§18.5 Case-Control Sampling

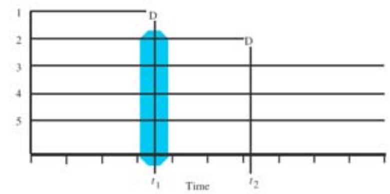
- Identify all cases in population
- Randomly select non-cases (controls) from same population
- Ascertain exposure status of cases and controls
- Cross-tabulate exposure & disease status

Efficient way to study rare outcomes

Case-Control

Incidence Density Sampling

Select non-case at random from source population when case occurs



Miettinen. *Am J Epidemiol 1976; 103*, 226-235.

Odds Ratio

Cross-tabulate exposure & disease status:

	D+	D−
E+	a_1	b_1
E−	a_2	b_2

Calculate

$$\hat{OR} = \frac{a_1 b_2}{a_2 b_1}$$

cross-product ratio

OR interpreted as a RR

BD1 Data

- Cases: esophageal cancer
- Controls: noncases selected at random from electoral lists
- Exposure: alcohol consumption dichotomized at 80 gms/day

Alcohol grams/day	Esophageal cancer Cases	Noncases
≥80	96	109
80	104	666
Total	200	775

$$\hat{OR} = \frac{a_1 b_2}{a_2 b_1} = \frac{96 \cdot 666}{104 \cdot 109} = 5.64$$

Relative risk associated with exposure

(1– α)100% CI for the *OR*

$$e^{\ln \hat{OR} \pm z_{1-\frac{\alpha}{2}} \cdot SE_{\ln \hat{OR}}}$$

$$\text{where } SE_{\ln \hat{OR}} = \sqrt{\frac{1}{a_1} + \frac{1}{b_1} + \frac{1}{a_2} + \frac{1}{b_2}}$$

90% CI for *OR* – Example

	D+	D−
E+	96	109
E−	104	666

$$\ln \hat{OR} = \ln(5.640) = 1.7229$$

$$SE_{\ln \hat{OR}} = \sqrt{\frac{1}{96} + \frac{1}{104} + \frac{1}{109} + \frac{1}{666}} = 0.1752$$

For 90% confidence use $z = 1.645$

$$e^{1.7299 \pm (1.645)(0.1752)} = e^{1.7229 \pm 0.2882}$$

$$= e^{1.4417, 2.0181}$$

$$= (4.23, 7.52)$$

WinPEPI > Compare2 > A.

Data entry

	Box 1 Yes (number)	Box 2 No (number)
A:	96	109
B:	104	666

Output

```
ODDS RATIO  = 5.64  [reciprocal = 0.18]
  Fisher's exact confidence intervals:
    90%:  4.16 to 7.62
    95%:  3.94 to 8.06
    99%:  3.54 to 8.99
  Mid-P exact confidence intervals:
    90%:  4.22 to 7.52
    95%:  3.99 to 7.95
    99%:  3.59 to 8.87
  Cornfield's confidence intervals:
    90%:  4.17 to 7.64
    95%:  3.94 to 8.07
    99%:  3.54 to 8.99
```

WinPEPI's Mid-P interval similar to ours

36

Ordinal Exposure

Alcohol (g/day)	Esophageal cancer Cases	Controls
1 (0–39)	29	386
2 (40–79)	75	280
3 (80–119)	51	87
4 (120+)	45	22
Total	200	775

Break data up into multiple tables, using the least exposed level as baseline each time

March 09 37

Ordinal Exposure

Alcohol grams/day	Esophageal cancer Yes	No
0-39	29	386
0-39	29	386

$\widehat{OR}_1 = \frac{29 \cdot 386}{386 \cdot 29} = 1.00$ (reference)

Alcohol grams/day	Esophageal cancer Yes	No
40-79	75	280
0-39	29	386

$\widehat{OR}_2 = \frac{a_1 b_2}{b_1 a_2} = \frac{75 \cdot 386}{280 \cdot 29} = 3.57$

Alcohol grams/day	Esophageal cancer Yes	No
80-119	51	87
0-39	29	386

$\widehat{OR}_3 = \frac{51 \cdot 386}{87 \cdot 29} = 7.80$

Alcohol grams/day	Esophageal cancer Yes	No
120+	45	22
0-39	29	386

$\widehat{OR}_4 = \frac{45 \cdot 386}{22 \cdot 29} = 27.23$

Dose-response

38

18.6 Matched Pairs

- Cohort matched pairs: each exposed individual uniquely matched to non-exposed individual
- Case-control matched pairs: each case uniquely matched to a control
- Controls for matching (confounding) factor
- Requires special matched-pair analysis

March 09 39

Matched-Pairs, Cohort

	Exposed D+	Exposed D−
Non-exp D+	*a*	*b*
Non-exp D−	*c*	*d*

$$\hat{OR} = \frac{c}{b}$$

$$95\% \text{ CI for } OR = e^{\ln \hat{OR} \pm z_{1-\frac{\alpha}{2}} \cdot SE_{\ln \hat{OR}}}$$

$$\text{where } SE_{\ln \hat{OR}} = \sqrt{\frac{1}{b} + \frac{1}{c}}$$

March 09 40

Matched-Pairs, Case-Control

	Case E+	Case E−
Control E+	*a*	*b*
Control E−	*c*	*d*

$$\hat{OR} = \frac{c}{b}$$

$$95\% \text{ CI for } OR = e^{\ln \hat{OR} \pm z_{1-\frac{\alpha}{2}} \cdot SE_{\ln \hat{OR}}}$$

$$\text{where } SE_{\ln \hat{OR}} = \sqrt{\frac{1}{b} + \frac{1}{c}}$$

March 09 41

Matched-Pairs Case-Cntl Example

Cases = colon polyps; Controls = no polyps
Exposure = low fruit & veg consumption

	Case E+	Case E−
Cntl E+	?	24
Cntl E−	45	?

$$\hat{OR} = \frac{c}{b} = \frac{45}{24} = 1.88$$

88% higher risk w/ low fruit/veg consumption

March 09 42

Matched-Pairs - Example

$\widehat{OR} = \frac{c}{b} = \frac{45}{24} = 1.88.$

$\ln\widehat{OR} = \ln(1.875) = 0.6286$ (by calculator)

$SE_{\ln\widehat{OR}} = \sqrt{\frac{1}{c} + \frac{1}{b}} = \sqrt{\frac{1}{45} + \frac{1}{24}} = 0.2528$

The 95% confidence interval for $OR = e^{0.6286 \pm (1.96)(0.2528)} = e^{0.6286 \pm 0.4959} = e^{(0.1331,\ 1.1241)} = (1.14 \text{ to } 3.07)$.

March 09 43

WinPEPI > PairEtc > A.

Input

		Sample A Yes	Sample A No
Sample B	Yes	0	24
	No	45	0

Total pairs = 69

Output

```
Odds ratio (odds A : odds B) = 1.875
  Fisher's confidence intervals:
     90% confidence interval = 1.207 to 2.956
     95% confidence interval = 1.118 to 3.218
     99% confidence interval = 0.964 to 3.809
  Mid-P confidence intervals:
     90% confidence interval = 1.240 to 2.867
     95% confidence interval = 1.148 to 3.121
     99% confidence interval = 0.988 to 3.695
```

March 09 44

Hypothesis Test
Matched Pairs

A. H_0: $OR = 1$

B. **McNemar's** test (z or chi-square)

$$z_{stat,McN} = \sqrt{\frac{(c-b)^2}{c+b}}$$

C. *P*-value from z stat

Avoid if fewer than 5 discordancies expected

March 09 45

Smoking Twins example

	Smoker D+	Smoker D−
Non-smoker D+	0	5
Non-smoker D−	17	0

(A) **Hypotheses.** H_0: no association between smoking and earlier mortality in population ($OR = 1$) vs. H_a: "association" ($OR \neq 1$).

(B) **Test statistic.** $z_{stat,McN} = \sqrt{\frac{(c-b)^2}{c+b}} = \sqrt{\frac{(17-5)^2}{17+5}} = 2.56$

(C) ***P*-value.** $P = 0.01047$ [from Table F]. With continuity correction,

$z_{stat,McN,c} = \sqrt{\frac{(|c-b|-1)^2}{c+b}} = \sqrt{\frac{(|17-5|-1)^2}{17+5}} = 2.35$ and $P = 0.019$. By convention, the association is said to be significant.