ANALYSIS OF AGE, PERIOD AND COHORT EFFECTS IN LONG-TERM FOLLOW-UP STUDIES

Simo Näyhä

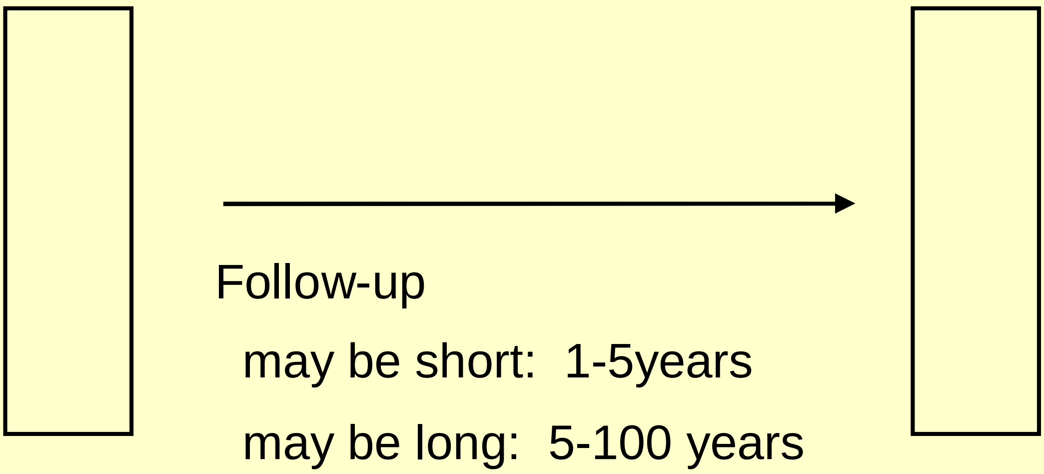
May 20, 2014

Starting point:

People at different ages are being followed up for a lengthy period

A TYPICAL DESIGN

A group of people aged 0-100 yr

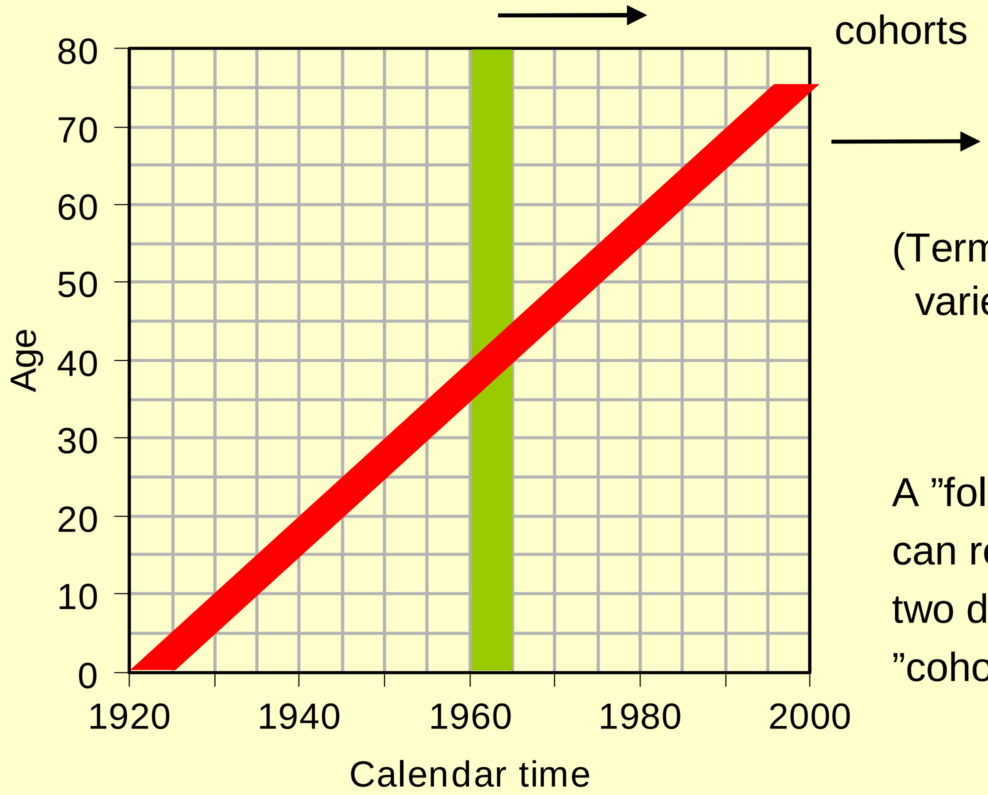


During the follow-up, not only calendar time goes on

THE LEXIS DIAGRAM

Cross-sectional
"cohorts"

True (synthetic)
cohorts

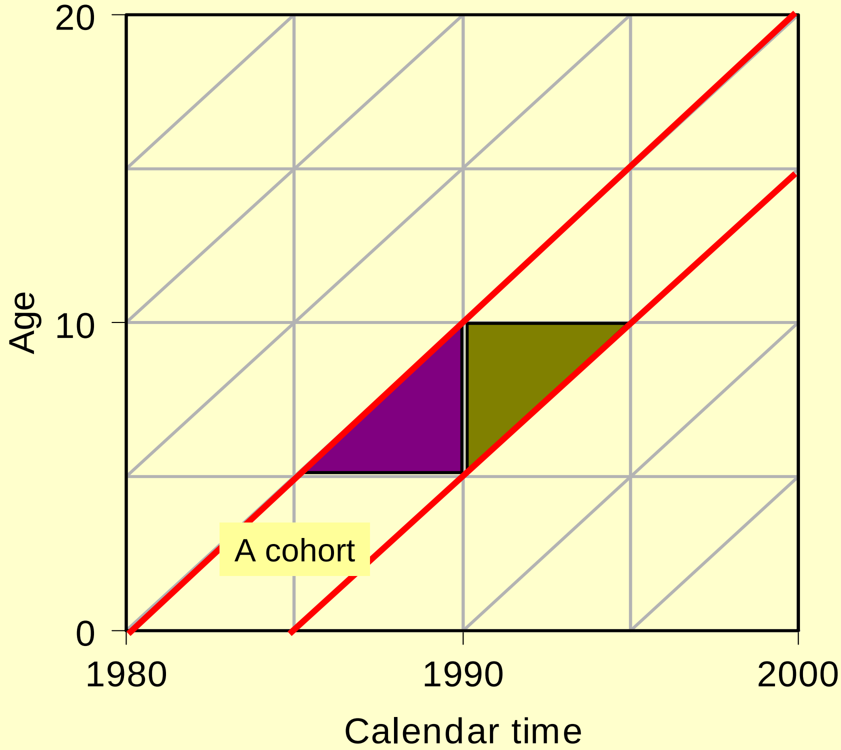


(Terminology
varies)

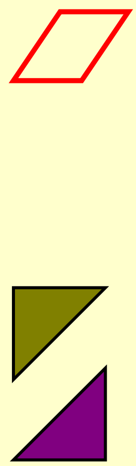
A "follow-up"
can refer to
two different
"cohorts"

THE LEXIS DIAGRAM

A close-up: 5 x 5 year cells



Each **cohort** is composed of two-triangle sets belonging to different **periods**

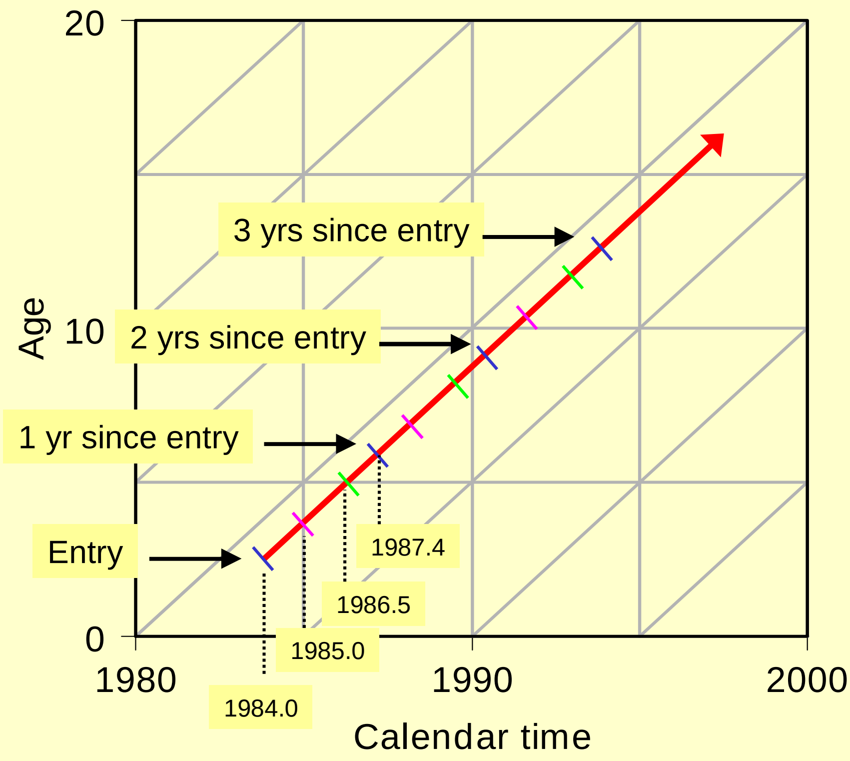


- Upper triangle

- Lower triangle

THE LEXIS DIAGRAM

Each individual's course through the age and period axes can be divided to time slices



THE PROBLEM

What really happens as "time" goes on ?

Three things happen:	"Effects"
1. People age	Age (A)
2. Calendar time goes on	Time Period (P)
3. Generations change	Generation or Cohort (C)

Obviously, these are difficulty to define as separate effects, since:

$$A = C + P \qquad P = C + A \qquad C = P - A$$

However, **A**, **P** and **C** are different effects

For example, THE EFFECT OF AGE

What is really meant by ”age pattern”?

1. Purely a biologically-based concept

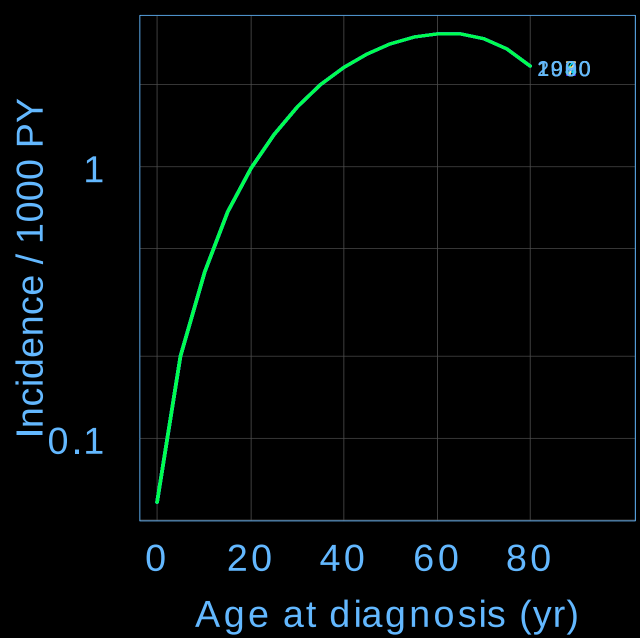
2. In an epidemiological setting, can be directly observed only if the temporal trend in disease incidence is unchanged over a lengthy period

3. Otherwise, the effect of age is inseparable from those of period / cohort

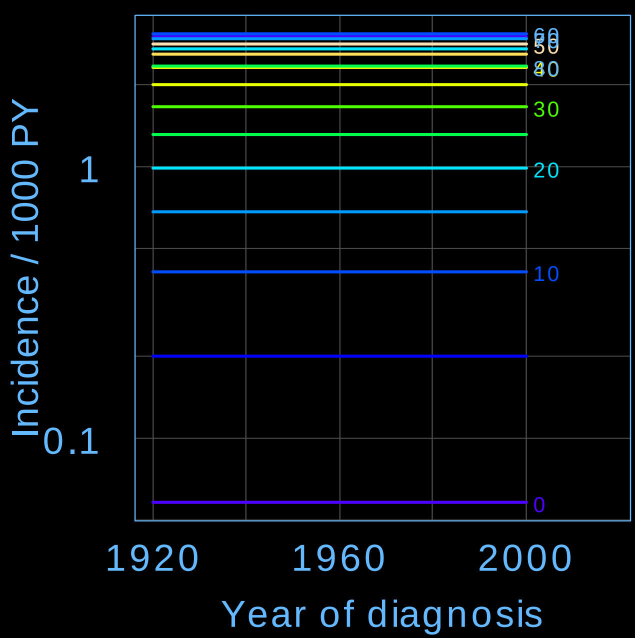
THE AGE PATTERN

Hypothetical data, assuming no change of incidence over time

Age pattern: unique
(here: a 3rd degree polynomial)



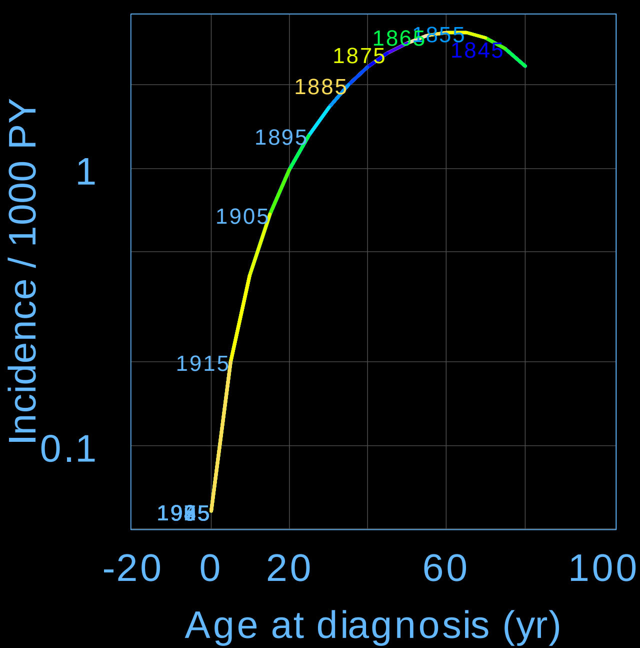
No change of incidence
at any age over time periods



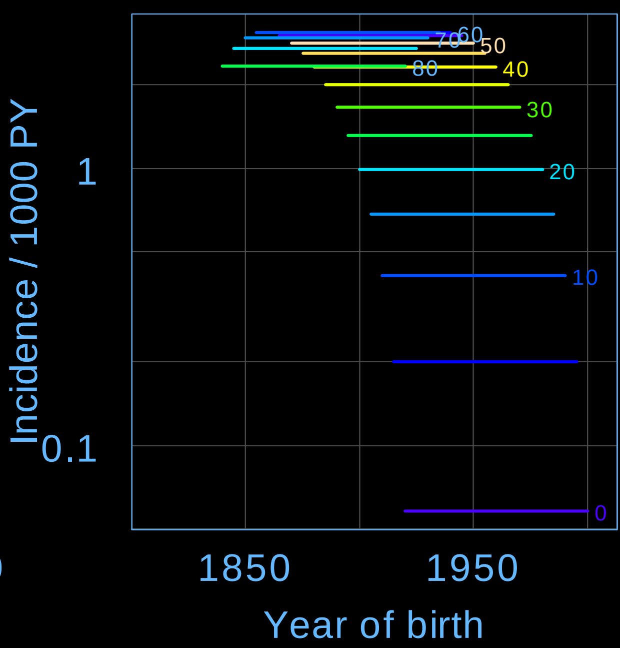
THE AGE PATTERN

Hypothetical data, assuming no change of incidence over time

Age pattern: unique
(a 3rd degree polynomial)



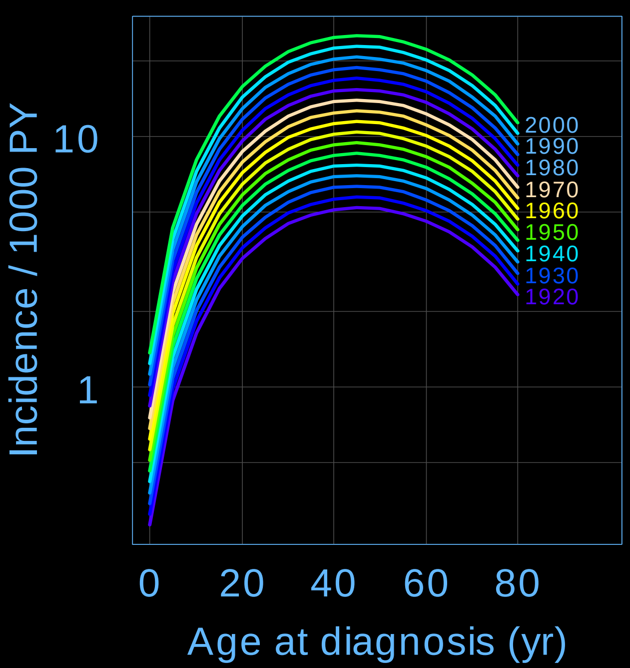
No change of incidence
at any age over cohorts



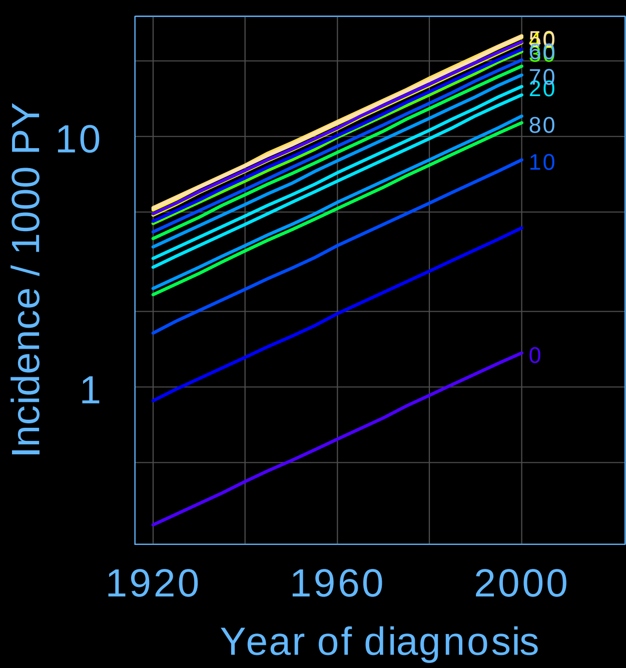
THE AGE PATTERN

Hypothetical data, assuming a 2% annual increase of incidence

Age pattern: not unique



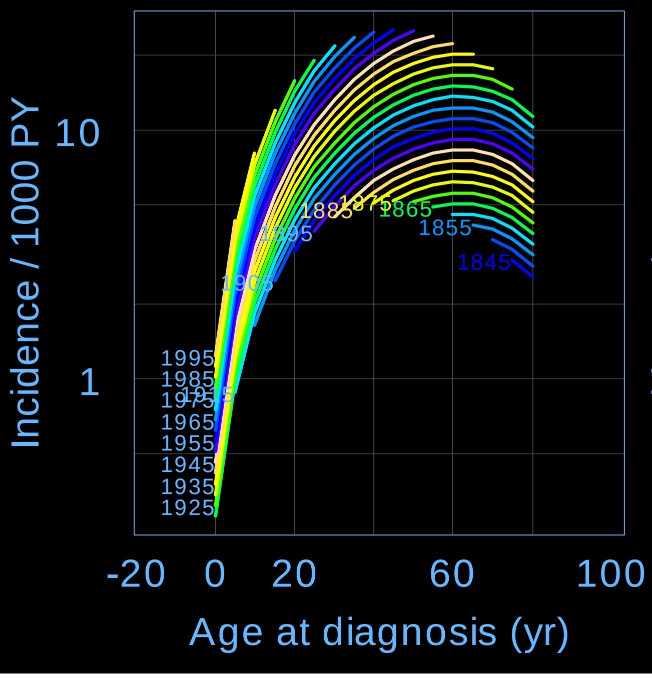
Constant increase of
incidence over periods



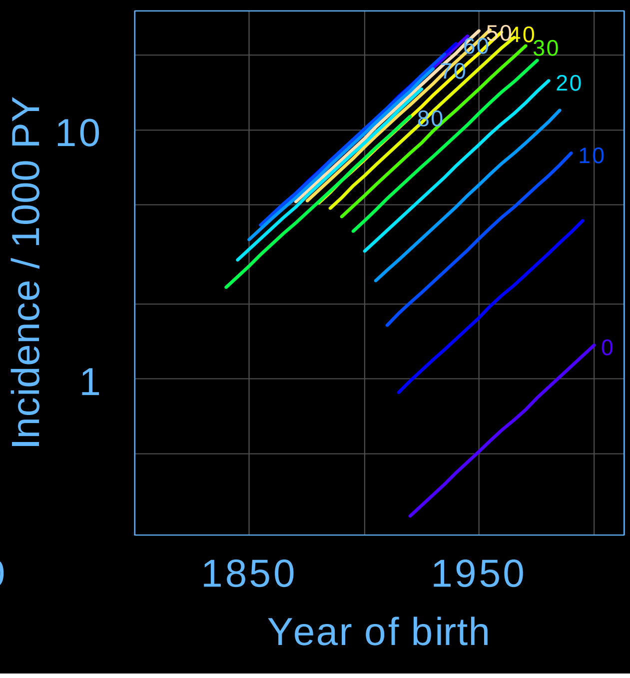
THE AGE PATTERN

Hypothetical data, assuming a 2% annual increase of incidence

Age pattern: not unique



Constant increase of incidence over cohorts



THE EFFECT OF "TIME"

Mostly perceived in terms of calendar time

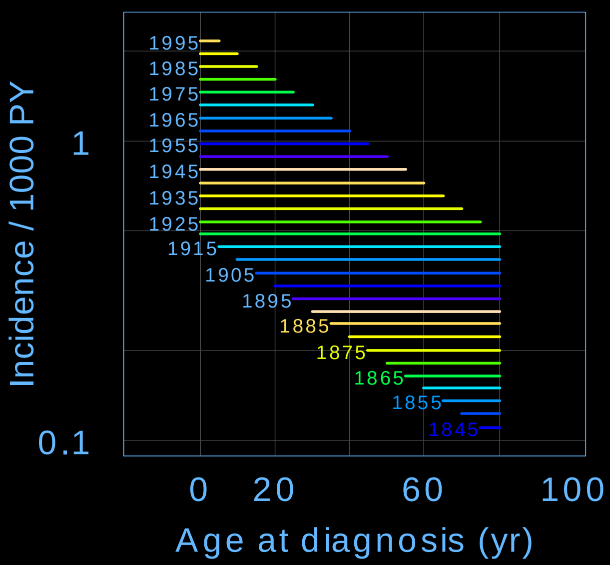
What is meant by an increasing / decreasing incidence ?

1. Can be understood in terms of a
   - ✓ change in calendar time (period effect)
   - ✓ change between cohorts (cohort / generation effect)
2. Not a unique concept when age is taken into account
3. Can be uniquely determined only in the absence of any age effect

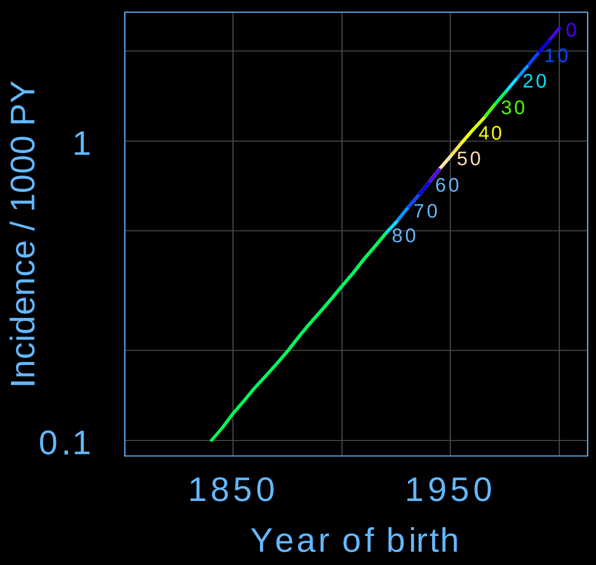
THE TEMPORAL CHANGE   P, C

Hypothetical data, assuming a 2% annual increase of incidence but no age effect

Age pattern constant in each cohort



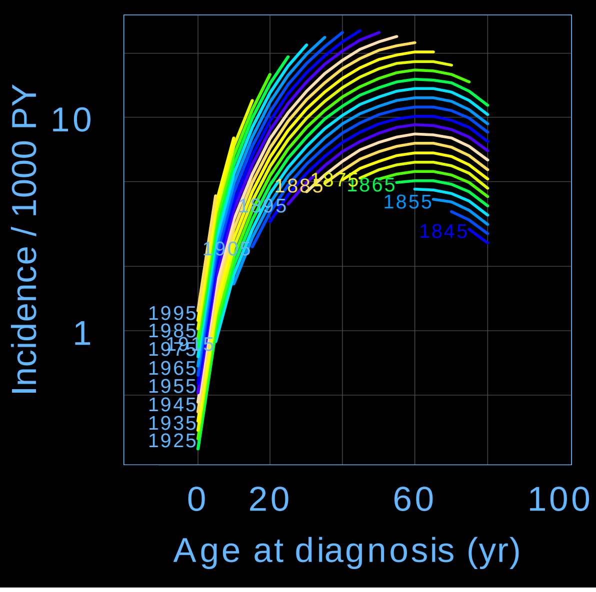
Unique temporal change



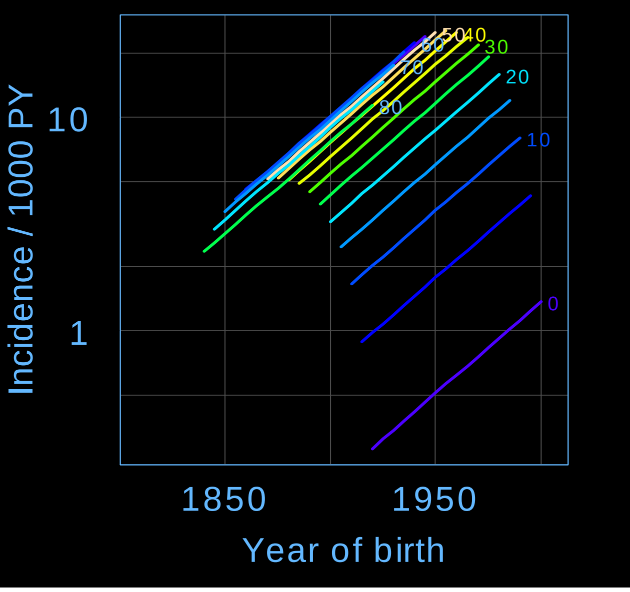
THE TEMPORAL CHANGE   P, C

Hypothetical data, assuming a 2% annual increase of incidence with an age effect (a 3rd degree polynomial of age)

Age pattern constant although proportional over cohorts



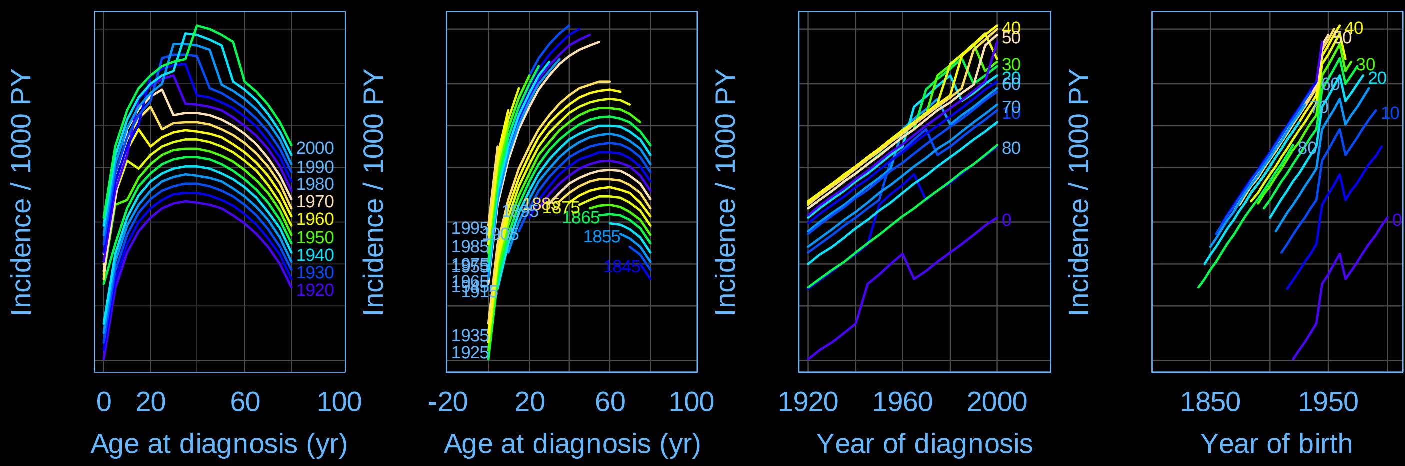
Incidence change constant but levels vary by cohorts



THE "GENERATION" EFFECT

How it looks like in a age-period setting? A simulated example

Age pattern by periods	Age pattern by cohorts	Incidence trend by periods	Incidence trend by cohorts
"Moving" peak age	Peak age unequivocal		Cohorts 1945-65



The effects of A, P, and C ?

1. In real-life situations, empirical data alone cannot tell the effects of age and "time", rather you need external information. The APC analysis can quantify the effects.

2. What are the effects of A, P and C ?

   - ✓ Age really affects morbididy on a biological basis => A
   - ✓ Interventions may affect all age groups => P
   - ✓ Causal factors which change over calendar time may be limited to a narrow age range => cohort effect => C

3. How to quantify the effects of A, P and C ?

Modelling A, P and C effects

1) Two factor model   A + P or A + C

   ✓ Useful if you know in advance that P or C is important

2) Three factor model A + P + C

   ✓ Problem: how to identify the effects of P and C

$Y = A + P + C$ is equivalent to $Y = A + P + (P - A)$

Makes no sense

Several solutions suggested how to overcome the problem

Suggested approaches to APC modelling

Instead of

$A + P + C$

introduce a ”drift” parameter such as

$A + P + C + d\,(c - c_0)$, or c_0 baseline cohort

$A + P + C + d\,(p - p_0)$ p_0 baseline period

This removes the ”common” linear trend in C / P

Detrended residuals (often curved) interpretable as effects of C / P

Parametrization of a drift model

Age function A — Age specific rates in a reference cohort c_0

Cohort function C — Interpretable as a risk ratio (RR) relative to the reference cohort c_0

Period function P — Interpretable as RR relative to the age-cohort prediction

= ”residual RR”

Drift parameter — Can be incorporated in C, or can be extracted as a separate parameter

C and P can be interchanged (they are equally valid)

Some aspect of modelling

Factor models (categorical explanatory factors)

- **Flexible, but power ↓**

Continuous functions (polynomials)

- **Retain continuity => power ↑**
- **A ”regular” shape**
- **May be unstable at the edges**

Generalized additive models (GAM)

- **Loess / lowess**
- **Splines (usually cubic, fitted between ”knots”)**
  - **Retain continuity => power ↑**
  - **No assumption of regularity, ”conforms” to data**

`Apc.fit` function

Available in the R software (http://www.r-project.org)

Several options to parametrize APC models

Reference:

Carstensen B, Keiding N. Age-Period-Cohort models.
Statistical inference in the Lexis diagram.
Available from: www.biostat.ku.dk/~bxc/APC

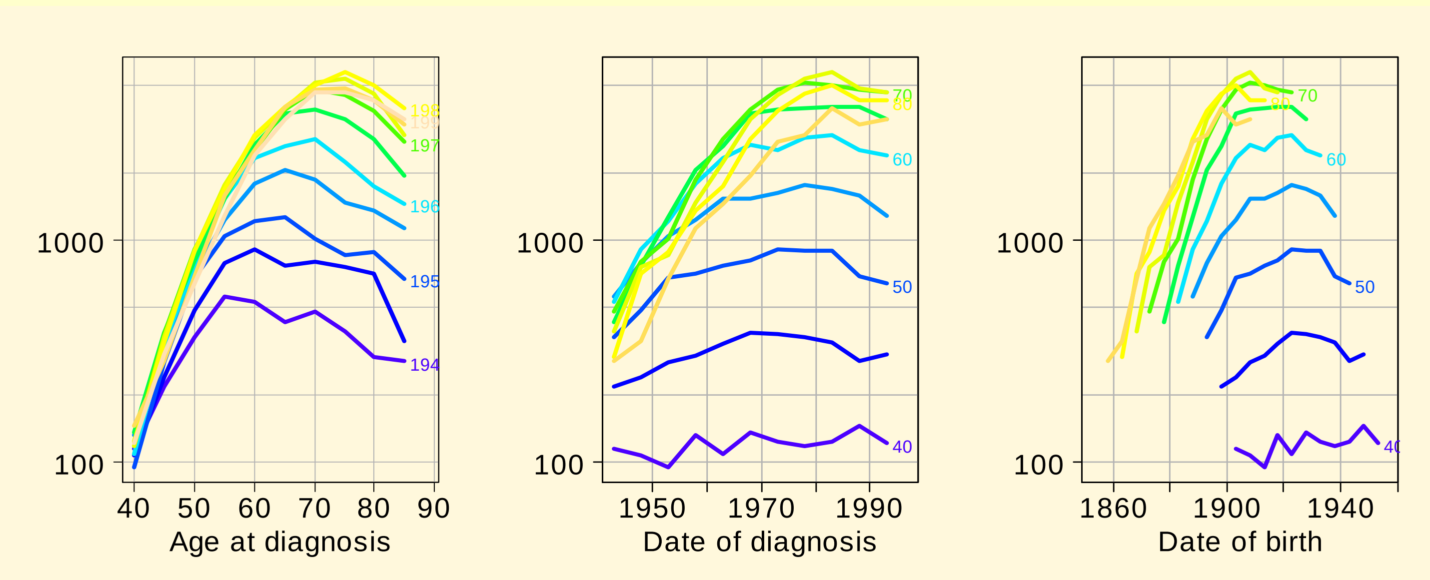
Recommended reading:

Carstensen B. Age-period-cohort models for the Lexis diagram.
Statistics in Medicine 2007; 26: 3018-45

Lung cancer incidence in Denmark, 1943-93

First plot empirical data: any suggestion for P or C ?

Incidence / age | **Incidence / periods** | **Incidence / cohorts**



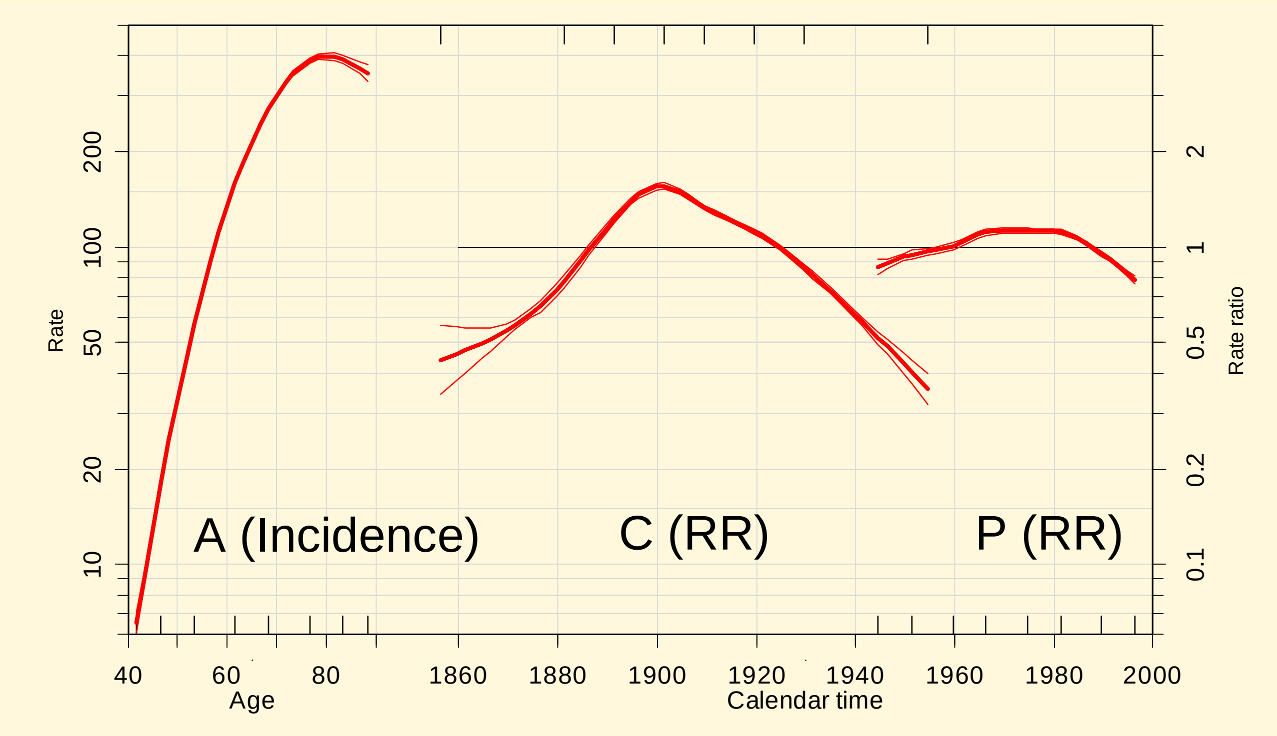
from: www.biostat.ku.dk/~bxc/APC

Lung cancer incidence in Denmark, 1943-93

Cubic splines fitted between 7 knots

AdCP model: drift not included (C, P detrended)

Incidence by age in baseline cohort | **Risk ratios vs age-specific baseline incidence**



Example: starting and quitting of smoking

Outcomes	
- starting of smoking	**A cohort effect assumed**
- quitting of smoking	**No cohort effect assumed**
Data	**A smoking survey 2003** **University of Tartu staff**
Design	**Cross sectional survey** **Cohort constructed retrospectively**
Questionnaire	**age of starting regular smoking** **age of quitting**

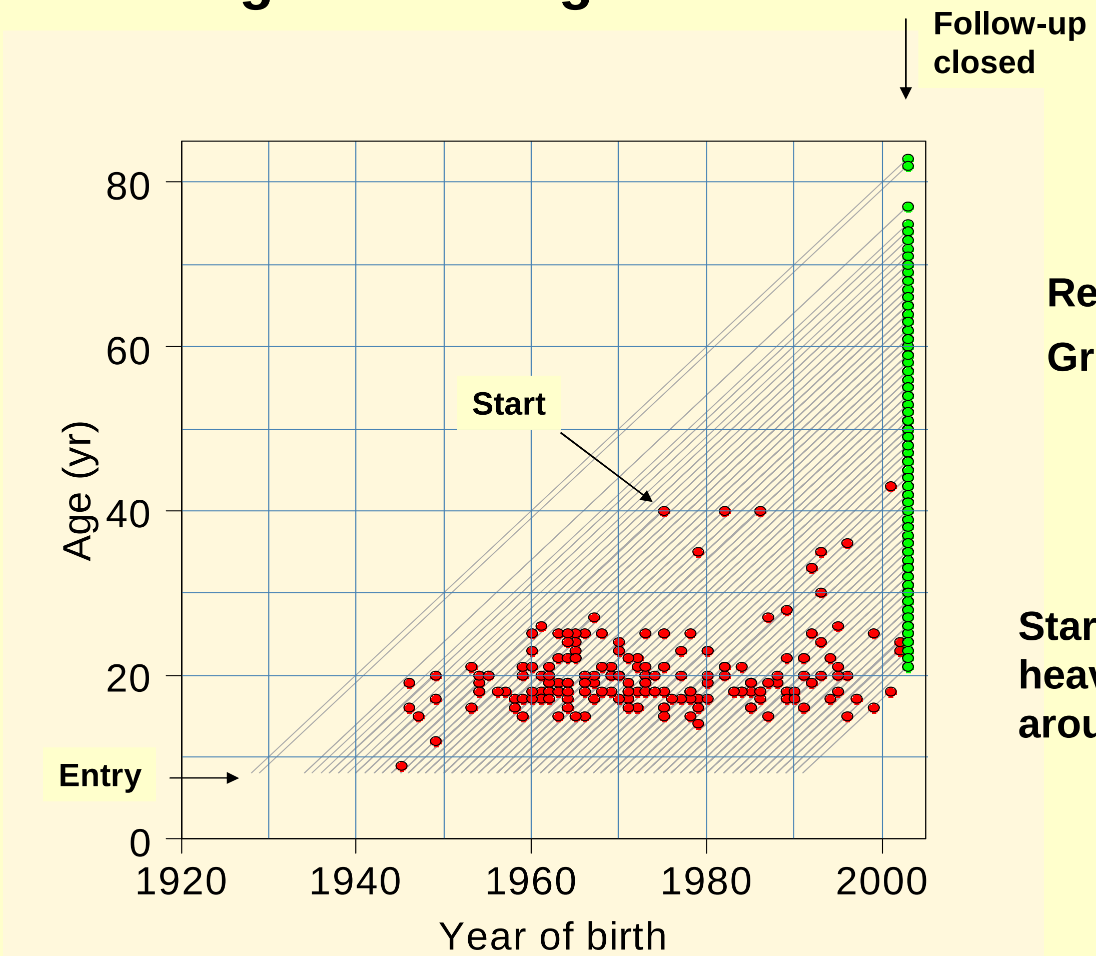
Starting of smoking: a cohort effect assumed

Kaplan-Meier cumulative incidence proportion

(= Years since birth)

Most people start before the age of 25, if they ever start

Starting of smoking: ”lifelines”



Red: ”event”

Green: censored

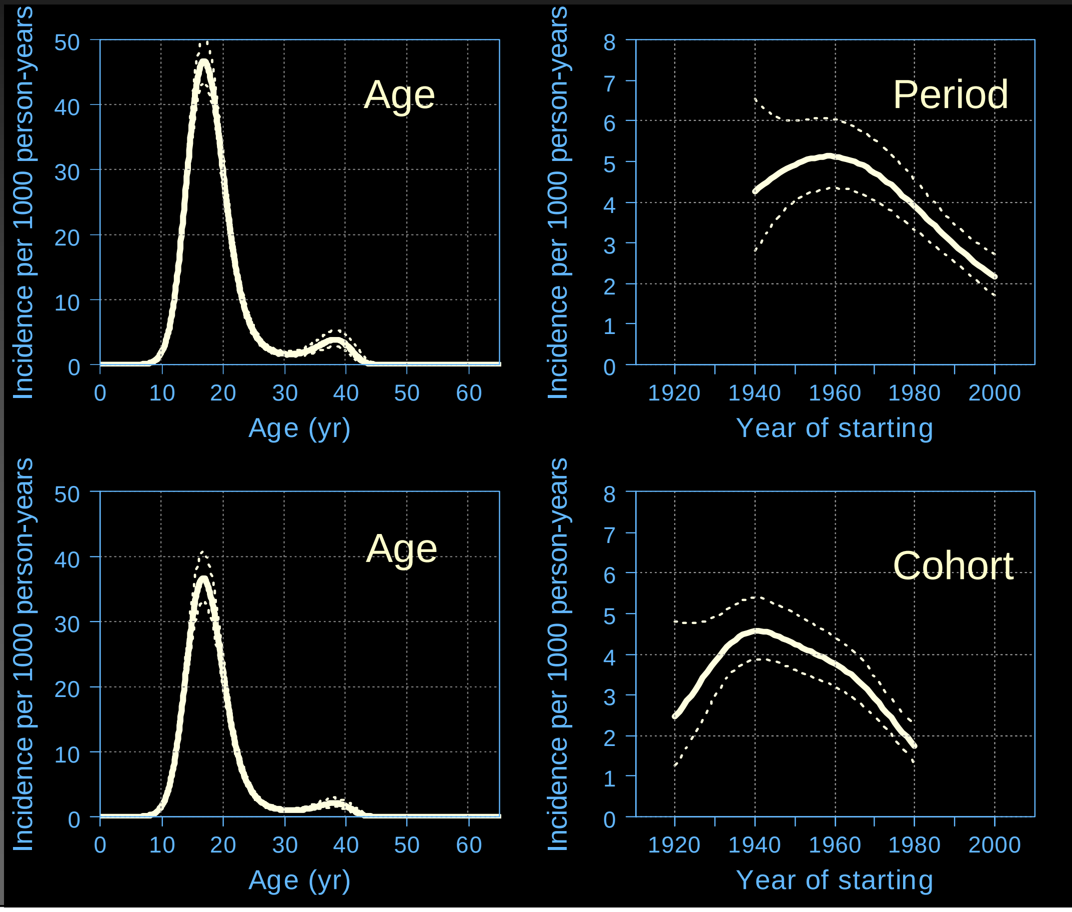
Starting of smoking heavily concentrated around the age of 20

Starting of smoking: AP / AC model

Incidence fitted by GAM with 5 knots

AP model

AC model



Starting of smoking: a drift model AdCP

Incidence fitted by GAM with 4 knots

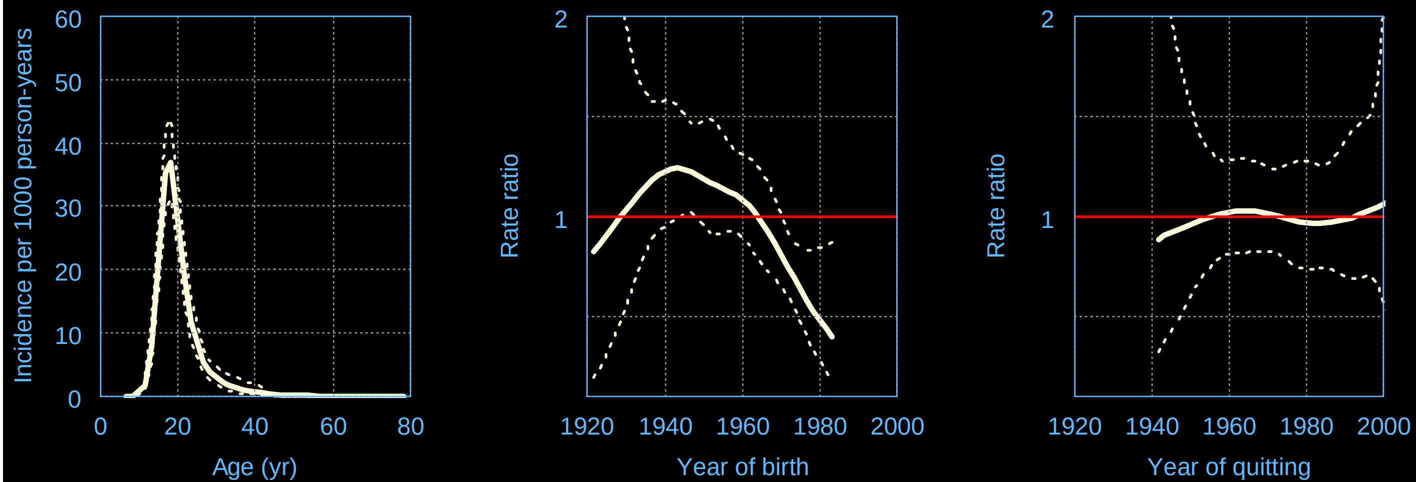
C, P constrained to 0, drift – 1.5 % / year (not included)

Now A, C, P ”detrended”

A

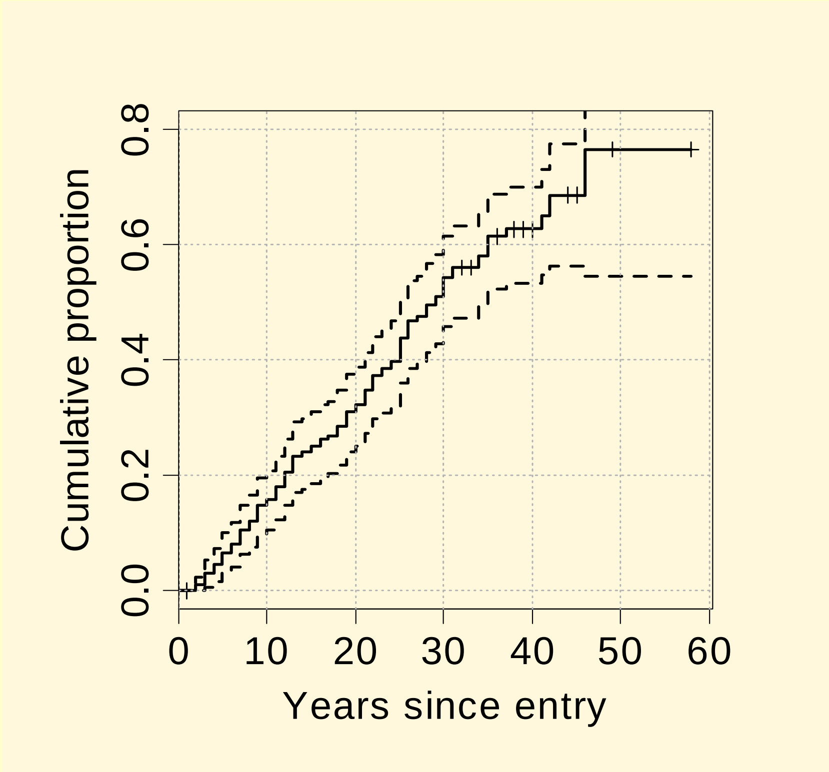
C

P



Quitting of smoking: a period effect assumed

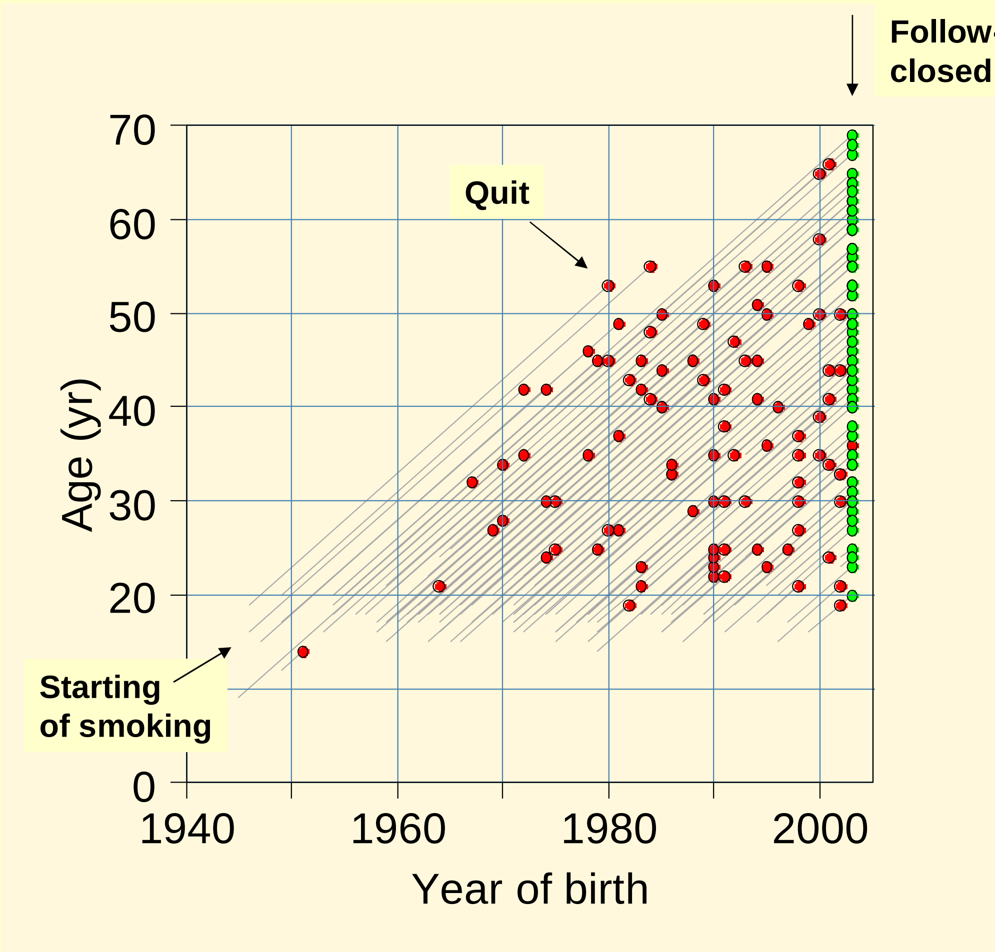
Kaplan-Meier cumulative incidence proportion



(= Years since starting smoking)

Proportion of quitters increases by time in a linear fashion

Quitting of smoking: "lifelines"



Red: "event"

Green: censored

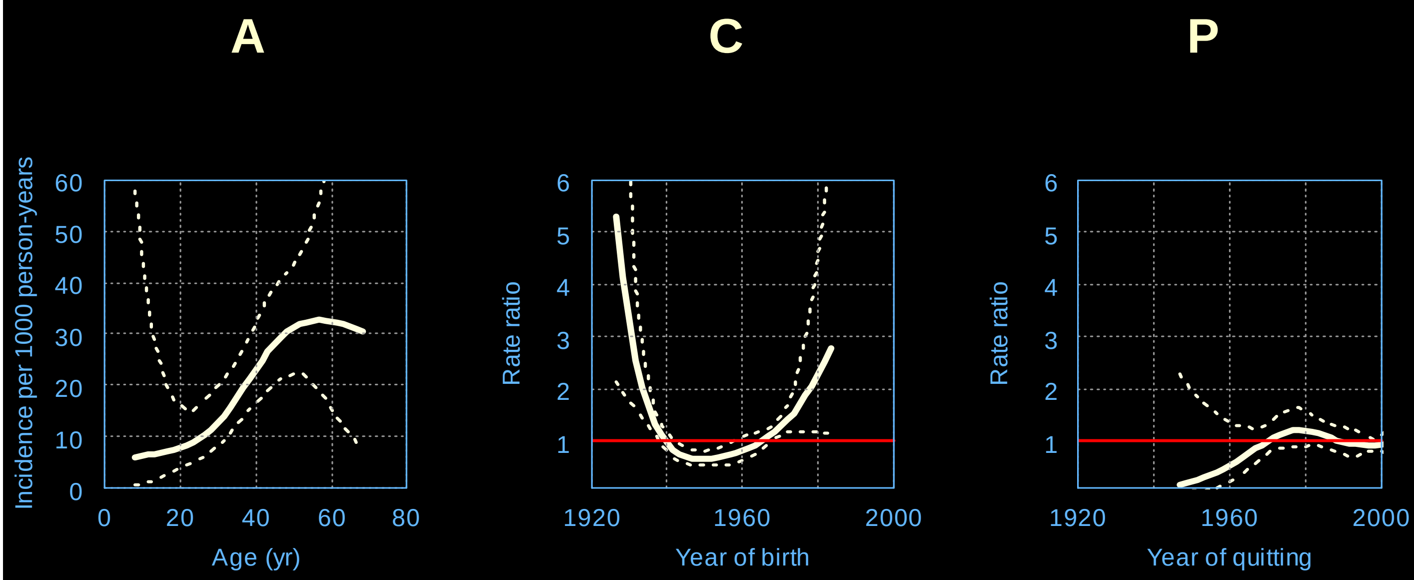
No predilection to any particular age

Quitting of smoking: drift model AdCP

Incidence fitted by GAM with 4 knots

C, P constrained to 0, drift + 4 % / year (not included)

A, C, P ”detrended”



Some recommendations for analysis of long-term follow-up data

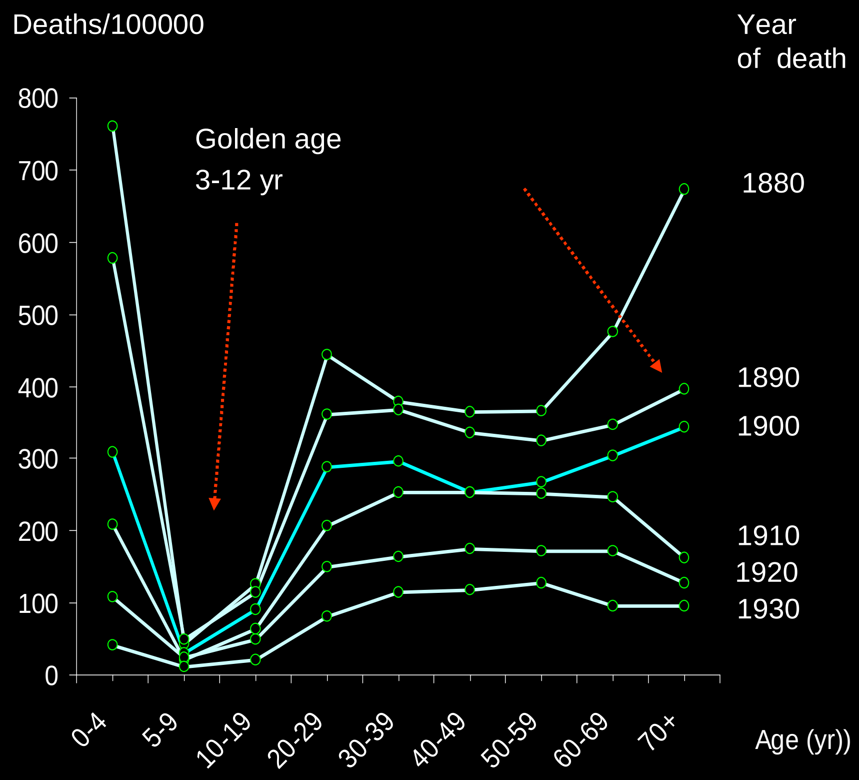
- ✓ **Arrange data to form a Lexis diagram (allows different time scales)**
- ✓ **Compute cases and person-times**
  (how to do it, see e.g. Carstensen 2007)
- ✓ **Use Poisson regression with age, period and cohort as <u>continuous</u> variables; specify the drift parameter**
- ✓ **Report age-specific incidence figures and relative rates (RR) versus the pertinent baseline**
- ✓ **P values can be calculated for A, P and C but are rarely useful: rather use confidence intervals**

from: www.biostat.ku.dk/~bxc/APC

AGE PATTERN OF TUBERCULOSIS

USA Mass, 1880-1930, Men (Frost 1939)

- Previously tbc typical of the young
- The age peak has shifted towards the older ages
- Assumed cause: impared resistance & lowered physiological reserves among the elderly



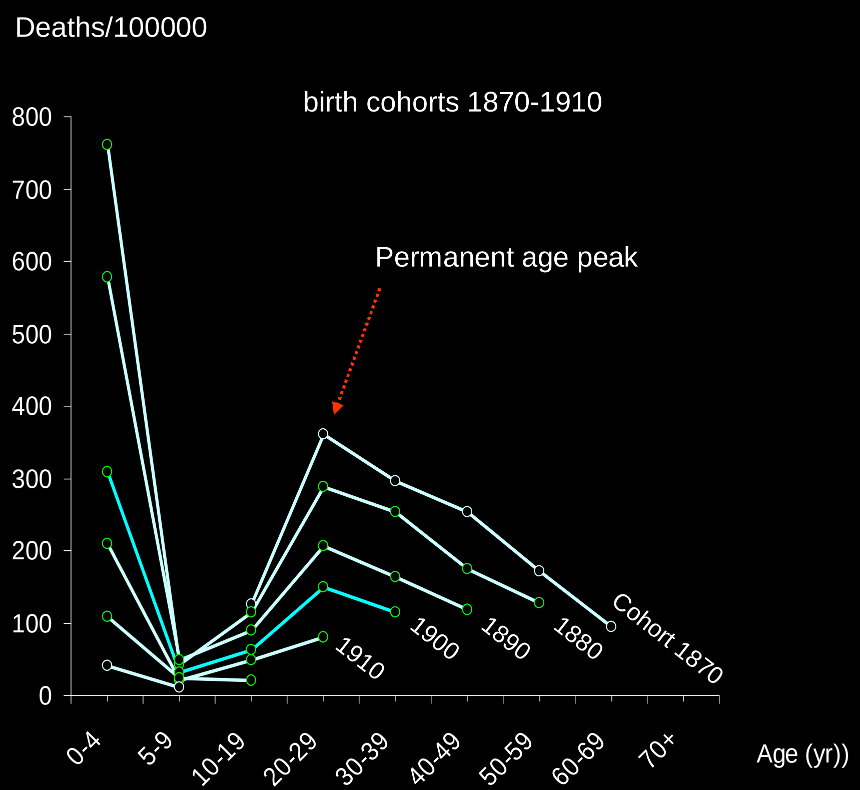
Frost W 1939

AGE PATTERN OF TUBERCULOSIS

USA Mass, 1880-1930 (Frost 1939)

Frost's observation

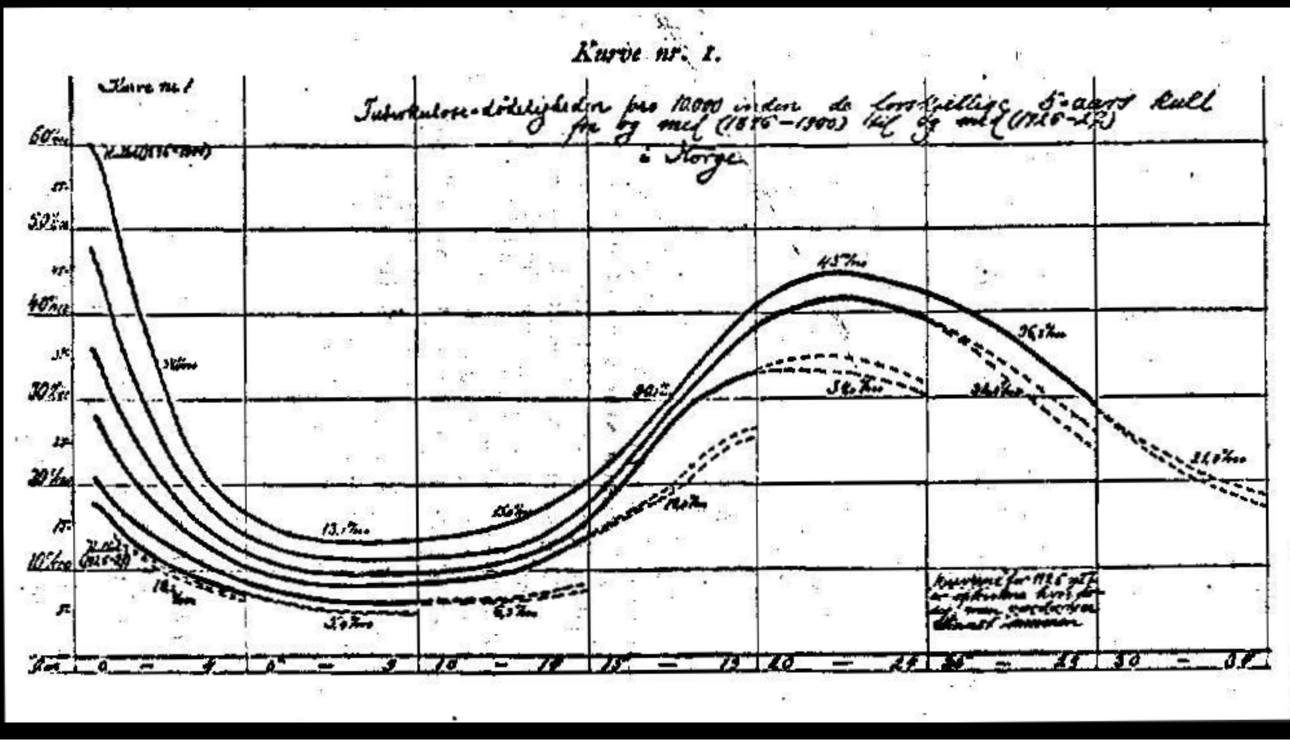
- Age pattern constant in successive generations
- Tbc declined similarly in all age groups
- Generation determines the entire lifetime risk



Frost W 1939

TUBERKULOSEDÖDELIGHED PRO 10000 INDEN DE FORSKJELLIGE 5-AARS KULL I SVERIGE, 1896- 1926

Anvord Kr. Hvad kan vi laera ved å folge tuberkulosens gang fra generasjon til generasjon? Norsk Magasin for Laegevidenskaben 1930; 91: 642-660



DISEASES WITH A SUSPECTED OR CONFIRMED COHORT EFFECT

Pulmonary tuberculosis	Anvord 1930, Frost 1939
Coronary disease & stroke	Feinleib 1993
Suicides	Åsgard *et al.* 1987
Duodenal ulcer & helicobacteria	Susser & Stein 2002
Chronic gastritis	Sipponen 1996
Ulcerative colitis	Sonnenberg 2002
Stomach cancer	Aragones 1997

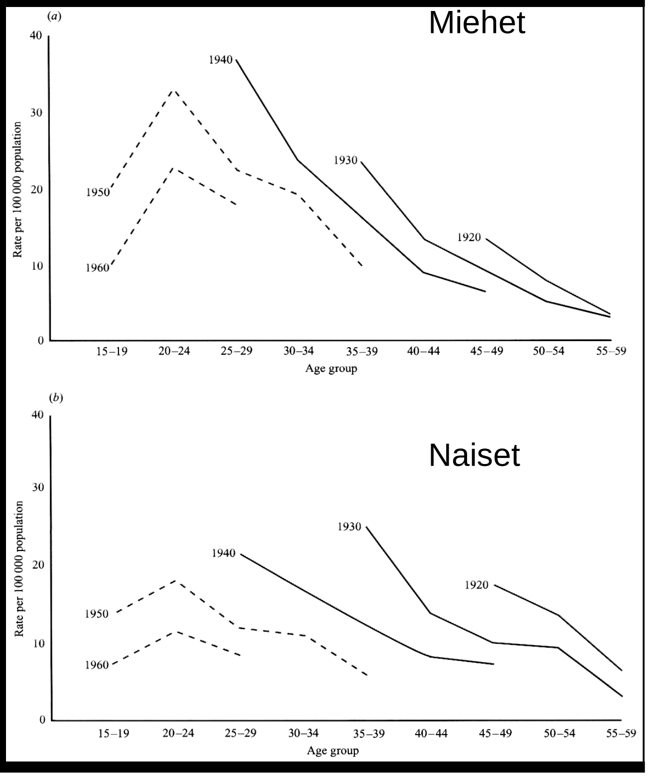
THE INCIDENCE OF SCHIZOPHRENIA

Takei N, Lewis G, Sham P, Murray RM. Age-period-cohort analysis of the incidence of schizophrenia in Scotland. Psychological Medicine 1996;26:963-73

- A cohort effect estimated at +10%
- The causative factor decreases in intensity over generations

Suggested explanations

- Mothers' nutrition improved
- Better control of infections



Takei N 1996

FINNISH GENERATIONS

Generation	Generations possibly affected	
Wars 1939-44	1920s	Smoking
Postwar time 1945-50	1940s	Adverse living conditions
Urbanization 1960s	1940s	Depopulation of countryside New life in cities
Economic depression in early 1990s	1980s ->	Widening of social gaps Marginalized people = a new social class

Backman G 1988.1