

# Studying Complex Problems in Intact Human Beings

Cohort and Case-Control Studies, and Measures of Association



# Study Designs

# Many different types...

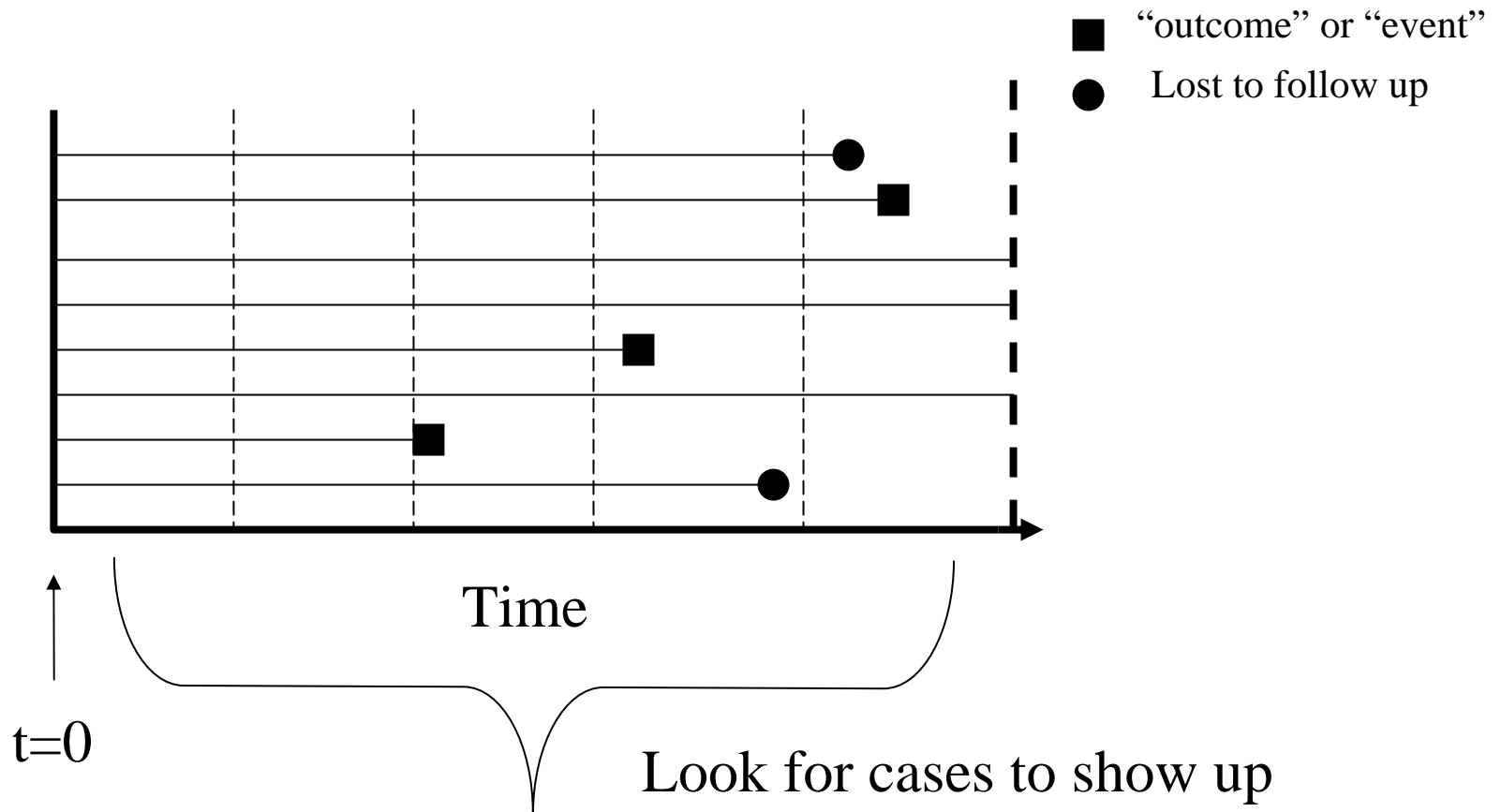
- Cohort
- Case-Control
- Nested Case-Control
- Case Cohort
- Case Cross-Over
- Randomized Controlled Trial (RCT)
- Ecological
- Etc...



# Two Examples

- Cohort
  - Longitudinal, Prospective
- Case Control
  - Retrospective

# Cohort Studies





American Journal of Epidemiology  
Copyright © 2006 by the Johns Hopkins Bloomberg School of Public Health  
All rights reserved; printed in U.S.A.

Vol. 163, No. 12  
DOI: 10.1093/aje/kwj148  
Advance Access publication May 3, 2006

---

## Original Contribution

---

### Sibship Characteristics and Risk of Multiple Sclerosis: A Nationwide Cohort Study in Denmark

**Peter Bager<sup>1</sup>, Nete Munk Nielsen<sup>1</sup>, Kristine Bihrmann<sup>1</sup>, Morten Frisch<sup>1</sup>, Jan Wohlfart<sup>1</sup>, Nils Koch-Henriksen<sup>2,3</sup>, Mads Melbye<sup>1</sup>, and Tine Westergaard<sup>1</sup>**

<sup>1</sup> Department of Epidemiology Research, Danish Epidemiology Science Centre, Statens Serum Institut, Copenhagen, Denmark.

<sup>2</sup> Department of Neurology, Aarhus University Hospital in Aalborg, Aalborg, Denmark.

<sup>3</sup> National Institute of Public Health, Copenhagen, Denmark.

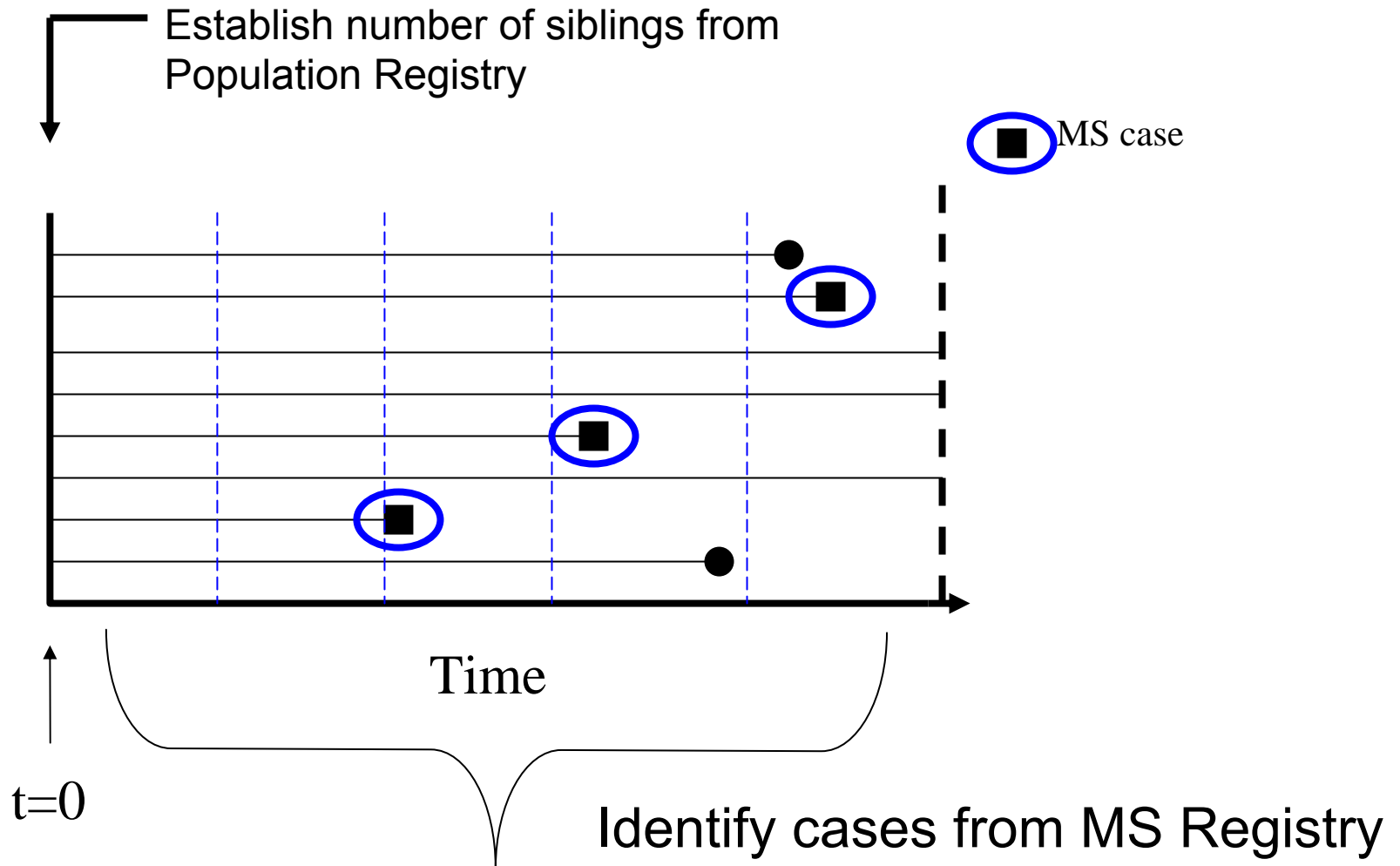
*Received for publication April 5, 2005; accepted for publication January 6, 2006.*

---

# How did this study work?

- Exposure or risk factor
  - Number of older siblings
- Outcome
  - Diagnosis with MS
- Population
  - Entire population of Denmark
- How??
  - Access to registries, one containing all MS cases, and one containing entire Danish population

# Cohort Studies





# Challenges of Cohort Studies

- Losses to Follow-up
  - People move, lose interest, etc
- Expensive!
  - Especially for a rare outcome
- Latency Period

# Case Control

- Begin with identification of cases
- “Create” sample of controls with which to compare the cases
  - Select ratio cases:controls = 1:1 or 1:2 or 1:4
- Ask both cases and controls about their histories regarding the exposure of interest

# Exposure to Infant Siblings During Early Life and Risk of Multiple Sclerosis

---

Anne-Louise Ponsonby, PhD

*JAMA. 2005;293:463-469*

---

Ingrid van der Mei, PhD

---

Terence Dwyer, MD

---

Leigh Blizzard, PhD

---

Bruce Taylor, MD

---

Andrew Kemp, PhD

---

Rex Simmons, PhD

---

Trevor Kilpatrick, PhD

---

# How did this study work?

- Exposure or risk factor
  - Number of older siblings

- Outcome

- Diagnosis with MS



Same as for cohort study!

- BUT begin with **cases**

- Controls were recruited from voter registry for the region

- 1:2 ratio

# Challenges of Case Control Studies

- Exposure information sometimes based on memory
  - What did you have for lunch a week ago?  
What about two weeks ago?
- Who are the controls?
  - Cases: all people who are diagnosed with MS in the city of Montreal between January 2005 and December 2006



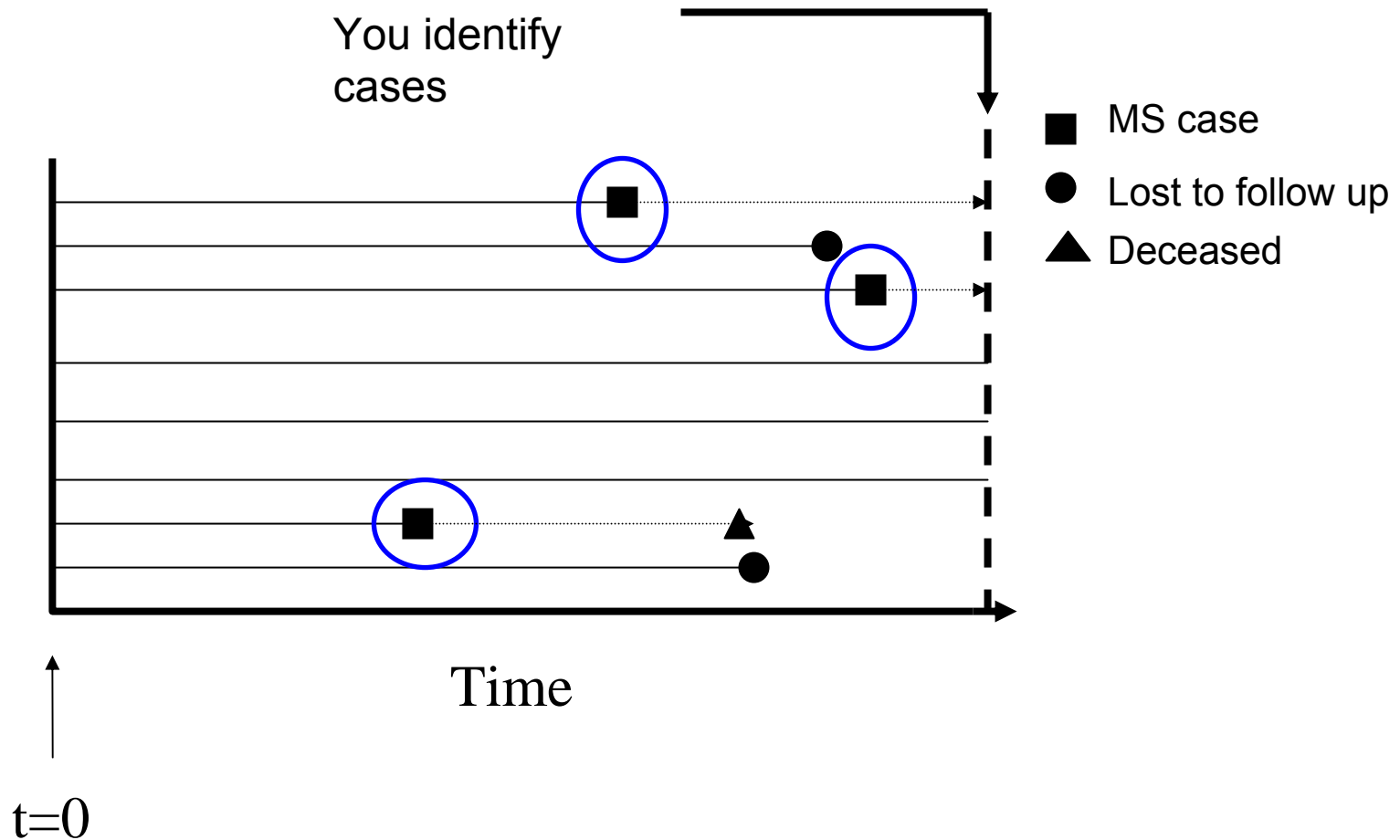
# Challenges of Case Control Studies (2)

- Who are the cases?
  - From previous example of MS: Diagnosed by whom? Diagnosed how? How can a researcher get at this information?
- Time

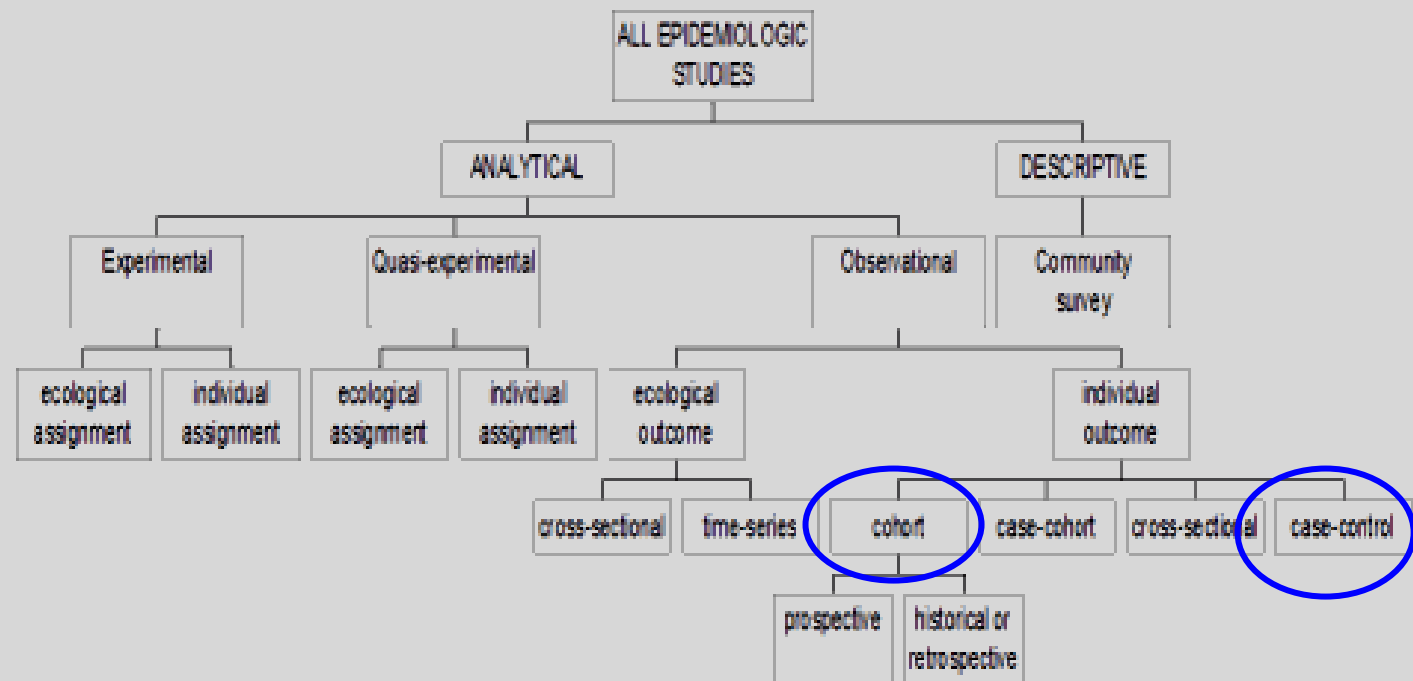
# “Rules”

- Controls represent the exposure distribution in the population that is the source of the cases
- Controls differ from cases ONLY in that they do not have the outcome
  - They would have been cases had they developed the disease

# Remember the Underlying Cohort...




## TYOLOGY OF STUDY DESIGN





# Measures of Association

- 
- Association?
  - Measure of Association for Cohort Study
  - Measure of Association for Case-Control Study
  - Testing Significance of Associations
  - Examples from the MS literature
  - Adjusting for Risk Modifiers
  - Summary of Study Design and Measures of Association

# Association?

Association between exposure (E) and disease (D)

⇒ *Statistical* dependence between E and D

⇒ Change in frequency of E *likely* to result in change in frequency of D

# Cohort Studies

Participants classified according to E status  
D status recorded during “follow-up”

Status	D+	D-	Total
E+	a	b	a+b
E-	c	d	c+d
Total	a+c	b+d	a+b+c+d

■ determined during follow-up

# The Relative Risk (RR)

Status	D+	D-	Total
E+	a	b	a+b
E-	c	d	c+d
Total	a+c	b+d	a+b+c+d

$p_{E+}$  = proportion exposed that develop disease

$p_{E-}$  = proportion unexposed that develop disease

$$RR = \frac{p_{E+}}{p_{E-}} = \frac{a/(a + b)}{c/(c + d)}$$

# Interpreting the RR

- $RR = 1$

⇒ exposure not associated with disease

- $RR > 1$

⇒ exposure is a 'risk factor'

- $RR < 1$

⇒ exposure 'protective'

# Measures of Association

## Case-Control Studies

Participants selected according to D status

- A proportion ( $p_1$ ) of the population of D+ is selected
- A proportion ( $p_2$ ) of the population of D- is selected

Population Counts (**unobservable!**)

	D+	D-
E+	A	B
E-	C	D
	A+C	B+D

Sample Counts

	D+	D-
E+	$p_1A$	$p_2B$
E-	$p_1C$	$p_2D$
	$p_1(A+C)$	$p_2(B+D)$

# Relative Risk?

$$\text{In population: } RR_{\text{population}} = \frac{A/(A + B)}{C/(C + D)}$$

$$\text{In sample: } RR_{\text{case-control}} = \frac{p_1A/(p_1A + p_2B)}{p_1C/(p_1C + p_2D)}$$

**\*\***  $RR_{\text{case-control}} \neq RR_{\text{population}}$  unless  $p_1 = p_2$

# Odds Ratio (OR)

Status	D+	D-	Total
E+	a	b	a+b
E-	c	d	c+d
Total	a+c	b+d	a+b+c+d

$$\text{OR} = \frac{\text{odds of E + in cases}}{\text{odds of E + in controls}} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

# Interpreting the OR

- $OR = 1$

⇒ exposure not associated with disease

- $OR > 1$

⇒ exposure is a 'risk factor'

- $OR < 1$

⇒ exposure 'protective'

# Testing Significance

$H_0$ :  $RR=1$  (or  $OR=1$ )

$H_1$ :  $RR>1$  (or  $<1$  or  $\neq 1$ , depending on study)

We start by assuming no association and then we look to our data to see if there is evidence to the contrary.

# Methods for Testing Significance

## 1. p-value

- The probability of observing an association as extreme or more extreme than that observed in the study given that there is truly no association between E and D
- $p\text{-value} < 0.05 \Rightarrow$  significant association
- p-value reflects both strength of association AND sample size!!

# Methods for testing significance

## 2. Confidence Interval

- Gives a plausible range of values for the measure of association
- As informative as p-value but width of interval reflects sample size
- $\uparrow$  sample size,  $\downarrow$  width of CI
- CI contains 1  $\Rightarrow$  no association

# MS Example Historical Cohort

- MS and infectious mononucleosis  
-Lindberg et al (1991)

	<b>D+</b>	<b>D-</b>
<b>E+</b>	16	6837
<b>E-</b>	12	12874

$$RR = \frac{16 / (16 + 6837)}{12 / (12 + 12874)} = 2.51$$

95% CI for RR = (1.19, 5.30)

# MS example Case-Control

## Occupational Solvent Exposure and Multiple Sclerosis

-Landtblom et al (1993)

	<b>D+</b>	<b>D-</b>
<b>E+</b>	24	62
<b>E-</b>	67	286

$$OR = (24 * 286) / (62 * 67) = 1.65$$

$$95\% \text{ CI for OR} = (0.96, 2.83)$$



# Adjusting for Known Risk Modifiers

- Age and Sex/Gender are strongly related to the risk of developing MS
- More sophisticated statistical techniques (e.g. regression) are required to adjust for these variables

# SOLVENT STUDY-a closer look

MALES

	D+	D-
E+	14	53
E-	10	119

$$OR_{\text{Male}}=3.14$$

FEMALES

	D+	D-
E+	10	9
E-	57	167

$$OR_{\text{Female}}=3.26$$

# Summary

- MANY, many study designs
- Cohort study - classify by E status
  - use RR
- Case-control study
  - most appropriate for MS
  - sampled according to D status
  - use OR
- CI to test significance
- Adjust for variables that are associated to the D and independently to the exposure

# References

- Landtblom AM, Flodin U, Karlsson M, Palhagen S, Axwolson O. Multiple Sclerosis and exposure to solvents, ionizing radiation and animals. Scand J Work Environ Health 1993; 19:399-404.
- Haahr S, Koch-Henriksen N, Moller-Larsen A, Eriksen LS, Andersen HMK. Increased risk of multiple sclerosis after late Epstein-Barr virus infection: a historical prospective study. Multiple Sclerosis 1995; 1:73-77.



# References – Epi Methods for MS

- Riise T, Wolfson C. The Epidemiologic Study of Exogenous Factors in the Etiology of Multiple Sclerosis. *Neurology* 1997. (Supplement 2 to *Neurology* 49)

# Reference – Very Basic Biostatistics

- Glantz S. Primer of Biostatistics, McGraw-Hill, 2001.

\*This covers the topics covered in the statistics segment and more. Very good for those without a strong background in statistics.

# References- Study Design and more...

- Hennekens C, Buring J.(1987) Epidemiology in Medicine, 1<sup>st</sup> Ed. Lippincott Williams and Wilkins: USA
  - This is a great overview of many epi concepts. I think there might be a 2<sup>nd</sup> edition that is more recent.
- Szklo M, Nieto F.(2000) Epidemiology: Beyond the basics. Aspen Publishers: USA.
  - Is really “beyond the basics”, gives a bit more depth