What does the odds ratio estimate in a case-control study?

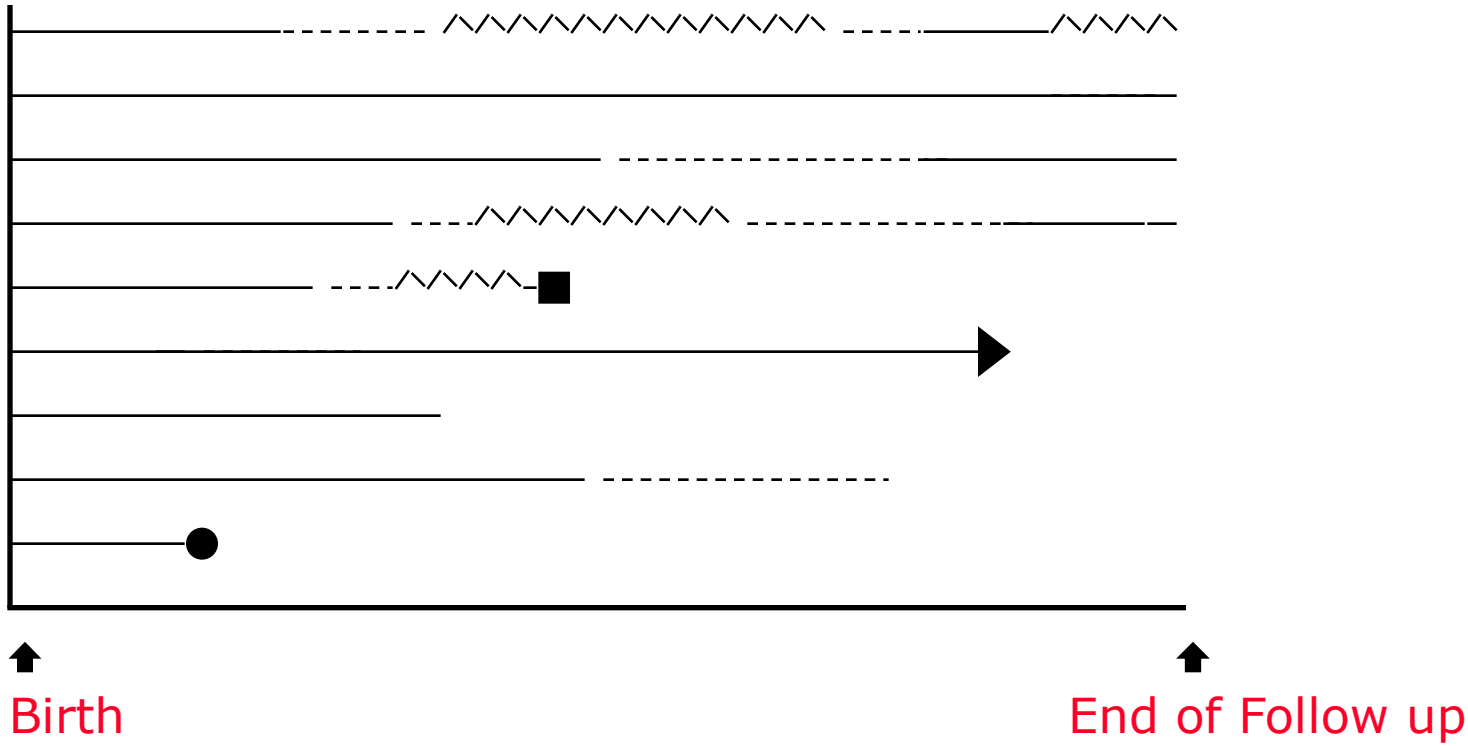
Neil Pearce
Department of Medical Statistics
London School of Hygiene and Tropical Medicine

How to design a case-control study

- Design the ideal randomised trial that you would really like to do
- Design the cohort study that corresponds to this randomised trial (except that exposure has not been randomised)
- Design the case-control study that corresponds to that cohort study

What does the odds ratio estimate in a case-control study?

- Cohort studies
- Selection of controls for case-control studies
- Implications



■ asthma death

▶ other death

● lost to follow up

— "non-asthmatic"

--- asthma symptoms

/\/\ severe asthma

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

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

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

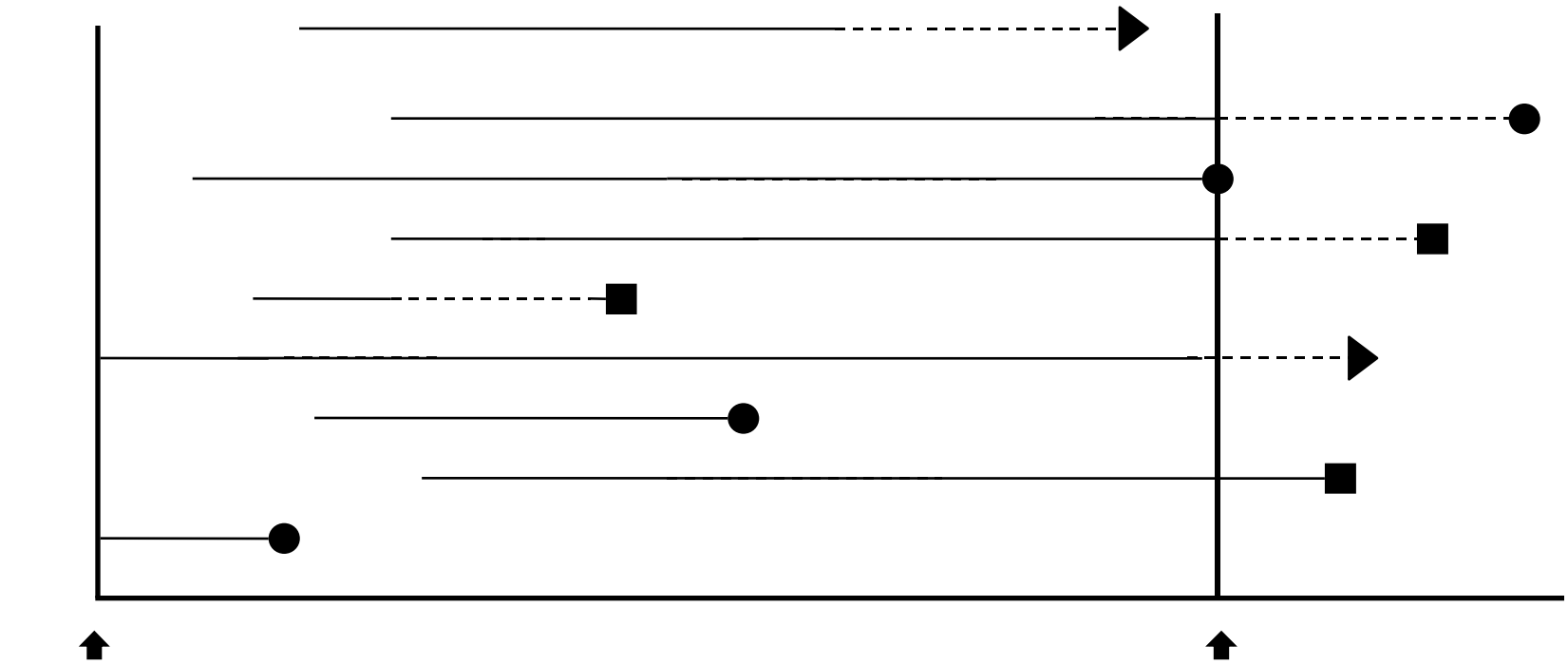
Death has long been considered humanity's number one health concern. Responsible for 100 percent of all recorded fatalities worldwide, the condition has no cure.

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)



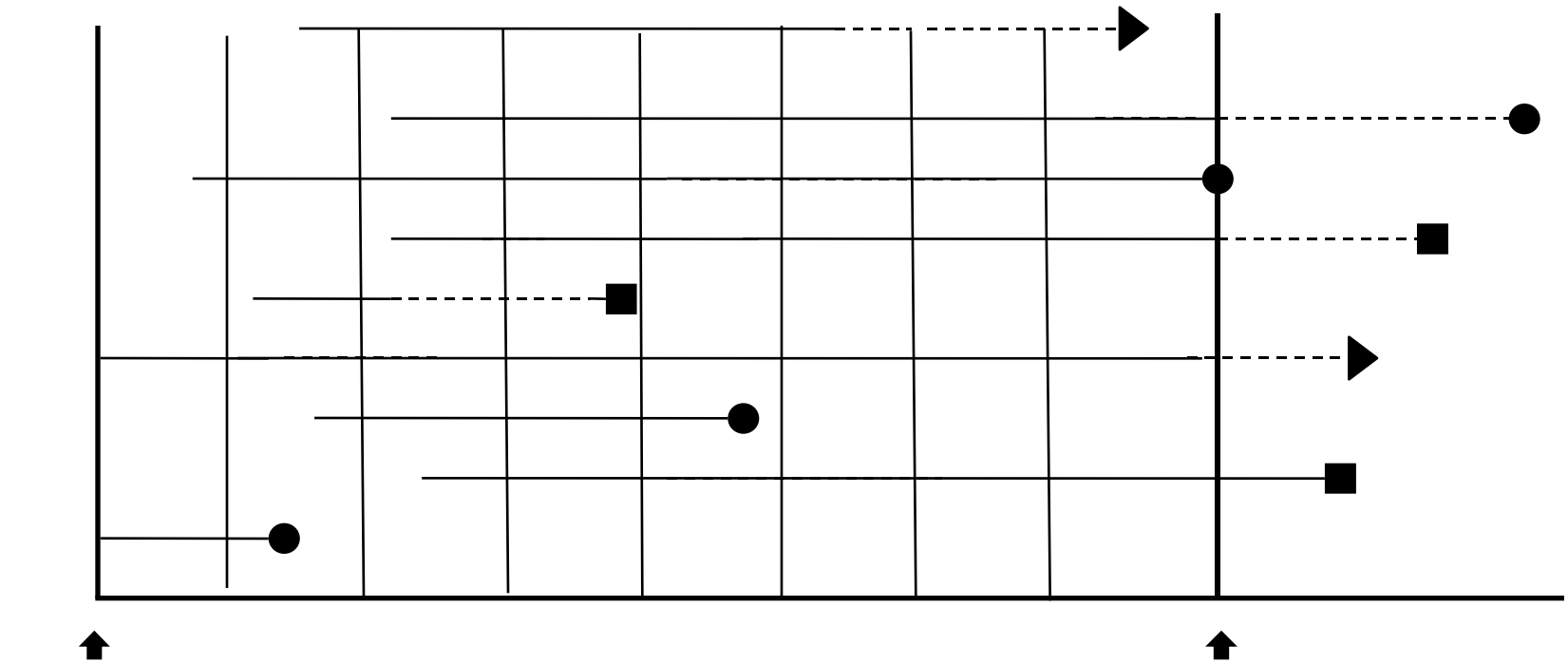
■ cancer death

▶ other death

● lost to follow up

— actively employed

--- not actively employed



■ cancer death

▶ other death

● lost to follow up

— actively employed

--- not actively employed

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

A Hypothetical Incidence Study: odds

	Exposed	Non-exposed	Ratio
Cases	1,813	952	
Non-cases	1,187	2,048	
Total	3,000	3,000	
Incidence odds	1.53	0.46	3.29

Example of non-collapsibility of odds ratios

	Young		Old		Total	
	Exposed	Non-exposed	Exposed	Non-exposed	Exposed	Non-exposed
Cases	575	183	1,238	769	1,813	952
Non-cases	925	1,317	262	731	1,187	2,048
Total	1,500	1,500	1,500	1,500	3,000	3,000
Odds ratio	4.47		4.49		3.29	

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 proportion (risk)	**0.1813**	**0.0952**	**1.90**
Incidence rate	**0.0200**	**0.0100**	**2.00**
Incidence odds	**0.2214**	**0.1052**	**2.11**

What does the odds ratio estimate in a case-control study?

- Cohort studies
- Selection of controls for case-control studies
- Implications

Control Sampling Strategies

- Rodrigues L, Kirkwood BR. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. Int J Epidemiol 1990; 19: 205-213.
- Pearce N. What does the odds ratio estimate in a case-control study? Int J Epidemiol 1993; 22: 1189-1192.

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)

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

A Hypothetical Case-Control Study

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

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

Cumulative Sampling

- “Traditional” method of control selection in nested case-control studies
- Controls are sampled from the “non-cases”, those free of disease at the *end* of the follow-up period, I.e. the *survivors*
- I.e. controls are sampled from the denominators for (cohort) *odds ratio* analyses

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

			Odds ratio
Incidence odds	1813/8187	952/9048	2.11

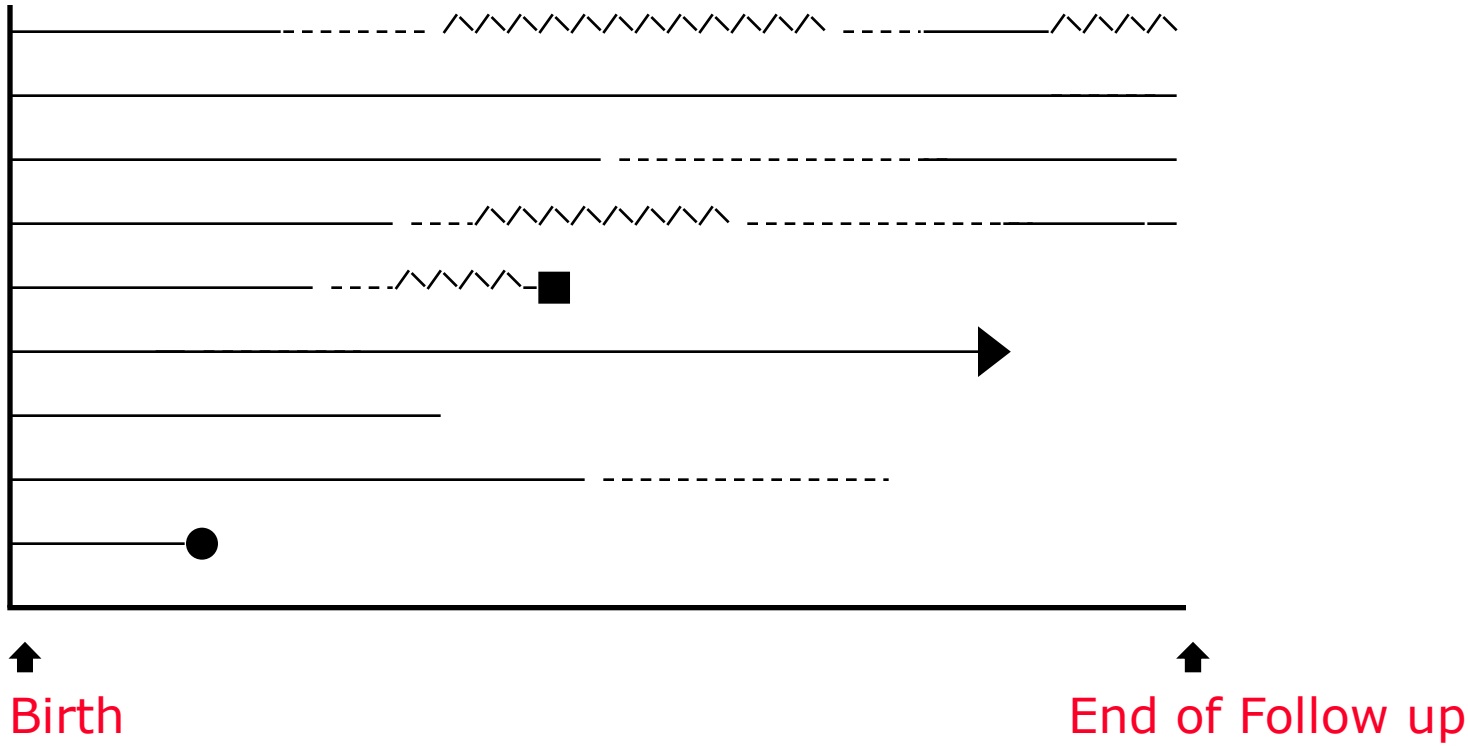
A Hypothetical Case-control Study

	Exposed	Non-exposed	Total
Cases	**1,812**	**952**	**2,765**
Controls	**1,313**	**1,452**	**2,765**

			Odds ratio
Odds	**1813/1313**	**952/1452**	**2.11**

Cumulative Sampling

- Estimates the (cohort) odds ratio (without any rare disease assumption)
- Estimates the risk ratio and rate ratio approximately (with a rare disease assumption)
- May involve matching on age, etc
- Exposure is usually only considered up to the “time” that the case occurred
- Cornfield (1951)



■ asthma death

▶ other death

● lost to follow up

— "non-asthmatic"

--- asthma symptoms

/\/\ severe asthma

Case-cohort Sampling

- Controls can be selected from those at risk at the *beginning* of the follow-up period, I.e. from the entire *source population*
- I.e. controls are selected from the denominators for (cohort) *risk ratio* analyses
- Also known as case-base sampling

A Hypothetical Incidence Study: risk

	Exposed	Non-exposed	Total
Cases	1,813	952	2,765
Total	10,000	10,000	20,000

			Odds ratio
Odds	1813/10000	952/10000	1.90

A Hypothetical Case-Control Study

	Exposed	Non-exposed	Total
Cases	**1,813**	**952**	**2,765**
Controls	**1,383**	**1,383**	**2,766**

			Odds ratio
Odds	**1813/1383**	**952/1383**	**1.90**

Case-Cohort Sampling

- Estimates the risk ratio (without any rare disease assumption)
- Requires minor modifications to the standard formulas for confidence intervals and p-values
- May involve matching on age, etc
- Once again, exposure is usually only considered up until the “time” that the case occurred
- Thomas (1972), Kupper et al (1975), Miettinen (1982), Smith et al (1984), Prentice (1986)

Density Sampling

- Controls are selected longitudinally throughout the course of the study, i.e. from the *person-time* of the study base
- I.e. controls are sampled from the denominators for the *rate ratio* analyses, i.e. from the person-years
- In general, controls are selected from the *"risk set"* of persons at risk at the "time" that each case occurred

A Hypothetical Incidence Study

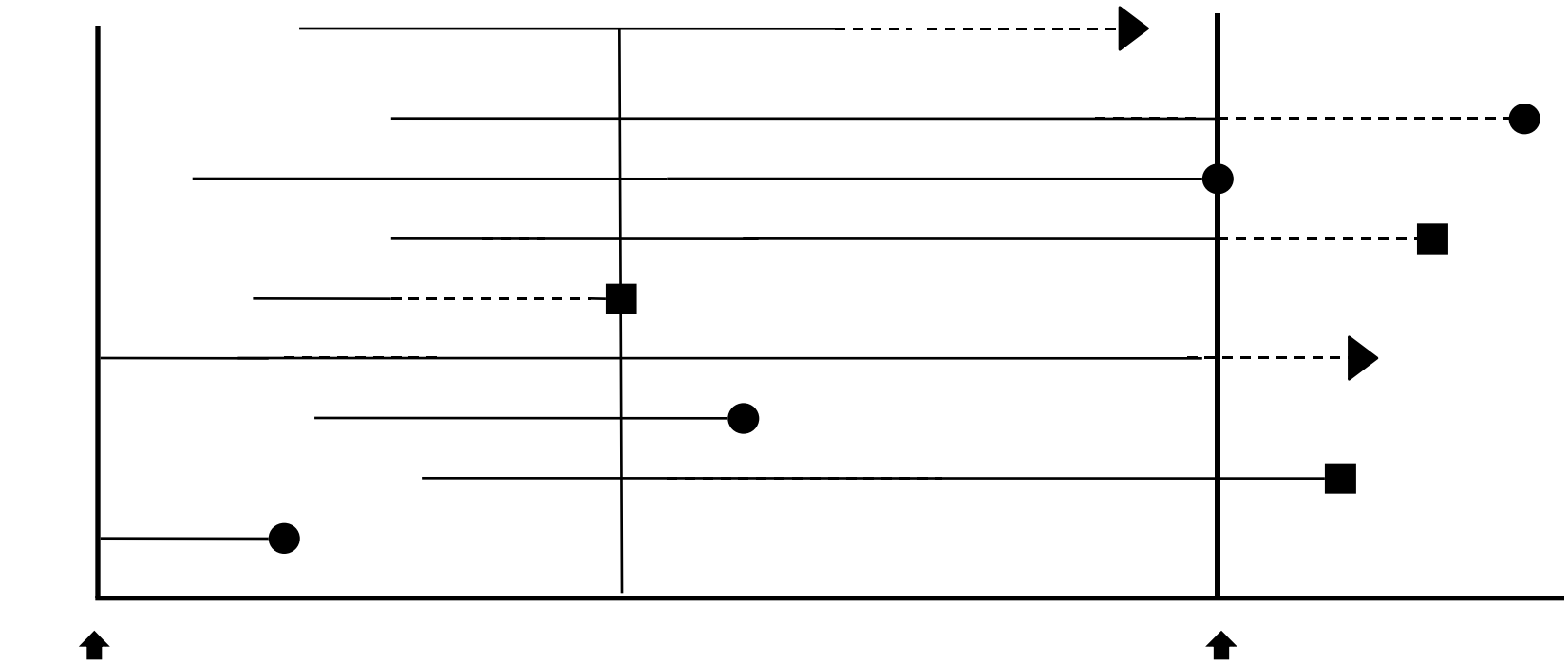
	Exposed	Non-exposed	Total
Cases	**1,813**	**952**	**27,643**
Person-years	**90,635**	**95,163**	**185,798**

			Rate ratio
Incidence rate	**1813/90635**	**952/95163**	**2.00**

A Hypothetical Case-Control Study

	Exposed	Non-exposed	Total
Cases	**1,813**	**952**	**2,765**
Controls	**1,349**	**1,416**	**2,765**

			Odds ratio
Odds	**1813/1349**	**952/1416**	**2.00**



■ cancer death

▶ other death

● lost to follow up

— actively employed

--- not actively employed

Density Sampling

- Estimates the rate ratio (without any rare disease assumption)
- The “time” variable can be:
  - survival time (proportional hazards analysis)
  - age (nested case-control study)
  - calendar time (nested or population-based case-control study)
- Matching may also be done on other time-related factors
- Sheehe (1962), Miettinen (1976)

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

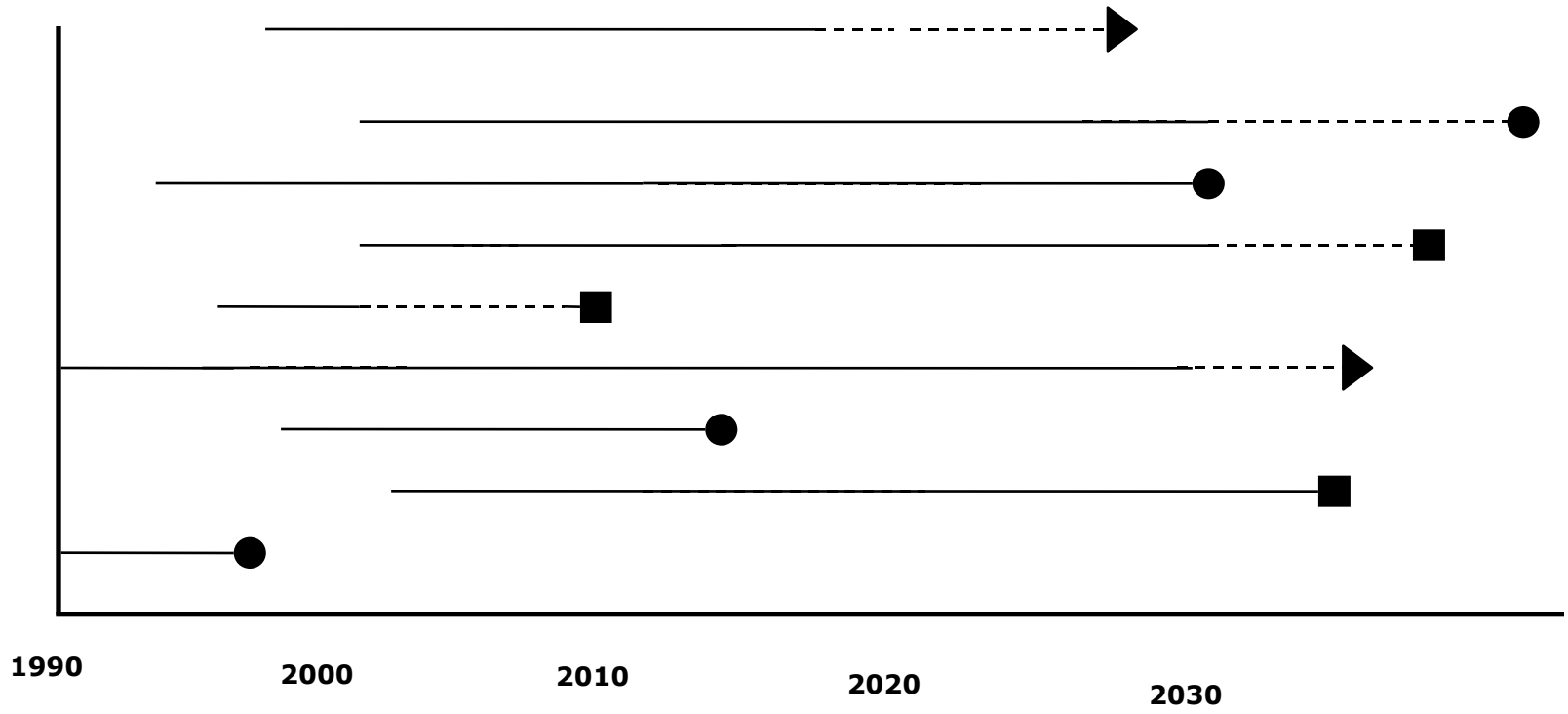
What does the odds ratio estimate in a case-control study?

- Cohort studies
- Selection of controls for case-control studies
- Implications

Control Sampling Strategies

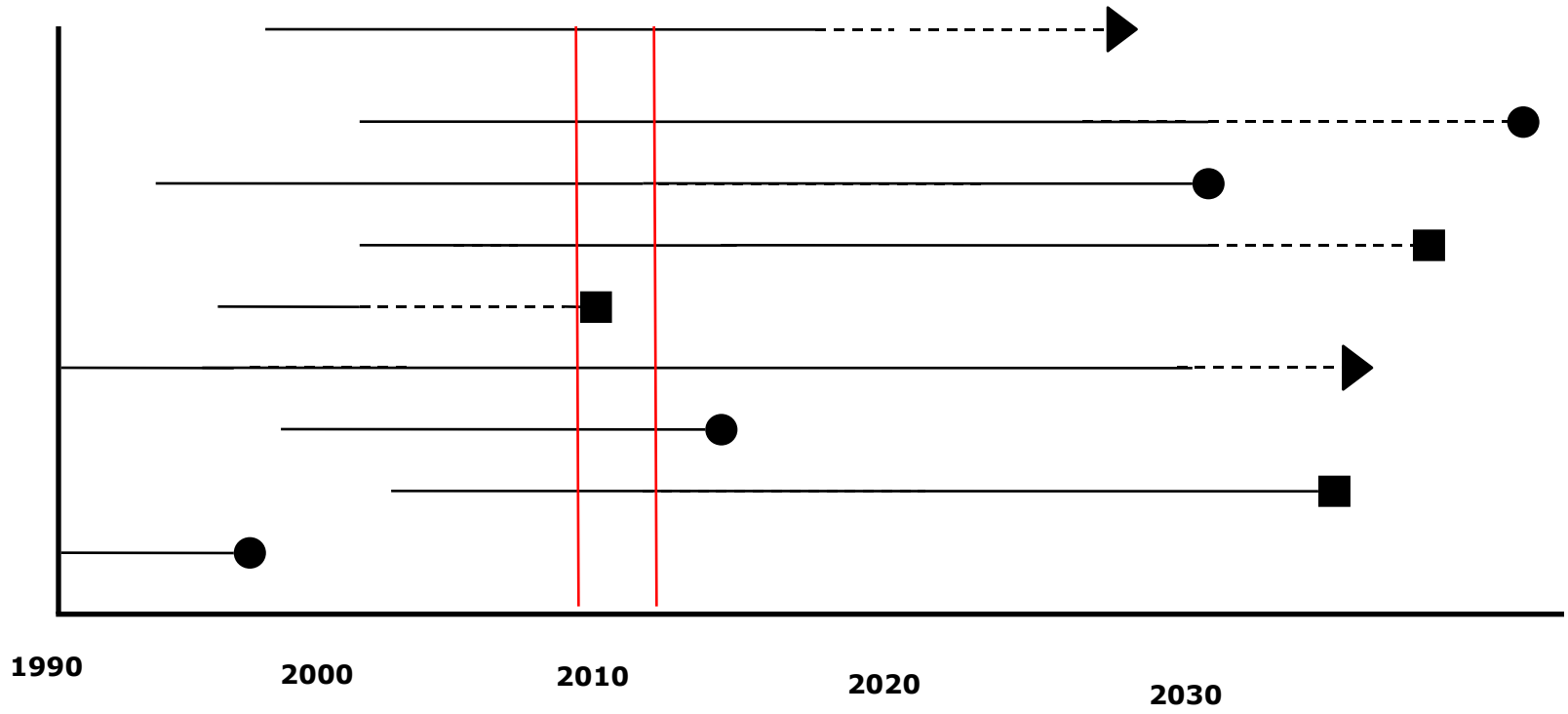
- Cumulative sampling (exclusive sampling)
- Case-cohort sampling (inclusive sampling)
- Density sampling (concurrent sampling)
- Which method do we use in population-based case-control studies?

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

Sampling options in case-control studies

		Nested case-control studies	Population-based case-control studies	What does the case-control odds ratio estimate?
Cumulative sampling (Exclusive)	Cornfield (1951)	No	No	Odds ratio
Case-cohort sampling (Inclusive)	Thomas (1972) Kupper et al (1975) Miettinen (1982) Smith et al (1984) Prentice (1986)	**Yes**	No	Risk ratio
Density sampling (Concurrent)	Sheehe (1962) Miettinen (1976)	**Yes**	**Yes**	Rate ratio

Common misconceptions about case-control studies

- Fundamentally different type of study design that proceeds from disease to exposure (retrospective)
- Inherently less valid (more biased) than cohort studies
- Require a rare-disease assumption
- Odds ratio only approximates the rate ratio or risk ratio (under the rare disease assumption)

Why do these misconceptions persist?

- Inertia – researchers (and their parents and grandparents) were taught case-control studies the Cornfield way, and it's hard to change
- "It's easier" – but only if you have been taught the Cornfield way yourself

Why do these misconceptions matter?

- It’s always important to get the theory right
- These misconceptions fuel the incorrect belief that case-control studies are inherently inferior and prone to bias
- If we teach the wrong theory it makes it harder to teach the right theory in more depth later on
- Getting the theory right helps guide our choice of control selection - if someone is contributing person-years at the “time” that a case occurs, then they are eligible to be selected as a control

What does the odds ratio estimate in a case-control study?

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