NATIONAL INSTITUTE OF PUBLIC HEALTH

NordForsk PhD course in Register-Based Epidemiology

Time trends analyses
Age-period-cohort models
Projections

Lau Caspar Thygesen

Introduction

Descriptive epidemiology

Monitoring

Plan health care resources for the future

Plan actions aimed at improving health

Observe sudden or unexpected changes in disease risks

Incidence and prevalence core indicators of public health

Necessary input to make projections of future population health

Modig 2017

Introduction

The Nordic countries special situation with comprehensive nationwide registers

Hospital-based data, mortality statistics, health-care contacts

Size of the relevant population

Prevalence and incidence

Sensitivity important

Modig 2017

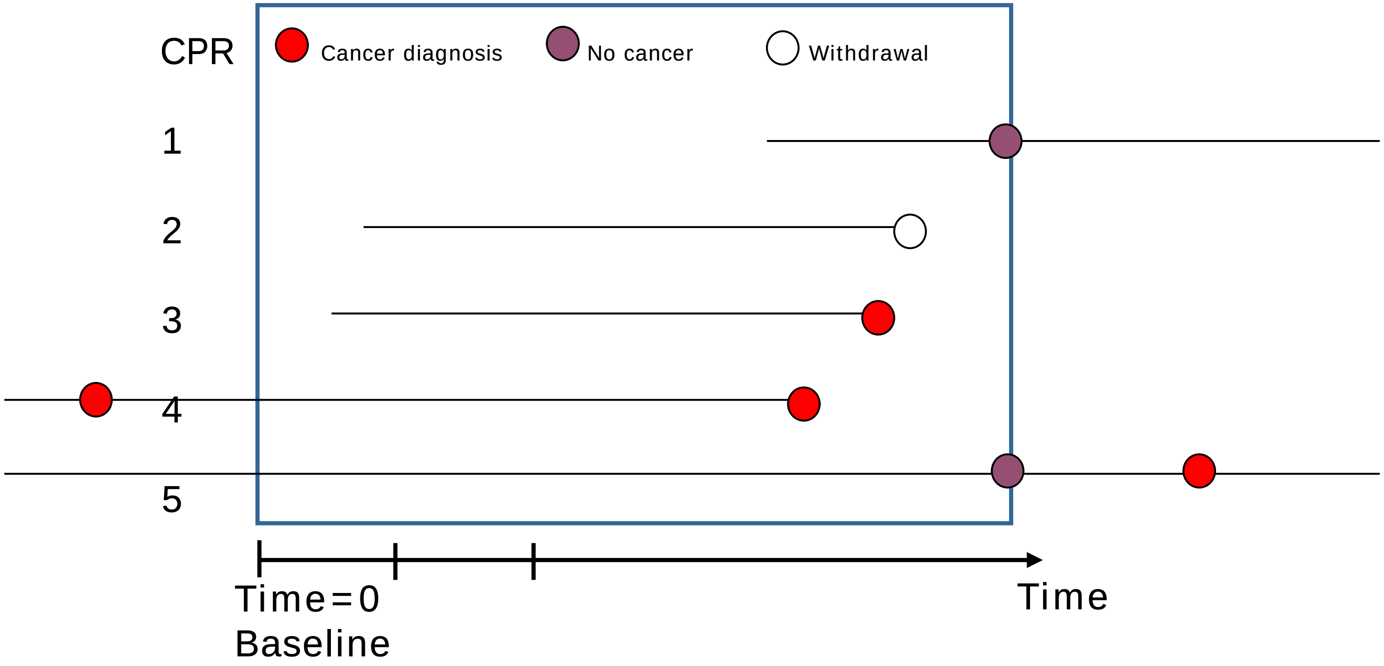
Prevalence proportion
Incidence rate

PP = Number diseased / N

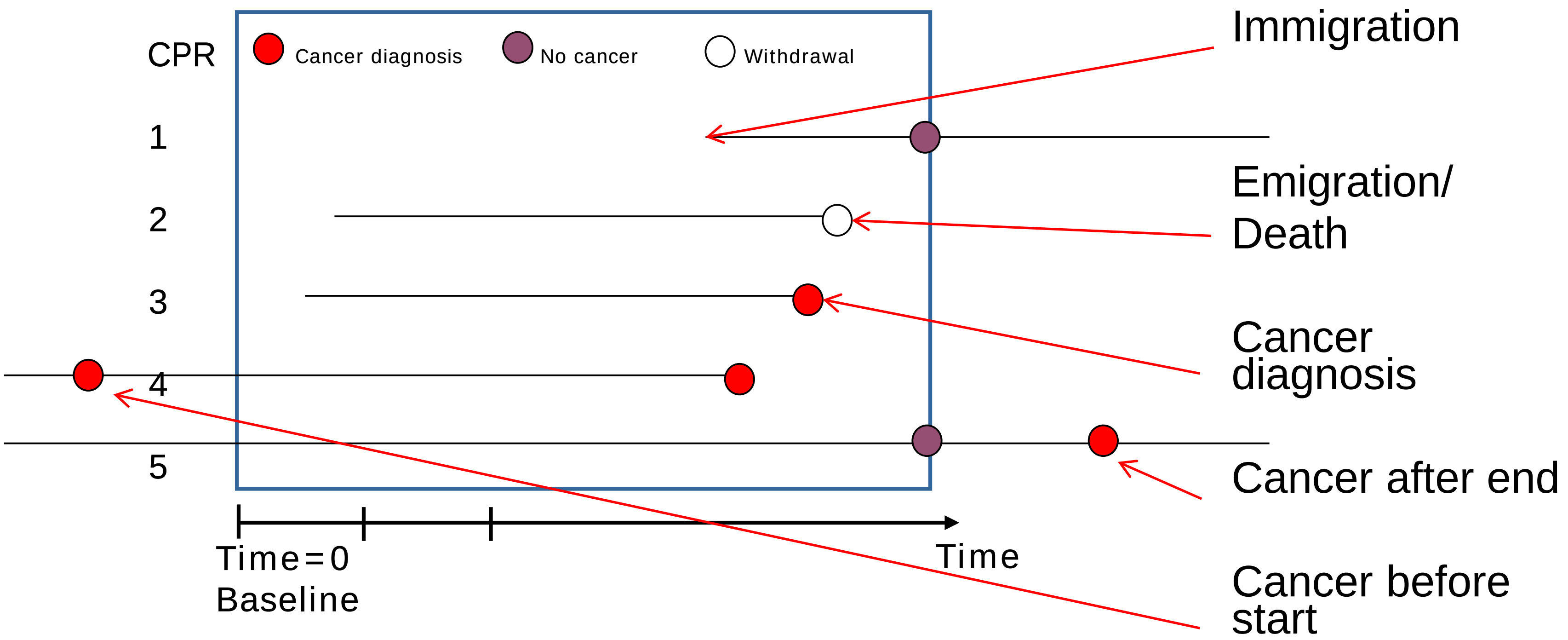
IP = New diseased / N

IR = New diseased / risk time

Dynamic cohort



Dynamic cohort



Incidence and prevalence

Need information about the entire disease history of individuals

Left truncation (register start)

Not possible to definitely define a first occurrence of disease

Wash-out period

How long? Modig (2017) 7 years

Moving 7 years or just 7 years before register-start?

Biological considerations

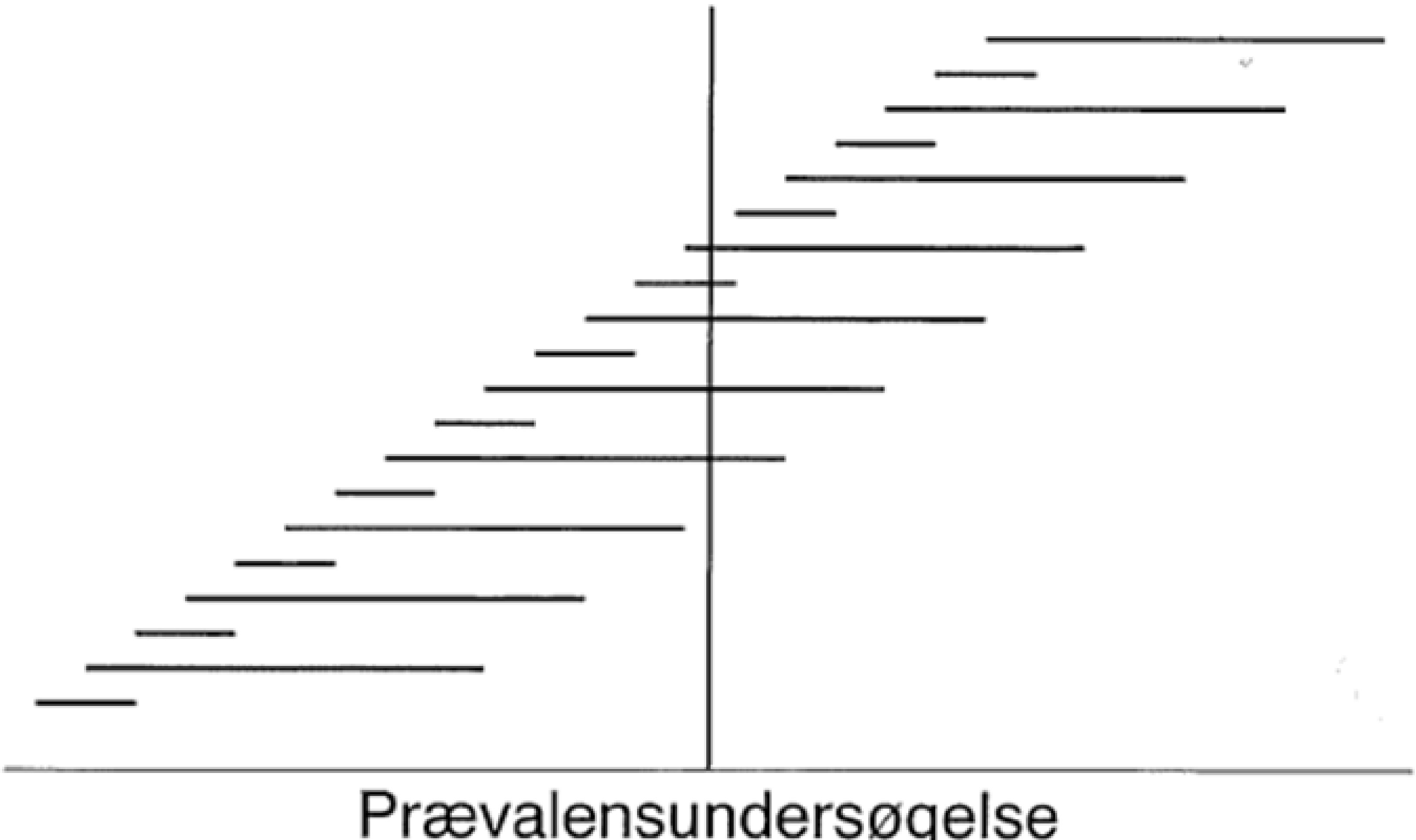
Information about entire disease history

Left truncation

Duration of disease

Point or period prevalence

Prevalence bias



Mid-year population estimate for risk time (person-years)

Calculation of precise person-time often not possible (or very cumbersome)

Movement in and out of population (mortality, birth, migration, disappearance)

Alternative use mid-year population as estimate of risk time

Approximation

Statistics Denmark publishes population size 1 July each year
OR
Mean value of population size at start and end of year

Assumptions

No large changes in population structure

Mortality happens throughout the year

Mean risk time for people who die after half year. And risk time for births is half year

Some people with the outcome of interest will contribute with risk time – assumes it is a small proportion

Probably reasonable

Less reasonable

- seasonality in mortality (and births)
- infant mortality (much higher just after birth)

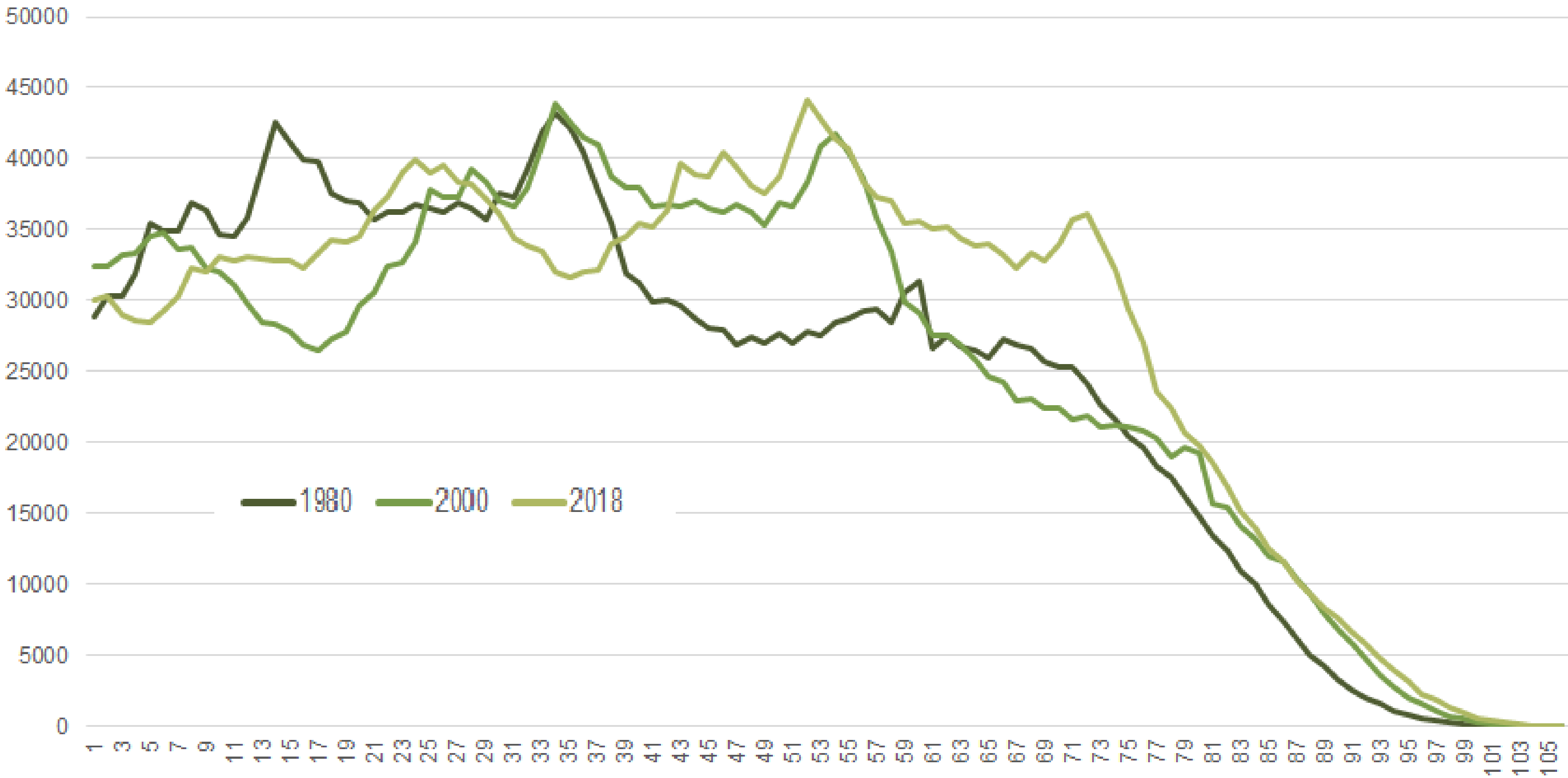
Exercise 1

You are interested in describing lung cancer prevalence and incidence in one Nordic country from 1980 to 2018

Please consider the following elements:

- Inclusion/exclusion criteria for prevalence and incidence studies
- Will you introduce any wash-out period?
- Are you interested in point or period prevalence?
- How do you estimate population size / risk time?
- How would you include changes in age-distribution over time?

Age distribution Denmark women



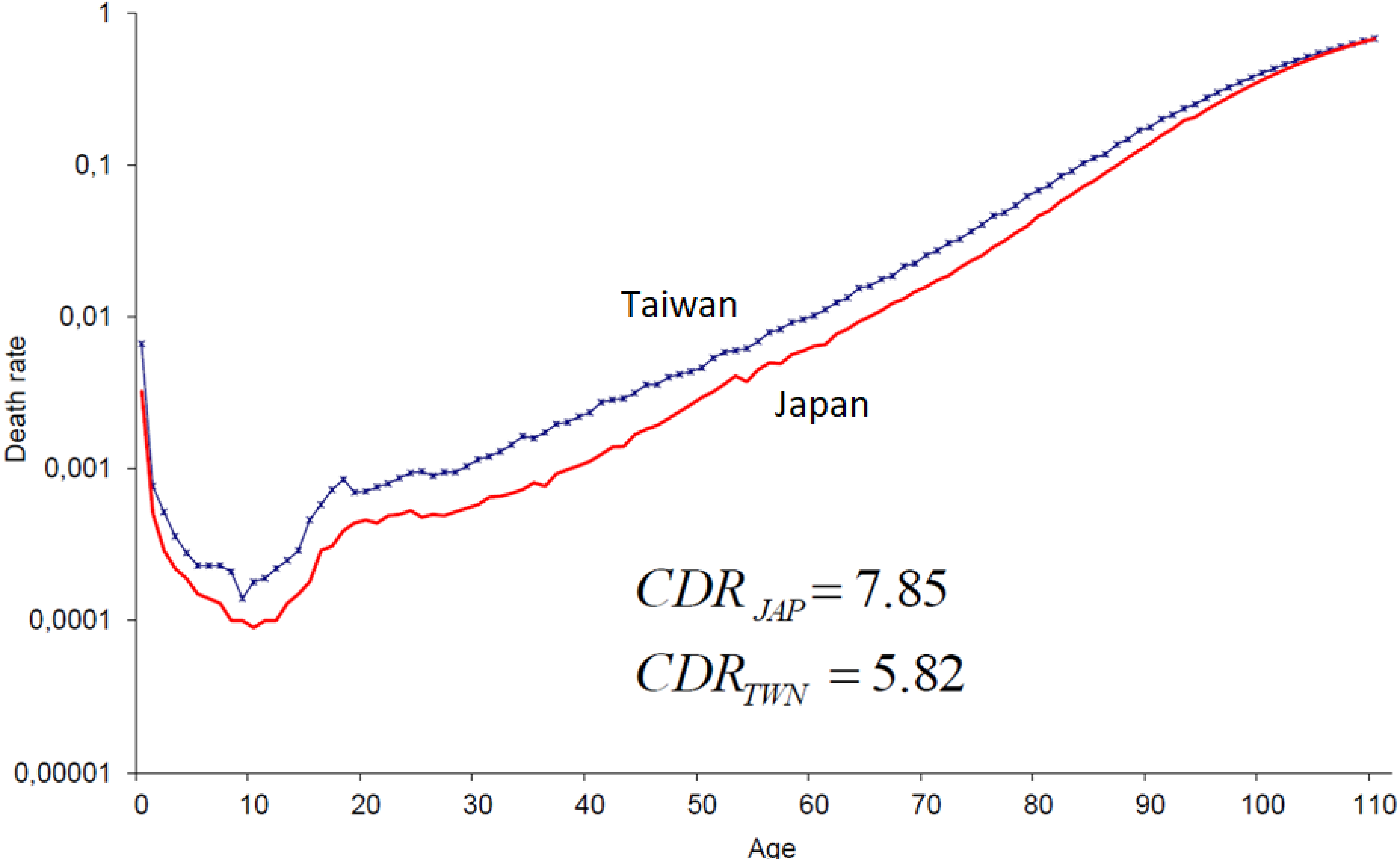
Age-specific death rates for the total population of Japan and Taiwan in 2000

$$CDR_{JAP} = 7.85$$

$$CDR_{TWN} = 5.82$$

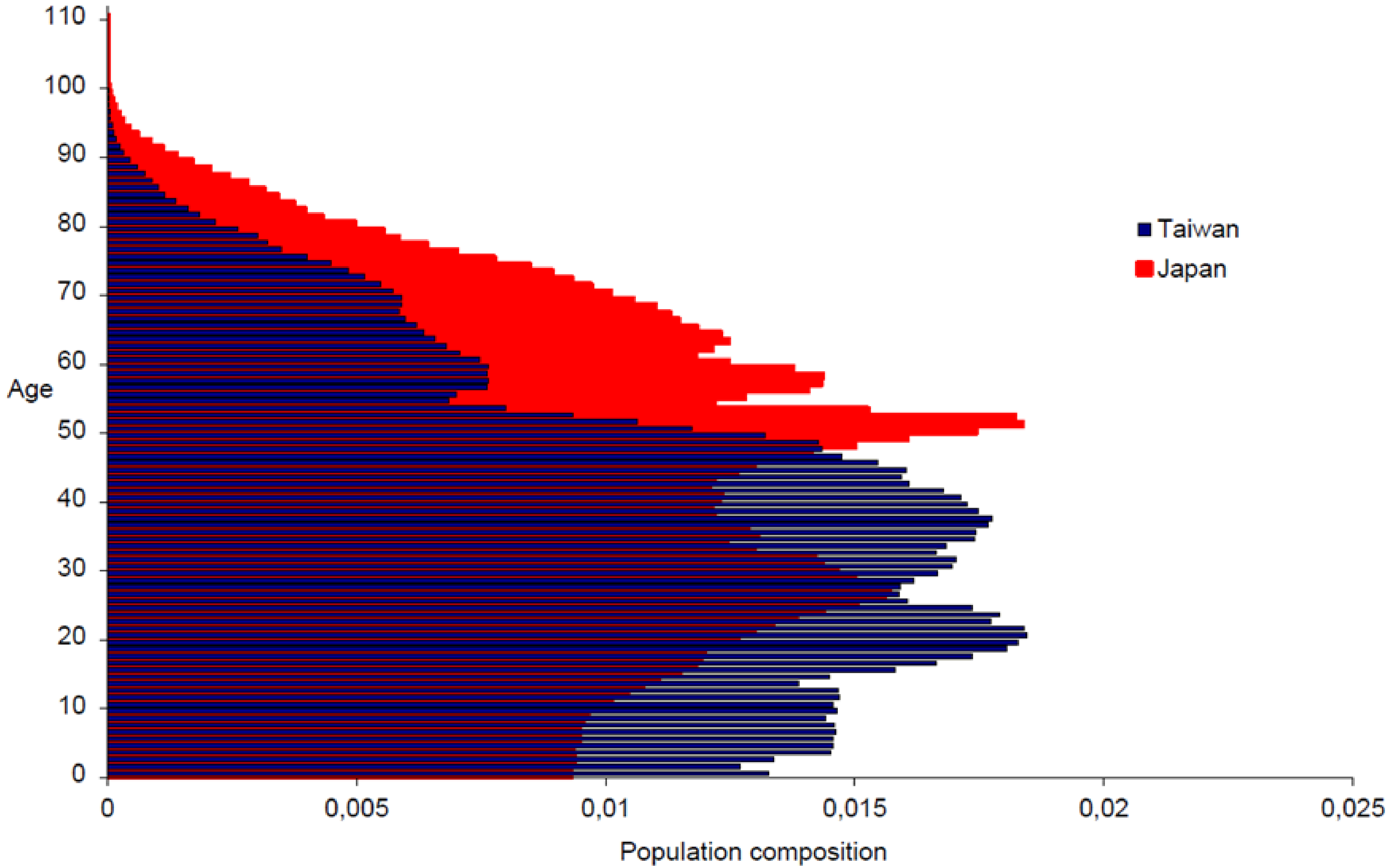
Source: Human Mortality Database (2007)

Age-specific death rates for the total population of Japan and Taiwan in 2000



Source: Human Mortality Database (2007)

Population composition for the total population of Japan and Taiwan in 2000



Source: Human Mortality Database (2007)

Standardization as one solution

Crude death rate (CDR) weighted average of age-specific rates
Weight is the proportion of population in age group
Only populations with same age dist have comparable CDR
Effect of age distribution should be removed when comparing
Two standardization approaches

Mortality in Denmark and Greenland, males, 1975

	Greenland			Denmark			
Age year	Death D_i	Observation years	Death per 1,000	Death D_i	Observation years	Death per 1,000	Ratio Denmark/ Greenland
<1	26	429	60.6	434	35,625	12.2	5.0
1–4	4	2,044	2.0	101	1,49,186	0.7	2.9
5–14	11	7,194	1.5	175	4,01,597	0.4	3.7
15–44	37	13,572	2.7	1,494	1,076,842	1.4	1.9
45–64	35	2,949	11.9	6,166	5,52,133	11.2	1.1
65+	47	640	73.4	19,204	2,88,834	66.5	1.1
Total	160	26,828	6.0	27,574	2,504,217	11.0	0.55

Interpret the table

SDU

Direct standardization

		Greenland			Denmark		
Age year		w_i	MR	Sum	w_i	MR	Σ
<1	429/26,828	= 0.016	X 60.6	= 0.970	0.014 X	12.2	= 0.174
1–4	2,044/26,828	= 0.076	X 2.0	= 0.152	0.060 X	0.7	= 0.042
5–14	7,194/26,828	= 0.268	X 1.5	= 0.402	0.160 X	0.4	= 0.064
15–44	13,572/26,828	= 0.506	X 2.7	= 1.366	0.430 X	1.4	= 0.602
45–64	2,949/26,828	= 0.110	X 11.9	= 1.308	0.220 X	11.2	= 2.469
65+	640/26,828	= 0.024	X 73.4	= 1.751	0.115	66.5	= 7.670
Total		1.0		6.0	1.0		11.0

MR(Danmark-standardized to the Greenlandic age distribution)
= 0.016*12.2+0.076*0.7+0.268*0.160+0.506*1.4+0.110*11.2+0.024*66.5
= 3.8

SDU

Choice of standard population

Mean between two populations

Standard populations

- African standard
- European standard
- Scandinavian standard
- World standard

Easier to compare populations

SDU

Age	African	World	Europe
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
40-44	5,000	6,000	7,000
45-49	5,000	6,000	7,000
50-54	3,000	5,000	7,000
55-59	2,000	4,000	6,000
60-64	2,000	4,000	6,000
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**

SDU

Indirect standardization

Calculate the expected number of cases if you use another countries rates

Indirect standardization

	Denmark	Greenland		
Age	Mortality rate per 1,000 observation years	Observation years	Observed number of deaths	Expected number of deaths[a]
<1 year	12.2	429	26	5.2
1–4	0.7	2,044	4	1.4
5–14	0.4	7,194	11	2.9
15–44	1.4	13,572	37	19.0
45–64	11.2	2,949	35	33.0
65+	66.5	640	47	42.6
Total	11.0	26,828	160	104.1

[a]If they had the same mortality rates as in the Danish population.

SDU

Standardized mortality ratio (SMR)

$$\text{SMR} = \frac{\text{observed number of deaths}}{\text{expected number of deaths}} = \frac{160}{104.1} = 1.54$$

54% higher number of deaths in Greenland than expected if the Greenlandic population had the same mortality rates as in Denmark

SDU

Time trends

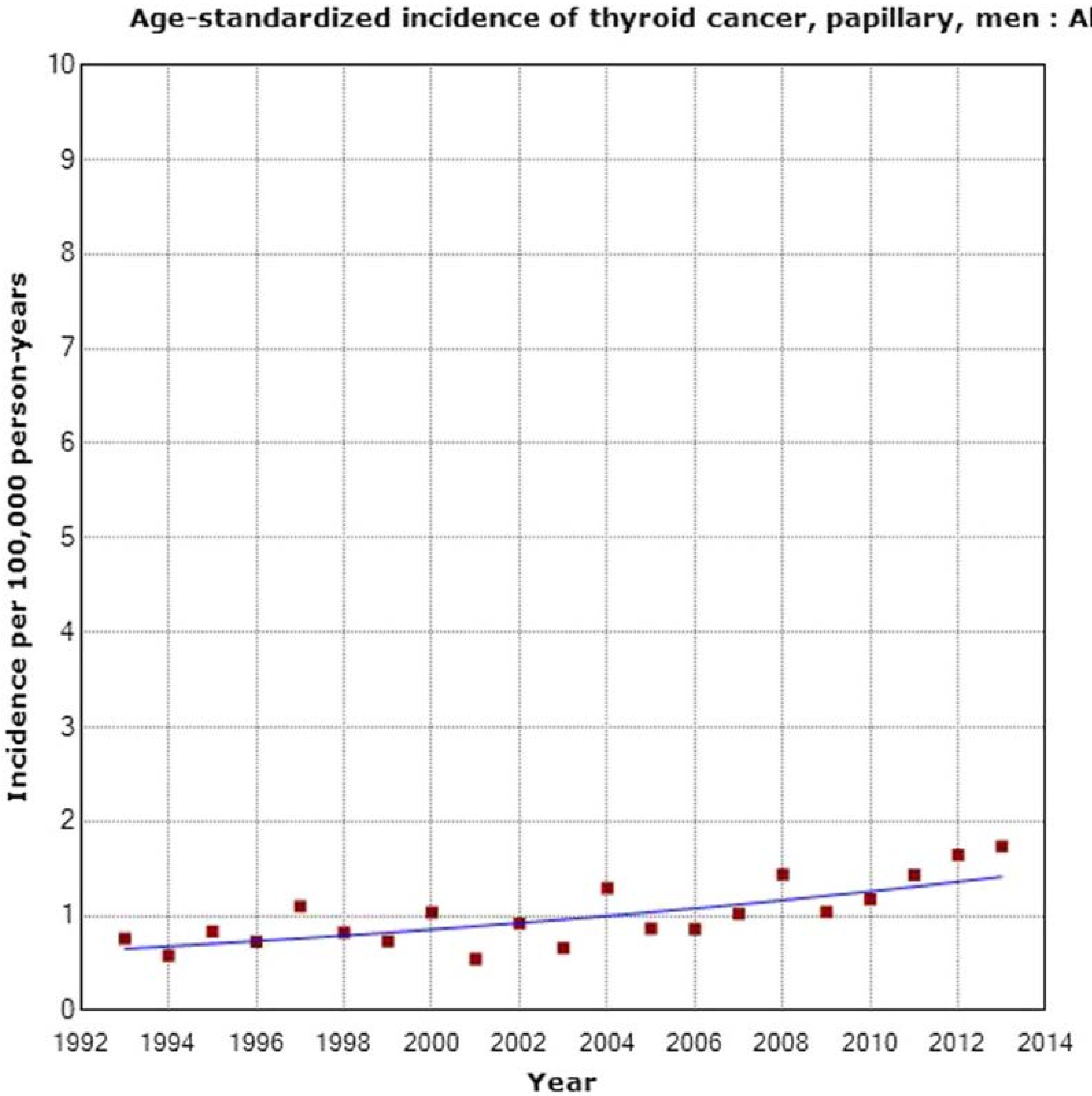
Annual percent change

Average annual percentage of change in the age-standardized rates

E.g. fitting a simple regression model to the log of the rate

Assumption of linearity on the log scale

Equivalent to a constant change assumption



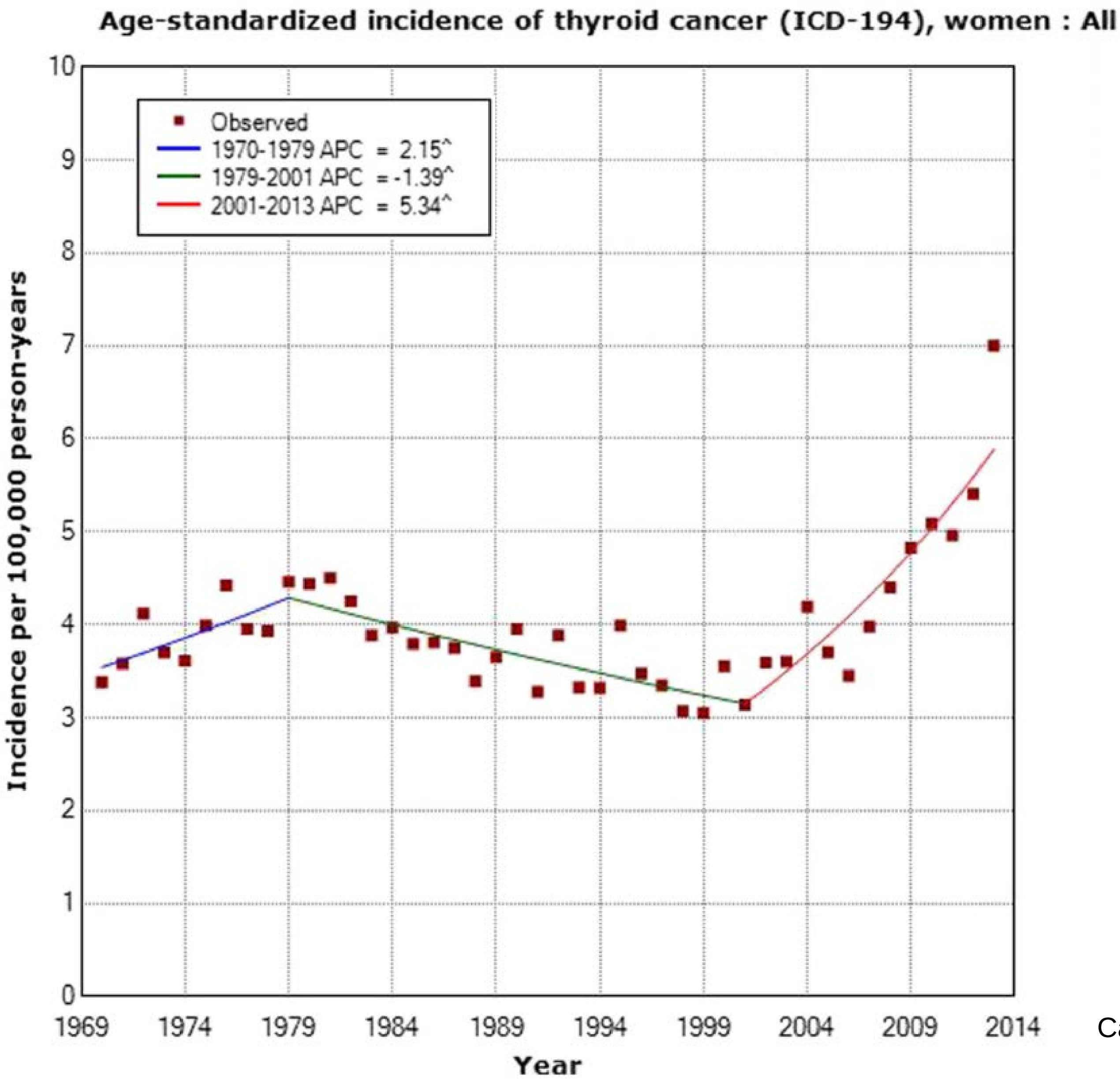
Carlberg et al. BMC Cancer (2016) 16:426

Joinpoint regression analysis

Extension

Trends in rates by fitting a model of 0–4 joinpoints

Annual percentage change for each linear segment



Carlberg et al. BMC Cancer (2016) 16:426

Exercise 2

You are interested in describing lung cancer prevalence and incidence in your Nordic country from 1980 to 2019

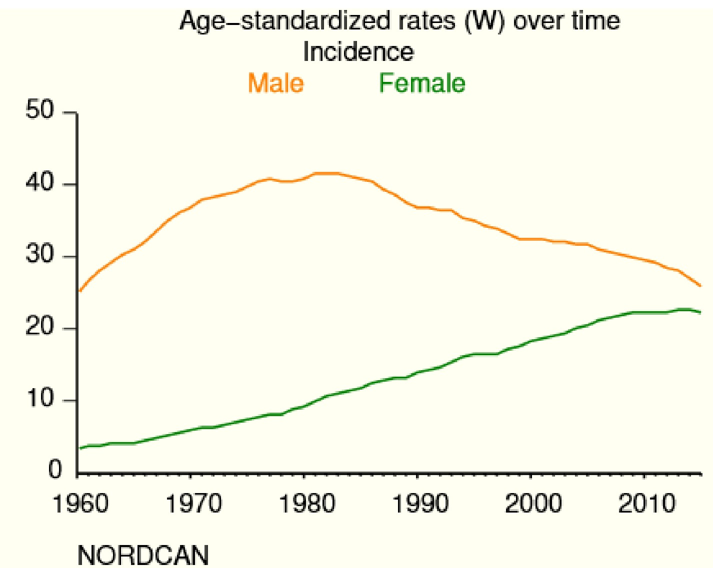
Now focus on lung cancer incidence

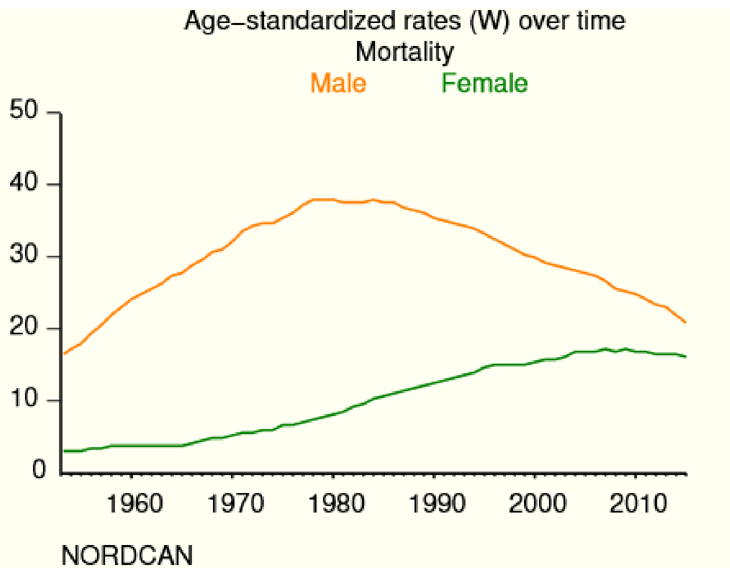
Please consider the following elements:

- How would you include change in age-distribution over time?
- Standardization? Direct or indirect? Standard population?
- How would you evaluate trends over this long time period?

Cancer stat fact sheets
Nordic countries - Lung

	Male	Female
Number of new cases per year (incidence 2011–2015)	7640	6622
Proportion of all cancers (%)	9.0	8.7
Proportion of all cancers except non–melanoma skin (%)	9.7	9.3
Risk of getting the disease before age 75 (%)	3.5	2.9
Age–standardized rate (W)	28.0	22.7
– Estimated annual change latest 10 years (%)	–1.7	+0.6
Number of deaths per year (2011–2015)	6503	5314
Proportion of all cancer deaths (%)	20.1	18.2
Risk of dying from the disease before age 75 (%)	2.7	2.1
Age–standardized rate (W)	22.9	16.7
– Estimated annual change latest 10 years (%)	–2.7	–0.5
Persons living with the diagnosis at the end of 2015 (prevalence)	15324	17912
Number of persons living with the diagnosis per 100 000	115	134



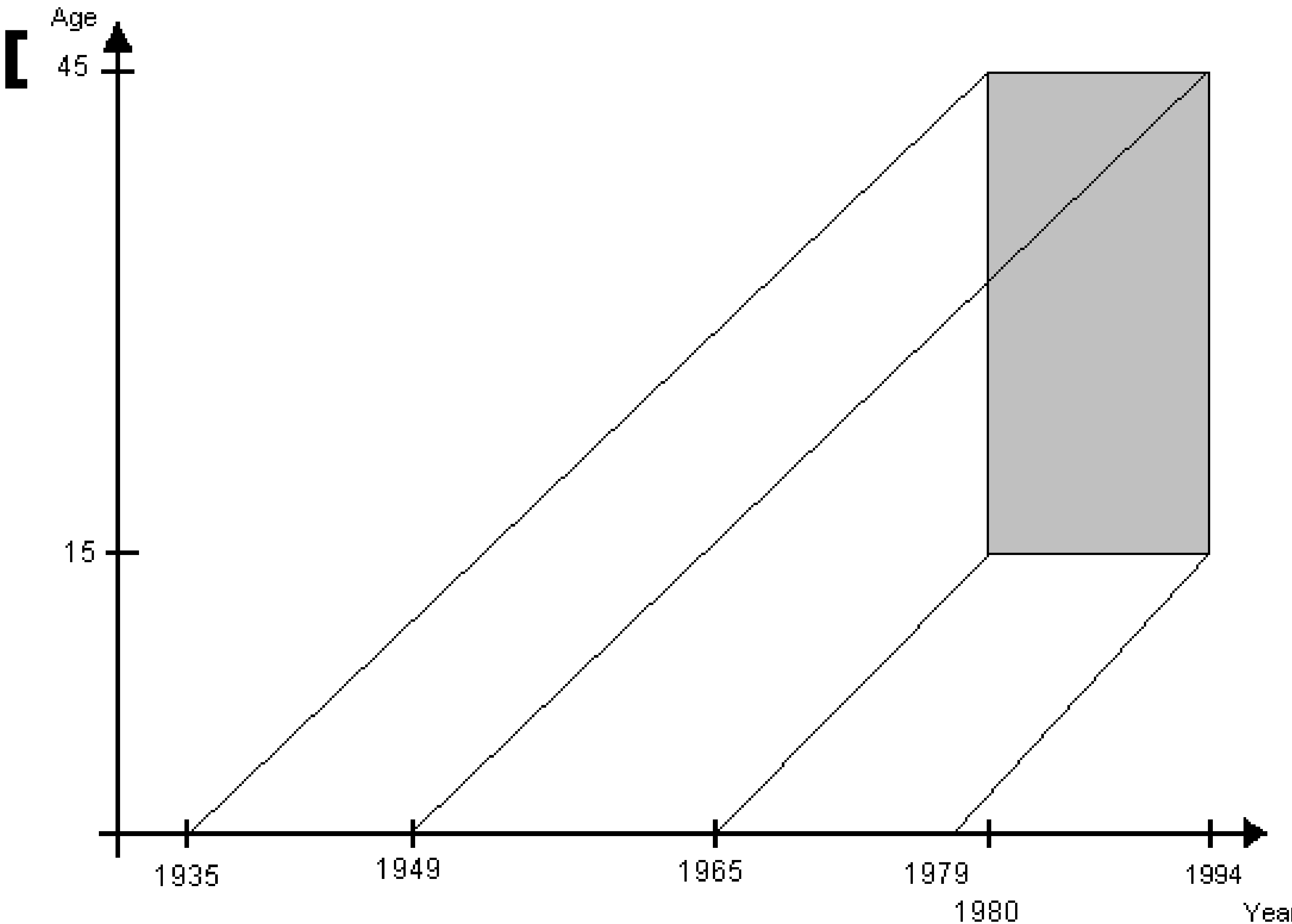


Age-Period-Cohort model

Class of models for demographic rates (mortality/ morbidity/ fertility/...) over a broad age range over a long time period

Lexis-diagram

A single person's life-trajectory is therefore a straight line with slope 1


Age
45
15
1935
1949
1965
1979
1980
1994
Year

APC-model

Describes the (log)rates as a sum of (non-linear) age- period- and cohort-effects

The three variables are related

$$a=p-c$$

Variables used to describe rates are linearly related

Model parametrized in different ways and still produce the same estimated rates

In popular terms you can say that it is possible to move a linear trend around between the three terms

The age-terms contains the linear effect of age

The period-terms contains the linear effect of period

The cohort effect contains the linear effect of cohort

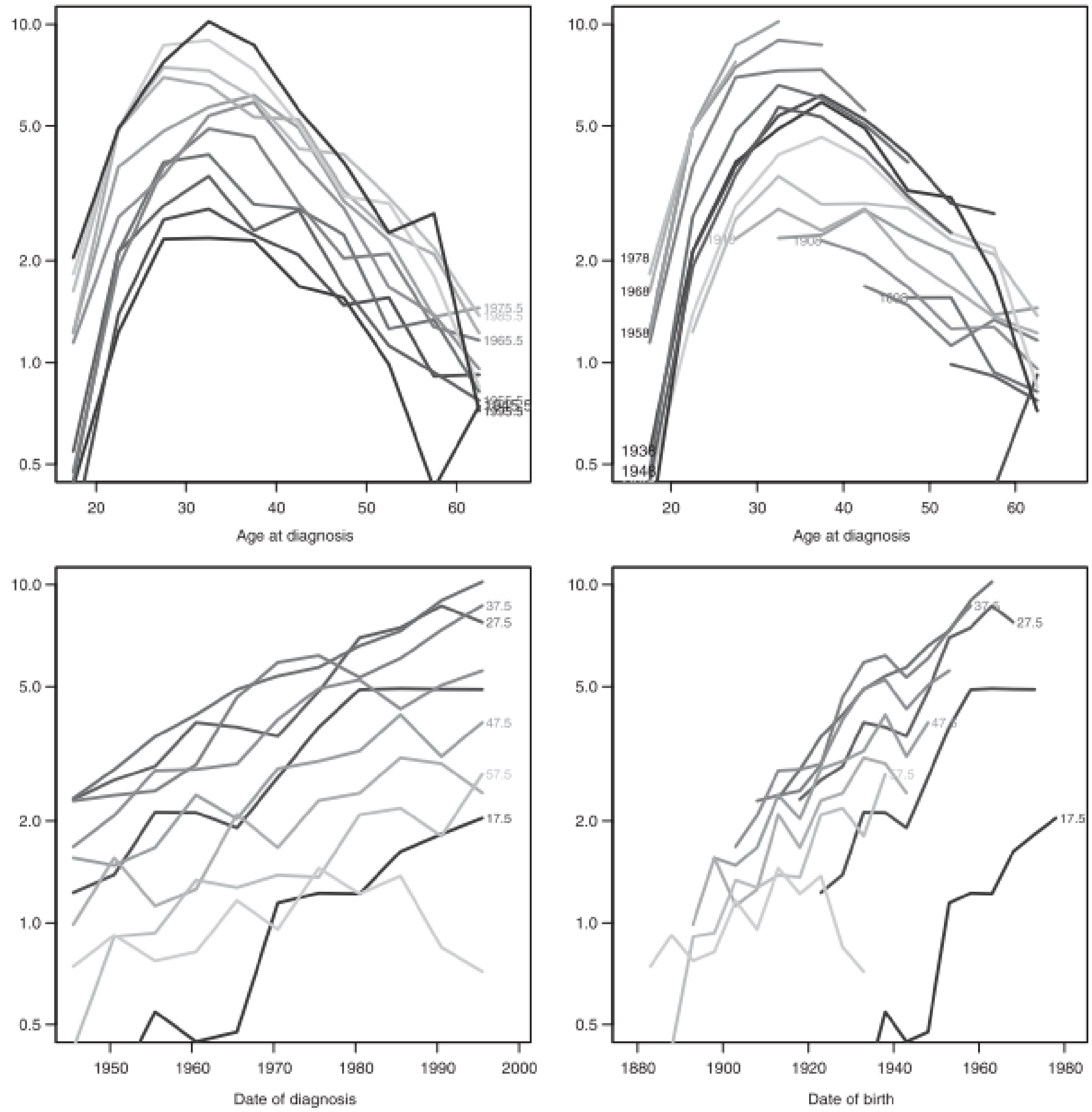
Drift parameter

Testis cancer

Denmark

1943-1997

15-64 yrs



Parametrization

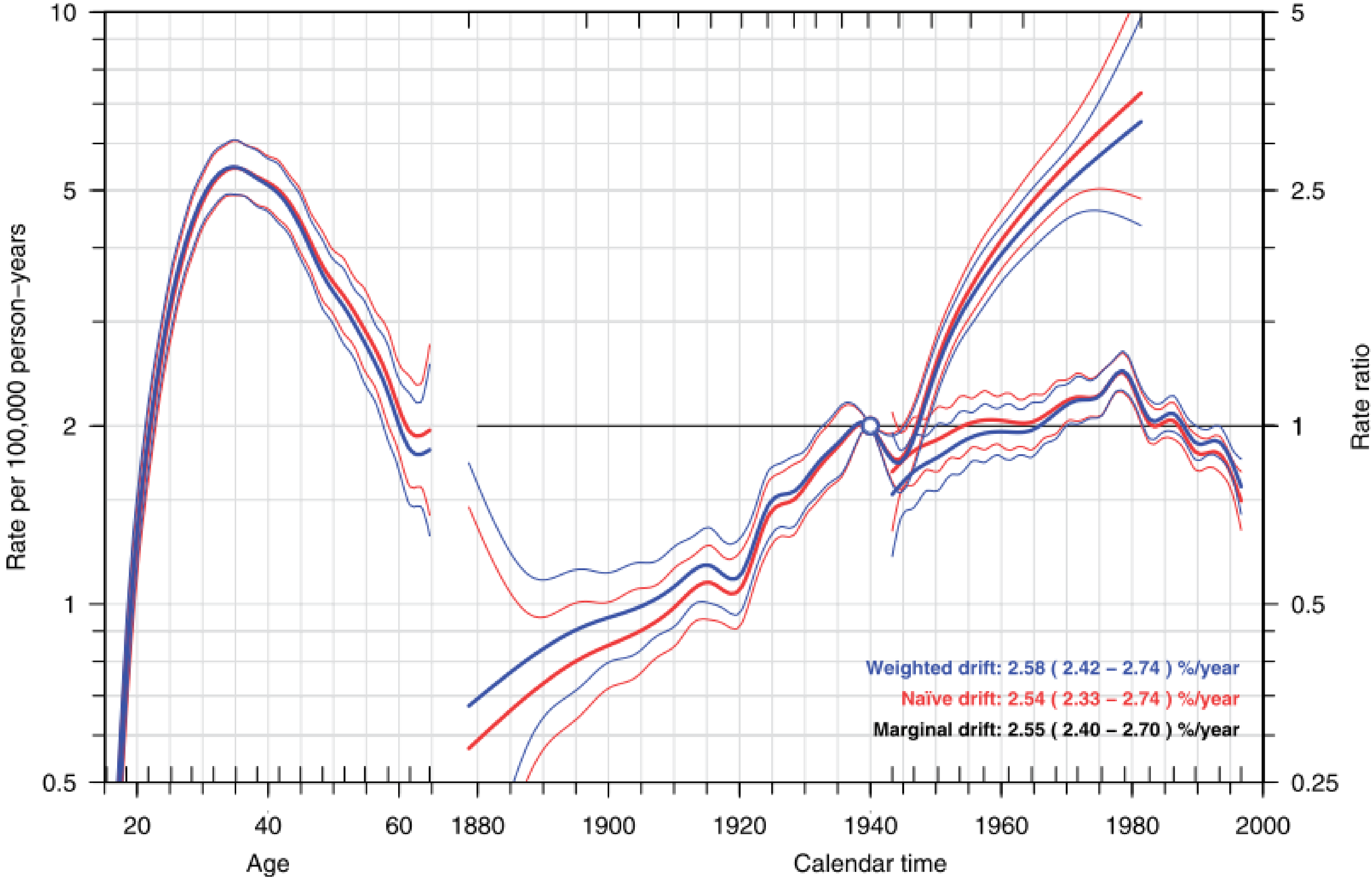
Normally one would choose a reference point for either period or cohort, and constrain the other to be 0 on average

Where should the drift (linear trend) be included?

Normally one would put this either with the cohort or the period effect, leaving the other one to have 0 slope on average

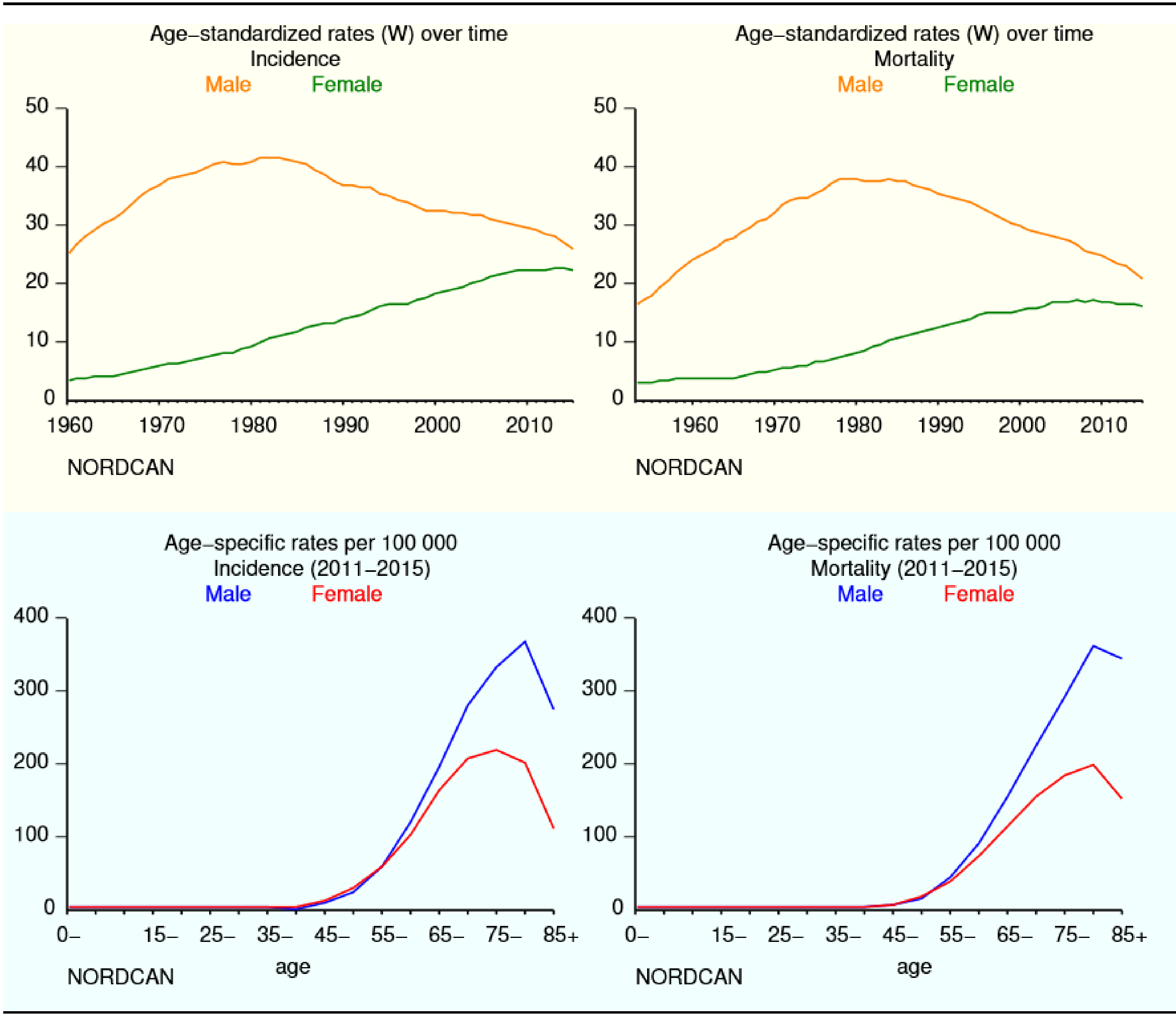
E.g. :

- The estimated age-specific rates in the 1940-cohort
- The cohort rate-ratio relative to the 1940-cohort
- The period rate-ratio taken as a residual RR (because it is constrained to be 0 on average with 0 slope)


10
5
2
1
0.5
Rate per 100,000 person−years
5
2.5
1
0.5
0.25
Rate ratio
20
40
60
Age
1880
1900
1920
1940
1960
1980
2000
Calendar time
Weighted drift: 2.58 (2.42 – 2.74) %/year
Naïve drift: 2.54 (2.33 – 2.74) %/year
Marginal drift: 2.55 (2.40 – 2.70) %/year

SDU

Lung cancer


Age–standardized rates (W) over time
Incidence
Male
Female
50
40
30
20
10
0
1960
1970
1980
1990
2000
2010
NORDCAN
Age–standardized rates (W) over time
Mortality
Male
Female
50
40
30
20
10
0
1960
1970
1980
1990
2000
2010
NORDCAN
Age–specific rates per 100 000
Incidence (2011–2015)
Male
Female
400
300
200
100
0
0–
15–
25–
35–
45–
55–
65–
75–
85+
age
NORDCAN
Age–specific rates per 100 000
Mortality (2011–2015)
Male
Female
400
300
200
100
0
0–
15–
25–
35–
45–
55–
65–
75–
85+
age
NORDCAN

Exercise 3

You are interested in describing lung cancer prevalence and incidence in your Nordic country from 1980 to 2019

Now focus on lung cancer incidence

Please consider the following elements:

- Age, period or cohort – what is of interest?
- If you observe period effect – how would you interpret that?
- If you observe cohort effect – interpretation?

Predictions
Projections
Forecasts

Forecasts are the basis for all forms of planning

Fundamental to social, economic or business planning

Human populations have two fundamental characteristics:

- Substantial overlap between current and future populations
- Every year we all get exactly one year older

Possible to predict future developments better than in other fields

Short and long-term

Long-term projections (25+ years): Planning of natural resources, provision of food, transportation and recreational facilities, etc.

Middle-range projections (10–25 years): Planning educational and medical facilities and services, housing needs, etc.

Definition - population projection

A computational procedure to calculate population size and structure at one time from population size and structure at another, together with a specification of how change takes place during the interim period

Two strategies

Total methods

Calculates size of total population using mathematical model

Distribute this total into sub-groups in ratio (current or extrapolated structure)

Ratio method of projection

Cohort component methods

Project each age group, sex and other categories separately

Aggregate to obtain total population

Cohort emphasis – people born at the same time go through life together

The size of a cohort at one age (and date) is strongly predictive of its size at other ages (and dates)

Total method of projection

Fitting a mathematical model to data on past trends in the size of the population

Use the fitted model to extrapolate forward

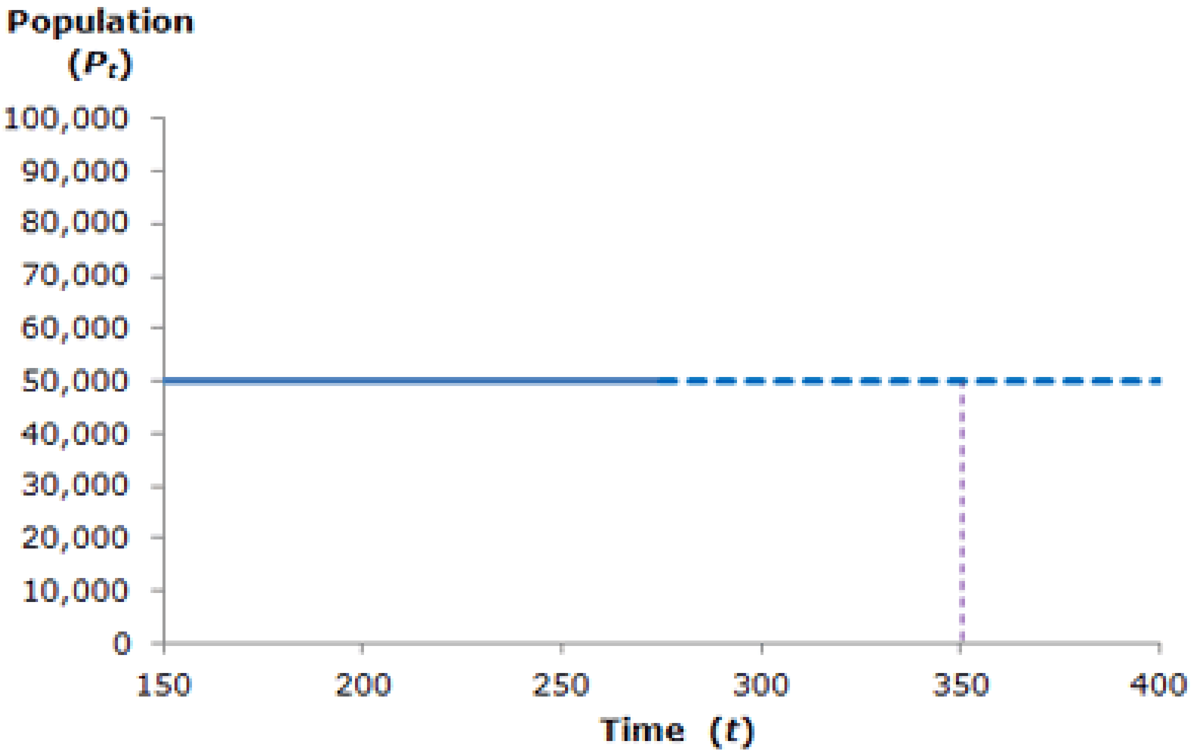
The main steps involve:

- Select an appropriate model of the growth process
- Estimate the parameters of the model from past estimates of the population
- Extrapolate the fitted curves and read off the projected population

Zero population growth

Simplest model

Size of population unchanging



Arithmetic growth

Linear growth

Assumes that a constant numeric change occurs in the size of the population in every period of the same length

Minimum two estimates of the population:

$P(t+n) = P(t) + a \times n$

Exponential growth

Exponential model assumes that the population is growing at a constant rate

Appropriate for expanding communities unaffected by constraints

Shrinking over time - exponential decay

Estimate growth rate

$P(t+n) = P(t) \times e^{rn}$

r is the constant annual growth rate

Fordoblingstiden vil aldrig ændre sig

Logistic growth

Assumes that the growth rate slows over time, eventually dropping to zero

Model assumes that p

In the logistic growth n

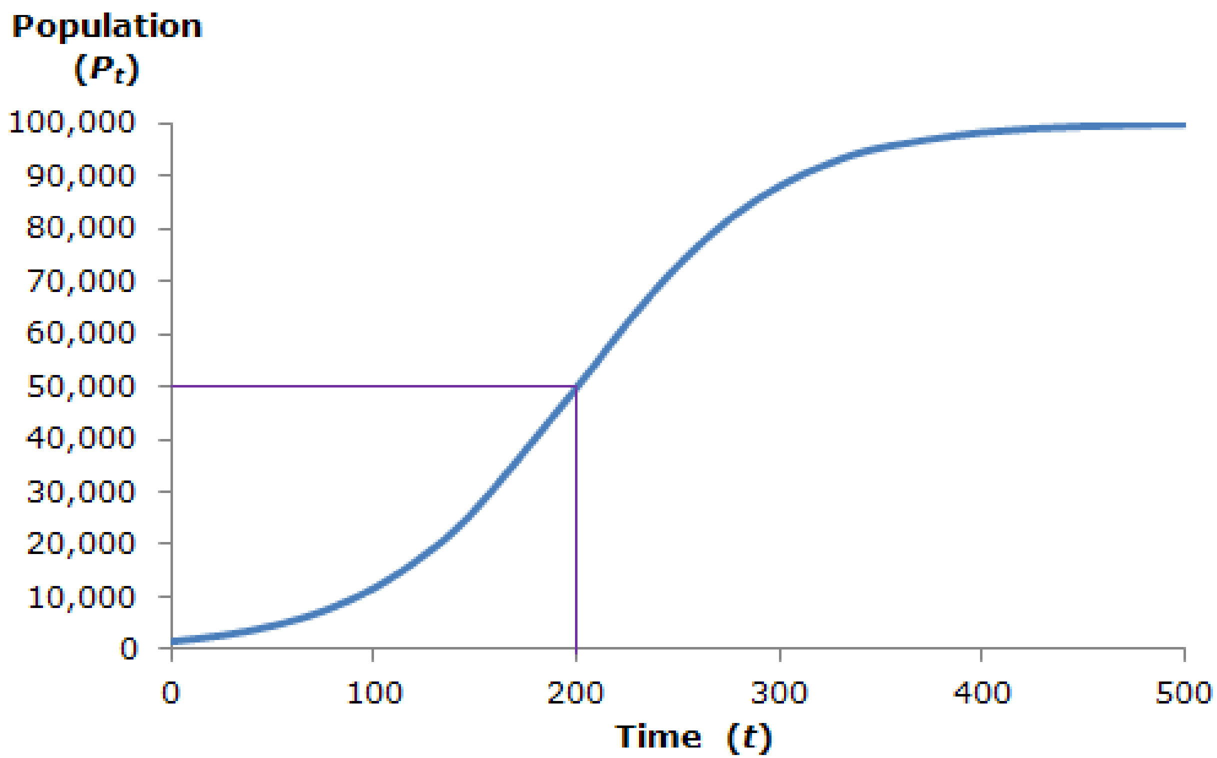
$P(t) = P(\infty) / (1 + e^{-s(t-h}$

P(∞) the final size of th

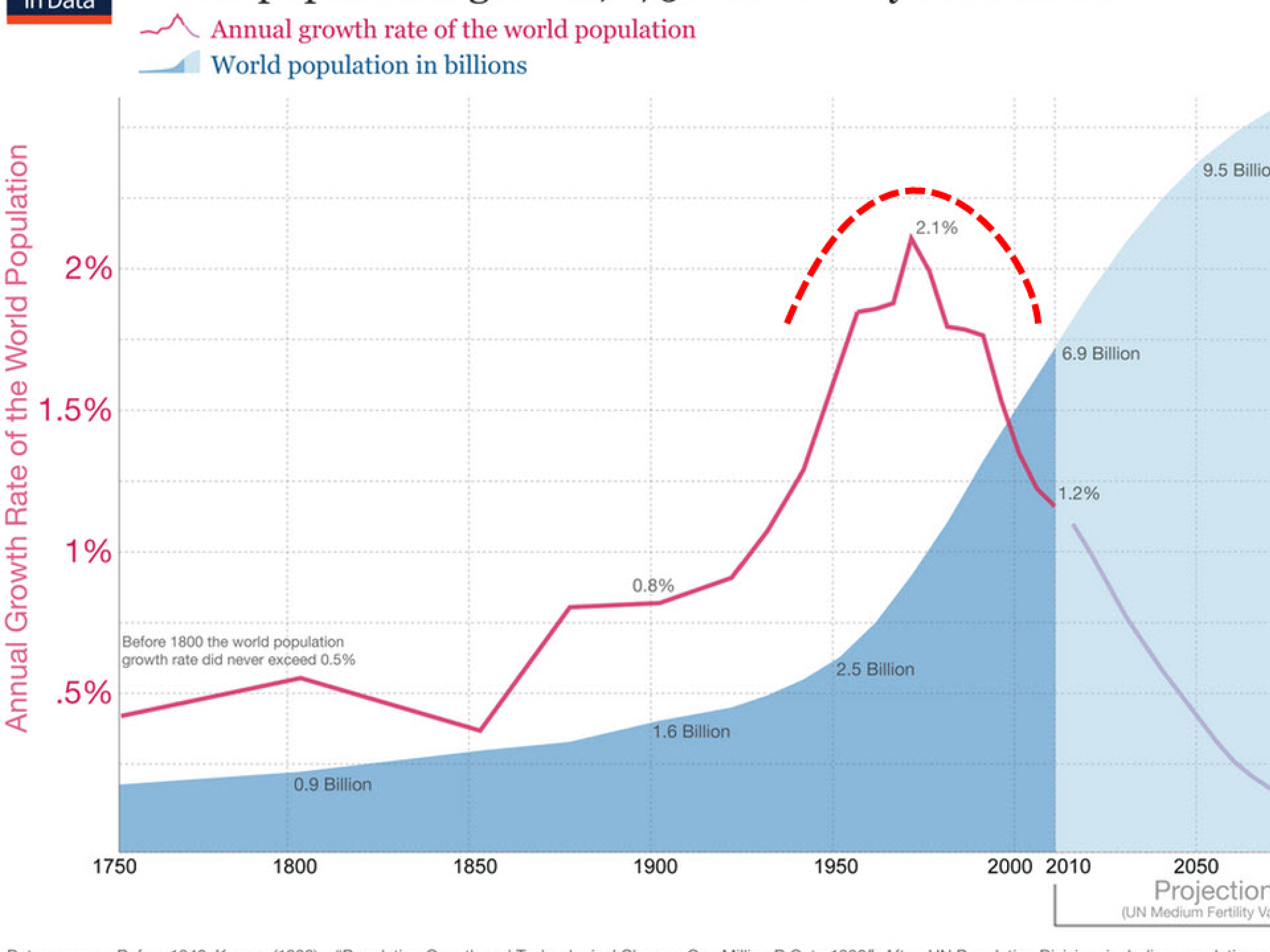
Time is measured rela the population reaches

s growth rate

$r = s(1 - P(t) / P(\infty))$



h



Annual growth rate of the world population
World population in billions
Annual Growth Rate of the World Population
2%
1.5%
1%
.5%
2.1%
9.5 Billio
6.9 Billion
1.2%
0.8%
Before 1800 the world population growth rate did never exceed 0.5%
2.5 Billion
1.6 Billion
0.9 Billion
1750
1800
1850
1900
1950
2000
2010
2050
Projection

Cohort-component population projections

Model age-sex structure of populations and not just the size

Model components of demographic change - fertility, mortality, and migration – and not just population growth

Demographic balancing equation:

$P(t+n) = P(t) + B(t) - D(t) + I(t) - E(t)$

Only two ways of joining a population

- Born into it
- Migrate into it

And two ways of leaving:

- Die
- Migrate out of

Cohort-component population projections

The steps of a cohort-component projection are:

- to project every age cohort for one projection interval at a time
- to calculate the births during this interval and add in the newly-born children
- to adjust for migration
- before moving on to repeat the procedure to project the population to the end of the next interval

To project a population in intervals of *n* years, one uses data on *n*-year age groups

SDU

Data required

Detailed assumptions

- Base year *population* subdivided by age and sex
- Sex-specific life tables for each projection interval in the projection period (*mortality*)
- Age-specific *fertility* rates for each projection interval in the projection period
- Age- and sex-specific net *migration* for each interval in the projection period (unless assuming closed population)

Exercise 4

You are interested in describing lung cancer prevalence and incidence in your Nordic country from 1980 to 2019

Now focus on lung cancer incidence

Please consider the following elements:

- How would you predict the future lung cancer incidence in your country?

Cancer prediction

Use population projections from statistical bureaus

Add cancer incidence rates for specific cancers

Include age-period-cohort models

Moller, B., et al. (2003). "Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches." Stat Med **22**(17): 2751-2766

SDU