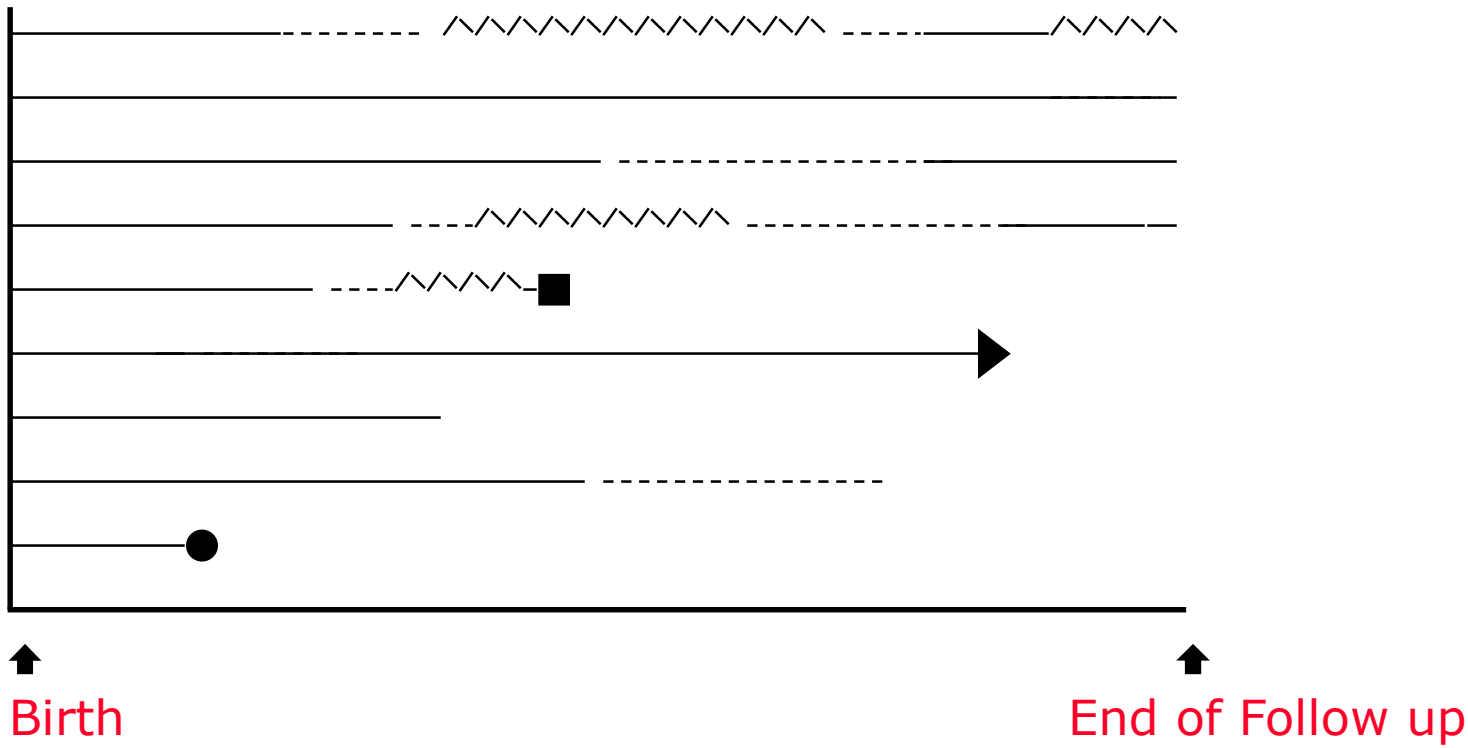
Study Design Options

Neil Pearce
Department of Medical Statistics
London School of Hygiene and
Tropical Medicine



■ Death

▶ other death

● lost to follow up

— "non-diseased"

--- symptoms

/\/\ severe disease

Study Design Options

- All epidemiological studies are (or should be) based on a particular population (the ***source population***) followed over a particular period of time (the ***risk period***)
- The different study design options differ only in how the source population is defined and how information is drawn from this population and time period

EDUCATION CORNER

Classification of epidemiological study designs

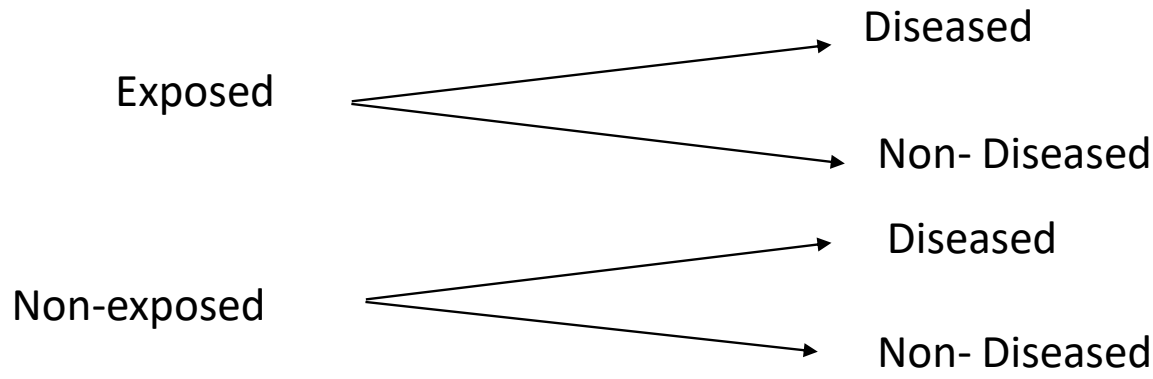
Neil Pearce[1,2]

I will argue that when the individual is the unit of analysis and the disease outcome under study is dichotomous, then epidemiological study designs can best be classified according to two criteria:

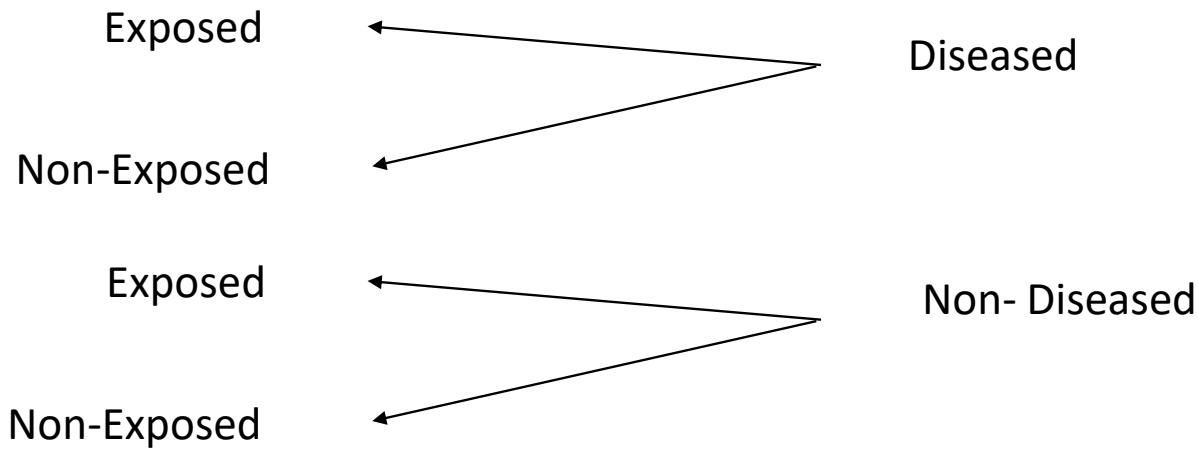
(i) the type of outcome under study (incidence or prevalence) and

(ii) whether there is sampling on the basis of the outcome.

Prospective (Cohort) studies



Case-control studies



Incidence and Prevalence

- ***Incidence*** is the number of new cases of the condition over a specified period of time
- ***Prevalence*** is the number of cases of the condition at a particular point in time

Incidence Studies

- Cohort studies, follow-up studies, longitudinal studies, prospective studies
- Uses all the information on the source population and risk period, i.e. collect information on exposure and outcome for everyone in the study
- Compares incidence in exposed and non-exposed

Incidence Studies

- Taussig et al (AJE 1989). Cohort study of 1246 infants born in Tucson during May 1980-October 1984
- ***Exposures:*** Questionnaire on demographic factors, parental history. Cord blood
- ***Outcomes:*** Wheezing lower respiratory tract illness (3 years), atopy (6 years)

A Hypothetical Incidence Study: risks

	Exposed	Non-exposed	Ratio
Cases	**1,813**	**952**	
Non-cases	**8,187**	**9,048**	
Total	**10,000**	**10,000**	
Incidence proportion	**0.1813**	**0.0952**	**1.90**

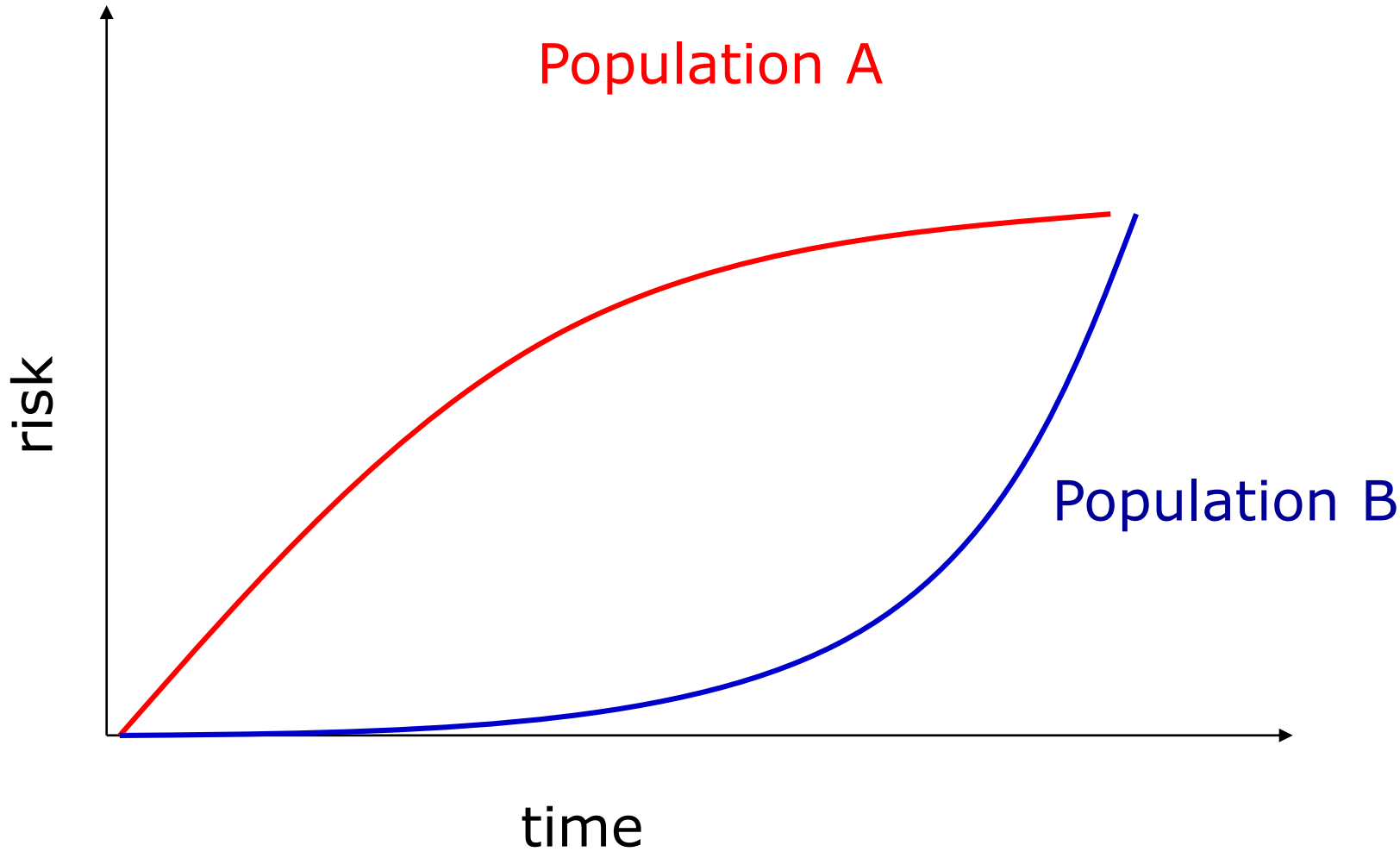
World Death Rate Holding Steady At 100 Percent

January 22, 1997 | Issue 31•02

GENEVA, SWITZERLAND—World Health Organization officials expressed disappointment Monday at the group's finding that, despite the enormous efforts of doctors, rescue workers and other medical professionals worldwide, the global death rate remains constant at 100 percent.

Death rates since 1992

Death has long been considered humanity's number one health concern. Responsible for 100 percent of all recorded fatalities worldwide, the condition has no cure.


Population A
risk
Population B
time

Incidence Proportion (Risk)

- Proportion of study participants who experience the outcome (for the first time)
- When there are significant “losses to follow-up” the incidence proportion cannot be estimated directly

A Hypothetical Incidence Study: rates

	Exposed	Non-exposed	Ratio
Cases	**1,813**	**952**	
Non-cases	**8,187**	**9,048**	
Total	**10,000**	**10,000**	
Person-years	**90,635**	**95,163**	
Incidence rate	**0.0200**	**0.0100**	**2.00**

Incidence Rate

- Number of new cases per unit time (e.g. per 100,000 person-years)

A Hypothetical Incidence Study: odds

	Exposed	Non-exposed	Ratio
Cases	**1,813**	**952**	
Non-cases	**8,187**	**9,048**	
Total	**10,000**	**10,000**	
Incidence odds	**0.2214**	**0.1052**	**2.11**

Incidence Odds

- Ratio of number of people who experience the outcome to the number of people who do not experience the outcome
- Has unfortunate statistical properties (particularly non-collapsibility) which mean that it should not be used as an outcome measure in a cohort study

Effect Measures in Incidence Studies

- Risk ratio
- Rate ratio
- Odds ratio

A Hypothetical Incidence Study

	Exposed	Non-exposed	Ratio
Cases	**1,813**	**952**	
Non-cases	**8,187**	**9,048**	
Total	**10,000**	**10,000**	
Person-years	**90,635**	**95,163**	
Incidence rate	**0.0200**	**0.0100**	**2.00**
Incidence proportion (risk)	**0.1813**	**0.0952**	**1.90**
Incidence odds	**0.2214**	**0.1052**	**2.11**

A Hypothetical Incidence Study: odds

	Exposed	Non-exposed	Total
Cases	**1,813**	**952**	**2,765**
Non-cases	**8,187**	**9,048**	**17,235**
Total	**10,000**	**10,000**	**20,000**
Incidence odds	**0.2214**	**0.1052**	

A Hypothetical Case-Control Study

$$\text{Odds ratio} = \frac{1813/8187}{952/9048} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

$$= \frac{1813/952}{8187/9048} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

Odds Ratio

- We can therefore estimate the odds ratio by taking all of the ***cases*** and a ***control sample*** of the non-cases

A Hypothetical Case-Control Study

	Exposed	Non-exposed	Odds
Cases	**1,813**	**952**	**1,813/952**
Controls	**1,313**	**1,452**	**1,313/1,452**
Odds	**1,813/1,313**	**952/1,452**	
Odds ratio			**2.11**

Odds Ratio

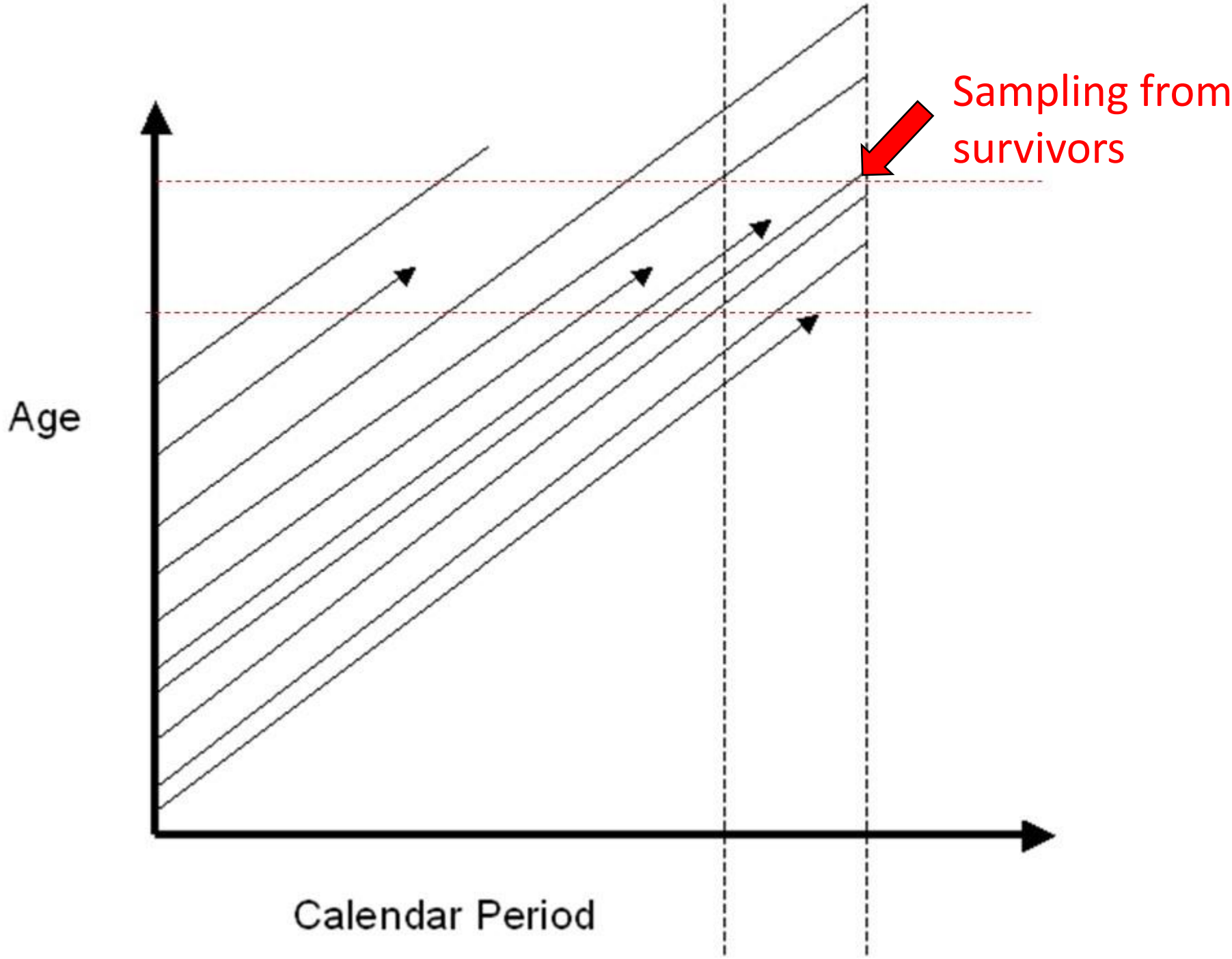
- OR=(1813/1313)/(952/1452) = 2.11
- This ***incidence case-control study*** yields the same estimate as would have been obtained by an incidence study but with a much smaller number of participants because we include ***all*** of the cases but only a ***sample*** of the non-cases

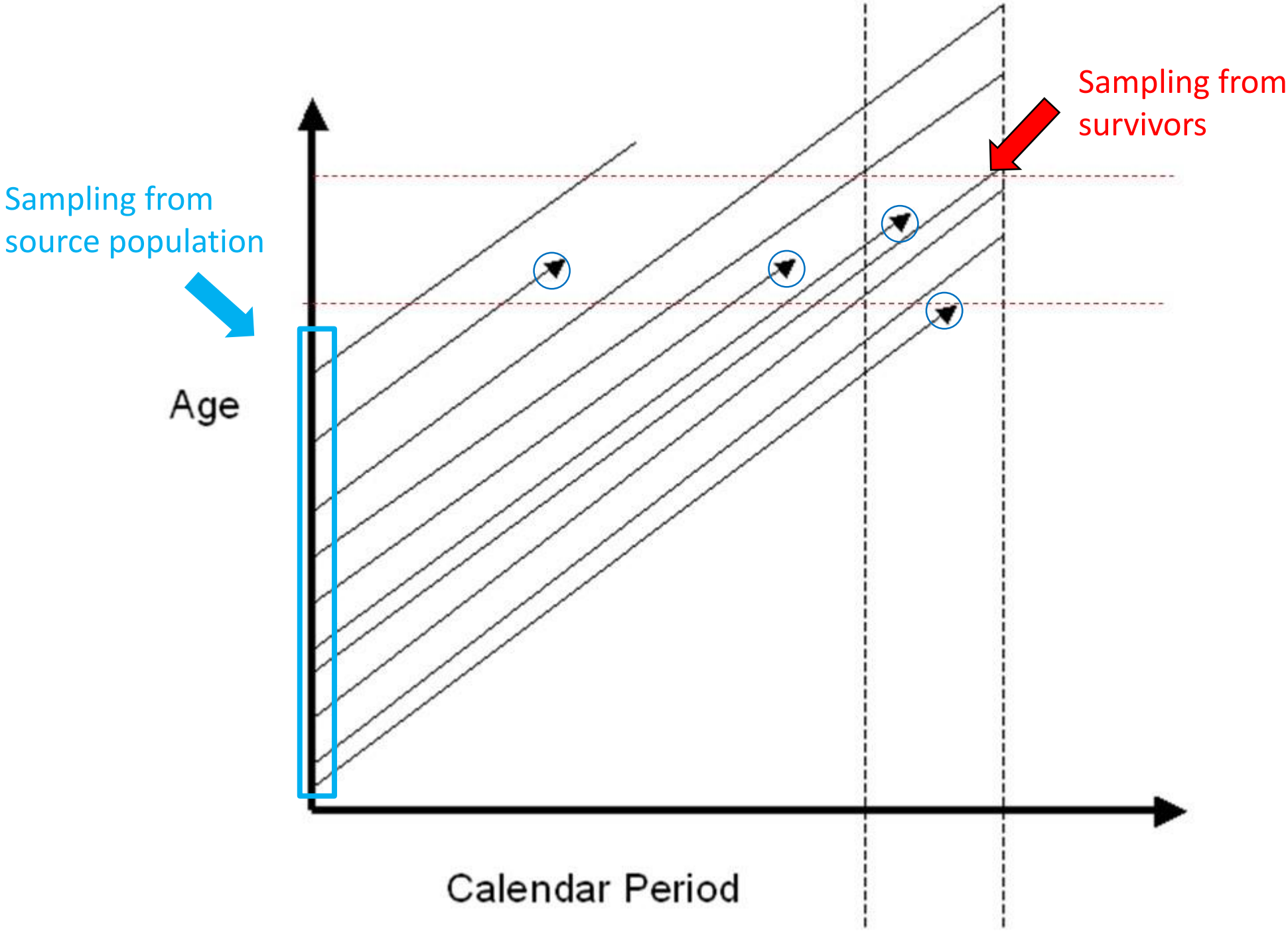
Control Sampling Strategies

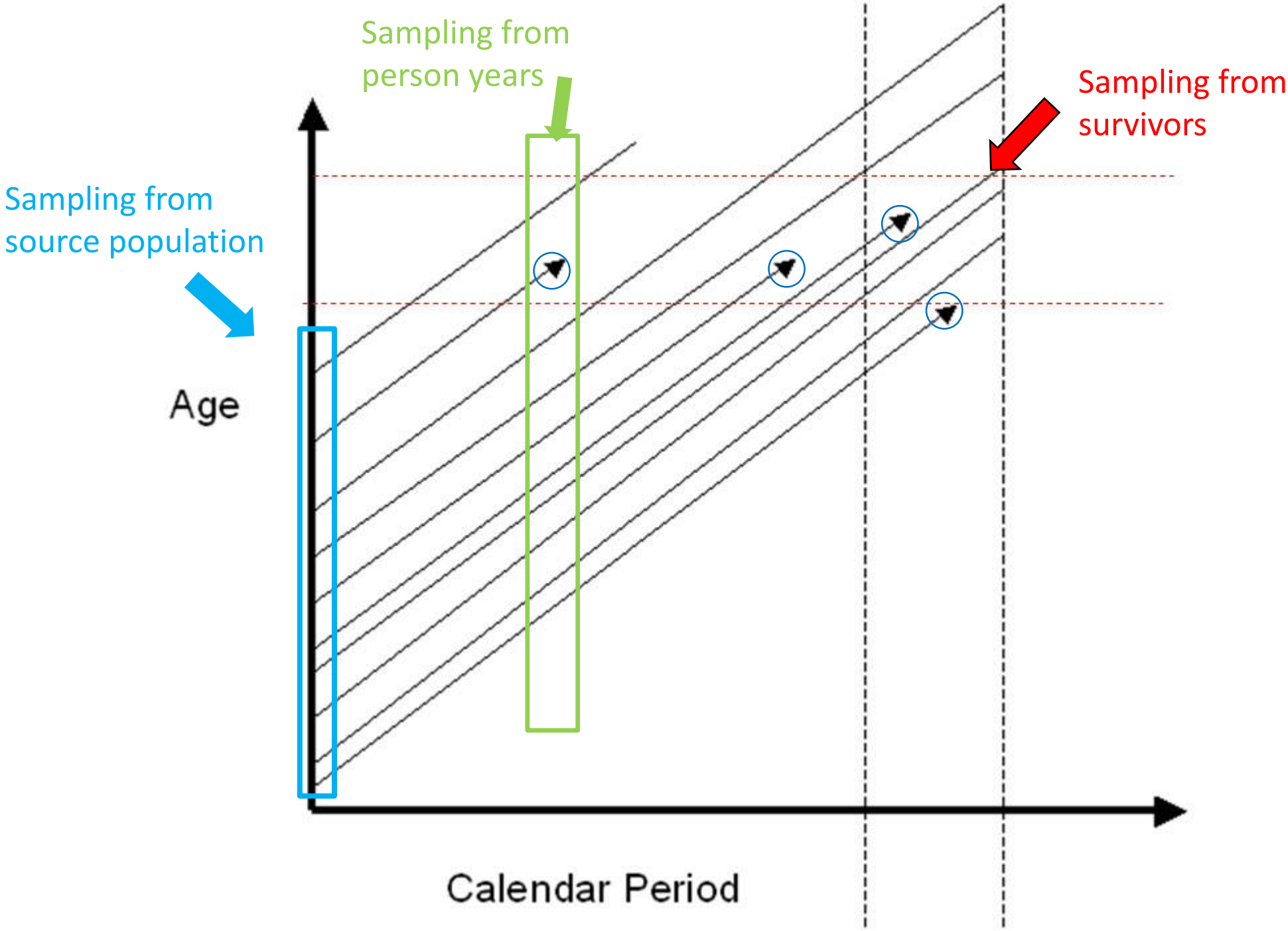
- Cumulative sampling (exclusive sampling)
  - Cornfield (1951)
- Case-cohort sampling (inclusive sampling)
  - Thomas (1972)
  - Kupper et al (1975)
  - Miettinen (1982)
  - Smith et al (1984)
  - Prentice (1986)
- Density sampling (concurrent sampling)
  - Sheehe (1962)
  - Miettinen (1976)

Methods of Sampling Controls

- From ***survivors*** (non-cases at end of follow-up) = cumulative sampling
- From ***source population*** = case-cohort sampling
- From ***person-years*** = density sampling


Sampling from survivors
Age
Calendar Period


Sampling from survivors
Sampling from source population
Age
Calendar Period


Sampling from person years
Sampling from survivors
Sampling from source population
Age
Calendar Period

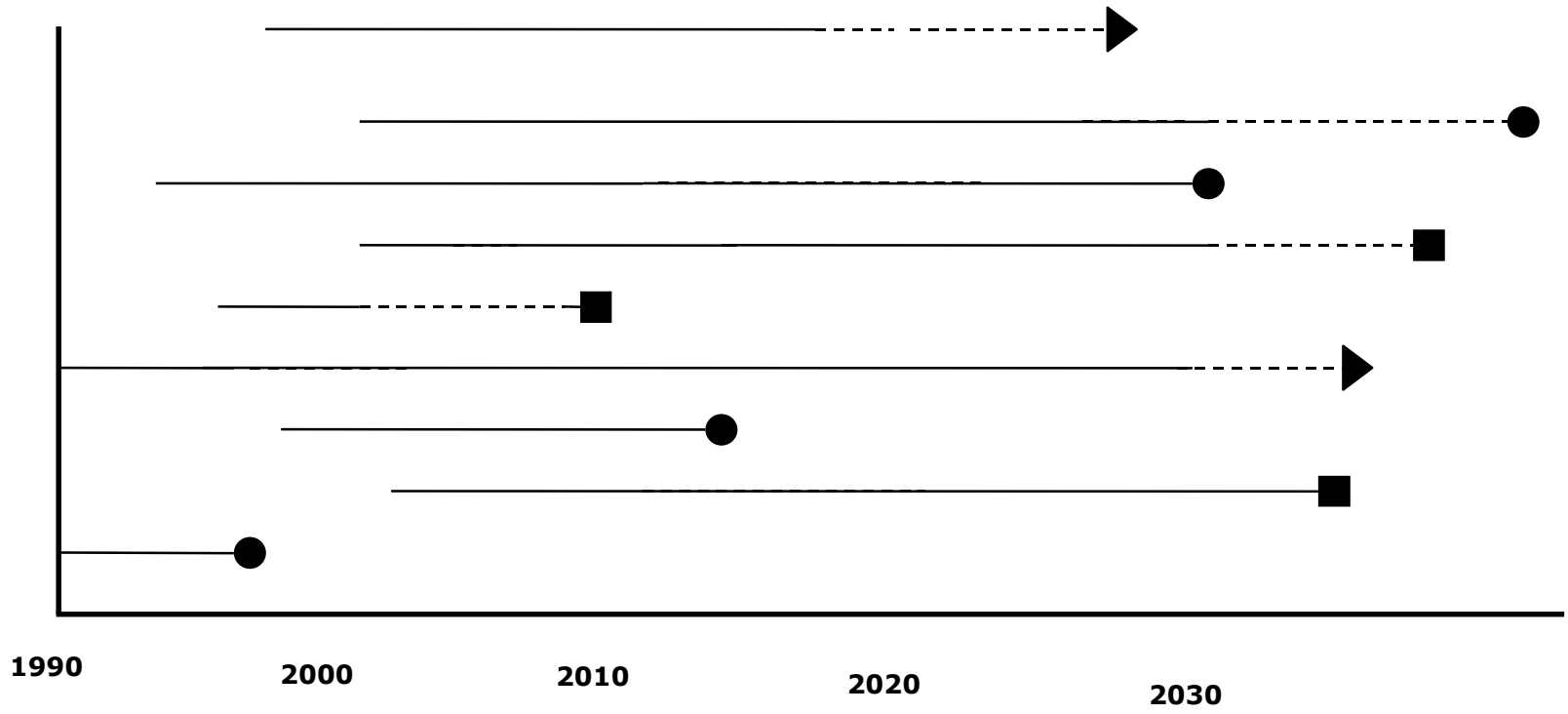
Methods of Sampling Controls

	Exposed	Non-exposed	Odds ratio
Cases	**1,813**	**952**	
Controls			
from survivors	**1,313**	**1,452**	**2.11**
from source population	**1,383**	**1,383**	**1.90**
from person-year	**1,349**	**1,416**	**2.00**

Sampling options in case-control studies

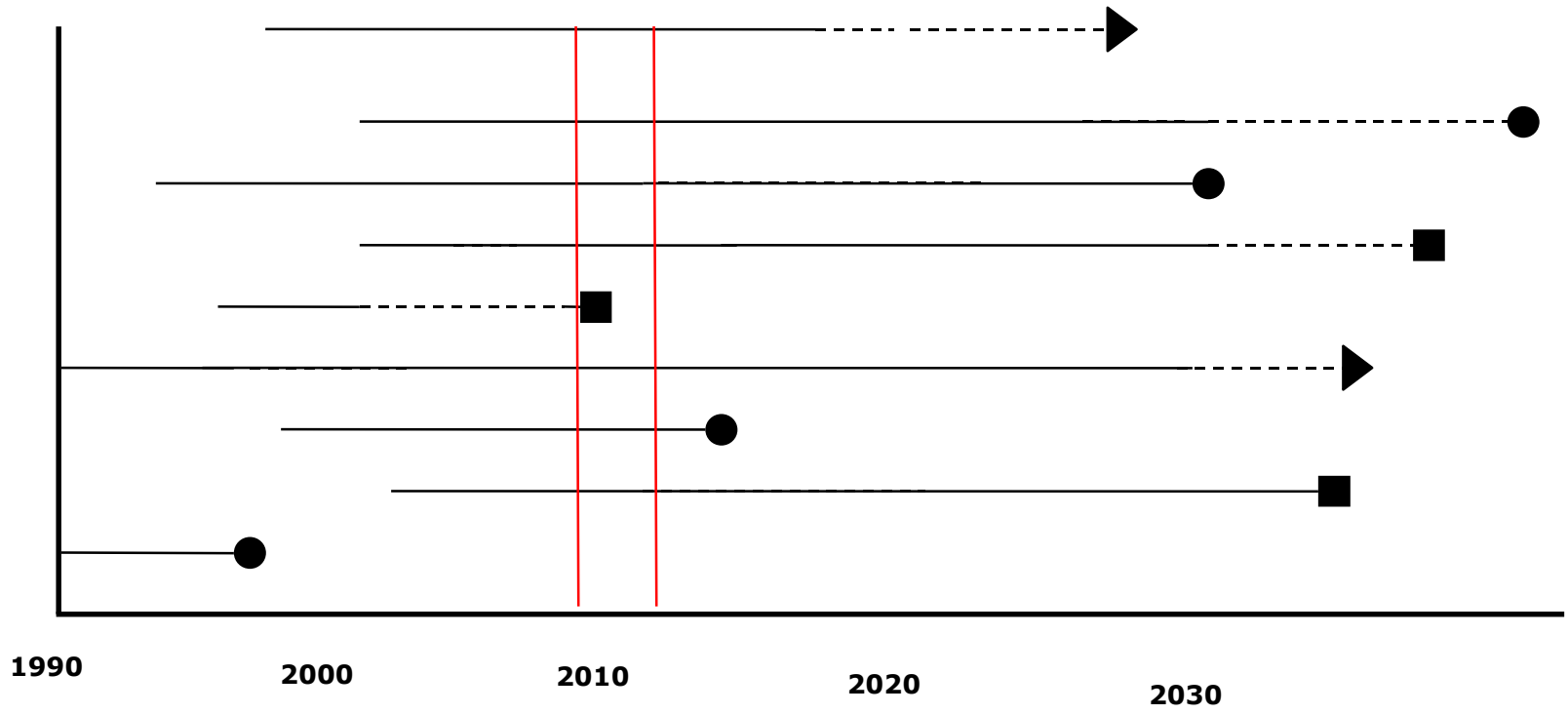
		What does the case-control odds ratio estimate?
Cumulative sampling (Exclusive)	Cornfield (1951)	Odds ratio
Case-cohort sampling (Inclusive)	Thomas (1972) Kupper et al (1975) Miettinen (1982) Smith et al (1984) Prentice (1986)	Risk ratio
Density sampling (Concurrent)	Sheehe (1962) Miettinen (1976)	Rate ratio

A hypothetical case-control study of lung cancer in London



- ■ cancer death
- ▶ other death
- ● lost to follow up

A hypothetical case-control study of lung cancer in London



- ■ cancer death
- ▶ other death
- ● lost to follow up

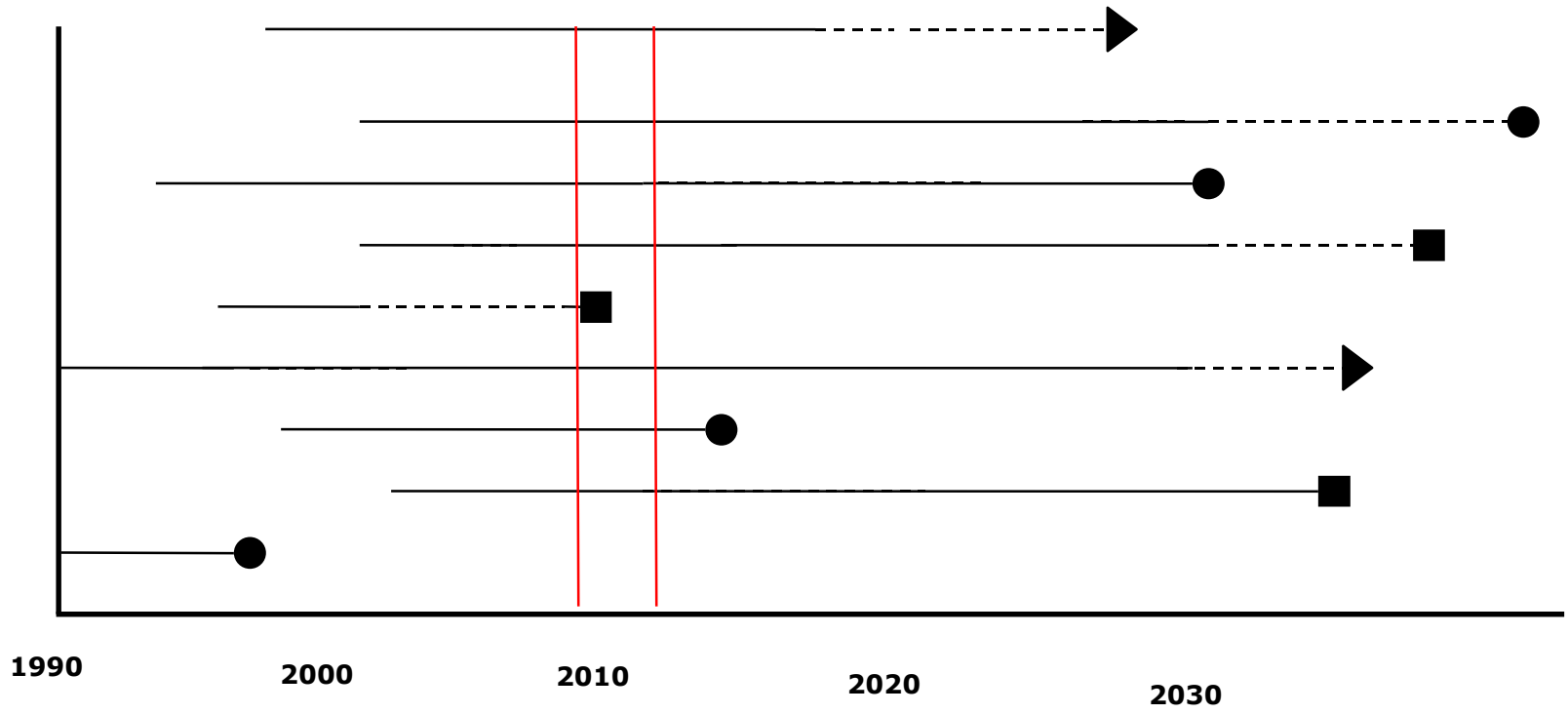
Misconceptions About Case-Control Studies

- Proceeds from effect (disease) to cause (exposure), i.e. reverse causality
- Inherently more prone to bias than cohort studies
- Odds ratio only approximately estimates the relative risk
- Depends on a “rare disease” assumption

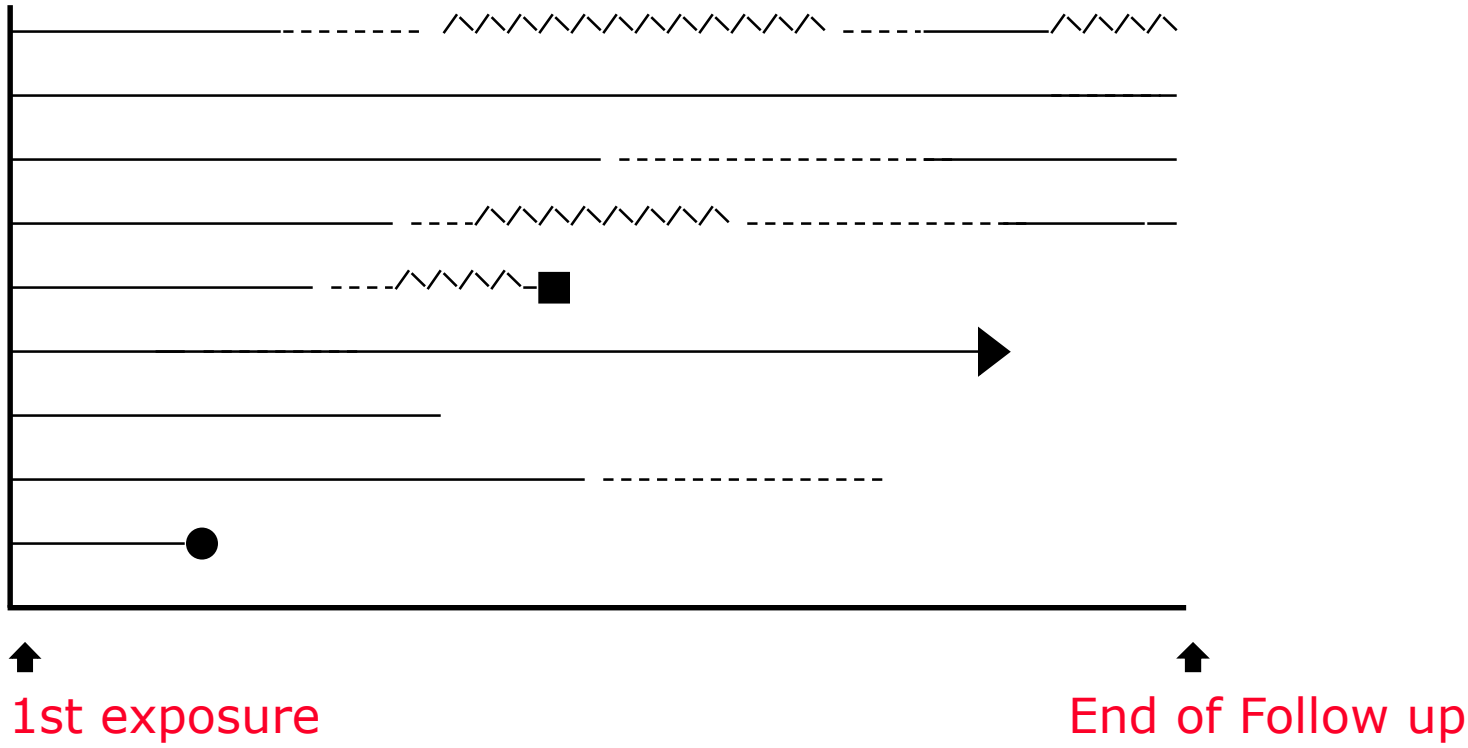
Dynamic populations involve ‘samples’ of what would be obtained from a fixed population

- In a fixed population we follow people from first exposure
- In a study in a dynamic population, we follow people (usually) starting some time after first exposure
- If there is effect modification by time since exposure, the two studies will not (automatically) yield the same findings

A hypothetical case-control study of lung cancer in London



- ■ cancer death
- ▶ other death
- ● lost to follow up



■ Death

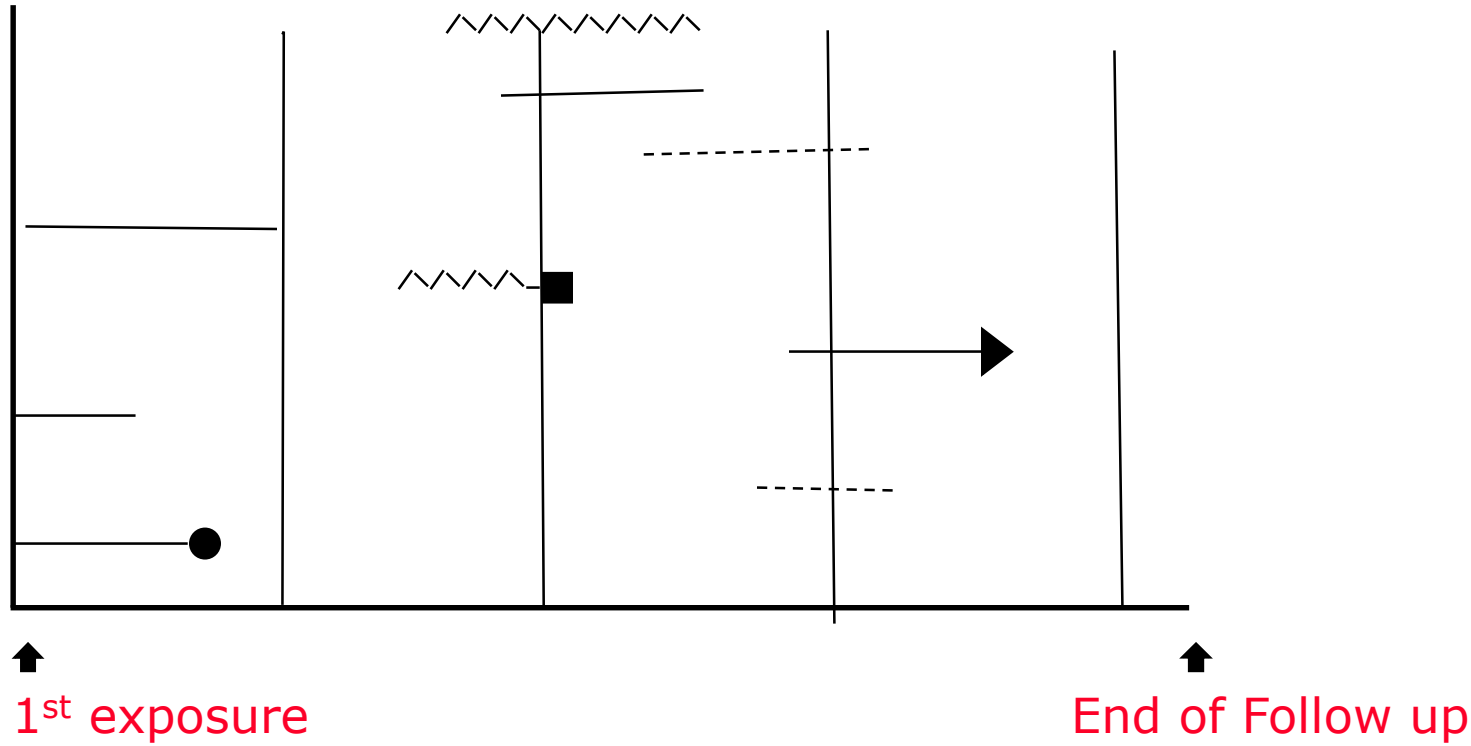
▶ other death

● lost to follow up

— "non-diseased"

--- symptoms

/\/\ severe disease



■ Death

▶ other death

● lost to follow up

— "non-diseased"

--- symptoms

/\/\ severe disease

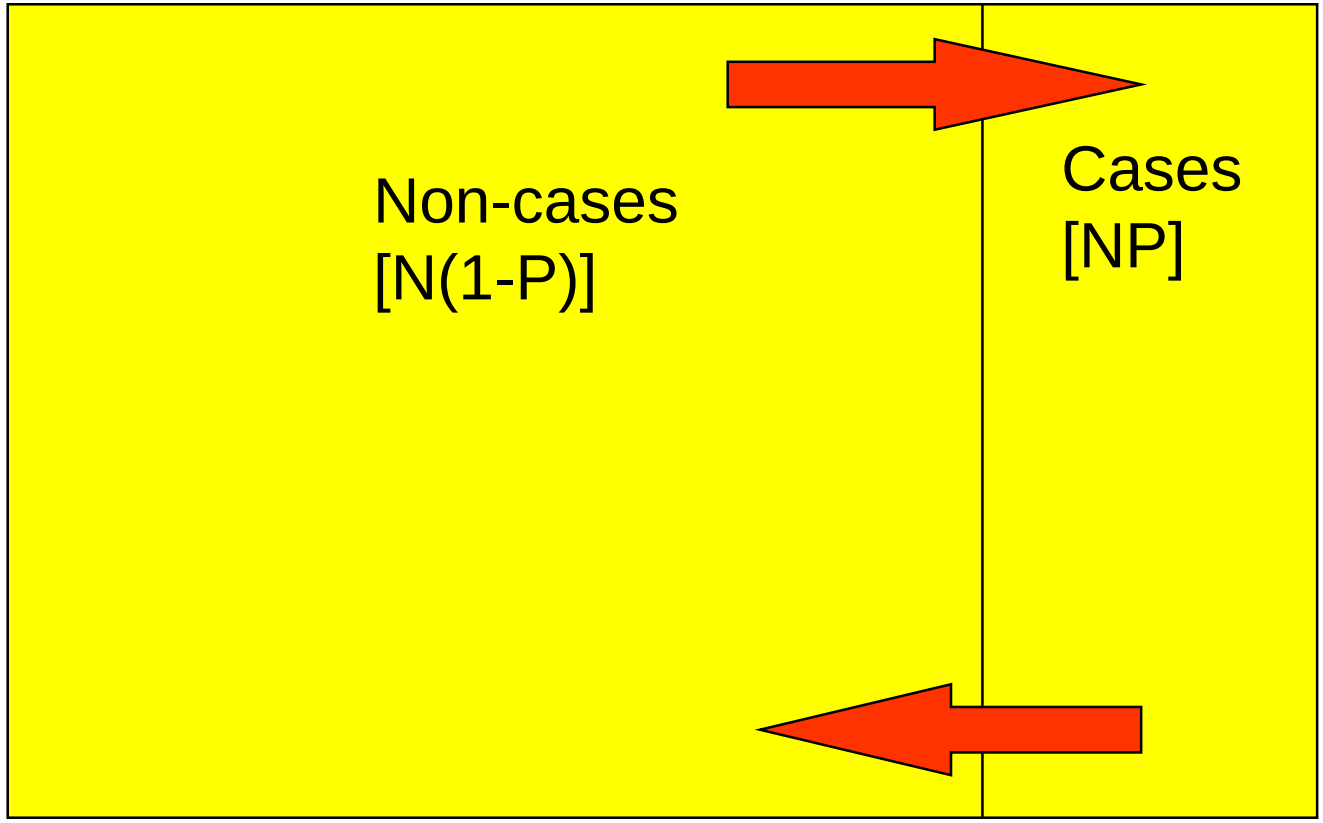
Example: HRT and CHD

- The Nurses Health (cohort) Study found a protective effect of HRT on CHD
- The Women's Health Initiative randomized trial found an increased risks
- Hernan et al (2008) showed that there were increased risks in the first few years after starting HRT and reduced risks subsequently
- When the observational study was analysed like an RCT, the two studies yielded very similar results

P=prevalence
I=incidence
D=duration
N=population

P/(1-P) = I x D

N(1-p) x I



NP/D

Prevalence Studies

Exposed group: $P_1/(1-P_1) = I_1D_1$

Non-exposed group: $P_0/(1-P_0) = I_0D_0$

$$\frac{P_1/(1-P_1)}{P_0/(1-P_0)} = \frac{I_1D_1}{I_0D_0}$$

Prevalence Studies

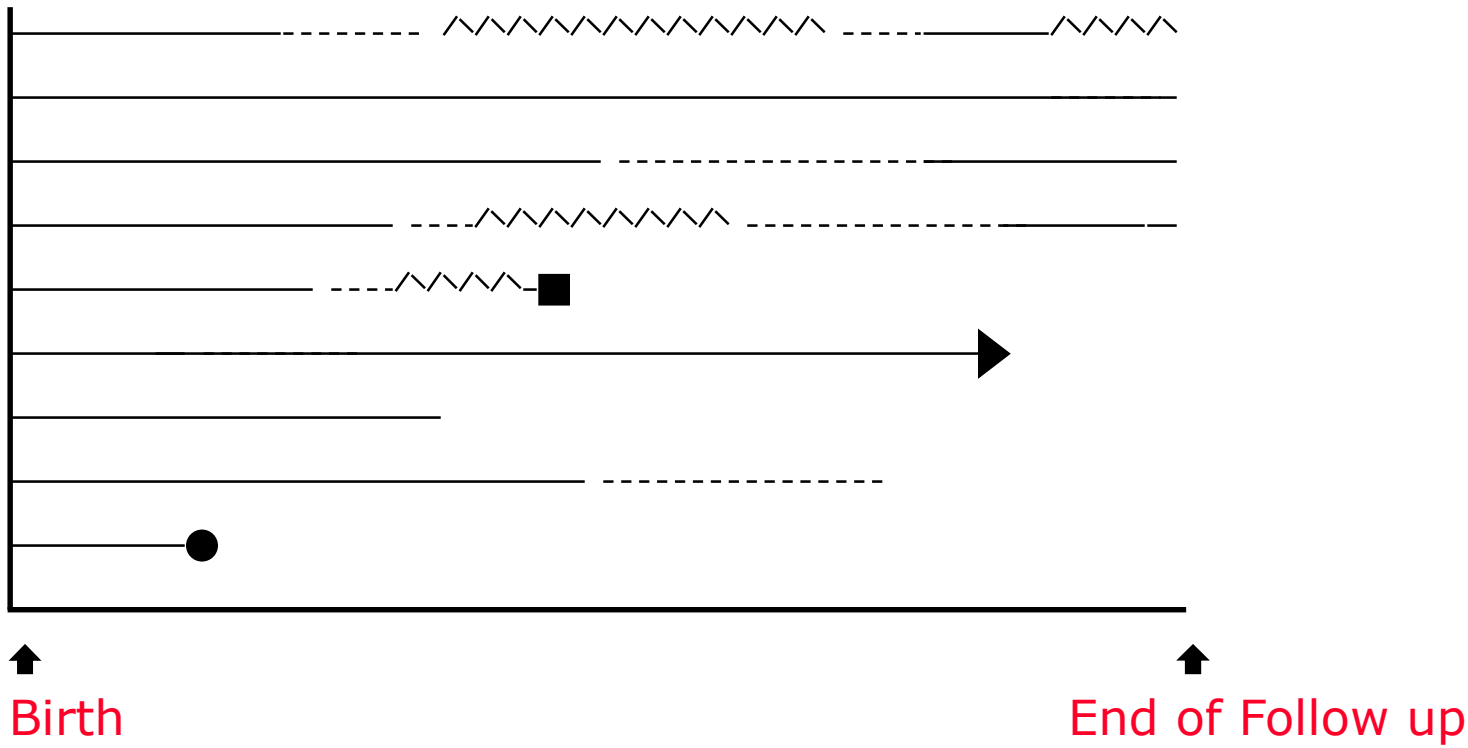
- Number (proportion) of people with disease at a particular point in time
- Under certain assumptions: P/(1-P)=ID
- $POR=I_1D_1/I_0D_0$
- Therefore differences in prevalence may be due to differences in incidence, differences in duration, or both

A Hypothetical Prevalence Study

	Exposed	Non-exposed	Ratio
Cases	**909**	**476**	
Non-cases	**9,091**	**9,524**	
Total	**10,000**	**10,000**	
Prevalence	**0.0909**	**0.0476**	**1.91**
Prevalence odds	**0.1000**	**0.0500**	**2.00**

Prevalence Studies

- Charpin et al (1988), asthma prevalence study in two French towns
- ***Exposure:*** Altitude - comparison of Marseille and Briancon
- ***Outcomes:*** asthma diagnosis (3.4% and 2.1%), asthma attacks (3.8% and 2.4%), house dust mite sensitivity (27.5% and 10.2%)



■ Death

▶ other death

● lost to follow up

— "non-diseased"

--- symptoms

/\/\ severe disease

A Hypothetical Prevalence Study

	Exposed	Non-exposed	Ratio
Cases	**909**	**476**	
Non-cases	**9,091**	**9,524**	
Total	**10,000**	**10,000**	
Prevalence	**0.0909**	**0.0476**	**1.91**
Prevalence odds	**0.1000**	**0.0500**	**2.00**

A Hypothetical Prevalence Case-Control Study

	Exposed	Non-exposed	Ratio
Cases	**909**	**476**	
Non-cases	**676**	**709**	
Total	**1,585**	**1,185**	
Prevalence odds	**1.345**	**0.671**	**2.00**

Prevalence Case-Control Studies

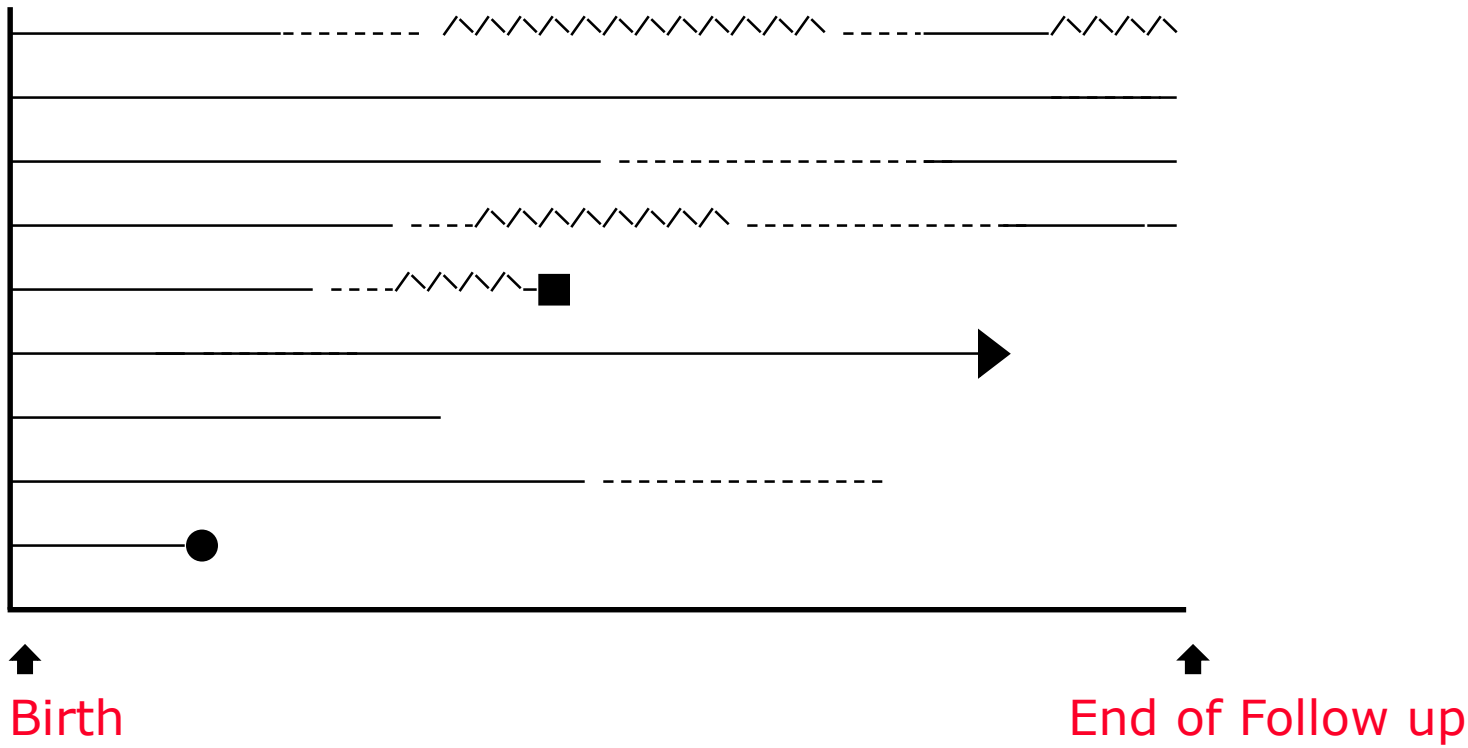
This ***prevalence case-control study*** yields the same estimate as would have been obtained by a prevalence study but with a much smaller number of participants because we include ***all*** of the prevalent cases but only a ***sample*** of the non-cases

Prevalence Case-Control Studies

- Oliveti et al (1996). Prevalence case-control study of asthma in inner city African-American children
- Cases: physician diagnosed asthma with recent symptoms
- Controls: non-asthmatics using the same hospital-based clinic
- Exposures: perinatal factors

Study Design Options

		Sampling on outcome: No	Sampling on outcome: Yes
Study outcome	Incidence	Incidence studies	Incidence case-control studies
	Prevalence	Prevalence studies	Prevalence case-control studies



■ Death

▶ other death

● lost to follow up

— "non-diseased"

--- symptoms

/\/\ severe disease

Continuous Outcome Measures

- Lung function in a cross-sectional study (a prevalence study is a cross-sectional study with a dichotomous outcome measure)
- Changes in lung function over time in a longitudinal study (an incidence study is a longitudinal study with a dichotomous outcome measure)

Continuous Outcome Measures

- Tager et al (1983), longitudinal study of pulmonary function in children aged 5-9 years, followed for 7 years
- ***Exposures:*** maternal smoking
- ***Outcomes:*** annual increase in FEV1 (this was 28mL lower in children exposed to maternal smoking)

Study Design Options

- Incidence studies
- Incidence case-control studies
- Prevalence studies
- Prevalence case-control studies
- Cross-sectional studies (with continuous outcome measure)
- Longitudinal studies (with continuous outcome measure)