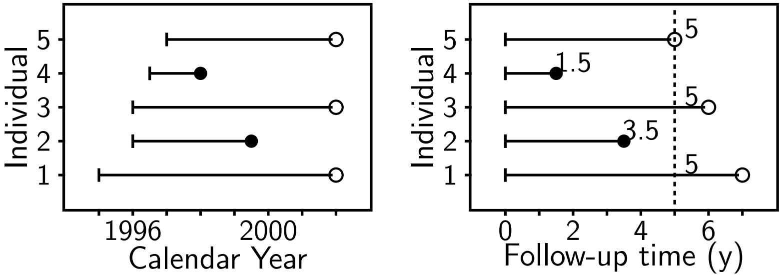
$\bullet$ = exit with outcome event (disease onset) observed



Example: Follow-up of a small cohort

| = entry, ○ = exit with censoring; outcome not observed,

• = exit with outcome event (disease onset) observed

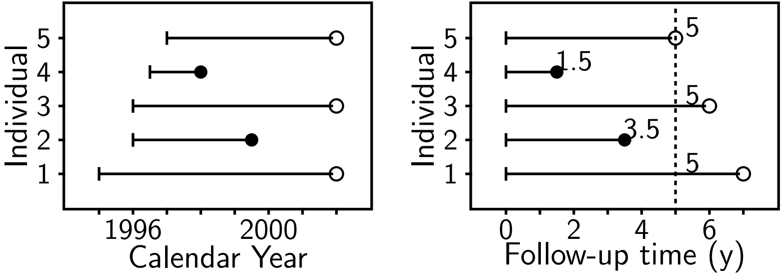


Complete follow-up in the 5-year risk period

Example: Follow-up of a small cohort

| = entry, ∘ = exit with censoring; outcome not observed,

• = exit with outcome event (disease onset) observed



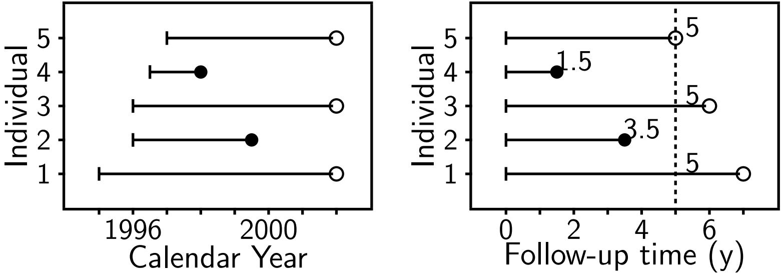
Complete follow-up in the 5-year risk period

⇒ can calculate both measures:

Example: Follow-up of a small cohort

$|$ = entry, $\circ$ = exit with censoring; outcome not observed,

$\bullet$ = exit with outcome event (disease onset) observed



Complete follow-up in the 5-year risk period

$\Rightarrow$ can calculate both measures:

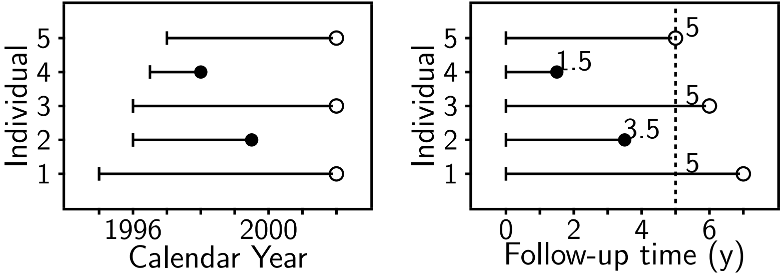
$$\text{Inc. rate} = \frac{2 \text{ cases}}{5 + 3.5 + 5 + 1.5 + 5 \text{ years}} = 10 \text{ per } 100 \text{ years}$$

$$= 2/5 = 0.4 \; (40 \,\%)$$

Example: Follow-up of a small cohort

$|$ = entry, $\circ$ = exit with censoring; outcome not observed,

$\bullet$ = exit with outcome event (disease onset) observed



Complete follow-up in the 5-year risk period

$\Rightarrow$ can calculate both measures:

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

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

Properties of incidence proportion

Properties of incidence proportion

- Dimensionless quantity ranging from 0 to 1 (0% to 100%) = *relative frequency*,

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,

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**

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,

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),

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 ⇒ Calculations need to be corrected using special methods of survival analysis.

Properties of incidence rate

Properties of incidence rate

- Like *a frequency* quantity in physics; measurement unit: *e.g.* Hz $= 1/$second, $1/$year, or $1/1000$ y.

Properties of incidence rate

- Like *a frequency* quantity in physics; measurement unit: *e.g.* Hz $= 1/$second, $1/$year, or $1/1000$ y.
- Estimates the average underlying **intensity** or **hazard rate** of the outcome in a population,

Properties of incidence rate

- Like *a frequency* quantity in physics; measurement unit: *e.g.* Hz = 1/second, 1/year, or 1/1000 y.
- Estimates the average underlying **intensity** or **hazard rate** of the outcome in a population,
- Estimation accurate in the **constant hazard model**,

Properties of incidence rate

- Like *a frequency* quantity in physics; measurement unit: *e.g.* Hz = 1/second, 1/year, or 1/1000 y.
- 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.

Properties of incidence rate

- Like *a frequency* quantity in physics; measurement unit: *e.g.* Hz $= 1/$second, $1/$year, or $1/1000$ y.
- 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)
  $\Rightarrow$ rates *specific to age group etc.* needed,

Properties of incidence rate

- Like *a frequency* quantity in physics; measurement unit: *e.g.* Hz $= 1/$second, $1/$year, or $1/1000$ y.
- 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) $\Rightarrow$ rates *specific to age group etc.* needed,
- Incidence proportions can be estimated from rates.

Properties of incidence rate

- Like *a frequency* quantity in physics; measurement unit: *e.g.* Hz = 1/second, 1/year, or 1/1000 y.
- 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) $\Rightarrow$ rates *specific to age group etc.* needed,
- Incidence proportions can be estimated from rates. In the constant hazard model with no competing risks:

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

Competing events and censoring

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.

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.

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.

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,

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

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**

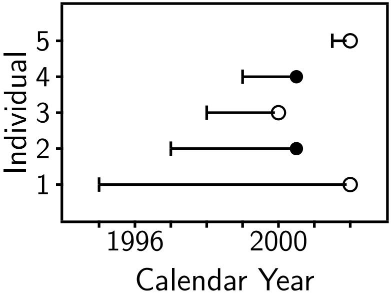
Competing events and censoring

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

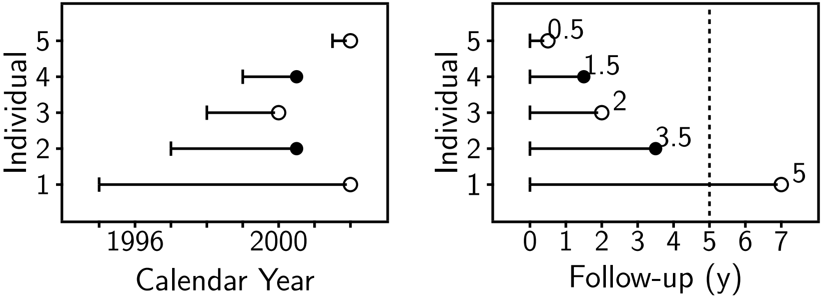
- Some subjects die (from other causes) before the event.
  - ⇒ Death is a **competing event** after which the outcome cannot occur any more.
- Others emigrate and escape national disease registration, or the whole study is closed "now", which prematurely interrupts the follow-up of some individuals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

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

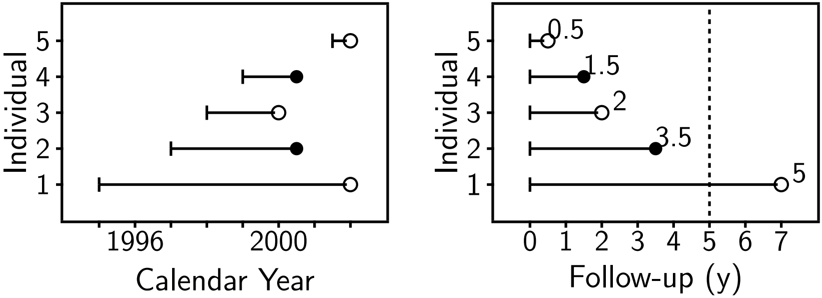
Follow-up of another small cohort



Follow-up of another small cohort

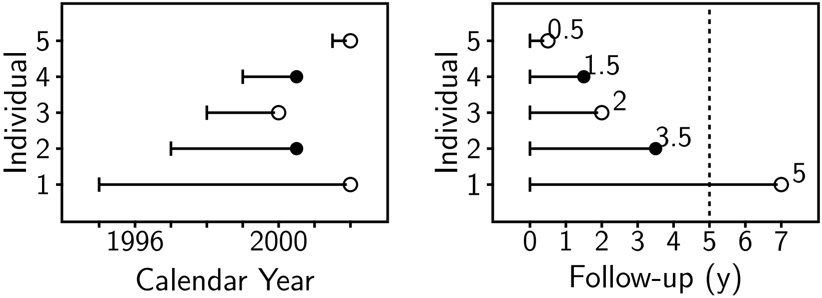


Follow-up of another small cohort



Two censored observations ⇒ can calculate the rate:

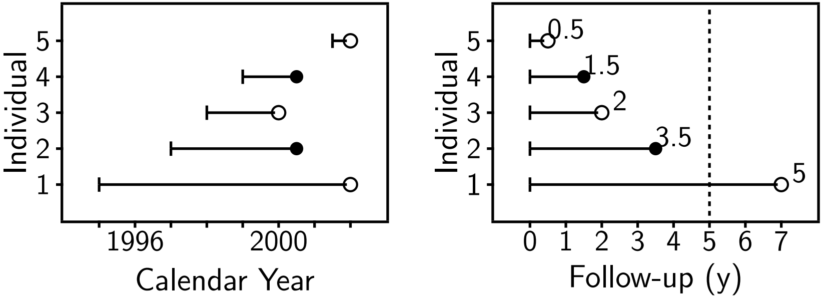
Follow-up of another small cohort



Two censored observations ⇒ can calculate the rate:

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

Follow-up of another small cohort

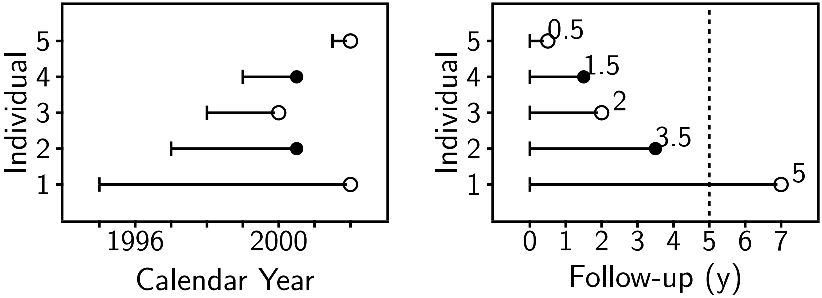


Two censored observations ⇒ can calculate the rate:

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

but the 5-year Q **IS NO MORE** $2/5$!

Follow-up of another small cohort

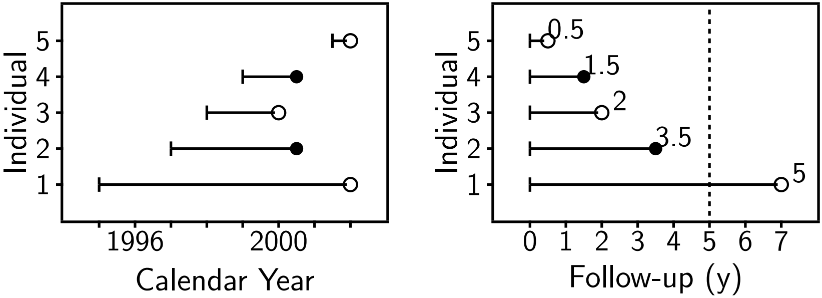


Two censored observations $\Rightarrow$ can calculate the rate:

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

but the 5-year Q **IS NO MORE** $2/5$!
However, under constant rate model and in the absence of competing risks:

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 and in the absence of competing risks:

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

Person-years in dynamic populations

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

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:

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:

(1) 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,

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:

(1) 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,

(2) Mid-population for the period: $\bar{N}_t = \frac{1}{2} \times (N_{t-1} + N_t)$.

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:

(1) 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,

(2) Mid-population for the period: $\bar{N}_t = \frac{1}{2} \times (N_{t-1} + N_t)$.

(3) Approximate person-years: $Y_t \approx \bar{N}_t \times L_t$.

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:

(1) 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,

(2) Mid-population for the period: $\bar{N}_t = \frac{1}{2} \times (N_{t-1} + N_t)$.

(3) 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

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):

1992: $\frac{1}{2} \times (2443 + 2457) \times 1 =$ 2450

$\frac{1}{2} \times (2457 + 2482) \times 2 =$ 4937

$\frac{1}{2} \times (2431 + 2492) \times 5 =$ 12307.5

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):

1992:	$\frac{1}{2} \times (2443 + 2457) \times 1 =$	2450
1993-94:	$\frac{1}{2} \times (2457 + 2482) \times 2 =$	4937
	$\frac{1}{2} \times (2431 + 2492) \times 5 =$	12307.5

Male person-years in Finland 1991-95

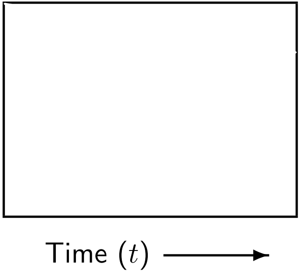
Total male population (1000s) on 31 December by year:

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

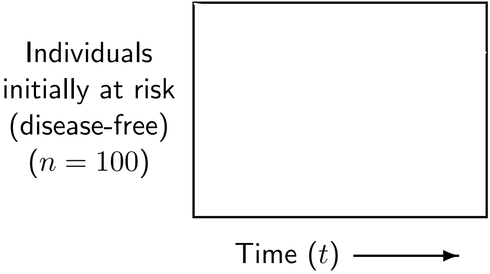
Approximate person-years (1000s):

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

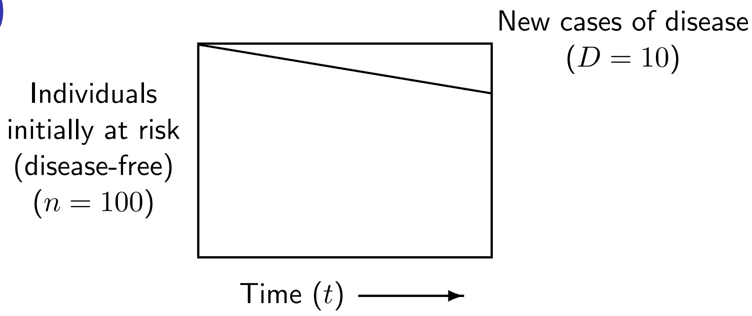
Incidence proportion, rate, and odds (IS, Ex 4.5)



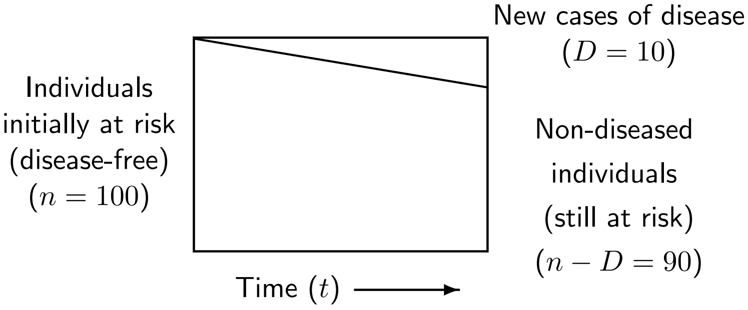
Incidence proportion, rate, and odds (IS, Ex 4.5)



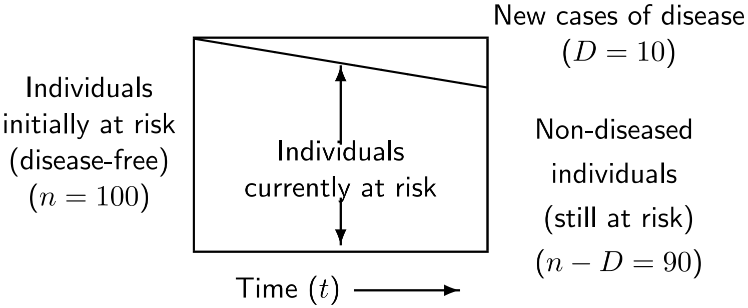
Incidence proportion, rate, and odds (IS, Ex 4.5)



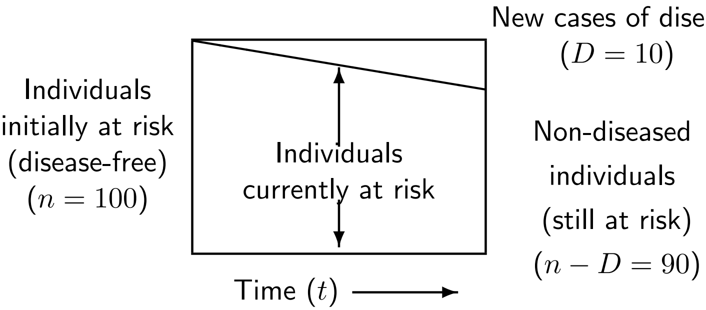
Incidence proportion, rate, and odds (IS, Ex 4.5)



Incidence proportion, rate, and odds (IS, Ex 4.5)

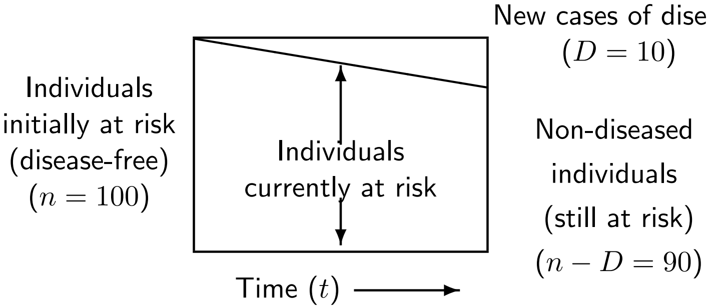


Incidence proportion, rate, and odds (IS, Ex 4.5)



Assuming a study period of 1 year with complete follow-up:

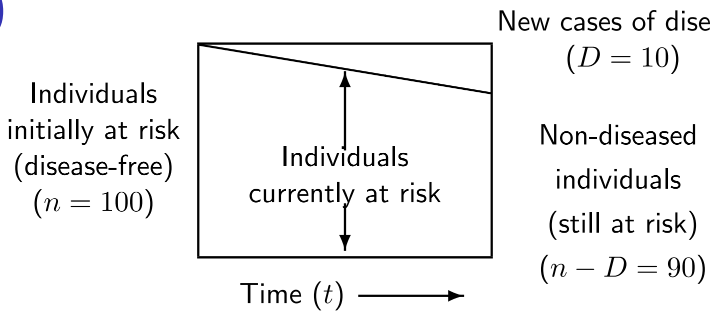
Incidence proportion, rate, and odds (IS, Ex 4.5)



Assuming a study period of 1 year with complete follow-up:

$$\begin{aligned} \text{Incidence proportion } Q &= 10/100 = 0.10 = 10\% \\ &= 10/95 \text{ y} = 10.5 \text{ per } 100 \text{ y} \\ &= 10/90 = 0.11 = 11 \text{ per } 100 \end{aligned}$$

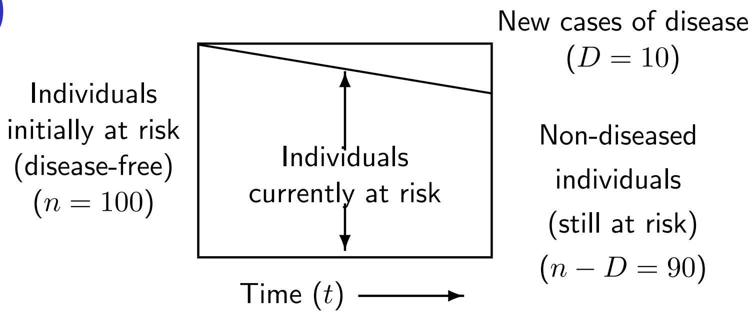
Incidence proportion, rate, and odds (IS, Ex 4.5)



Assuming a study period of 1 year with complete follow-up:

$$
\begin{aligned}
\text{Incidence proportion } Q &= 10/100 = 0.10 = 10\% \\
\text{Incidence rate } I &= 10/95 \text{ y} = 10.5 \text{ per } 100 \text{ y} \\
&= 10/90 = 0.11 = 11 \text{ per } 100
\end{aligned}
$$

Incidence proportion, rate, and odds (IS, Ex 4.5)



Assuming a study period of 1 year with complete follow-up:

$$\begin{aligned} \text{Incidence proportion } Q &= 10/100 = 0.10 = 10\% \\ \text{Incidence rate } I &= 10/95 \text{ y} = 10.5 \text{ per } 100 \text{ y} \\ \textbf{Incidence odds } Q/(1-Q) &= 10/90 = 0.11 = 11 \text{ per } 100 \end{aligned}$$

Approximate relations btw measures

Approximate relations btw measures

With sufficiently

- "short" length Δ of risk period and

the incidence proportion Q, rate I and odds are approximately related:

Approximate relations btw measures

With sufficiently

- "short" length Δ of risk period and
- "low" risk (say $Q < 10\%$)

the incidence proportion Q, rate I and odds are approximately related:

Approximate relations btw measures

With sufficiently

- "short" length Δ of risk period and
- "low" risk (say $Q < 10\%$)

the incidence proportion Q, rate I and odds are approximately related:

$$\frac{Q}{1-Q} \approx Q \approx I \times \Delta$$

Approximate relations btw measures

With sufficiently

- "short" length Δ of risk period and
- "low" risk (say $Q < 10\%$)

the incidence proportion Q, rate I and odds are approximately related:

$$\frac{Q}{1-Q} \approx Q \approx I \times \Delta$$

The "**rare disease assumption**".

Mortality

Mortality

Cause-specific mortality from disease S is described by **mortality rates** defined like I but

Mortality

Cause-specific mortality from disease S is described by **mortality rates** defined like I but

- cases are *deaths* from S, and

Cause-specific **mortality proportions** must be corrected for the incidence of **competing causes of death**

Mortality

Cause-specific mortality from disease S is described by **mortality rates** defined like I but

- cases are *deaths* from S, and
- follow-up is extended until death or censoring.

Cause-specific **mortality proportions** must be corrected for the incidence of **competing causes of death**

Mortality

Cause-specific mortality from disease S is described by **mortality rates** defined like I but

- cases are *deaths* from S, and
- follow-up is extended until death or censoring.

Cause-specific **mortality proportions** must be corrected for the incidence of **competing causes of death**

Total mortality:

- cases are deaths from any cause.

Mortality

Cause-specific mortality from disease S is described by **mortality rates** defined like I but

- cases are *deaths* from S, and
- follow-up is extended until death or censoring.

Cause-specific **mortality proportions** must be corrected for the incidence of **competing causes of death**

Total mortality:

- cases are deaths from any cause.

Mortality depends on the incidence and the **prognosis** or **case fatality** of the disease, *i.e.* the **survival** of those affected by it.

Mathematical concepts describing risks

Mathematical concepts describing risks

Analysis of risks

Mathematical concepts describing risks

Analysis of risks = analysis of **times to event** or

Mathematical concepts describing risks

Analysis of risks = analysis of **times to event** or **failure times** or

Mathematical concepts describing risks

Analysis of risks = analysis of **times to event** or **failure times** or **survival** data.

Mathematical concepts describing risks

Analysis of risks = analysis of **times to event** or **failure times** or **survival** data.

$$
\begin{aligned}
T &= \text{ \textbf{time} to outcome event – random variable,} \\
&\quad \text{which has a probability distribution with} \\
&= P(T \leq t) = \text{\textbf{risk} function (cumul. distrib. f.)} \\
&= \text{probability of the outcome to occur before } t, \\
&= P(T > t) = 1 - F(t) = \text{ \textbf{survival} function of } T, \\
&= \text{ probability of avoiding the event up to given time } t, \\
&= F'(t) = \text{ \textbf{density} function of } T, \\
&= -\frac{S'(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \text{ \textbf{intensity} or \textbf{hazard} function,} \\
&= \int_0^t \lambda(u) du = -\log S(t) = \text{\textbf{cumulative hazard},} \\
\Leftrightarrow &\quad F(t) = 1 - \exp\{-\Lambda(t)\}, \quad f(t) = \lambda(t) S(t).
\end{aligned}
$$

Mathematical concepts describing risks

Analysis of risks = analysis of **times to event** or **failure times** or **survival** data.

T = **time** to outcome event – random variable, which has a probability distribution with

$F(t)$ = $P(T \leq t)$ = **risk** function (cumul. distrib. f.)

= probability of the outcome to occur before t,

= $P(T > t) = 1 - F(t) =$ **survival** function of T,

= probability of avoiding the event up to given time t,

= $F'(t) =$ **density** function of T,

= $-\dfrac{S'(t)}{S(t)} = \dfrac{f(t)}{1 - F(t)}$ **intensity** or **hazard** function,

= $\displaystyle\int_0^t \lambda(u)du = -\log S(t) =$ **cumulative hazard**,

$\Leftrightarrow$ $F(t) = 1 - \exp\{-\Lambda(t)\}, \quad f(t) = \lambda(t)S(t).$

Mathematical concepts describing risks

Analysis of risks = analysis of **times to event** or **failure times** or **survival** data.

$$
\begin{aligned}
T &= \text{ \textbf{time} to outcome event -- random variable,} \\
  & \quad \text{which has a probability distribution with} \\
F(t) &= P(T \le t) = \textbf{risk} \text{ function (cumul. distrib. f.)} \\
  &= \text{probability of the outcome to occur before } t, \\
S(t) &= P(T > t) = 1 - F(t) = \textbf{ survival} \text{ function of } T, \\
  &= \text{ probability of avoiding the event up to given time } t, \\
  &= F'(t) = \textbf{ density} \text{ function of } T, \\
  &= -\frac{S'(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \textbf{ intensity} \text{ or } \textbf{hazard} \text{ function,} \\
  &= \int_0^t \lambda(u) du = -\log S(t) = \textbf{cumulative hazard}, \\
\Leftrightarrow & \quad F(t) = 1 - \exp\{-\Lambda(t)\}, \quad f(t) = \lambda(t) S(t).
\end{aligned}
$$

Mathematical concepts describing risks

Analysis of risks = analysis of **times to event** or **failure times** or **survival** data.

$$
\begin{aligned}
T &= \text{time to outcome event – random variable,} \\
& \quad \text{which has a probability distribution with} \\
F(t) &= P(T \le t) = \textbf{risk} \text{ function (cumul. distrib. f.)} \\
&= \text{probability of the outcome to occur before } t, \\
S(t) &= P(T > t) = 1 - F(t) = \textbf{survival} \text{ function of } T, \\
&= \text{probability of avoiding the event up to given time } t, \\
f(t) &= F'(t) = \textbf{density} \text{ function of } T, \\
&= -\frac{S'(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \ \textbf{intensity} \text{ or } \textbf{hazard} \text{ function,} \\
&= \int_0^t \lambda(u) du = -\log S(t) = \textbf{cumulative hazard}, \\
\Leftrightarrow & \quad F(t) = 1 - \exp\{-\Lambda(t)\}, \quad f(t) = \lambda(t) S(t).
\end{aligned}
$$

Mathematical concepts describing risks

Analysis of risks = analysis of **times to event** or **failure times** or **survival** data.

$$
\begin{aligned}
T &= \textbf{time} \text{ to outcome event – random variable,} \\
&\quad \text{which has a probability distribution with} \\
F(t) &= P(T \le t) = \textbf{risk} \text{ function (cumul. distrib. f.)} \\
&= \text{probability of the outcome to occur before } t, \\
S(t) &= P(T > t) = 1 - F(t) = \textbf{survival} \text{ function of } T, \\
&= \text{probability of avoiding the event up to given time } t, \\
f(t) &= F'(t) = \textbf{density} \text{ function of } T, \\
\lambda(t) &= -\frac{S'(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \ \textbf{intensity} \text{ or } \textbf{hazard} \text{ function,} \\
&= \int_0^t \lambda(u)du = -\log S(t) = \textbf{cumulative hazard}, \\
&\Leftrightarrow F(t) = 1 - \exp\{-\Lambda(t)\}, \quad f(t) = \lambda(t)S(t).
\end{aligned}
$$

Mathematical concepts describing risks

Analysis of risks = analysis of **times to event** or **failure times** or **survival** data.

$$
\begin{aligned}
T &= \text{\textbf{time} to outcome event -- random variable,} \\
&\quad\ \text{which has a probability distribution with} \\
F(t) &= P(T \le t) = \text{\textbf{risk} function (cumul. distrib. f.)} \\
&= \text{probability of the outcome to occur before } t, \\
S(t) &= P(T > t) = 1 - F(t) = \text{ \textbf{survival} function of } T, \\
&= \text{ probability of avoiding the event up to given time } t, \\
f(t) &= F'(t) = \text{ \textbf{density} function of } T, \\
\lambda(t) &= -\frac{S'(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \text{ \textbf{intensity} or \textbf{hazard} function,} \\
\Lambda(t) &= \int_0^t \lambda(u)du = -\log S(t) = \text{\textbf{cumulative hazard}}, \\
&\Leftrightarrow F(t) = 1 - \exp\{-\Lambda(t)\}, \quad f(t) = \lambda(t)S(t).
\end{aligned}
$$

Mathematical concepts describing risks

Analysis of risks = analysis of **times to event** or **failure times** or **survival** data.

$$
\begin{aligned}
T &= \text{\textbf{time} to outcome event – random variable,} \\
&\quad \text{which has a probability distribution with} \\
F(t) &= P(T \leq t) = \text{\textbf{risk} function (cumul. distrib. f.)} \\
&= \text{probability of the outcome to occur before } t, \\
S(t) &= P(T > t) = 1 - F(t) = \text{ \textbf{survival} function of } T, \\
&= \text{ probability of avoiding the event up to given time } t, \\
f(t) &= F'(t) = \text{ \textbf{density} function of } T, \\
\lambda(t) &= -\frac{S'(t)}{S(t)} = \frac{f(t)}{1 - F(t)} \text{ \textbf{intensity} or \textbf{hazard} function,} \\
\Lambda(t) &= \int_0^t \lambda(u) du = -\log S(t) = \text{\textbf{cumulative hazard},} \\
&\Leftrightarrow F(t) = 1 - \exp\{-\Lambda(t)\}, \quad f(t) = \lambda(t) S(t).
\end{aligned}
$$

Hazard and risk

Hazard and risk

Hazard or intensity can be viewed as **theoretical incidence rate**.

Hazard and risk

Hazard or intensity can be viewed as **theoretical incidence rate**. Formally defined

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

Hazard and risk

Hazard or intensity can be viewed as **theoretical incidence rate**. Formally defined

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

$\approx$ 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.

Hazard and risk

Hazard or intensity can be viewed as **theoretical incidence rate**. Formally defined

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

$\approx$ 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.

This is equivalent to saying that over a short interval

risk $\approx$ intensity $\times$ length of interval

Hazard and risk

Hazard or intensity can be viewed as **theoretical incidence rate**. Formally defined

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

$\approx$ 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.

This is equivalent to saying that over a short interval

risk $\approx$ intensity $\times$ length of interval

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

Exponential or constant hazard model

Simplest probability model for time to event:

Exponential or constant hazard model

Simplest probability model for time to event:

Exponential distribution, Exp(λ), in which

Exponential or constant hazard model

Simplest probability model for time to event:

Exponential distribution, Exp(λ), in which

$$\begin{array}{rrcl} & \text{rate at any time } t: & \lambda(t) & = & \lambda, \text{ constant over time} \\ \Rightarrow & \text{risk over period }]0,t]: & F(t) & = & 1 - \exp(-\lambda t) \end{array}$$

Exponential or constant hazard model

Simplest probability model for time to event:

Exponential distribution, Exp(λ), in which

$$\begin{aligned} \text{rate at any time } t: \quad \lambda(t) &= \lambda, \text{ constant over time} \\ \Rightarrow \text{ risk over period }]0,t]: \quad F(t) &= 1 - \exp(-\lambda t) \end{aligned}$$

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

Exponential or constant hazard model

Simplest probability model for time to event:

Exponential distribution, Exp(λ), in which

$$\begin{aligned} \text{rate at any time } t: \quad \lambda(t) &= \lambda, \text{ constant over time} \\ \Rightarrow \text{ risk over period }]0,t]: \quad F(t) &= 1 - \exp(-\lambda t) \end{aligned}$$

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

$$\begin{aligned} y_i &= \text{time to event or censoring, total:} \quad Y = \sum y_i \\ &= \text{1/0-indicator for observing event, total:} \quad D = \sum d_i \end{aligned}$$

Exponential or constant hazard model

Simplest probability model for time to event:

Exponential distribution, Exp(λ), in which

$$\begin{aligned} \text{rate at any time } t: \quad \lambda(t) &= \lambda, \text{ constant over time} \\ \Rightarrow \text{ risk over period }]0,t]: \quad F(t) &= 1 - \exp(-\lambda t) \end{aligned}$$

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

$$\begin{aligned} y_i &= \text{time to event or censoring, total:} \quad Y = \sum y_i \\ d_i &= 1/0\text{-indicator for observing event, total:} \quad D = \sum d_i \end{aligned}$$

Exponential or constant hazard model

Simplest probability model for time to event:

Exponential distribution, Exp(λ), in which

$$\begin{aligned} \text{rate at any time } t: \quad \lambda(t) &= \lambda, \text{ constant over time} \\ \Rightarrow \text{ risk over period }]0,t]: \quad F(t) &= 1 - \exp(-\lambda t) \end{aligned}$$

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

$$\begin{aligned} y_i &= \text{time to event or censoring, total: } \quad Y = \sum y_i \\ d_i &= 1/0\text{-indicator for observing event, total: } \quad D = \sum d_i \end{aligned}$$

Exp(λ) model $\Rightarrow$ **Likelihood function** of λ is equivalent to that when number of cases D would be **Poisson**-distributed

Basic statistical analysis of rates

Asymptotic statistical inference based on likelihood:

Basic statistical analysis of rates

Asymptotic statistical inference based on likelihood:

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

Basic statistical analysis of rates

Asymptotic statistical inference based on likelihood:

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

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

Basic statistical analysis of 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 $I \times 1/\sqrt{D}$

Basic statistical analysis of 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 $I \times 1/\sqrt{D}$

$\Rightarrow$ The more cases, the greater is **precision** in rate!

Basic statistical analysis of 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 $I \times 1/\sqrt{D}$

$\Rightarrow$ The more cases, the greater is **precision** in rate!

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

Basic statistical analysis of 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 $I \times 1/\sqrt{D}$

$\Rightarrow$ The more cases, the greater is **precision** in rate!

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

$$\text{estimator} \ \pm \ 1.96 \times \text{standard error}$$

Basic statistical analysis of 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 $I \times 1/\sqrt{D}$

$\Rightarrow$ The more cases, the greater is **precision** in rate!

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

$$\text{estimator} \pm 1.96 \times \text{standard error}$$

More about these issues in Bendix's lectures next week.

Prevalence measures

Prevalence measures

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

Prevalence measures

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

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

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.

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

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

Prevalence measures

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

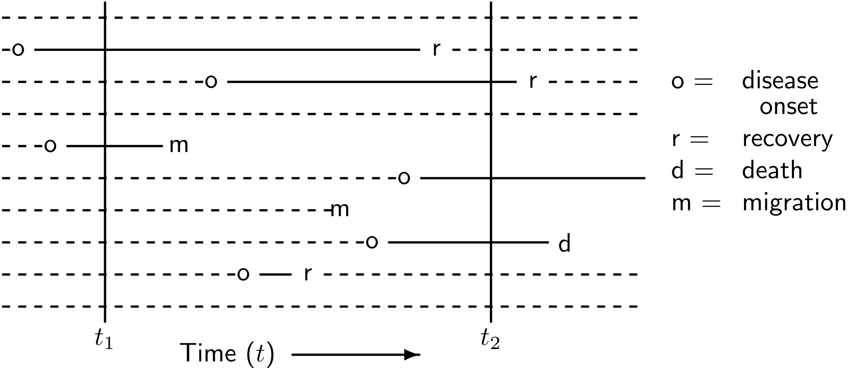
$$P = \frac{\text{number of existing or prevalent cases of } C}{\text{size of the whole population}}$$

This is calculable from a cross-sectional study base.

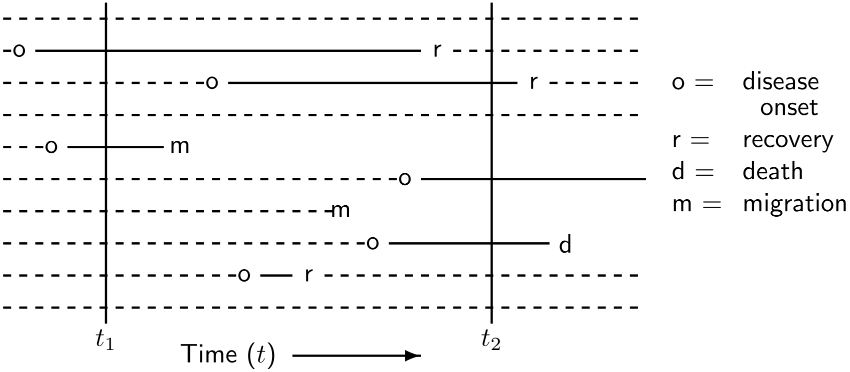
Period prevalence for period from t_1 to t_2 is like P but

- 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)

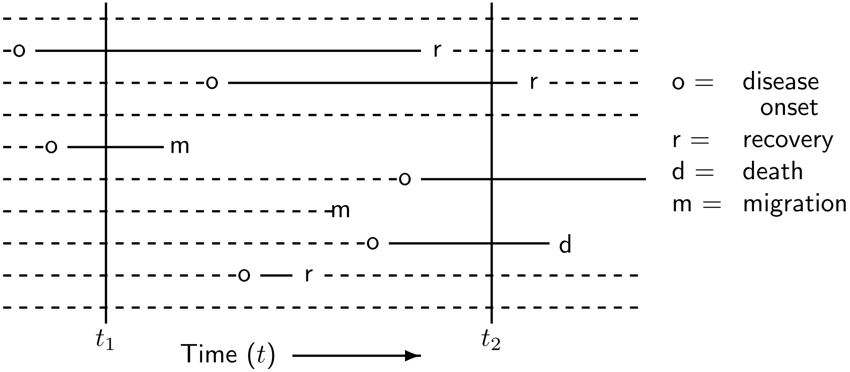


Example 4.1 (IS: p. 59)



Prevalence at time t_1 : $2/10 = 0.2 = 20\%$

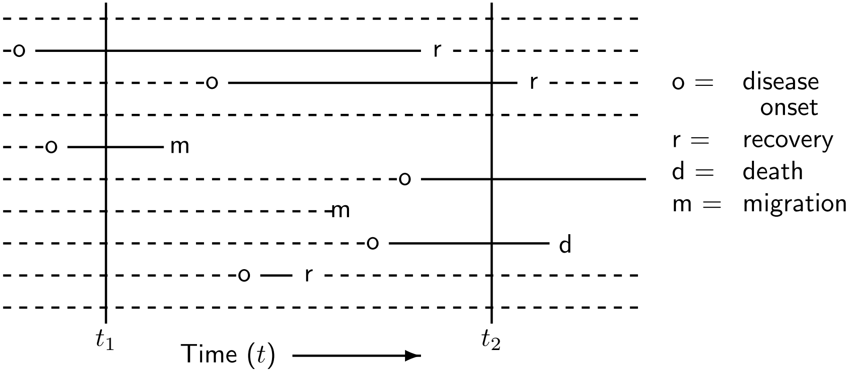
Example 4.1 (IS: p. 59)



Prevalence at time t_1 : $2/10 = 0.2 = 20\%$

Prevalence at time t_2 : $3/8 = 0.38 = 38\%$

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: $5/8 = 0.62 = 62\%$

Prevalence and incidence are related

Prevalence and incidence are related

Point prevalence of S at given time point t depends on

Prevalence and incidence are related

Point prevalence of S at given time point t depends on

- *incidence* of new cases of S before t

in complicated ways.

Prevalence and incidence are related

Point prevalence of S at given time point t depends on

- *incidence* of new cases of S before t
- *duration* of S, depending in turn on the probability of

in complicated ways.

Prevalence and incidence are related

Point prevalence of S at given time point t depends on

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

in complicated ways.

Prevalence and incidence are related

Point prevalence of S at given time point t depends on

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

in complicated ways.

Simple special case: In a **stationary** population

Prevalence and incidence are related

Point prevalence of S at given time point t depends on

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

in complicated ways.

Simple special case: In a **stationary** population prevalence (P), incidence (I), and average duration ($\bar{d}$) of S are related:

Prevalence and incidence are related

Point prevalence of S at given time point t depends on

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

in complicated ways.

Simple special case: In a **stationary** population prevalence (P), incidence (I), and average duration ($\bar{d}$) of S are related:

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

Prevalence and incidence are related

Point prevalence of S at given time point t depends on

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

in complicated ways.

Simple special case: In a **stationary** population prevalence (P), incidence (I), and average duration ($\bar{d}$) of S are related:

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

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

Prevalence of cancer?

Prevalence of cancer?

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

Prevalence of cancer?

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

$\Rightarrow$ Existing or prevalent cancer case problematic to define.

Prevalence of cancer?

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

$\Rightarrow$ 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

Prevalence of cancer?

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

$\Rightarrow$ 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

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

Prevalence of cancer?

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

$\Rightarrow$ 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.

Prevalence of cancer?

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

$\Rightarrow$ 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

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
	123	8
	120	1337
	36	*97*
– " – diagnosed 1-$<$ 5 y ago	*53*	*291*
– " – diagnosed 5-$<$ 10 y ago	*17*	*304*
– " – diagnosed $>$ 10 y ago	*14*	*642*

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
	120	1337
	36	*97*
– " – diagnosed 1-$<$ 5 y ago	*53*	*291*
– " – diagnosed 5-$<$ 10 y ago	*17*	*304*
– " – diagnosed $>$ 10 y ago	*14*	*642*

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
	36	*97*
– " – diagnosed 1-$<$ 5 y ago	*53*	*291*
– " – diagnosed 5-$<$ 10 y ago	*17*	*304*
– " – diagnosed $>$ 10 y ago	*14*	*642*

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*

COMPARISON OF FREQUENCIES

COMPARISON OF FREQUENCIES

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

COMPARISON OF FREQUENCIES

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

COMPARISON OF FREQUENCIES

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)

COMPARISON OF FREQUENCIES

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)

COMPARISON OF FREQUENCIES

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.

$\Rightarrow$ **comparative measures $\approx$ effect measures**

Relative comparative measures

Relative comparative measures

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

Relative comparative measures

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

- incidence rate ratio I_1/I_0,

Relative comparative measures

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

- incidence rate ratio I_1/I_0,
- incidence proportion ratio Q_1/Q_0,

Relative comparative measures

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

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

Relative comparative measures

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

- 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

Relative comparative measures

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

- 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 $[P_1/(1-P_1)]/[P_0/(1-P_0)]$,

Relative comparative measures

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

- 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

Absolute comparative measures

Generic **"excess risk"** or **"risk difference"** (RD) btw exposed and unexposed can refer to

Absolute comparative measures

Generic **"excess risk"** or **"risk difference"** (RD) btw exposed and unexposed can refer to

- incidence rate difference $I_1 - I_0$,

Absolute comparative measures

Generic **"excess risk"** or **"risk difference"** (RD) btw exposed and unexposed can refer to

- incidence rate difference $I_1 - I_0$,
- incidence proportion difference $Q_1 - Q_0$, or

Absolute comparative measures

Generic **“excess risk”** or **“risk difference”** (RD) btw exposed and unexposed can refer to

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

Absolute comparative measures

Generic **"excess risk"** or **"risk difference"** (RD) btw exposed and unexposed can refer to

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

Use of relative and absolute comparisons

Absolute comparative measures

Generic **"excess risk"** or **"risk difference"** (RD) btw exposed and unexposed can refer to

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

Use of relative and absolute comparisons

- Ratios – describe the **biological strength** of the exposure

Absolute comparative measures

Generic **"excess risk"** or **"risk difference"** (RD) btw exposed and unexposed can refer to

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

Use of relative and absolute comparisons

- Ratios – describe the **biological strength** of the exposure
- Differences – inform 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.

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
	5	40
	4.0	2.0
	15	40

[a] Rates per 100 000 pyrs.

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
	4.0	2.0
	15	40

[a] Rates per 100 000 pyrs.

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
	15	40

[a] Rates per 100 000 pyrs.

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.

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)

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

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

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

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

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

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

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

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

$$\text{Inc. odds ratio} = \frac{30/(4\,000-30)}{60/(16\,000-60)} = \frac{0.00756}{0.00376} = 2.0076$$

Measures of potential impact

Measures of potential impact

Combine absolute and relative comparisons.

Measures of potential impact

Combine absolute and relative comparisons.

When incidence is higher for the exposed, we can calculate

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

Measures of potential impact

Combine absolute and relative comparisons.

When incidence is higher for the exposed, we can calculate

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

also called **attributable fraction** (or "attributable risk").

Measures of potential impact

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** (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

Next time: Graphics of impact measures

Apply Bendix's R script on how to draw pictures to illustrate the concepts of excess fraction and population excess fraction with given RRs and prevalences of exposure.

Measures of potential impact (cont'd)

Measures of potential impact (cont'd)

When the exposed have a lower incidence, we can calculate

$$\textbf{Prevented fraction, PF} = \frac{Q_0 - Q_1}{Q_0} = 1 - \text{RR}$$

Measures of potential impact (cont'd)

When the exposed have a lower incidence, we can calculate

$$\textbf{Prevented fraction, PF} = \frac{Q_0 - Q_1}{Q_0} = 1 - \text{RR}$$

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

Measures of potential impact (cont'd)

When the exposed have a lower incidence, we can calculate

$$\textbf{Prevented fraction, PF} = \frac{Q_0 - Q_1}{Q_0} = 1 - \text{RR}$$

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).

Measures of potential impact (cont'd)

When the exposed have a lower incidence, we can calculate

$$\textbf{Prevented fraction, PF} = \frac{Q_0 - Q_1}{Q_0} = 1 - \text{RR}$$

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).

Population EF and **population PF** combine these further with the *prevalence of exposure* in target population.

Effect of smoking on mortality by cause

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

Underlying cause of death	Never smoked regularly Rate[b] (1)	Current cigarette smoker Rate[b] (2)	Rate ratio (2)/(1)	Rate difference[b] $(2)-(1)$	Excess fraction (%) $\frac{(2)-(1)}{(2)} \times 100$
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.

RATES BY VARIOUS TIME AXES

RATES BY VARIOUS TIME AXES

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

RATES BY VARIOUS TIME AXES

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

- age (starting point = birth),

RATES BY VARIOUS TIME AXES

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

- age (starting point = birth),
- exposure time (first exposure),

RATES BY VARIOUS TIME AXES

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

- age (starting point = birth),
- exposure time (first exposure),
- follow-up time (entry to study),

RATES BY VARIOUS TIME AXES

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

- age (starting point = birth),
- exposure time (first exposure),
- follow-up time (entry to study),
- duration of disease (diagnosis).

RATES BY VARIOUS TIME AXES

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

- age (starting point = 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.

RATES BY VARIOUS TIME AXES

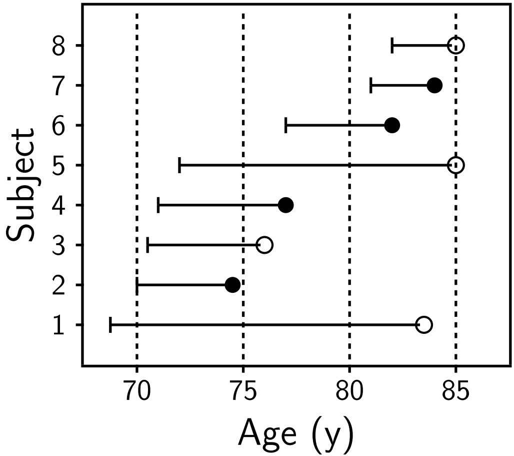
Incidence can be studied on various time scales, *e.g.*

- age (starting point = birth),
- exposure time (first exposure),
- follow-up time (entry to study),
- duration of disease (diagnosis).

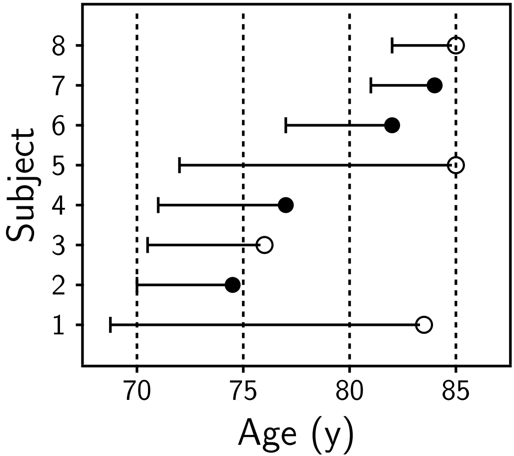
Age is usully the strongest time-dependent determinant of health outcomes.

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

Follow-up of a small geriatric cohort

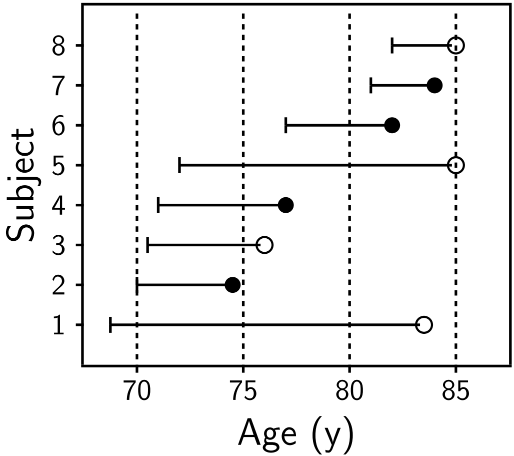


Follow-up of a small geriatric cohort



Overall rate: 4 cases/53.5 person-years = 7.5 per 100 y

Follow-up of a small 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.

Splitting follow-up into agebands

Splitting follow-up into agebands

- To describe, how incidence varies by age, individual person-years from age of entry to age of exit must first be split or divided into narrower agebands.

Splitting follow-up into agebands

- To describe, how incidence varies by age, individual person-years from age of entry to age of exit must first be split or divided into narrower agebands.
- Usually these are based on common 5-year age grouping.

Splitting follow-up into agebands

- To describe, how incidence varies by age, individual person-years from age of entry to age of exit must first be split or divided into narrower agebands.
- Usually these are based on common 5-year age grouping.
- Numbers of cases are equally divided into same agebands.

Splitting follow-up into agebands

- To describe, how incidence varies by age, individual person-years from age of entry to age of exit must first be split or divided into narrower agebands.
- Usually these are based on common 5-year age grouping.
- Numbers of cases are equally divided into same agebands.
- **Age-specific incidence rate** for age group k is

Splitting follow-up into agebands

- To describe, how incidence varies by age, individual person-years from age of entry to age of exit must first be split or divided into narrower agebands.
- Usually these are based on common 5-year age grouping.
- Numbers of cases are equally divided into same agebands.

- **Age-specific incidence rate** for age group k is

$$I_k = \frac{\text{number of cases observed in ageband}}{\text{person-years contained in ageband}}$$

- Underlying assumption: **piecewise constant rates model**

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

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

	Ageband			
Subject	70-74	75-79	80-84	Total
	5.0	5.0	3.5	13.5
	4.5	-	-	4.5
	4.5	1.0	-	5.5
	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
	21.0	16.0	16.5	53.5
	1	1	2	4
	4.8	6.2	12.1	7.5
	Age-specific rates			overall

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
	4.5	-	-	4.5
	4.5	1.0	-	5.5
	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
	21.0	16.0	16.5	53.5
	1	1	2	4
	4.8	6.2	12.1	7.5
	Age-specific rates			overall

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
	4.5	1.0	-	5.5
	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
	21.0	16.0	16.5	53.5
	1	1	2	4
	4.8	6.2	12.1	7.5
	Age-specific rates			overall

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
	4.0	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
	21.0	16.0	16.5	53.5
	1	1	2	4
	4.8	6.2	12.1	7.5
	Age-specific rates			overall

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
	21.0	16.0	16.5	53.5
	1	1	2	4
	4.8	6.2	12.1	7.5
	Age-specific rates			overall

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
	1	1	2	4
	4.8	6.2	12.1	7.5
	Age-specific rates			overall

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

Subject	Ageband 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
	4.8	6.2	12.1	7.5
	Age-specific rates			overall

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

Ex. Lung cancer incidence in Finland by age and period (compare IS, Table 4.1)

Ex. Lung cancer incidence in Finland by age and period (compare IS, Table 4.1)

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

Ex. Lung cancer incidence in Finland by age and period (compare IS, Table 4.1)

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.

Ex. Lung cancer incidence in Finland by age and period (compare IS, Table 4.1)

Calendar period	Age group (y) 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.

Lung cancer rates by age and period

Lung cancer rates by age and period

- Age-incidence curves: overall level and peak age variable across periods.

Lung cancer rates by age and period

- Age-incidence curves: overall level and peak age variable across periods.
- Time trends inconsistent across age groups.

Incidence by age, period & cohort

Incidence by age, period & cohort

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

Incidence by age, period & 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.

Incidence by age, period & 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 – may reflect differences in risk factor levels.

Incidence by age, period & 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 – may reflect differences in risk factor levels.
- Often more informative about "true" age-incidence pattern than age-specific incidences of single calendar period.

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	119	209	276	340	295	279	
1958-62	22	65	135	243	360	405	429	368	
1963-67	24	61	143	258	395	487	509	479	A
1968-72	21	61	134	278	424	529	614	563	
1973-77	16	50	134	251	413	541	629	580	
1978-82	13	36	115	234	369	514	621	653	B
1983-87	11	31	74	186	347	450	566	635	
1988-92	9	25	57	128	262	411	506	507	
1993-97	7	22	48	106	188	329	467	533	C
1998-02	5	14	46	77	150	239	358	445	
			E: 1947/48			D: 1932/33			

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

Diagonals reflect age-incidence pattern in birth cohorts.

Age-specific rates by birth cohort

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

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

Diagonals reflect age-incidence pattern in birth cohorts.

Age-incidence curves in 5 birth cohorts

Variable overall levels but fairly consistent form and similar peak age across different 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)

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)

$\rightarrow$ often compared to incidence in a *reference* or “general” population

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)

$\rightarrow$ often compared to incidence in a *reference* or "general" population

For examples, see Laufey's lecture on cohort studies (*e.g.* atomic bomb survivors, rubber workers, and those exposed to dyestaff)

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)

$\rightarrow$ often compared to incidence in a *reference* or "general" population

For examples, see Laufey's lecture on cohort studies (*e.g.* atomic bomb survivors, rubber workers, and those exposed to dyestaff)

Adjustment for age and calendar time needed, *e.g.* by comparing *observed* to *expected* cases with SIR (see p. 76-79).

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)

$\rightarrow$ often compared to incidence in a *reference* or "general" population

For examples, see Laufey's lecture on cohort studies (*e.g.* atomic bomb survivors, rubber workers, and those exposed to dyestaff)

Adjustment for age and calendar time needed, *e.g.* by comparing *observed* to *expected* cases with SIR (see p. 76-79).

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

Example (C&H, Tables 6.2 & 6.3, p. 54)

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

Example (C&H, Tables 6.2 & 6.3, p. 54)

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

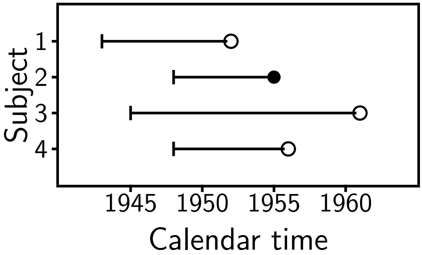
Example: (C&H, Figures 6.1 & 6.2, p. 55)

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)



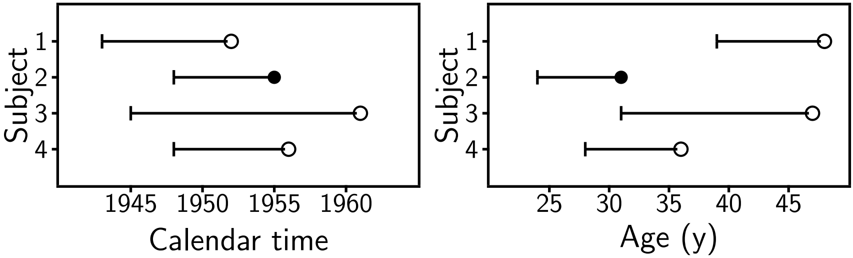
Example: (C&H, Figures 6.1 & 6.2, p. 55)

Follow-up of cohort members by calendar time and age

| entry

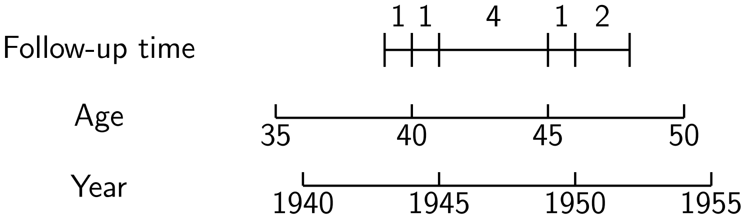
• exit because of disease onset (outcome of interest)

○ exit due to other reason (censoring)



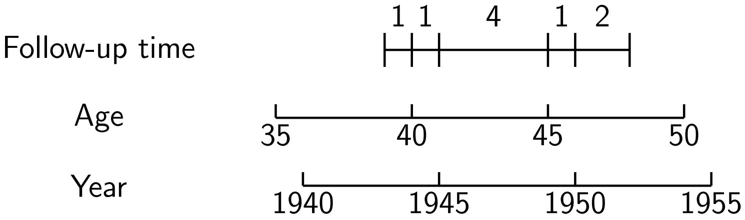
Person-years by age and period
(C&H, Figure 6.4)

Subject 1: Follow-up jointly split by age and calendar time:



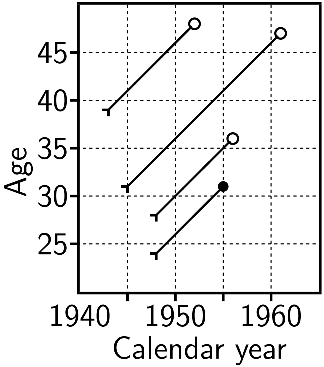
Person-years by age and period
(C&H, Figure 6.4)

Subject 1: Follow-up jointly split by age and calendar time:

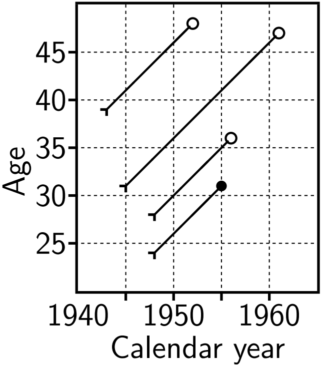


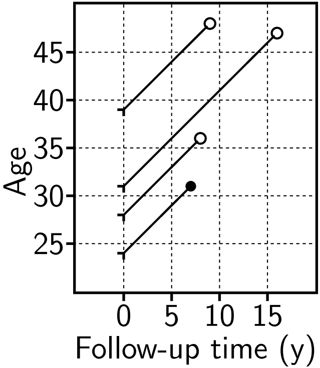
This subject contributes person-time into 5 different cells defined by ageband & calendar period

Follow-up in Lexis-diagrams
(C&H, pp. 58-59)

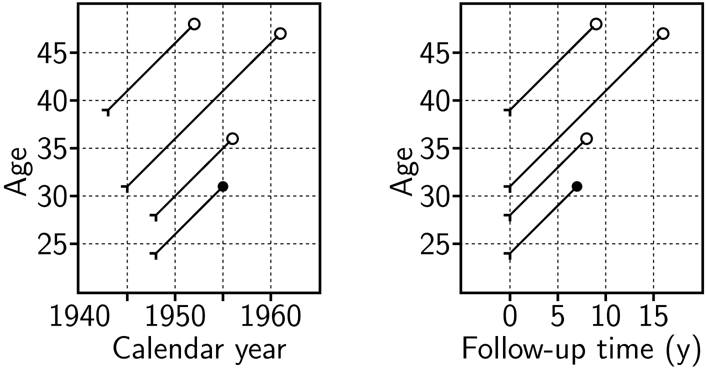


Follow-up in Lexis-diagrams
(C&H, pp. 58-59)





Follow-up in Lexis-diagrams
(C&H, pp. 58-59)



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

See also Laufey's lecture on cohort studies, slide 4.

STANDARDIZATION OF RATES

STANDARDIZATION OF RATES

- Incidence of most cancers (and many other diseases) increases strongly by age in all populations.

STANDARDIZATION OF RATES

- Incidence of most cancers (and many other diseases) increases strongly by age in all populations.

  $\Rightarrow$ Most of the caseload comes from older age groups.

STANDARDIZATION OF RATES

- Incidence of most cancers (and many other diseases) increases strongly by age in all populations.

  $\Rightarrow$ Most of the caseload comes from older age groups.

- **Crude incidence rate** is a rate in which:

STANDARDIZATION OF RATES

- Incidence of most cancers (and many other diseases) increases strongly by age in all populations.

  $\Rightarrow$ 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,

STANDARDIZATION OF RATES

- Incidence of most cancers (and many other diseases) increases strongly by age in all populations.

  $\Rightarrow$ 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.

STANDARDIZATION OF RATES

- Incidence of most cancers (and many other diseases) increases strongly by age in all populations.

  $\Rightarrow$ 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**.

STANDARDIZATION OF RATES

- Incidence of most cancers (and many other diseases) increases strongly by age in all populations.

  $\Rightarrow$ 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.

STANDARDIZATION OF RATES

- Incidence of most cancers (and many other diseases) increases strongly by age in all populations.

  $\Rightarrow$ 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.
- **Adjustment** or **standardization** for age needed!

Ex. Male stomach cancer in Cali and Birmingham (IS, Table 4.2, p. 71)

	Cali			Birmingham			
Age (y)	Male cases 1982 -86	Male Population 1984 ($\times 10^3$)	Incid. Rate ($/10^5$y) 1982 -86	Male cases 1983 -86	Male Population 1985 ($\times 10^3$))	Incid. Rate ($/10^5$y) 1983 -86	Rate ratio
	39	524.2	**1.5**	79	1 683.6	**1.2**	*1.25*
	266	76.3	**69.7**	1037	581.5	**44.6**	*1.56*
	315	22.4	**281.3**	2352	291.1	**202.0**	*1.39*
	620	622.9	**19.9**	3468	2 556.2	**33.9**	*0.59*

Ex. Male stomach cancer in Cali and Birmingham (IS, Table 4.2, p. 71)

	Cali			Birmingham			
Age (y)	Male cases 1982 -86	Male Population 1984 ($\times 10^3$)	Incid. Rate ($/10^5$y) 1982 -86	Male cases 1983 -86	Male Population 1985 ($\times 10^3$))	Incid. Rate ($/10^5$y) 1983 -86	Rate ratio
0–44	39	524.2	**1.5**	79	1 683.6	**1.2**	*1.25*
	266	76.3	**69.7**	1037	581.5	**44.6**	*1.56*
	315	22.4	**281.3**	2352	291.1	**202.0**	*1.39*
	620	622.9	**19.9**	3468	2 556.2	**33.9**	*0.59*

Ex. Male stomach cancer in Cali and Birmingham (IS, Table 4.2, p. 71)

	Cali			Birmingham			
Age (y)	Male cases 1982 -86	Male Population 1984 ($\times 10^3$)	Incid. Rate ($/10^5$y) 1982 -86	Male cases 1983 -86	Male Population 1985 ($\times 10^3$))	Incid. 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*
	315	22.4	**281.3**	2352	291.1	**202.0**	*1.39*
	620	622.9	**19.9**	3468	2 556.2	**33.9**	*0.59*

Ex. Male stomach cancer in Cali and Birmingham (IS, Table 4.2, p. 71)

	Cali			Birmingham			
Age (y)	Male cases 1982 -86	Male Population 1984 ($\times 10^3$)	Incid. Rate ($/10^5$y) 1982 -86	Male cases 1983 -86	Male Population 1985 ($\times 10^3$))	Incid. 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*
	620	622.9	**19.9**	3468	2 556.2	**33.9**	*0.59*

Ex. Male stomach cancer in Cali and Birmingham (IS, Table 4.2, p. 71)

	Cali			Birmingham			
Age (y)	Male cases 1982 -86	Male Population 1984 ($\times 10^3$)	Incid. Rate ($/10^5$y) 1982 -86	Male cases 1983 -86	Male Population 1985 ($\times 10^3$))	Incid. 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*

Ex. Male stomach cancer in Cali and Birmingham (IS, Table 4.2, p. 71)

	Cali			Birmingham			
Age (y)	Male cases 1982 -86	Male Population 1984 ($\times 10^3$)	Incid. Rate ($/10^5$y) 1982 -86	Male cases 1983 -86	Male Population 1985 ($\times 10^3$))	Incid. 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.

Ex. Male stomach cancer in Cali and Birmingham (IS, Table 4.2, p. 71)

	Cali			Birmingham			
Age (y)	Male cases 1982 -86	Male Population 1984 ($\times 10^3$)	Incid. Rate ($/10^5$y) 1982 -86	Male cases 1983 -86	Male Population 1985 ($\times 10^3$))	Incid. 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

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.

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.

$\Rightarrow$ Crude rates are **confounded** by age.

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.

$\Rightarrow$ Crude rates are **confounded** by age.

$\Rightarrow$ Any summary rate must be **adjusted for age**.

Age-adjustment by standardisation

Age-adjustment by standardisation

Age-standardised incidence rate (ASR):

Age-adjustment by standardisation

Age-standardised incidence rate (ASR):

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

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 $k = 1, \ldots, K$.

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**.

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)

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**

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.

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**

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

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 cause C or all causes).

Cumulative rate and “cumulative risk”

Cumulative rate and “cumulative risk”

- Choice of standard somewhat arbitrary.

Cumulative rate and “cumulative risk”

- Choice of standard somewhat arbitrary.
- Alternative and maybe more natural method for age-adjustment is provided by **cumulative rate**:

$$\mathsf{CR} = \sum_{k=1}^{K} \mathsf{width}_k \times \mathsf{rate}_k,$$

Cumulative rate and “cumulative risk”

- Choice of standard somewhat arbitrary.
- Alternative and maybe more natural method for age-adjustment is provided by **cumulative rate**:

$$\mathsf{CR} = \sum_{k=1}^{K} \mathsf{width}_k \times \mathsf{rate}_k,$$

- Weights are widths of the agebands to be included.

Cumulative rate and "cumulative risk"

- Choice of standard somewhat arbitrary.
- Alternative and maybe more natural method for age-adjustment is provided by **cumulative rate**:

$$\text{CR} = \sum_{k=1}^{K} \text{width}_k \times \text{rate}_k,$$

- Weights are widths of the agebands to be included.
- Usually computed up to 65 or 75 y with 5-y bands.

Cumulative rate and “cumulative risk”

- Choice of standard somewhat arbitrary.
- Alternative and maybe more natural method for age-adjustment is provided by **cumulative rate**:

$$\text{CR} = \sum_{k=1}^{K} \text{width}_k \times \text{rate}_k,$$

- Weights are widths of the agebands to be included.
- Usually computed up to 65 or 75 y with 5-y bands.
- Often interpreted as approximating the average **“cumulative risk”** (incidence proportion) to get the disease by 65 or 75 years, given survival until then.

Cumulative rate and "cumulative risk"

- Choice of standard somewhat arbitrary.
- Alternative and maybe more natural method for age-adjustment is provided by **cumulative rate**:

$$\text{CR} = \sum_{k=1}^{K} \text{width}_k \times \text{rate}_k,$$

- Weights are widths of the agebands to be included.
- Usually computed up to 65 or 75 y with 5-y bands.
- Often interpreted as approximating the average **"cumulative risk"** (incidence proportion) to get the disease by 65 or 75 years, given survival until then.
- Based on relation btw risk $F(t)$ and hazard $\lambda(t)$, or

$$\text{Inc. prop'n } = 1 - \exp(-\text{cum. rate}) \;\approx\; \text{cum. rate}$$

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

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

$$\text{Cali:}\quad 45\ y \times \frac{1.5}{10^5 y} + 20\ y \times \frac{69.7}{10^5 y} = 0.0146 = \mathbf{1.46}\ \text{per}\ 100$$

$$45\ y \times \frac{1.2}{10^5 y} + 20\ y \times \frac{44.6}{10^5 y} = 0.0095 = \mathbf{0.95}\ \text{per}\ 100$$

$$1.46/0.95 = \mathbf{1.54}$$

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

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

$$1.46/0.95 = \mathbf{1.54}$$

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

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

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

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

Cumulative "risks" & their ratio up to 65 y:

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" & their ratio up to 65 y:

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

NB. For more appropriate estimates of cumulative risks, correction for total mortality (competing event) needed.

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

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

Cumulative "risks" & their ratio up to 65 y:

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

NB. For more appropriate estimates of cumulative risks, correction for total mortality (competing event) needed.

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" & their ratio up to 65 y:

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

NB. For more appropriate estimates of cumulative risks, correction for total mortality (competing event) needed.

Cumulative measures using 5-y groups

(IS, Fig 4.11, p. 77)

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

Cumulative measures using 5-y groups

(IS, Fig 4.11, p. 77)

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

$$\text{Cum. rate 0-75 y} = 5\text{ y} \times \frac{524.9}{10^5\text{ y}} = 0.0262 = \mathbf{2.6} \text{ per 100}$$
$$= 1 - \exp(-0.0262) = 0.0259 = \mathbf{2.6}\%.$$

Cumulative measures using 5-y groups

(IS, Fig 4.11, p. 77)

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

$$\text{Cum. rate 0-75 y} = 5\text{ y} \times \frac{524.9}{10^5\text{ y}} = 0.0262 = \mathbf{2.6}\text{ per 100}$$

$$\text{Cum. “risk” 0-75 y} = 1 - \exp(-0.0262) = 0.0259 = \mathbf{2.6}\%.$$

Cumulative and life-time risks

Interesting and relevant question

> "*What are my chances of getting cancer* C *in the next 10 years, between ages 50 to 75 years, or during the whole lifetime?*"

Cumulative and life-time risks

Interesting and relevant question

> "*What are my chances of getting cancer* C *in the next 10 years, between ages 50 to 75 years, or during the whole lifetime?*"

Difficult to answer.

- Fully individualized risks are unidentifiable.

Cumulative and life-time risks

Interesting and relevant question

> "*What are my chances of getting cancer C in the next 10 years, between ages 50 to 75 years, or during the whole lifetime?*"

Difficult to answer.

- Fully individualized risks are unidentifiable.
- Age-specific and standardized rates are not very informative as such.

Cumulative and life-time risks

Interesting and relevant question

> "*What are my chances of getting cancer C in the next 10 years, between ages 50 to 75 years, or during the whole lifetime?*"

Difficult to answer.

- Fully individualized risks are unidentifiable.
- Age-specific and standardized rates are not very informative as such.
- Average cumulative risks are often estimated from cumulative rates.

Cumulative and life-time risks

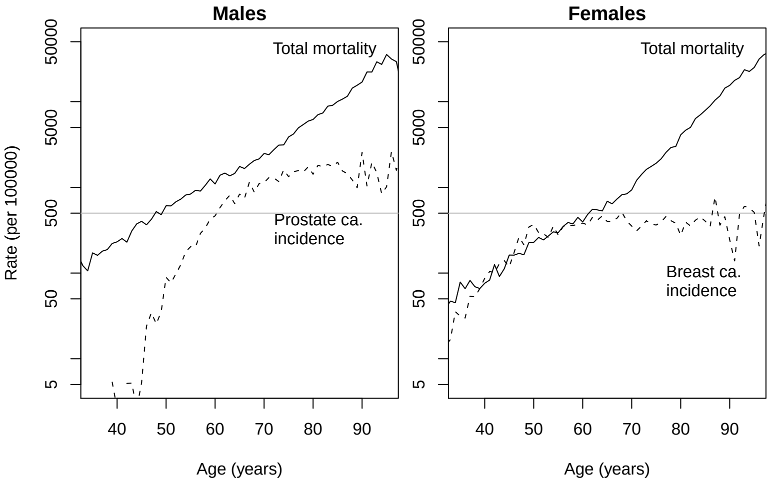
Interesting and relevant question

> "*What are my chances of getting cancer C in the next 10 years, between ages 50 to 75 years, or during the whole lifetime?*"

Difficult to answer.

- Fully individualized risks are unidentifiable.
- Age-specific and standardized rates are not very informative as such.
- Average cumulative risks are often estimated from cumulative rates.
- Yet, these estimates fictitiously presume that a person would not die from any cause before cancer hits him/her, but could even survive forever!

Total mortality and incidence of two common cancers by age, Finland 2005



Estimation of cumulative risks

Estimation of cumulative risks

- The probability of contracting cancer during realistic lifespan or in any age range depends not only on age-specific hazard rates of cancer itself but also of probabilities of overall survival up to relevant ages,

Estimation of cumulative risks

- The probability of contracting cancer during realistic lifespan or in any age range depends not only on age-specific hazard rates of cancer itself but also of probabilities of overall survival up to relevant ages,
- Hence, the dependence of total mortality by age in the population at risk must be incorporated in estimation of cumulative risks of cancer.

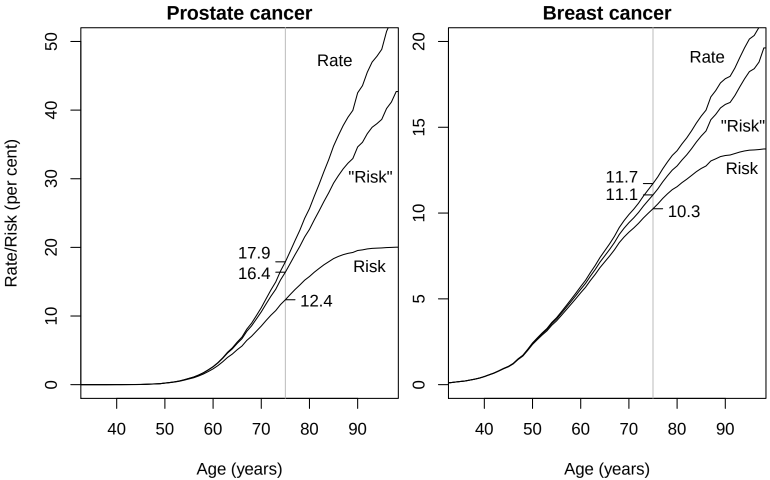
Estimation of cumulative risks

- The probability of contracting cancer during realistic lifespan or in any age range depends not only on age-specific hazard rates of cancer itself but also of probabilities of overall survival up to relevant ages,
- Hence, the dependence of total mortality by age in the population at risk must be incorporated in estimation of cumulative risks of cancer.
- When this is properly done, the corrected estimates of cumulative risk will always be lower than the uncorrected "risks".

Estimation of cumulative risks

- The probability of contracting cancer during realistic lifespan or in any age range depends not only on age-specific hazard rates of cancer itself but also of probabilities of overall survival up to relevant ages,
- Hence, the dependence of total mortality by age in the population at risk must be incorporated in estimation of cumulative risks of cancer.
- When this is properly done, the corrected estimates of cumulative risk will always be lower than the uncorrected "risks".
- The magnitude of bias in the latter grows by age, but is reduced with increased life expectancy.

Cumulative measures, Finland 2005



Greater differences in males reflect shorter life expectancy and relatively high rates of prostate ca. in old ages.

Special cohorts of exposed subjects

- Occupational cohorts, exposed to potentially hazardous agents (*e.g.* rubber workers, see Laufey's lecture on cohort studies, slides 19-20)
- Cohorts of patients on chronic medication, which may have harmful long-term side-effects
- No internal comparison group of unexposed subjects.

Question: Do incidence or mortality rates in the *exposed* target cohort differ from those of a roughly comparable *reference* population?

Reference rates obtained from:

- population statistics (mortality rates)
- disease & hospital discharge registers (incidence)

Observed and expected cases – SIR

- Compare rates in a study cohort with a standard set of age–specific rates from the reference population.
- Reference rates normally based on large numbers of cases, so they are assumed to be "known" without error.
- Calculate **expected** number of cases, E, if the standard age-specific rates had applied in our study cohort.
- Compare this with the **observed** number of cases, D, by the **standardized incidence ratio** SIR
(or st'zed mortality ratio SMR with death as outcome)

$$\text{SIR} = D/E, \qquad \text{SE}(\log[\text{SIR}]) = 1/\sqrt{D}$$

Example: HT and breast ca.

- A cohort of 974 women treated with hormone (replacement) therapy were followed up.
- $D = 15$ incident cases of breast cancer were observed.
- Person-years (Y) and reference rates (λ_a^*, per 100000 y) by age group (a) were:

Age	Y	λ_a^*	E
40–44	975	113	1.10
45–49	1079	162	1.75
50–54	2161	151	3.26
55–59	2793	183	5.11
60–64	3096	179	5.54
$\sum$			16.77

Ex: HT and breast ca. (cont'd)

- "Expected" cases at ages 40–44:

$$975 \times \frac{113}{100\,000} = 1.10$$

- Total "expected" cases is $E = 16.77$
- SIR $= 15/16.77 = 0.89$.
- Error-factor: $\exp(1.96 \times \sqrt{1/15}) = 1.66$
- 95% confidence interval is:

$$0.89 \overset{\times}{\div} 1.66 = (0.54, 1.48)$$

SIR for Cali with B'ham 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
	=**3 114 610**	**Total expected (E) 427.82**

SIR for Cali with B'ham 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**

SIR for Cali with B'ham 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 $O = 620$.

SIR for Cali with B'ham 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 rates compared

(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*
		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*
		14.6	9.5	*1.54*
		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

Crude and adjusted rates compared

(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*
		14.6	9.5	*1.54*
		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

Crude and adjusted rates compared

(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*
		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

Crude and adjusted rates compared

(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

Crude and adjusted rates compared

(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 ANALYSIS OF CANCER

SURVIVAL ANALYSIS OF CANCER

Prognosis of cancer:

- what are the patients' chances to **survive** 1 year, 5 years,*etc.* since diagnosis?

SURVIVAL ANALYSIS OF CANCER

Prognosis of cancer:

- what are the patients' chances to **survive** 1 year, 5 years,*etc.* since diagnosis?

Survival analysis: In principle like incidence analysis but

SURVIVAL ANALYSIS OF CANCER

Prognosis of cancer:

- what are the patients' chances to **survive** 1 year, 5 years,*etc.* since diagnosis?

Survival analysis: In principle like incidence analysis but

- population at risk = patients with cancer,

SURVIVAL ANALYSIS OF CANCER

Prognosis of cancer:

- what are the patients' chances to **survive** 1 year, 5 years,*etc.* since 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,

SURVIVAL ANALYSIS OF CANCER

Prognosis of cancer:

- what are the patients' chances to **survive** 1 year, 5 years,*etc.* since 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,

SURVIVAL ANALYSIS OF CANCER

Prognosis of cancer:

- what are the patients' chances to **survive** 1 year, 5 years,*etc.* since 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)	Stage[a]	Date of diagnosis	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
	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

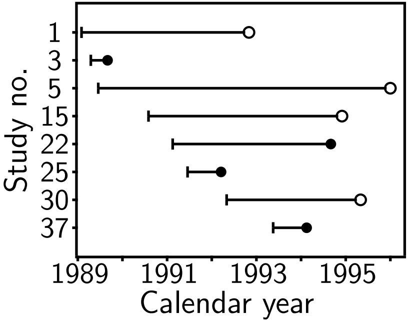
Follow-up of 8 out of 40 breast cancer patients (from IS, table 12.1., p. 264)

No.	Age (y)	Stage[a]	Date of diagnosis	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 ca. pts (cont'd)

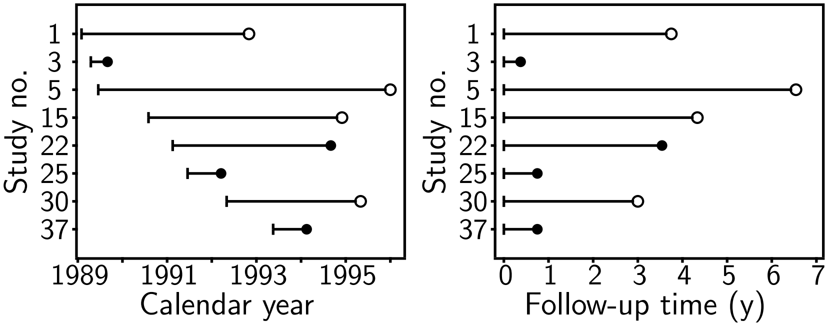
| entry = diagnosis; • exit = death; ∘ exit = censoring



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

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

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



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

Life table or “actuarial” method

Life table or "actuarial" method

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

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

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,

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,

Life table or "actuarial" method

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

(2) Tabulate from original data for each interval

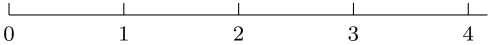
$N_k =$ size of the **risk set**, *i.e.* the no. of subjects still alive and under follow-up at the start of interval,

$D_k =$ no. of **cases**, *i.e.* deaths observed in the interval,

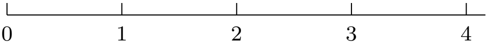
$L_k =$ no. of **losses**, *i.e.* individuals **censored** during the interval before being observed to die.

Life table items in a tree diagram

Life table items in a tree diagram

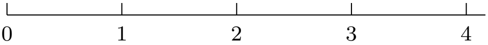


Life table items in a tree diagram



Follow-up time (years), divided into 1-y subintervals

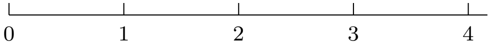
Life table items in a tree diagram



Follow-up time (years), divided into 1-y subintervals

N_k = population at risk at the start of the kth subinterval

Life table items in a tree diagram

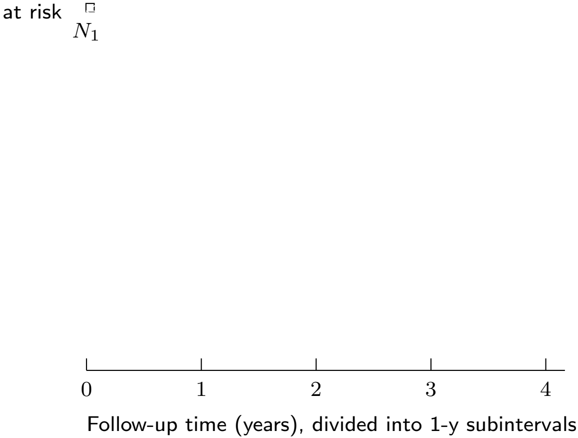


Follow-up time (years), divided into 1-y subintervals

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

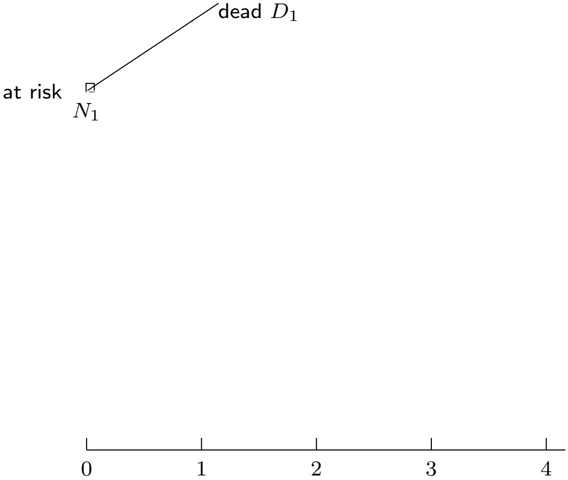
Life table items in a tree diagram



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 in a tree diagram

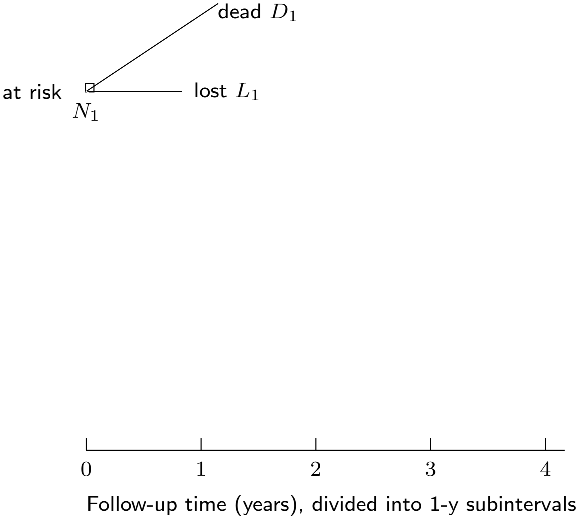


Follow-up time (years), divided into 1-y subintervals

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

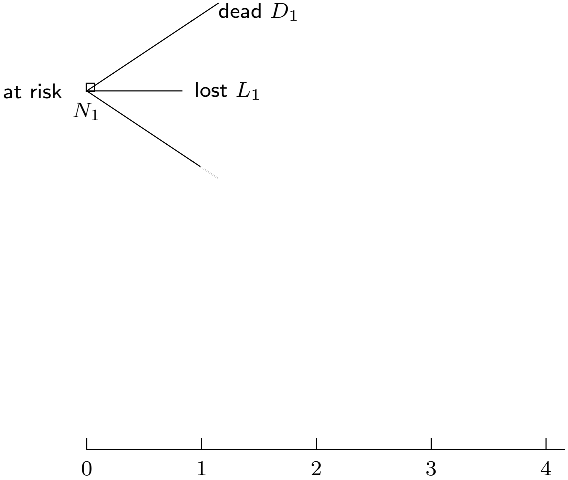
Life table items in a tree diagram



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 in a tree diagram

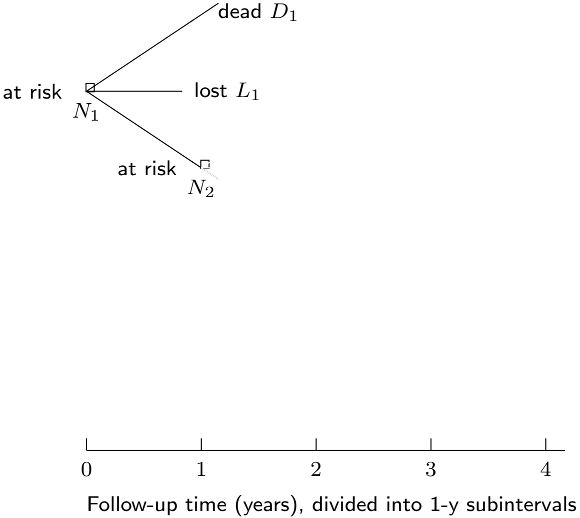


Follow-up time (years), divided into 1-y subintervals

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

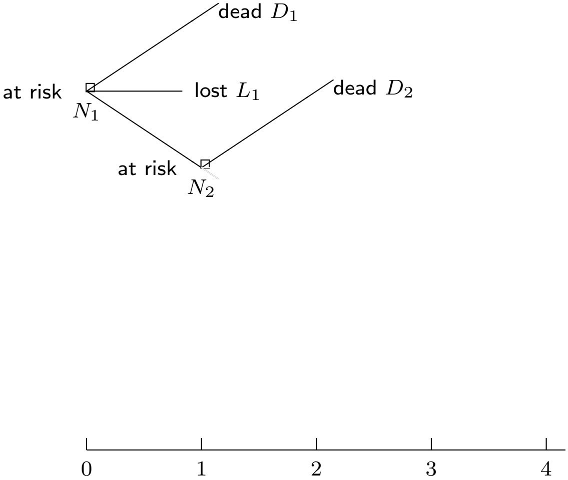
Life table items in a tree diagram



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 in a tree diagram

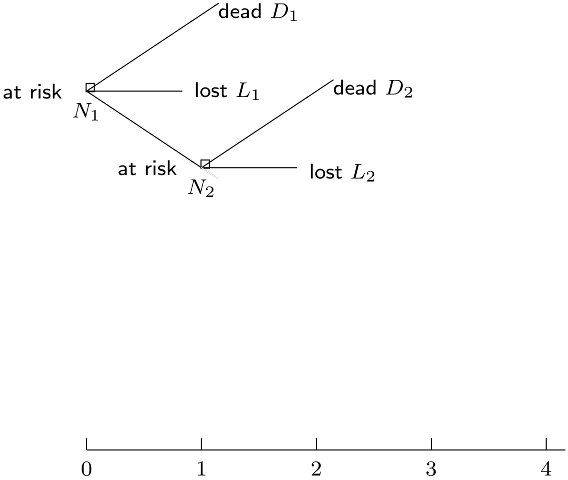


Follow-up time (years), divided into 1-y subintervals

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 in a tree diagram

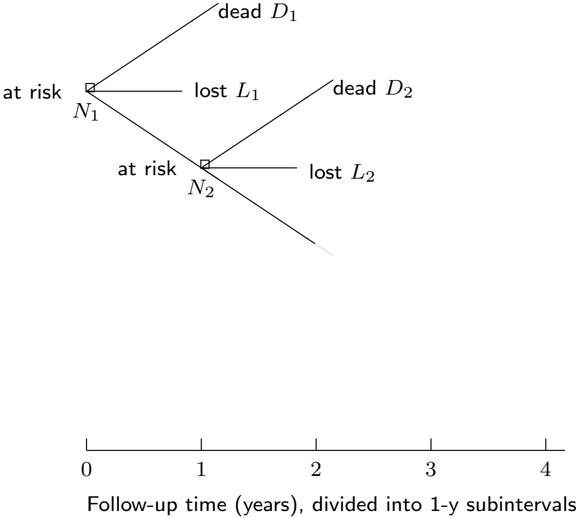


Follow-up time (years), divided into 1-y subintervals

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

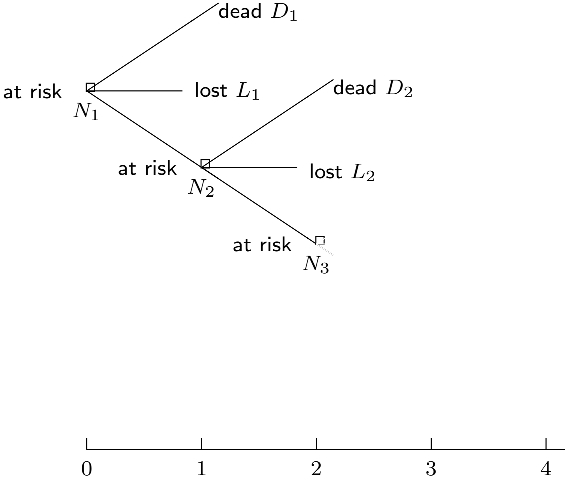
Life table items in a tree diagram



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 in a tree diagram

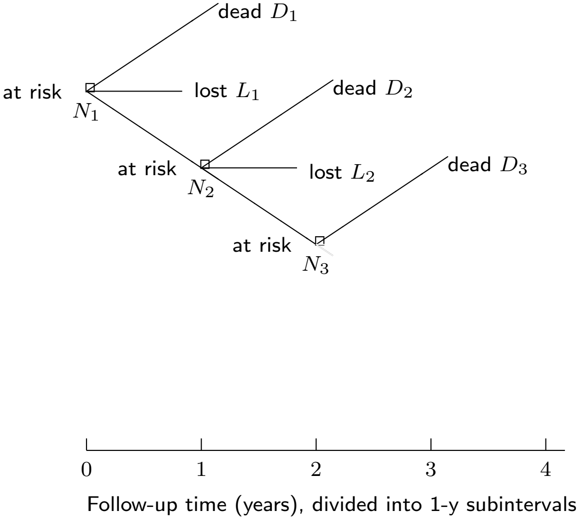


Follow-up time (years), divided into 1-y subintervals

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

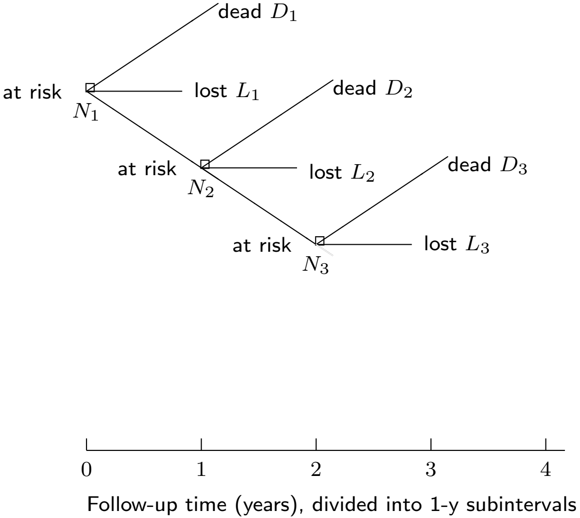
Life table items in a tree diagram



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

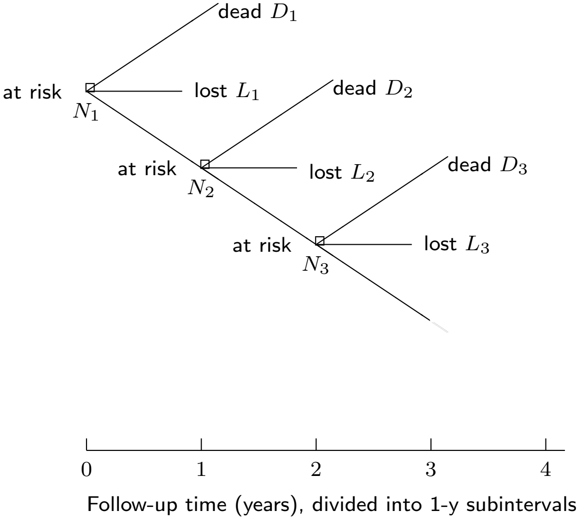
Life table items in a tree diagram



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

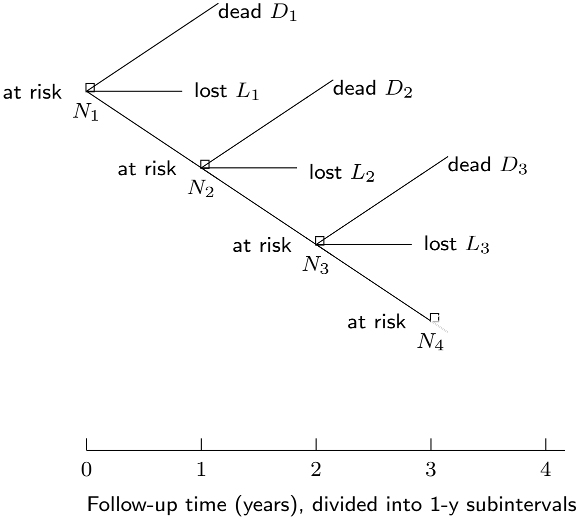
Life table items in a tree diagram



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

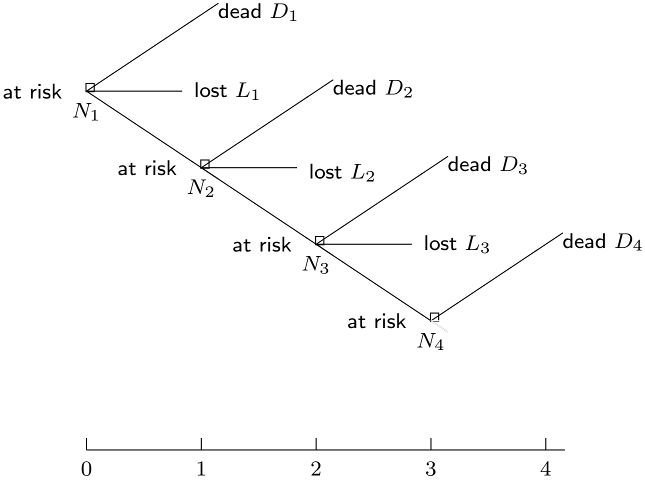
Life table items in a tree diagram



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 in a tree diagram

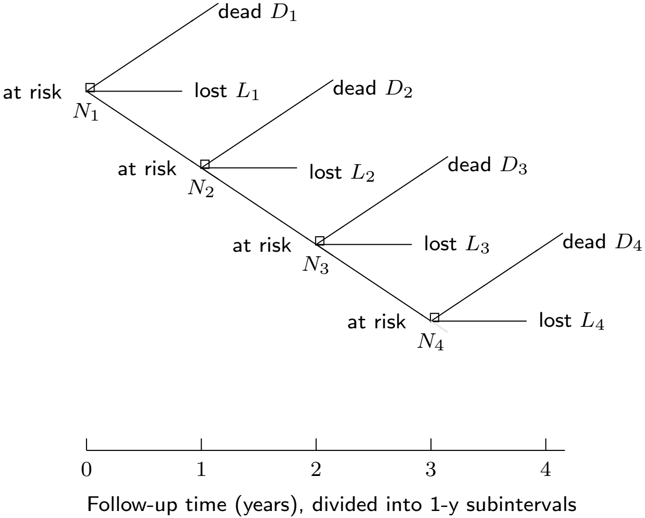


Follow-up time (years), divided into 1-y subintervals

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

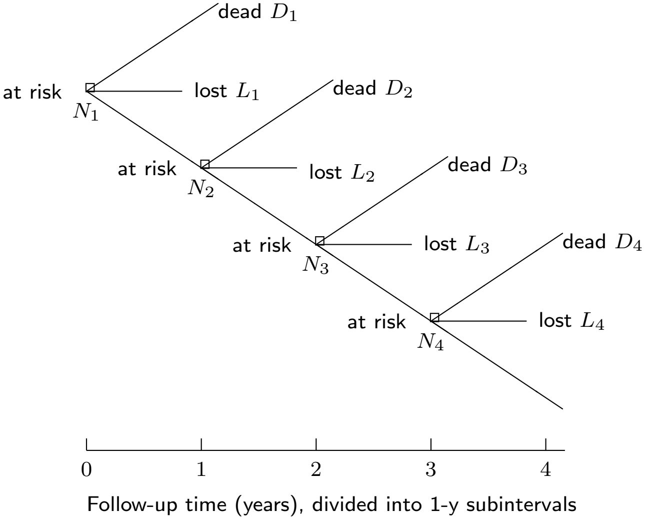
Life table items in a tree diagram



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 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)

Interval (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

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,

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,

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,

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

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)

Interval (k)	Years since diagnosis	No. at start of interval (N_k)	No. of deaths (D_k)	No. of losses (L_k)	Effective denominator (N'_k)	Cond'l prop'n of deaths during int'l (q_k)	Survival prop'n over int'l (p_k)	Cumul. survival; est'd survival prob'ty (S_k)
1	0– < 1	40	7	0	40.0	0.175	0.825	0.825
2	1– < 2	33	3	6	30.0	0.100	0.900	0.743
3	2– < 3	24	4	3	22.5	0.178	0.822	0.610
4	3– < 4	17	4	4	15.0	0.267	0.733	0.447
5	4– < 5	9	2	3	7.5	0.267	0.733	0.328
6	5– < 6	4	1	2	3.0	0.333	0.667	0.219
7	6– < 7	1	0	1	0.5	0.0	1.0	0.219

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

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

Interval (k)	Years since diagnosis	No. at start of interval (N_k)	No. of deaths (D_k)	No. of losses (L_k)	Effective denominator (N'_k)	Cond'l prop'n of deaths during int'l (q_k)	Survival prop'n over int'l (p_k)	Cumul. survival; est'd survival prob'ty (S_k)
1	0– <1	40	7	0	40.0	0.175	0.825	0.825
2	1– <2	33	3	6	30.0	0.100	0.900	0.743
3	2– <3	24	4	3	22.5	0.178	0.822	0.610
4	3– <4	17	4	4	15.0	0.267	0.733	0.447
5	4– <5	9	2	3	7.5	0.267	0.733	0.328
6	5– <6	4	1	2	3.0	0.333	0.667	0.219
7	6– <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 methods

Comparison to previous methods

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

Comparison to previous methods

- Complement of survival proportion $Q_k = 1 - S_k =$ incidence proportion of deaths. Estimates the cumulative risk of death from start of follow-up till end of kth interval.
- "Actuarial" indidence rate in the kth interval:

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

Comparison to previous methods

- Complement of survival proportion $Q_k = 1 - S_k =$ incidence proportion of deaths. Estimates the 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

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

  The dead and censored thus contribute half of the interval length.

Survival curve and other measures

Survival curve and other measures

Line diagram of survival proportions through interval endpoints

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.*:

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

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.

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 and relative survival

(A) **Cause-specific** survival analysis:

Cause-specific and relative survival

(A) **Cause-specific** survival analysis:

- outcome event: death from the disease C itself,

Cause-specific and relative survival

(A) **Cause-specific** survival analysis:

- outcome event: death from the disease C itself,
- deaths from other causes $\rightarrow$ counted as losses,

Cause-specific and relative survival

(A) **Cause-specific** survival analysis:

- outcome event: death from the disease C itself,
- deaths from other causes $\rightarrow$ counted as losses,

– problems with cause of death & competing causes.

Cause-specific and relative survival

(A) **Cause-specific** survival analysis:

- outcome event: death from the disease C itself,
- deaths from other causes $\rightarrow$ counted as losses,

– problems with cause of death & competing causes.

(B) **Relative survival** analysis:

Cause-specific and relative survival

(A) **Cause-specific** survival analysis:

- outcome event: death from the disease C itself,
- deaths from other causes $\rightarrow$ counted as losses,

– problems with cause of death & competing causes.

(B) **Relative survival** analysis: Compute

$$R_k = S_k^{\text{obs}} / S_k^{\text{exp}},$$

the **relative survival proportion** $=$ ratio of

- **observed** survival proportion S_k^{obs} in cancer patients,

Cause-specific and relative survival

(A) **Cause-specific** survival analysis:

- outcome event: death from the disease C itself,
- deaths from other causes $\rightarrow$ counted as losses,

– problems with cause of death & competing causes.

(B) **Relative survival** analysis: Compute

$$R_k = S_k^{\text{obs}} / S_k^{\text{exp}},$$

the **relative survival proportion** $=$ ratio of

- **observed** survival proportion S_k^{obs} in cancer patients,
- **expected** survival proportion S_k^{exp}

Cause-specific and relative survival

(A) **Cause-specific** survival analysis:

- outcome event: death from the disease C itself,
- deaths from other causes $\rightarrow$ counted as losses,

– problems with cause of death & competing causes.

(B) **Relative survival** analysis: Compute

$$R_k = S_k^{\text{obs}} / S_k^{\text{exp}},$$

the **relative survival proportion** = ratio of

- **observed** survival proportion S_k^{obs} in cancer patients,
- **expected** survival proportion S_k^{exp} based on age-specific mortalities in a reference population (*cf.* SIR!)

Cause-specific and relative survival

(A) **Cause-specific** survival analysis:

- outcome event: death from the disease C itself,
- deaths from other causes $\rightarrow$ counted as losses,

– problems with cause of death & competing causes.

(B) **Relative survival** analysis: Compute

$$R_k = S_k^{\text{obs}} / S_k^{\text{exp}},$$

the **relative survival proportion** $=$ ratio of

- **observed** survival proportion S_k^{obs} in cancer patients,
- **expected** survival proportion S_k^{exp} based on age-specific mortalities in a reference population (*cf.* SIR!)

$+$ no information on causes of death needed.

Ex. 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":

NB. Meaning of "'cause-specific survival"?

CONCLUSION

CONCLUSION

Measuring and comparing disease frequencies

CONCLUSION

Measuring and comparing disease frequencies

- not a trivial task but

CONCLUSION

Measuring and comparing disease frequencies

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

CONCLUSION

Measuring and comparing disease frequencies

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

Major challenges:

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,

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,

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,

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.