Week 3 – Tuesday session

Stratified Case-control Analysis

EPI202 – Epidemiologic Methods II

Murray A. Mittleman, MD, DrPH
Department of Epidemiology, Harvard TH Chan School of Public Health

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Week 3: Discussion Topics

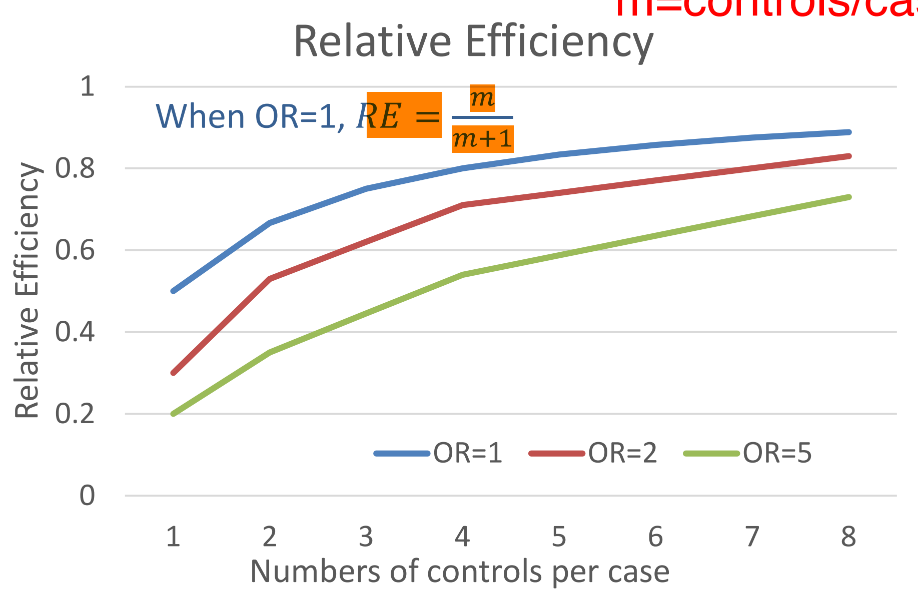
1. Stratified case-control data data
   - Hypothesis tests
   - Point and interval estimates – weighted averages
   - Assumption of homogeneity
   - Effect measure modification
2. Matching
   - Matched cohort design
   - Matched case-control design
   - Appropriate, unnecessary and overmatching
   - Matching ratio – impact on precision
   - Matched case-control analysis
     - McNemar test and odds ratio estimator
     - Relationship to Mantel-Haenszel methods

Matching Ratios
Relative Efficiency of the Test of the Null Hypothesis

		Number of Controls/Case						
H_A: RR =	Pr(A=1|Y=0)	1	2	4	8	16	32	∞
1	0-100%	50%	66%	80%	88%	95%	97%	100%
2	5%	30%	53%	71%	83%	92%	95%	100%
5	5%	20%	35%	54%	73%	84%	93%	100%

m=controls/cases

- Relative efficiency is the ratio of the variance of a particular design relative to having infinite controls (equivalent to conducting the full cohort analysis)
- Under the null, the efficiency of a particular matching ratio relative to the full cohort design is proportional to m/(m+1), where m is the ratio of controls:cases.
- Pr(A=1|Y=0) is the exposure prevalence in the controls



Hypothesis Test for Stratified Data (1)

- We first stratify the data on all confounding variables to form I strata and then calculate the **Mantel-Haenszel χ^2 test statistic**:

$$Z^2 = \frac{\left[\sum_{i=1}^{I} X_i - \sum_{i=1}^{I} E_i\left(X_i \mid H_0\right)\right]^2}{\sum_{i=1}^{I} \mathrm{Var}_i\left(X_i \mid H_0\right)} \sim \chi_1^2$$

Summary Odds Ratio
Point Estimate

- To calculate the estimate of the summary odds ratio, we use a weighted sum of the stratum-specific estimates:

$$\boxed{\hat{O}R = \frac{\sum_{i=1}^{I} w'_i \, \hat{O}R_i}{\sum w'_i} = \sum_{i=1}^{I} w_i \, \hat{O}R_i}$$

$$Where \quad w_i = \frac{w'_i}{\sum w'_i} \quad and \quad \sum w_i = 1$$

- The exact value of the summary odds ratio depends upon the chosen weights.
- In the absence of effect measure modification, any weights will yield an unbiased estimate of the common odds ratio
- Weights are chosen to optimize the statistical precision of the estimate

Mantel-Haenszel Estimator of the Odds Ratio

- Computational formula for the **Mantel-Haenszel summary odds ratio**:

$$\boxed{\hat{OR}_{MH} = \frac{\sum_{i=1}^{I} w_i' \hat{OR}_i}{\sum_{i=1}^{I} w_i'} = \frac{\sum_{i=1}^{I} \frac{b_i c_i}{T_i} \frac{a_i d_i}{b_i c_i}}{\sum_{i=1}^{I} \frac{b_i c_i}{T_i}} = \frac{\sum_{i=1}^{I} \frac{a_i d_i}{T_i}}{\sum_{i=1}^{I} \frac{b_i c_i}{T_i}}}$$

- A given stratum may contribute to the numerator, the denominator, both or neither

95% CI for $ln(\mathrm{OR}_{MH})$

$$X \pm Z_{1-\alpha/2}\sqrt{\hat{V}ar\,(X)}$$

- Where:
  - □ $X = \ln(\hat{O}R_{MH})$
  - □ $Z_{1-\alpha/2}$ = $100(1-\alpha/2)^{th}$ percentile of the standard normal distribution
  - □ $\hat{V}ar\,(X)$ = estimated variance of $\ln(\hat{O}R_{MH})$

True or False? The Mantel-Haenszel odds ratio will always be a value between the stratum-specific odds ratio estimates.

$$\hat{OR}_{MH} = \frac{\sum_{i=1}^{I} \frac{a_i d_i}{T_i}}{\sum_{i=1}^{I} \frac{b_i c_i}{T_i}} = \frac{\frac{4*224}{292} + \frac{9*390}{444} + \frac{4*330}{393} + \frac{6*362}{442}}{\frac{2*62}{292} + \frac{12*33}{444} + \frac{33*26}{393} + \frac{65*9}{442}} = 3.99$$

- True ✓
- False
- Do not know

Because the OR(MH) is a weighted average of stratum-specific estimated, it will lie between them.

True or False? The crude odds ratio will always be a value between the stratum-specific odds ratio estimates.

	1995		1996		1997		Combined	
Derek Jeter	12/48	.250	183/582	.314	190/654	.291	385/1284	**.300**
David Justice	104/411	**.253**	45/140	**.321**	163/495	**.329**	312/1046	.298

True

False

Do not know

Test for Homogeneity
Odds Ratio

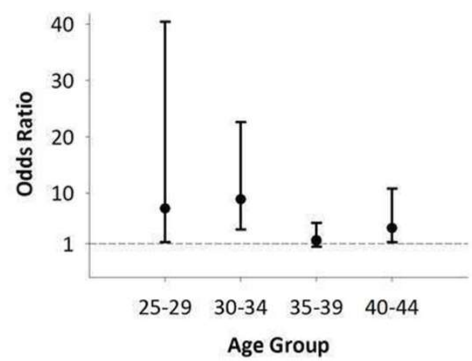
- The test for homogeneity of the odds ratio has the following form

$$H = \sum_{i=1}^{I} \frac{[\ln(\hat{O}R_i) - \ln \hat{O}R_{MH}]^2}{\hat{V}ar_i[\ln(\hat{O}R_i)]} \sim \chi^2_{I-1}$$

- Where:
  - $\hat{O}R_i$ is the stratum specific odds ratio
  - $\hat{O}R_{MH}$ is the Mantel-Haenszel summary odds ratio
  - $\hat{V}ar[ln(\hat{O}R_i)]$ is the (large-sample) variance of the stratum specific ln(OR);

- H is distributed χ^2 with I-1 degrees of freedom

Based on the point estimates and 95% confidence intervals from a case-control study, do you think there is clear evidence of effect measure modification of the odds ratio for the association between OC use and myocardial infarction?



- Yes
- No
- Maybe ✓

The confidence intervals are wide and tend to overlap with each other considerably.

It is not clear whether the observed heterogeneity is more than one would expect due to random variability.

If you reject the null hypothesis and conclude that there is statistically significant effect measure modification of the odds ratio, it is valid to report which of the following:

$$H = \sum_{i=1}^{I} \frac{[\mathit{ln}(\hat{OR}_i) - \mathit{ln}\,\hat{OR}_{MH}]^2}{\hat{V}ar_i[\mathit{ln}(\hat{OR}_i)]}$$

- The crude OR
- The Mantel Haenszel OR
- The stratum-specific ORs
- The standardized OR

HAVE A GOOD WEEK