



# Observational Designs

---

Journal Club

January 11, 2008

# Presented Article

---

Rebbeck, Troxel, Norman et al. (2007) A retrospective case-control study of the use of hormone-related supplements and association with breast cancer. *Int J Cancer*, 120, 1523-28.

Study Design: case-control study.

949 cases

1524 controls

Disease: breast cancer

Exposure: hormone-related supplements

# Design Types

---

- Experimental:
  - Clinical Trials
  - Randomized, controlled
- Observational:
  - Prospective Cohort study
  - Retrospective Cohort study
  - Case-Control

# Experimental Designs

---

- Exposure/treatments are controlled by design
  - dose levels fixed
  - time course fixed
  - systematic data collection
  - predefined sample size
  - usually randomized if comparative

# Observational Studies

---

- “Sit back and watch”
  - no “control” over doses, treatments, exposures
  - individuals self-select exposure
- Measurements
  - Exposures
  - Diagnoses
  - Often self-reported



# Prospective Cohort Studies

---

- E.g. Framingham study
- population followed forward in time
- assess exposures in the present tense
- watch for disease in the future
- usually a “representative” (random) sample, but sometimes sampling is based on exposure
- **goal is to compare exposed and unexposed individuals**

# Case-Control Studies

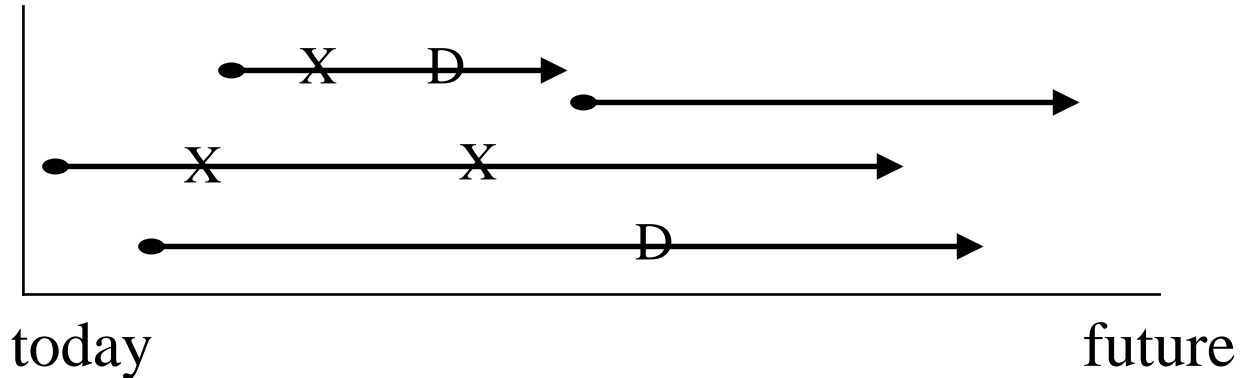
---

- E.g. HRS ~ breast cancer
- population followed backward in time
- assess disease status in the present tense
- look for exposure in the past
- designed so that sampling is based on disease status
- goal is to compare diseased and non-diseased individuals

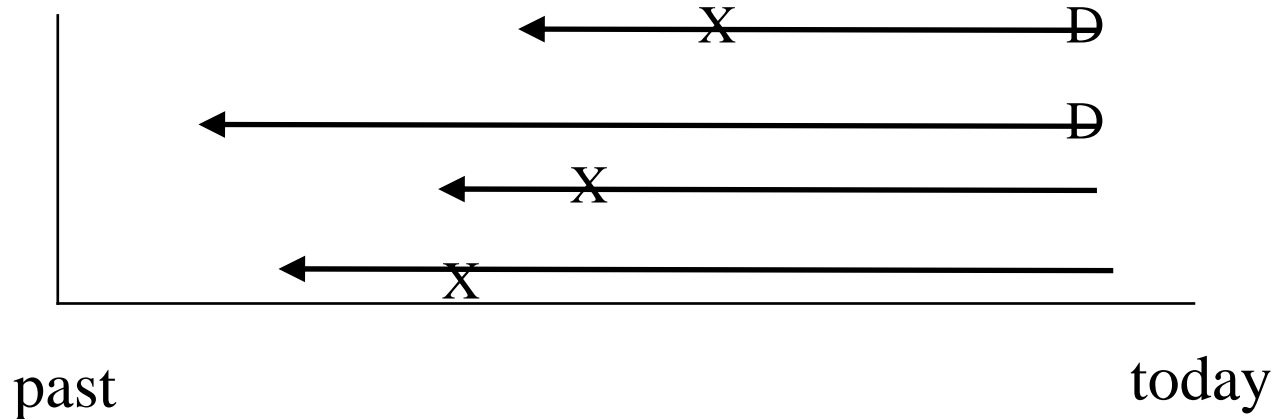
# Designs

## Prospective Cohort:

---



## Case-Control:



# One more to consider

---

- Retrospective cohort study
  - Similar to prospective cohort because sample tends to be “representative”
  - Sampling not based on case/disease status
  - uses historical data (“chart review”)
  - can be treated the same as prospective cohort study because we are comparing exposed and non-exposed populations

# Key difference

---

WHO IS BEING COMPARED?

**COHORT:**

EXPOSED VS. UNEXPOSED

**CASE-CONTROL:**

DISEASED VS. NON-DISEASED

# Pros & Cons

---

- Cohort studies are expensive
- Cohort studies can (usually) measure exposure precisely
- In cohort studies, disease prevalence can be measured
- Cohort studies are impractical for study of rare disease.
- Can assess temporal relationship
- Case control studies are cheap
- Case control studies tend to rely on recall for exposure measure
- Case control studies don't allow for measurement of disease prevalence
- Case control studies are efficient in rare diseases
- Can't always assess temporal relationship

# Case-Control and Cohort

---

- In both, inferences can be biased due to confounders
- Confounding would be protected against if we could randomize!
- Both allow for inference when randomized clinical trial would be **unethical**
  - Smoking?
  - Sun exposure?



# Measuring Risk

---

- Cohort Study:

- What is the probability of getting diseased if you are exposed as compared to unexposed?

- Case-Control Study:

- What is the probability of having been exposed if you have the disease compared to not having the disease?

# Risk in Cohort Studies

	Disease	Non-Diseased	
Exposed	A	B	A+B
Unexposed	C	D	C+D
	A+C	B+D	

○ Relative Risk (RR):

$$\begin{aligned} RR &= \frac{\text{probability of disease given exposed}}{\text{probability of disease given unexposed}} \\ &= \frac{A / (A + B)}{C / (C + D)} \end{aligned}$$

# Risk in Cohort Studies

---

	Disease	Non-Diseased	
Exposed	A	B	A+B
Unexposed	C	D	C+D
	A+C	B+D	

---

## ○ Odds Ratio (OR):

$$\begin{aligned} OR &= \frac{\text{probability of disease given exposed} / (1 - \text{probability of disease given exposed})}{\text{probability of disease given unexposed} / (1 - \text{probability of disease given unexposed})} \\ &= \frac{[A / (A + B)] / [B / (A + B)]}{[C / (C + D)] / [D / (C + D)]} \\ &= \frac{A / B}{C / D} \\ &= \frac{AD}{BC} \end{aligned}$$

# Risk in Case-Control Studies

	Disease	Non-Diseased	
Exposed	A	B	A+B
Unexposed	C	D	C+D
	A+C	B+D	

## ○ Odds Ratio (OR):

$$\begin{aligned} OR &= \frac{\text{probability of exposure given disease} / (1 - \text{probability of exposure given disease})}{\text{probability of exposure given non-diseased} / (1 - \text{probability of exposure given non-diseased})} \\ &= \frac{[A / (A + C)] / [C / (A + C)]}{[B / (B + D)] / [D / (B + D)]} \\ &= \frac{A / C}{B / D} \\ &= \frac{AD}{BC} \end{aligned}$$

# Take Home Point

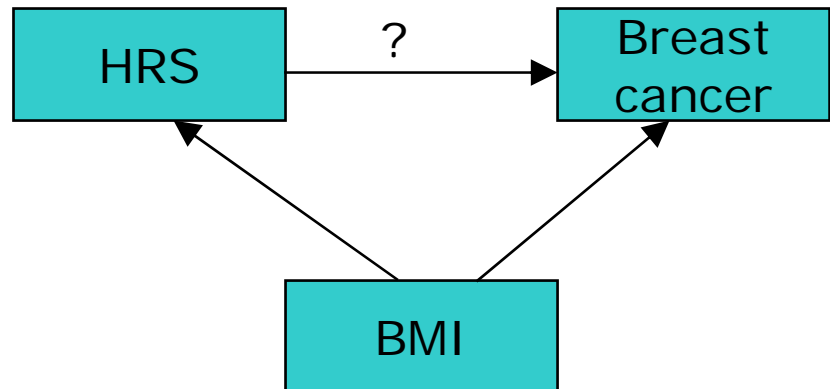
---

- Despite difference in design, the odds ratio is the SAME measure of risk in both types of studies.
- In the simplest analytic approach, we can easily calculate  $AD/BC$  from the 2x2 table of an observational study.
- But, things do tend to get more complicated:
  - what if exposure is not binary, like HRS?
  - what if we need to adjust for known, measured confounders, such as BMI, smoking, age, parity, etc?

# Logistic Regression

---

- Logistic regression allows us to do 2x2 table analysis, and much more
- We can account for 'confounders'
- example:
  - Assume BMI is associated with HRS use
  - We know BMI is associated with breast cancer risk
  - After adjusting for BMI, is HRS associated with breast cancer?



# Why is logistic regression SO important in observational studies?

---

- We see it in clinical trials, but it is not as omnipresent as in observational
- Big difference: in clinical trials, we often rely on *randomization* to ensure comparability of groups.
- In observational studies, individuals self-select treatment/exposure and that choice may be related to other factors.
- We MUST perform adjustment for confounding factors!
- Issues:
  - We need to know the confounders
  - We need to have measured the confounders

# Examples

---

1. Exercise and selenium: what if selenium is strongly associated with prostate cancer? People who exercise tend to eat better diets, rich in selenium. If we consider the association between exercise and prostate cancer without adjusting for selenium, then we may falsely conclude that exercise and prostate cancer are associated.
2. Coffee and lung cancer: A case-control study found a strong association between coffee and lung cancer. However, after adjusting for smoking, the association “went away.” Why? People who self-select smoking also tend to self-select coffee consumption

# implications

---

- Randomized clinical trials are the “gold standard”
- Many people don't put much stock in observational studies
- But we can't always do randomized trials due to
  - Ethics
  - Costs (time, money, etc.)
  - General feasibility
- Some observational studies have been enormously informative
  - Framingham
  - Nurses' Health Study
  - Physicians' Health Study
  - Olmsted County, Minnesota

# This study

---

- What about other health habits?
  - Diet?
  - Nutrition?
  - Exercise?
- These might be related to HRS use
- Argument: but then why black cohosh and not others?

1. Most others were not as prevalent
2. all others were in the same direction

Exposure	Use in European Americans		Use in African Americans		Ever use of specific herb vs. never use of any HRS, OR <sup>1</sup>
	Cases (N = 677)	Controls (N = 905)	Cases (N = 272)	Controls (N = 619)	
Any HRS	77 (11.4) <sup>2</sup>	155 (17.2)	46 (16.9)	125 (20.2)	0.65 [0.49–0.87]
Any phytoestrogen	20 (3.0)	44 (4.9)	20 (7.4)	46 (7.4)	0.69 [0.43–1.11]
Any Isoflavone or genistein	9 (1.3)	19 (2.1)	2 (0.7)	8 (1.3)	0.67 [0.29–1.53]
Isoflavone	9 (1.3)	17 (1.9)	0	5 (0.8)	ND <sup>4</sup>
Genistein	0	2 (0.2)	2 (0.7)	3 (0.5)	ND <sup>4</sup>
Red clover	2 (0.3)	8 (0.9)	13 (4.8)	29 (4.7)	0.70 [0.33–1.47]
Soy medications	11 (1.6)	21 (2.3)	6 (2.2)	14 (2.2)	0.69 [0.33–1.44]
Black cohosh or Remifemin	15 (2.2)	36 (4.0)	10 (3.7)	40 (6.5)	0.44 [0.25–0.77]
Black cohosh	13 (1.9)	34 (3.8)	9 (3.3)	39 (6.3)	0.37 [0.20–0.66]
Remifemin	3 (0.4)	6 (0.7)	2 (0.7)	2 (0.3)	ND <sup>4</sup>
Bioestrogen	0	1 (0.1)	0	0	ND <sup>4</sup>
DHEA	8 (1.2)	16 (1.8)	2 (0.7)	4 (0.7)	ND <sup>4</sup>
Dong quai	12 (1.8)	21 (2.3)	9 (3.3)	20 (3.2)	0.75 [0.39–1.45]
Estrovin	3 (0.4)	4 (0.4)	3 (1.1)	7 (1.1)	ND <sup>4</sup>
Ginseng	41 (6.1)	84 (9.3)	31 (11.4)	80 (12.9)	0.75 [0.53–1.06]
Promensil	1 (0.2)	4 (0.4)	0	2 (0.3)	ND <sup>4</sup>
Rejuvex	7 (1.0)	10 (1.1)	2 (0.7)	8 (1.3)	ND <sup>4</sup>
Steroid creams	6 (0.9)	13 (1.4)	1 (0.4)	3 (0.5)	ND <sup>4</sup>
Yam creams	5 (0.7)	10 (1.1)	1 (0.4)	4 (0.7)	ND <sup>4</sup>