EPID600 (Spring 2007) module on Confounding

Objectives:

- Define confounding bias.
- State the criteria to consider a covariate to be a confounder.
- Distinguish between potential confounders and real confounders.
- State the conditions under which a potential confounder would not be controlled for as a confounder.
- Distinguish between a crude and an adjusted measure of association.
- Use crude and adjusted measures to determine whether confounding is present.
- Describe methods to control for confounding in both the design and analysis of a study.
- Apply concepts to a case study.

Instructions:

1. **Read**: Aschengrau and Seage, ch. 11 - Confounding, and ch. 13 - Effect Measure Modification . Answer the practice questions at the end of the chapter or at http://publichealth.jbpub.com/aschengrau/student_resources.cfm and check your answers (recommended, but optional) (animated flashcards, weblinks, and Powerpoint slides from the authors] can also be found at that URL)

2. Look over the case study questions and then read the case study reading: (Slide presentation in confounding module in Blackboard and film [classroom course])

3. (Optional, but earns credit) Before lab, submit the answers to the starred case study questions (numbers 1, 2, 4, 6, and 9).

4. Read the lecture slides and attend the lecture (or read the speaker notes).

5. Work on the rest of the case study questions in **lab** and afterwards.

6. Agree on the answers, so the facilitator can submit the group's consensus answers by the following Sunday evening (EST).

Case Study Questions (NOTE: For some of these questions there may not be one "right answer".)

**1. What four types of factors (give an example for each) were hypothesized to be potential contributors to the development of esophageal cancer in Lin Xian? Can you think of other factors that might be relevant?

**2. What specific risk factors were presented as suspects in causing esophageal cancer? What is the rationale for suspecting each one?

The following questions are based on a hypothetical study of Barrett's esophagus (BE), a pre-cancerous neoplasia that often precedes esophageal cancer. The study enrolled 4,000 people age 40-49 years and collected baseline data on several aspects of their diet: moldy bread (MB), pickled cabbage (PC), and insufficient vitamin C (LVC). The cohort was followed for 10 years, during which 205 cases of Barrett's esophagus were detected. Although the analysis of a study like this would probably use incidence rates based on person-years of follow-up, for simplicity we will analyze incidence proportions and assume that there was no loss to follow-up.

After carefully reviewing and cleaning their data, the investigators carried out a crude analysis to examine the relation of each dietary risk factor (MB, PC, LVC) to Barrett's esophagus (BE). The crude analysis is shown in the first set of tables at www.unc.edu/epid600/classes/2007a/modules/casestudies/12/output1.htm#table1 [These tables are in the classroom coursepack and linked to the on-line copy of the case study.]

The bottom row of the first table [MB(Moldy bread) by BE (Barrett's esophagus)] shows that there were 205 cases of BE (incidence proportion 5.13%, shown in red) among the 4,000 members of the cohort. The right-most column shows that 770 of the total of 4,000 were exposed to moldy bread. The prevalence of exposure to moldy bread (19.25%) is shown in blue.

Among the 770 participants exposed to moldy bread, 73 (9.48%, in red) developed Barrett's esophagus, whereas only 4.09% (in red) of the 3,230 participants not exposed to moldy bread developed BE. Thus, the relative risk (cumulative incidence ratio, CIR, also called incidence proportion ratio, IPR) was 2.32 (shown in red in the subtable headed "Common Relative Risk" - the term "Common" appears because the subtable is most often used to present the results of a stratified analysis that controls for other variables; we will have an example of that presently). (Note: these tables were created with the Statistical Analysis System [SAS], though the first two have been slightly reformatted. Should you be interested, you can view the SAS program used to create the dataset and to generate the analyses for this case study, at www.unc.edu/epid600/classes/2007a/modules/casestudies/12/sasprogram.htm).

3. a. What was the prevalence of exposure to pickled cabbage (PC, see table 2, labelled Pickled cabbage [by] Barrett's esophagus)?

b. What were the incidence proportions of BE for, respectively, persons exposed to and not exposed to PC?

c. What was the relative risk (IPR) for BE in relation to exposure to PC?

4. a. **Calculate the prevalence of exposure to low vitamin C (LVC, see table 3) and compare it to the value from the table.

b. **Calculate** the incidence proportions of BE for, respectively, persons exposed to and not exposed to LVC and compare these IP's to the values in the table.

c. **Calculate** the relative risk (IPR) for BE in relation to exposure to LVC and compare your calculation to the value in the table. (Note: SAS uses double precision arithmetic, so you may see slight differences between your results and those in the SAS output.)

The conceptual model on which the study was based proposed that each of the three dietary risk factors was an independent contributor to esophageal cancer. However, as might be expected, the three dietary risk factors were associated with one another in the cohort, so the investigators were concerned about the possibility of confounding. In order to gain a better understanding of the possible relations among the three exposure variables, the investigators examined tables comparing the (crude) associations of the dietary factors with each other (see the three tables Moldy bread by Pickled cabbage, Pickled cabbage by Vitamin C, and Moldy bread by Vitamin C under the heading Associations among risk factors www.unc.edu/epid600/classes/2007a/modules/casestudies/12/output2.htm).

5. Determine whether or not moldy bread (MB) and pickled cabbage (PC) are associated and quantify the strength of their association. [Note: think about what measure(s) of association will be appropriate to quantify strength of association in this context.]

**6. a. Is there an association between PC and low vitamin C intake? Between MB and low vitamin C intake? Provide a relevant measure of the strength of each of these associations.

b. What are the implications of this descriptive analysis for the presence of confounding?

7. The primary strategies for avoiding or controlling confounding are restriction, balancing exposed and unexposed groups by randomization or matching, stratified analysis, and regression modeling? What are advantages and drawbacks of each method?

With stratified analysis we partition the data into subsets defined by the “covariables” and examine the exposure-outcome relation within each subset. Then we can calculate a measure of association that summarizes the associations found in the various subsets.

For this case study we have partitioned the dataset into 4 subsets, so that we can control for two dichotomous covariables simultaneously. So, for example, we will examine the association between BE and MB, controlling for both PC and LVC: i) both PC and LVC, ii) PC present and LVC absent (sufficient), iii) PC absent and LVC present (low), and iv) both PC and LVC absent. These four data subsets are presented in the first 4 tables under the page title “Stratified analysis of associations with Barrett’s esophagus” www.unc.edu/epid600/classes/2007a/modules/casestudies/12/output3.htm The four tables are labeled “Table x of MB by BE, Controlling for PC= ... LVC=”, where x is 1,2,3, or 4 for the different combinations of levels of PC and LVC.

At the end of the four tables are summary estimates of the “Common Relative Risk” across the four tables. (SAS uses the term “common relative risk” to refer to a summary of the relative risks across all of the subtables. This summary is based on the assumption that the true relative risk is the same for each stratum, so that any differences across the strata arise only from random variation. A more general term is “adjusted relative risk”, which refers to any relative risk estimate that shows the relationship between the exposure and disease variables after removing the effects of one or more covariables. A standardized relative risk is an example of an adjusted relative risk that does not involve the assumption of uniformity of relative risks across the strata.

Since we have cohort data and the disease category is column 1, we are interested in the rows labelled “Cohort (Col1 Risk)”, so the additional rows displayed by SAS have been deleted. “Mantel-Haenszel” and “Logit” refer to two different techniques for computing summary relative risks. In these data, where the numbers of observations are rather large, the two types of estimates are nearly identical. The Mantel-Haenszel estimate is often preferred when the data are sparse. We will use the Logit estimates.

8. a. Examine the association between BE and MB in each of the four tables and then the common relative risk estimate. Describe what you see.

b. Compare these results to the crude relative risk estimate (see text above question 5). Is there confounding?

c. Confounding arises from noncomparability of the exposed and unexposed groups, i.e., when the comparison group is not a good substitute for the counterfactual condition. In what way are the exposed (MB) and unexposed (not MB) groups not comparable?

Compare the joint distributions of the risk factors PC and LVC between those exposed to MB and those not exposed to MB by filling in the following table from the SAS output and commenting on the results.

PC	LVC	# MB	% MB	# No MB	% No MB
Yes (1)	Low (1)				
Yes (1)	Sufficient (0)				
No (0)	Low (1)				
No (0)	Sufficient (0)				
Total	Total		100%		100%

**9. a. Examine the stratified analyses for the associations between (a) BE and PC and (b) BE and LVC. Compare the common relative risks from the stratified analysis with the corresponding crude RR's and interpret what you find.

b. Which measure - the adjusted RR (from the stratified analysis) or the crude RR - is a better indicator of the independent association of the risk factor and BE? For example, if the association is causal, which RR could be used to estimate the benefit from eliminating one of the risk factors?

Although stratified analysis is a powerful and easy-to-understand method of controlling for potential confounding and obtaining adjusted estimates, it is somewhat cumbersome when there are several covariables and/or when one or more of these has multiple levels, rather than being dichotomous. In fact, since stratified analysis requires a table for each combination of covariables, controlling for 6 variables could easily require 64 tables to summarize! For these reasons it is common to construct a mathematical model that, given certain assumptions, provides a more efficient summary of the relations between each covariable and the outcome. For a cohort study estimating incidence proportion ratios, the preferred model is called relative risk regression. The analyst proposes the model structure, such as

ln(risk of BE) = ln(baseline risk) + ln(RR from MB) + ln(RR from PC) + ln(RR from LVC)

which is mathematically equivalent to:

risk of BE = (baseline risk) x (RR from MB) x (RR from PC) x (RR from LVC)

(*ln* signifies the natural logarithm - the base *e* logarithm; if you would like to refresh your memory of these concepts, have a look at Math Refresh, at www.mclph.umn.edu/mathrefresh/)

A statistical procedure is then used to estimate what the RR's would be given the data

and this model structure. Models for ratio measures are usually fit on the log scale, as indicated in this example (taking logarithms turns products into sums, and it's easier to work with sums than products). So the results of the analysis are then converted back to the natural scale by exponentiation (taking anti-logarithms)

10. A mathematical model was fit using the SAS procedure GENMOD to estimate the relative risks for each of the three dietary risk factors while controlling for the other two. (See the SAS output Relative risks controlled using relative risk regression (www.unc.edu/epid600/classes/2007a/modules/casestudies/12/output4.htm) The relative risk estimates (on the log scale) are shown in the table called "Parameter Estimates" (www.unc.edu/epid600/classes/2007a/modules/casestudies/12/output4.htm#parameters). Find the estimated ln(RR)'s for MB, PC, and LVC, compute their antilogarithms (the EXP function in MS Excel), and compare the resulting adjusted RR's to the adjusted ones from the stratified analysis.

11. Since the adjusted RR controls for confounding, and the crude RR does not, does the crude RR have any value when confounding is present?

Postscript: There has been extensive research on the problem of esophageal cancer in Linxian Province since production of "The Cancer Detectives of Lin Xian", including two large intervention trials of multivitamin and mineral supplementation (Blot *et al*., 1993; Li *et al*., 1993). The larger of these trials observed reductions of 9% in all-cause mortality and 13% in cancer mortality in the group receiving supplementation with selenium, β-carotene, and vitamin E. Follow-up to these trials continues, to assess long-term effects of vitamin / mineral supplements. A recent analysis of the relation between serum selenium and esophageal cancer (Mark *et a*l., 2000) found a 44% reduction in esophageal cancer risk (RR=0.56) among persons in the highest fourth of the distribution of serum selenium compared to the lowest fourth.

Blot WJ, Li JY, Taylor PR, *et al*. Nutrition intervention trials in Linxian, China: supplementation with specific vitamin / mineral combinations, cancer incidence, and disease specific mortality in the general population. *J National Cancer Institute* 1993;85(18):1483-1492.

Li JY, Taylor PR, Li B, *et al*. Nutrition intervention trials in Linxian, China: multiple vitamin / mineral supplementation, cancer incidence, and disease specific mortality among adults with esophageal dysplasia. *J National Cancer Institute* 1993;856(18):1492-1498.

Steven D. Mark, You-Lin Qiao, Sanford M. Dawsey, Yan-Ping Wu, Hormuzd Katki, Elaine W. Gunter, Joseph F. Fraumeni, Jr., William J. Blot, Zhi-Wei Dong, Phillip R. Taylor. Prospective Study of Serum Selenium Levels and Incident Esophageal and Gastric Cancers. *J National Cancer Institute*, 2000;92(21):1753-1763.

11/15,17,23,25/2005vs, 11/29/2006vs, 1/10/2007vs

EPID600 confounding case study – statistical analysis output

Vic Schoenbach, 7/11/2004, 11/15/2005

Crude association with Barrett's esophagus

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

TABLE 1

MB (Moldy bread: 1=yes, 0=no)	BE (Barrett's esophagus: 1=case, 0=noncase)		
	Case(1)	**Noncase(0)**	**Total**
Exposed(1)	73 1.83 **9.48** 35.61	697 17.43 90.52 18.37	770 **19.25**
Unexposed(0)	132 3.30 **4.09** 64.39	3,098 77.45 95.91 81.63	3,230 80.75
Total	205 **5.13**	3,795 94.88	4,000 100.00

Estimates of the Common Relative Risk				
Type of Study	Method	Value	95% Confidence Limits	
Cohort	Mantel-Haenszel	2.3199	1.7623	3.0537
(Col1 Risk)	Logit	**2.3199**	1.7623	3.0537

Frequency
Percent
Row Pct
Col Pct

TABLE 2

PC (Pickled cabbage: 1=yes, 0=no)	**BE (Barrett's esophagus: 1=case, 0=noncase)**		
	Case(1)	**Noncase(0)**	**Total**
Exposed(1)	114 2.85 **10.55** 55.61	967 24.18 89.45 25.48	1081 **27.03**
Unexposed(0)	91 2.28 **3.12** 44.39	2828 70.70 96.88 74.52	2919 72.98
Total	205 **5.13**	3795 94.88	4000 100.00

Estimates of the Common Relative Risk				
Type of Study	Method	Value	95% Confidence Limits	
Cohort	Mantel-Haenszel	3.3828	2.5913	4.4160
(Col1 Risk)	Logit	3.3828	2.5913	4.4160

TABLE 3

Frequency
Percent
Row Pct
Col Pct

Table of LVC by BE			
LVC(Vitamin C: 1=low, 0=adequate)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Low(1)	136 3.40 **7.85** 66.34	1597 39.93 92.15 42.08	1733 **43.33**
Sufficient(0)	69 1.73 **3.04** 33.66	2198 54.95 96.96 57.92	2267 56.68
Total	205 **5.13**	3795 94.88	4000 100.00

Estimates of the Common Relative Risk				
Type of Study	Method	Value	95% Confidence Limits	
Cohort	Mantel-Haenszel	2.5784	1.9431	3.4213
(Col1 Risk)	Logit	2.5784	1.9431	3.4213

Associations among risk factors

Frequency
Percent
Row Pct
Col Pct

Table of MB by PC

MB(Moldy bread: 1=yes, 0=no)	PC(Pickled cabbage: 1=yes, 0=no)		Total
	Exposed(1)	Unexposed(0)	
Exposed(1)	470 11.75 61.04 43.48	300 7.50 38.96 10.28	770 19.25
Unexposed(0)	611 15.28 18.92 56.52	2619 65.48 81.08 89.72	3230 80.75
Total	1081 27.03	2919 72.98	4000 100.00

Frequency
Percent
Row Pct
Col Pct

Table of PC by LVC

PC(Pickled cabbage: 1=yes, 0=no)	LVC(Vitamin C: 1=low, 0=adequate)		Total
	Low(1)	Sufficient(0)	
Exposed(1)	786 19.65 72.71 45.35	295 7.38 27.29 13.01	1081 27.03
Unexposed(0)	947 23.68 32.44 54.65	1972 49.30 67.56 86.99	2919 72.98
Total	1733 43.33	2267 56.68	4000 100.00

Frequency
Percent
Row Pct
Col Pct

Table of MB by LVC			
MB(Moldy bread: 1=yes, 0=no)	LVC(Vitamin C: 1=low, 0=adequate)		Total
	Low(1)	Sufficient(0)	
Exposed(1)	607 15.18 78.83 35.03	163 4.08 21.17 7.19	770 19.25
Unexposed(0)	1126 28.15 34.86 64.97	2104 52.60 65.14 92.81	3230 80.75
Total	1733 43.33	2267 56.68	4000 100.00

Stratified analysis of associations with Barrett's esophagus

Frequency
Percent
Row Pct
Col Pct

Table 1 of MB by BE			
Controlling for PC=Exposed(1) LVC=Low(1)			
MB(Moldy bread: 1=yes, 0=no)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Exposed(1)	53 6.74 12.27 60.23	379 48.22 87.73 54.30	432 54.96
Unexposed(0)	35 4.45 9.89 39.77	319 40.59 90.11 45.70	354 45.04
Total	88 11.20	698 88.80	786 100.00

Frequency
Percent
Row Pct
Col Pct

Table 2 of MB by BE

Controlling for PC=Exposed(1) LVC=Sufficient(0)

MB(Moldy bread: 1=yes, 0=no)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Exposed(1)	4 1.36 10.53 15.38	34 11.53 89.47 12.64	38 12.88
Unexposed(0)	22 7.46 8.56 84.62	235 79.66 91.44 87.36	257 87.12
Total	26 8.81	269 91.19	295 100.00

Frequency
Percent
Row Pct
Col Pct

Table 3 of MB by BE

Controlling for PC=Unexposed(0) LVC=Low(1)

MB(Moldy bread: 1=yes, 0=no)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Exposed(1)	11 1.16 6.29 22.92	164 17.32 93.71 18.24	175 18.48
Unexposed(0)	37 3.91 4.79 77.08	735 77.61 95.21 81.76	772 81.52
Total	48 5.07	899 94.93	947 100.00

Frequency Percent Row Pct Col Pct	**Table 4 of MB by BE**		
	Controlling for PC=Unexposed(0) LVC=Sufficient(0)		
MB(Moldy bread: 1=yes, 0=no)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Exposed(1)	5 0.25 4.00 11.63	120 6.09 96.00 6.22	125 6.34
Unexposed(0)	38 1.93 2.06 88.37	1809 91.73 97.94 93.78	1847 93.66
Total	43 2.18	1929 97.82	1972 100.00

Stratified analysis of associations with Barrett's esophagus
Summary Statistics for MB by BE
Controlling for PC and LVC

Estimates of the Common Relative Risk (Row1/Row2)				
Type of Study	Method	Value	95% Confidence Limits	
Case-Control	Mantel-Haenszel	1.3444	0.9579	1.8868
(Odds Ratio)	Logit	1.3593	0.9715	1.9018
Cohort	Mantel-Haenszel	1.3094	0.9642	1.7781
(Col1 Risk)	Logit	**1.3197**	**0.9718**	**1.7922**
Cohort	Mantel-Haenszel	0.9788	0.9540	1.0041
(Col2 Risk)	Logit	0.9799	0.9572	1.0032

Frequency
Percent
Row Pct
Col Pct

Table 1 of PC by BE

Controlling for LVC=Low(1) MB=Exposed(1)

PC(Pickled cabbage: 1=yes, 0=no)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Exposed(1)	53 8.73 12.27 82.81	379 62.44 87.73 69.80	432 71.17
Unexposed(0)	11 1.81 6.29 17.19	164 27.02 93.71 30.20	175 28.83
Total	64 10.54	543 89.46	607 100.00

Frequency
Percent
Row Pct
Col Pct

Table 2 of PC by BE

Controlling for LVC=Low(1) MB=Unexposed(0)

PC(Pickled cabbage: 1=yes, 0=no)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Exposed(1)	35 3.11 9.89 48.61	319 28.33 90.11 30.27	354 31.44
Unexposed(0)	37 3.29 4.79 51.39	735 65.28 95.21 69.73	772 68.56
Total	72 6.39	1054 93.61	1126 100.00

Frequency
Percent
Row Pct
Col Pct

Table 3 of PC by BE

Controlling for LVC=Sufficient(0) MB=Exposed(1)

PC(Pickled cabbage: 1=yes, 0=no)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Exposed(1)	4 2.45 10.53 44.44	34 20.86 89.47 22.08	38 23.31
Unexposed(0)	5 3.07 4.00 55.56	120 73.62 96.00 77.92	125 76.69
Total	9 5.52	154 94.48	163 100.00

Frequency
Percent
Row Pct
Col Pct

Table 4 of PC by BE

Controlling for LVC=Sufficient(0) MB=Unexposed(0)

PC(Pickled cabbage: 1=yes, 0=no)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Exposed(1)	22 1.05 8.56 36.67	235 11.17 91.44 11.50	257 12.21
Unexposed(0)	38 1.81 2.06 63.33	1809 85.98 97.94 88.50	1847 87.79
Total	60 2.85	2044 97.15	2104 100.00

Stratified analysis of associations with Barrett's esophagus
Summary Statistics for PC by BE
Controlling for LVC and MB

Estimates of the Common Relative Risk (Row1/Row2)				
Type of Study	Method	Value	95% Confidence Limits	
Case-Control	Mantel-Haenszel	2.6074	1.8942	3.5893
(Odds Ratio)	Logit	2.7605	2.0264	3.7607
Cohort	Mantel-Haenszel	2.4401	1.8262	3.2604
(Col1 Risk)	**Logit**	**2.5832**	**1.9380**	**3.4433**
Cohort	Mantel-Haenszel	0.9390	0.9173	0.9612
(Col2 Risk)	Logit	0.9389	0.9172	0.9611

Frequency
Percent
Row Pct
Col Pct

Table 1 of LVC by BE

Controlling for MB=Exposed(1) PC=Exposed(1)

LVC(Vitamin C: 1=low, 0=adequate)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Low(1)	53 11.28 12.27 92.98	379 80.64 87.73 91.77	432 91.91
Sufficient(0)	4 0.85 10.53 7.02	34 7.23 89.47 8.23	38 8.09
Total	57 12.13	413 87.87	470 100.00

Frequency
Percent
Row Pct
Col Pct

Table 2 of LVC by BE

Controlling for MB=Exposed(1) PC=Unexposed(0)

LVC(Vitamin C: 1=low, 0=adequate)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Low(1)	11 3.67 6.29 68.75	164 54.67 93.71 57.75	175 58.33
Sufficient(0)	5 1.67 4.00 31.25	120 40.00 96.00 42.25	125 41.67
Total	16 5.33	284 94.67	300 100.00

Frequency
Percent
Row Pct
Col Pct

Table 3 of LVC by BE

Controlling for MB=Unexposed(0) PC=Exposed(1)

LVC(Vitamin C: 1=low, 0=adequate)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Low(1)	35 5.73 9.89 61.40	319 52.21 90.11 57.58	354 57.94
Sufficient(0)	22 3.60 8.56 38.60	235 38.46 91.44 42.42	257 42.06
Total	57 9.33	554 90.67	611 100.00

Frequency
Percent
Row Pct
Col Pct

Table 4 of LVC by BE

Controlling for MB=Unexposed(0) PC=Unexposed(0)

LVC(Vitamin C: 1=low, 0=adequate)	BE(Barrett's esophagus: 1=case, 0=noncase)		Total
	Case(1)	Noncase(0)	
Low(1)	37 1.41 4.79 49.33	735 28.06 95.21 28.89	772 29.48
Sufficient(0)	38 1.45 2.06 50.67	1809 69.07 97.94 71.11	1847 70.52
Total	75 2.86	2544 97.14	2619 100.00

Stratified analysis of associations with Barrett's esophagus

Summary Statistics for LVC by BE

Controlling for MB and PC

Estimates of the Common Relative Risk (Row1/Row2)				
Type of Study	Method	Value	95% Confidence Limits	
Case-Control	Mantel-Haenszel	1.6810	1.2122	2.3312
(Odds Ratio)	Logit	1.7128	1.2409	2.3642
Cohort	Mantel-Haenszel	1.6268	1.2071	2.1925
(Col1 Risk)	**Logit**	**1.6412**	**1.2130**	**2.2204**
Cohort	Mantel-Haenszel	0.9752	0.9594	0.9912
(Col2 Risk)	Logit	0.9738	0.9589	0.9889

Relative risks controlled using relative risk regression

The GENMOD Procedure

Model Information		
Data Set	CONFCS.BARRETTS	
Distribution	Binomial	
Link Function	Log	
Dependent Variable	BE	Barrett's esophagus: 1=case, 0=noncase
Observations Used	4000	

Response Profile		
Ordered Value	BE	Total Frequency
1	Case(1)	205
2	Noncase(0)	3795

Analysis Of Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-3.6904	0.1254	-3.9362	-3.4446	866.05	<.0001
MB	1	0.2150	0.1574	-0.0934	0.5235	1.87	0.1718
PC	1	0.9344	0.1557	0.6292	1.2395	36.02	<.0001
LVC	1	0.5115	0.1632	0.1916	0.8314	9.82	0.0017
Scale	0	1.0000	0.0000	1.0000	1.0000		

The scale parameter was held fixed.

Principles of Epidemiology for Public Health (EPID600)

Confounding

Victor J. Schoenbach, PhD
www.unc.edu/~vschoenb/

Department of Epidemiology
School of Public Health
University of North Carolina at Chapel Hill

www.unc.edu/epid600/

This week and next (and the following)

- Data analysis and interpretation
  - Confounding (this week and next)
  - Causal inference (in 2 weeks)

Setting the scene

"The data speak for themselves."

versus

"Our data say nothing at all."*

*(Epidemiology guru Sander Greenland, Congress of Epidemiology 2001, Toronto)

Setting the scene

- Logically sound inferences involve (1) data + (2) assumptions
- No assumptions ⟶ no inference
- So always need a conceptual model

Sander Greenland, Congress of Epidemiology 2001, Toronto

Causal inference in everyday living

Does exercise make me feel better?

- Try getting exercise – how do I feel?
- Try not getting exercise – how do I feel?
- Try getting exercise again – do I feel better?

Causal inference in everyday living

Does getting too little sleep make me irritable?

- Try sleeping too little – ask my partner
- Try sleeping enough – ask my partner
- Try sleeping too little – ask my partner

Desirable attributes of crossover experiments

- Exposure is under investigator's control
- Comparison condition is a true control
- Can go back and forth, providing some control for secular changes

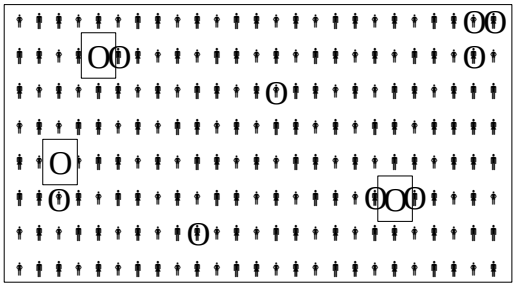
Constraints on cross-over experiments

- Exposures may be harmful or not under our control
- Effects may not be quickly reversible
- Experimental subjects or the environment may have changed

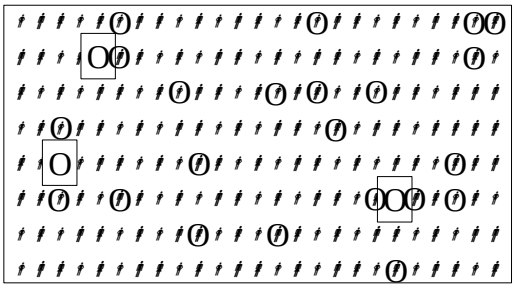
Key attribute of crossover experiments

Can compare what happens to people who are exposed to what happens to the same people when they are not exposed – almost at the same time

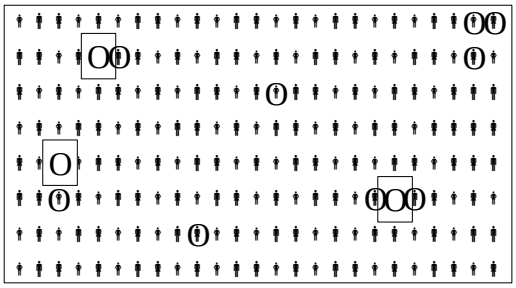
People



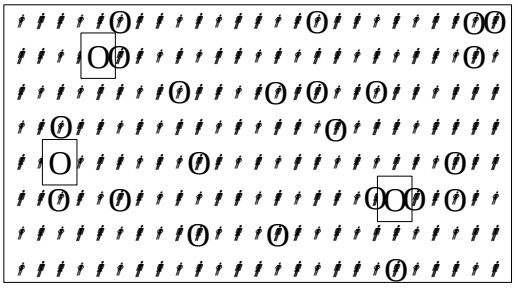
People – with an exposure



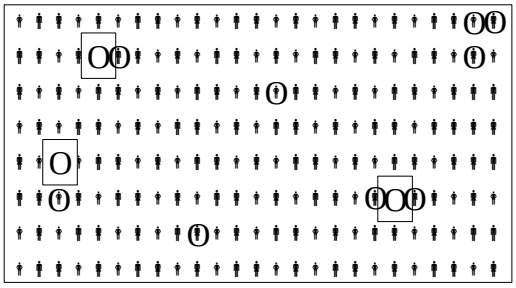
The same people – without the exposure



People – with an exposure



The same people – without the exposure



Modern formulation of causal inference

This comparison provides the best evidence that the exposure causes the outcome.

The modern formulation of causal inference and confounding is based on this "counterfactual model".

Problem of causal inference

Problem: cannot observe both conditions

Solution: observe a "substitute population", a population whose experience will represent that of the exposed population without the exposure

"Counterfactual" model

Conceptual model for causal inference:

- Compare experience of a population exposed to a factor with experience of the same population at the same time but without the exposure
- Since cannot do that, compare to experience of a substitute population.

Confounding

The substitute population is not equivalent to counterfactual condition

I.e., the substitute population does not show the "outcome in the exposed population without the exposure"

Problem of comparison

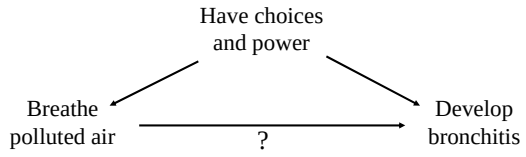
Confounding is a problem of comparison – we compare the exposed population to a substitute population, but the substitute population does not show the "outcome in the exposed population without the exposure"

Why worry about confounding?

- Does air pollution cause bronchitis ?

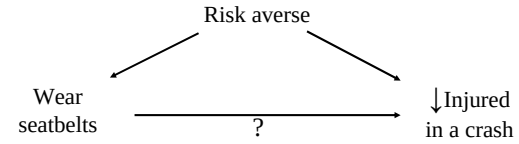
Why worry about confounding?

- Does air pollution cause bronchitis ?



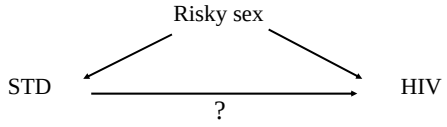
Why worry about confounding?

- Does air pollution cause bronchitis ?
- Do seatbelts reduce crash injuries?



Why worry about confounding?

- Does air pollution cause bronchitis ?
- Do seatbelts reduce crash injuries?
- Do STD's increase HIV transmission?

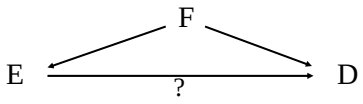


Why worry about confounding?

- Does air pollution cause bronchitis ?
- Do seatbelts reduce crash injuries?
- Do STD's increase HIV transmission?
- Does smoking lead to illicit drug use?

What comparison population?

- Does air pollution cause bronchitis ?
- Do seatbelts reduce crash injuries?
- Do STD's increase HIV transmission?
- Does smoking lead to illicit drug use?



LEARNING OBJECTIVES

- Understand (basic) **confounding**
- Recognize **potential** confounding
- Recognize **actual** confounding
- Know how to **control** confounding
- **Follow** discussions about confounding

LEARNING OBJECTIVES - 2

- Define and explain:
  - confounding
  - potential confounder
  - actual confounder
  - control of confounding

Conventional perspective

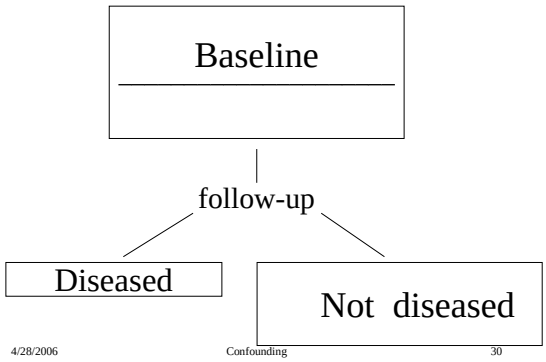
Confounding: "mixing of effects"

> Some other risk factor may be responsible for at least some of the association under investigation.

Common confounders

- Age -- e.g., exposed persons are older
- Sex -- e.g., more exposure in men
- Risk factors - more exposed persons (or unexposed) smoke(-), exercise(+), eat vegetables(+), use drugs(-), . . .

Example of confounding in a cohort



Cohort study – known risk factor

Risk factor absent

Risk factor present

Cohort study – known risk factor

Risk factor absent

Risk factor present

follow-up

Diseased

Not diseased

Cohort study for a new exposure

(Sleepy) Exposed

(Rested) Unexposed

follow-up

Diseased

Not Diseased

follow-up

Diseased

Not Diseased

Confounding in a cohort

(Sleepy) Exposed

(Rested) Unexposed

Unexposed population is the "**substitute population**" to tell us what would happen in the exposed population without its exposure - <u>but suppose</u> that the exposed population have another risk factor:

Confounding in a cohort

(Sleepy) Exposed

(Rested) Unexposed

Substitute population will not show us what would happen in the exposed population without its exposure

Confounding in a cohort

(Sleepy) Exposed

(Rested) Unexposed

follow-up

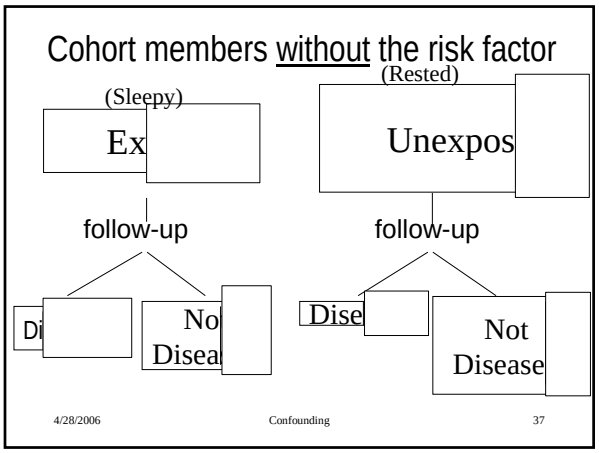
Diseased

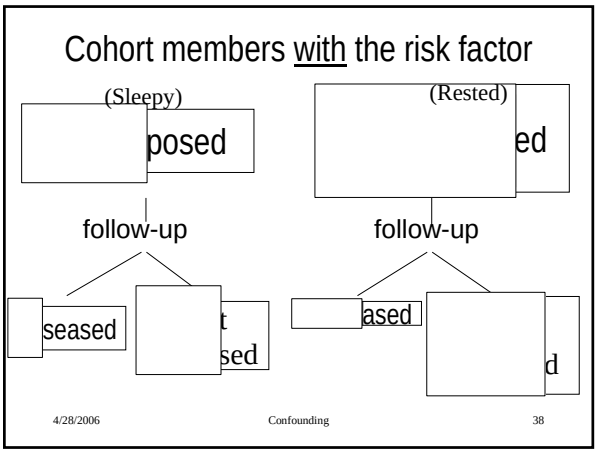
Not Diseased

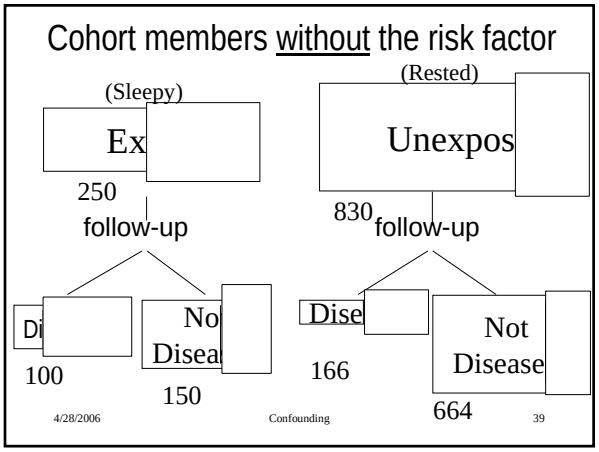
follow-up

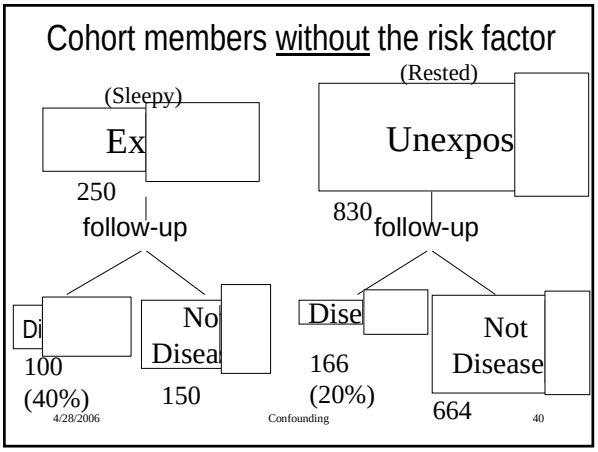
Diseased

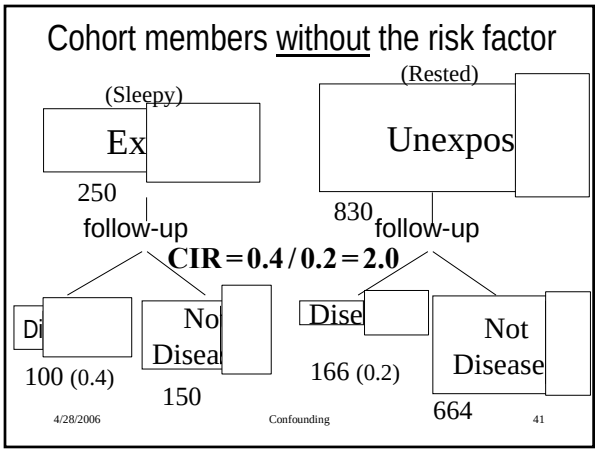
Not Diseased

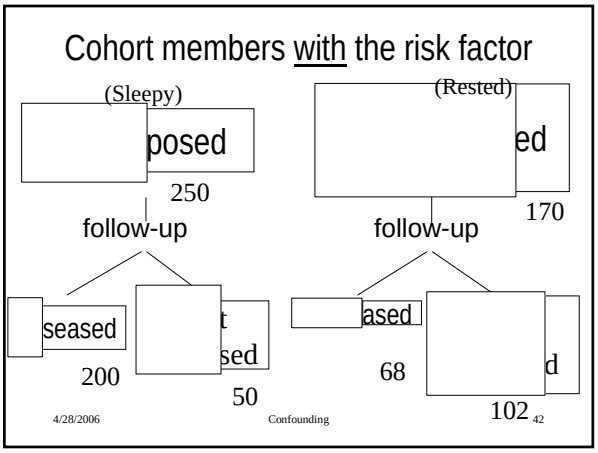

Cohort members without the risk factor
(Sleepy)
(Rested)
Ex
Unexpos
follow-up
follow-up
Di
No
Disea
Dise
Not
Disease
4/28/2006
Confounding
37

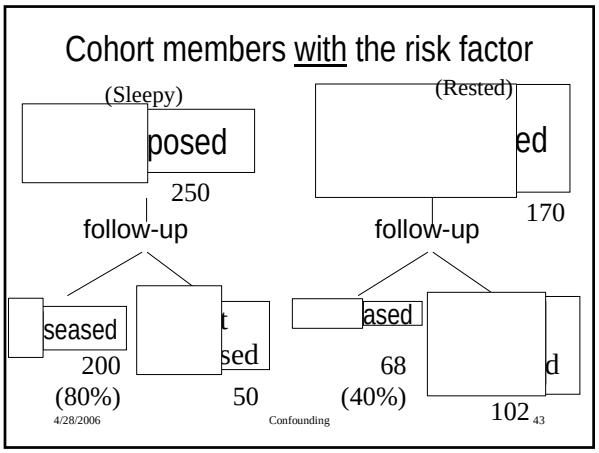

Cohort members with the risk factor
(Sleepy)
(Rested)
posed
ed
follow-up
follow-up
seased
t
sed
ased
d
4/28/2006
Confounding
38

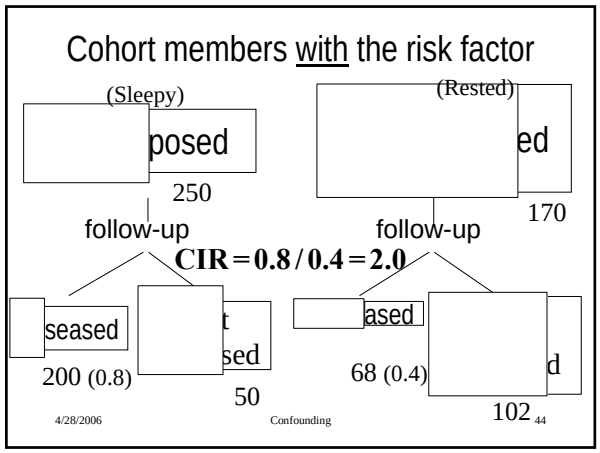

Cohort members without the risk factor
(Sleepy)
(Rested)
Ex
Unexpos
250
830
follow-up
follow-up
Di
No
Disea
Dise
Not
Disease
100
150
166
664
4/28/2006
Confounding
39

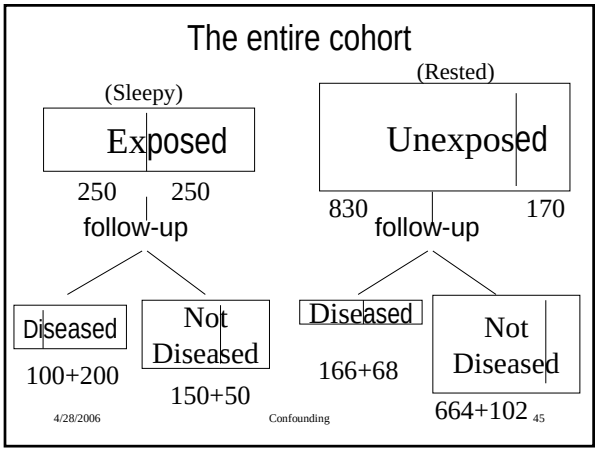

Cohort members without the risk factor
(Sleepy)
(Rested)
Ex
Unexpos
250
830
follow-up
follow-up
Di
No
Disea
Dise
Not
Disease
100
(40%)
150
166
(20%)
664
4/28/2006
Confounding
40

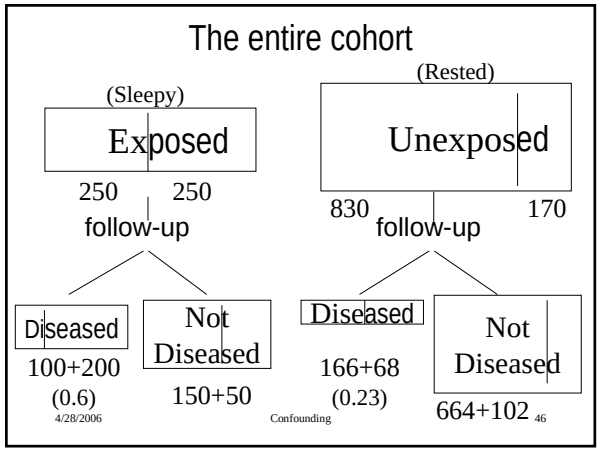

Cohort members without the risk factor
(Sleepy)
(Rested)
Ex
Unexpos
250
830
follow-up
follow-up
CIR = 0.4 / 0.2 = 2.0
Di
No
Disea
Dise
Not
Disease
100 (0.4)
150
166 (0.2)
664
4/28/2006
Confounding
41

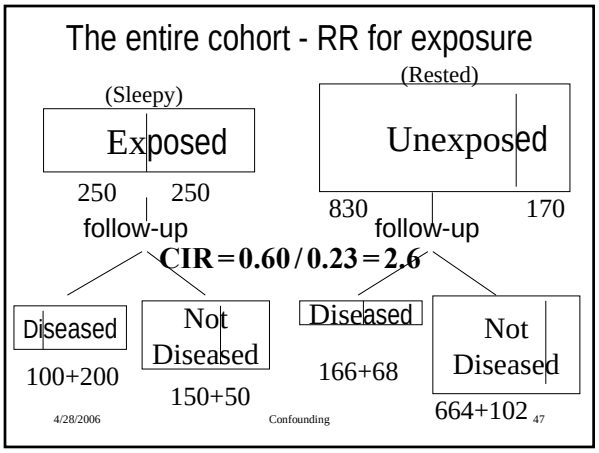

Cohort members with the risk factor
(Sleepy)
(Rested)
posed
ed
250
170
follow-up
follow-up
seased
t
sed
ased
d
200
50
68
102
4/28/2006
Confounding
42

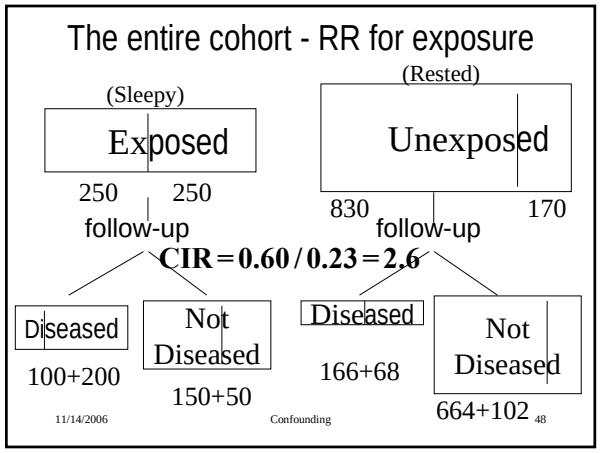

Cohort members with the risk factor
(Sleepy)
posed
250
follow-up
seased
200
(80%)
sed
50
(Rested)
ed
170
follow-up
ased
68
(40%)
d
102
4/28/2006
Confounding
43


Cohort members with the risk factor
(Sleepy)
posed
250
follow-up
CIR = 0.8 / 0.4 = 2.0
seased
200 (0.8)
sed
50
(Rested)
ed
170
follow-up
ased
68 (0.4)
d
102
4/28/2006
Confounding
44

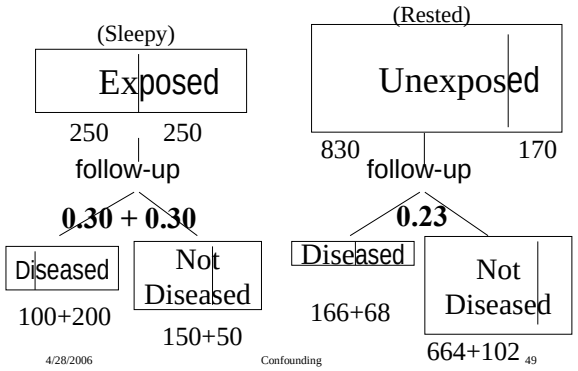

The entire cohort
(Sleepy)
Exposed
250 250
follow-up
Diseased
100+200
Not Diseased
150+50
(Rested)
Unexposed
830 170
follow-up
Diseased
166+68
Not Diseased
664+102
4/28/2006
Confounding
45


The entire cohort
(Sleepy)
Exposed
250 250
follow-up
Diseased
100+200
(0.6)
Not Diseased
150+50
(Rested)
Unexposed
830 170
follow-up
Diseased
166+68
(0.23)
Not Diseased
664+102
4/28/2006
Confounding
46

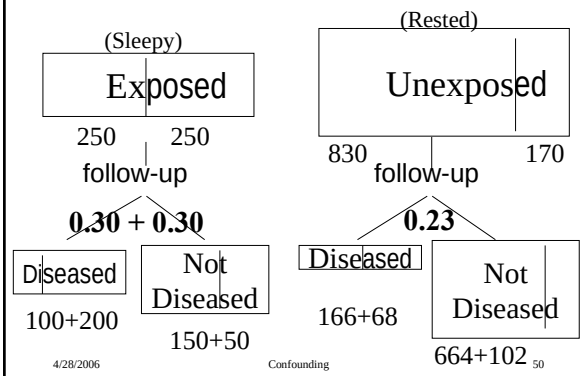

The entire cohort - RR for exposure
(Sleepy)
Exposed
250 250
follow-up
CIR = 0.60 / 0.23 = 2.6
Diseased
100+200
Not Diseased
150+50
(Rested)
Unexposed
830 170
follow-up
Diseased
166+68
Not Diseased
664+102
4/28/2006
Confounding
47


The entire cohort - RR for exposure
(Sleepy)
Exposed
250 250
follow-up
CIR = 0.60 / 0.23 = 2.6
Diseased
100+200
Not Diseased
150+50
(Rested)
Unexposed
830 170
follow-up
Diseased
166+68
Not Diseased
664+102
11/14/2006
Confounding
48

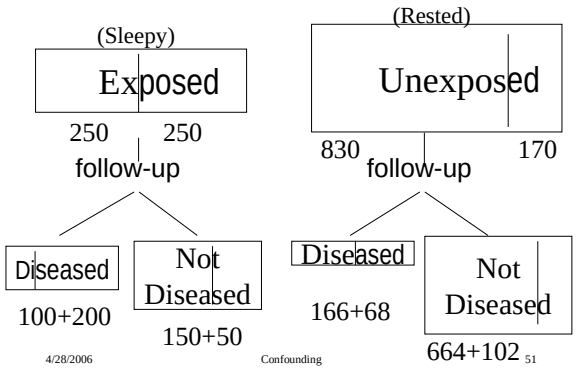
The entire cohort - RR for exposure



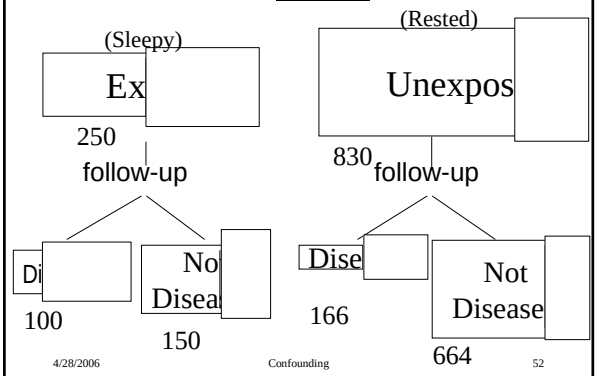
Mixing of effects and weighted averages



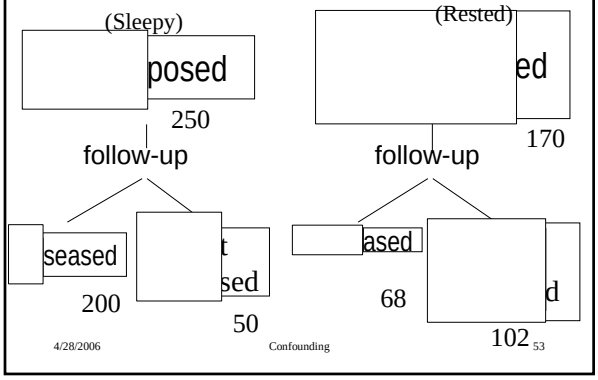
Confounding in a case-control study



Cohort members without the risk factor



Cohort members with the risk factor



Potential confounder

Determinant or risk factor for the outcome must be a potential alternative explanation for the association.

Actual confounder

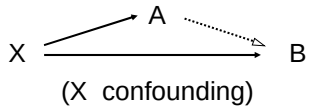
The potential confounder becomes an actual confounder when one exposure group has more of it than the other, so it's not fair to compare them

What is a confounder?

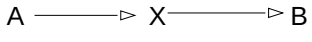
Gordis: X is a confounder of the association between factor A and disease B if both:

1. X is a known risk factor for disease B;
2. X is associated with factor A but is not a result of factor A (i.e., not intervening).

Causal models



(X confounding)



(X intervening)

What is a confounder - 2?

A confounder is:

1. "associated with the exposure and the disease" – it causes "guilt by association".
2. capable of being an "alternate explanation", i.e., the "real culprit".

Control of confounding

Controlling confounding means doing something to make comparison fair:

- Exclude people who have the risk factor ("restriction")
- Stratified analysis (adjustment, standardization)
- Mathematical modeling (e.g., regression)

Control of confounding – hard to control unknown risk factors

- These methods can control only <u>known</u> potential confounders.
- Only random assignment of exposure can control for <u>unknown</u> potential confounders.

Limitations in ability to control

Effective control of confounding requires:

- Knowing the causal pathways
- Knowing all relevant causal factors
- Measuring all relevant causal factors – accurately

Limitations in ability to control

Effective control of confounding requires assumptions, such as the mathematical form of relationships between covariables and outcome

Large, randomized experiments uniquely powerful for causal inference but . . .

Confounded confounding!

- Does overweight increase CHD risk independently of cholesterol, hypertension, and diabetes?

Confounding – key concepts

1. Interpreting data requires assumptions about causal relations (including what factors are potential confounders, i.e., what factors affect incidence and are not themselves caused by the exposure).

Confounding – key concepts

2. If exposed people and unexposed people differ on factors that affect disease incidence, then those factors may confound (distort) the observed relation between exposure and disease (i.e., actual confounding).

Confounding – key concepts

3. We can control confounding by study design if we can make the exposed and unexposed groups similar in respect to all disease determinants, though matching or randomized assignment of exposure.

Confounding – key concepts

4. We can control confounding in the analysis if we can stratify the data by disease determinants that are not themselves caused by the exposure (i.e., not causal intermediates).

Confounding – key concepts

5. The best way to understand a case-control study is to analyze it as a window into a cohort and to be aware that many books and teachings still follow the traditional and somewhat misleading perspective.

Keep hope alive!

Confounding can be confounding – do not be discouraged if you do not understand it yet.

Dietary advice

The Japanese eat very little fat and suffer fewer heart attacks than the British or Americans.

On the other hand, the French eat a lot of fat and also suffer fewer heart attacks than the British or Americans.

Dietary advice

The Japanese drink very little red wine and suffer fewer heart attacks than the British or Americans.

On the other hand, Italians drink excessive amounts of red wine and also suffer fewer heart attacks than the British or Americans.

Dietary advice - conclusion

Conclusion: Eat & drink what you like. It appears that speaking English is what kills you.

(submitted by Natasha Jamison, EPID 160 student)