In-class exercise on Confounding — **Instructor Guide**

Goal: To familiarize students with the factors associated with confounding and methods of stratified analysis for evaluation of confounding.

Background

In the small island nation of Epidoria, a team of reproductive epidemiologists has been studying the relationship between very low birth weight and risk of cognitive, motor, and behavioral problems. Five years ago these investigators initiated a cohort study. Using birth certificate files and delivery room entry logs, these investigators attempted to identify all full-term births in Epidoria over a 6-month period. The investigators enrolled all low birth weight babies and a representative sample of normal birth weight babies into their study. The investigators then examined the children every year until age 3 years. During the last examination, the investigators administered a standardized developmental screening test to assess personal-social, language, and motor-adaptive skills. Based on this test, the investigators classified the children into two groups: normal development and delayed development.

The results from the study were:

	Birth weight		
Development	Low	Normal	Total
Delayed	140	77	217
Normal	220	283	503
Total	360	360	720

1. Calculate the crude cumulative incidence ratio for the primary exposure (low birth weight).

   Crude CIR= (140/360)/(77/360)=1.82

2. To take account of the possibility that environmental lead exposure might confound the relationship between birth weight and developmental status, blood lead levels were determined from blood samples collected at the age 3-year visit. Elevated lead levels (> 10 μg/dL) were found in 173 of the low birth weight children (88 of whom had delayed development according to their screening test). Elevated lead levels were also found in 72 of the normal birth weight children (24 of whom had delayed development). Diagram several plausible sets of relationships among birth weight, blood lead level, and delayed development. In which ones could blood lead confound the association between low birth weight and delayed development.

3. Carry out a stratified analysis of birth weight and developmental delay, controlling for blood lead level. Create 2 x 2 tables for each stratum, estimate the CIR for each stratum, and interpret the results in comparison with the crude CIR from question 1.

	Low Lead			**High Lead**		
	Birth weight			Birth weight		
Development	Low	Normal	Total	Low	Normal	Total
Delayed	52	53	105	88	24	112
Normal	135	235	370	85	48	133
Total	187	288	475	173	72	245
CI	52/187	53/288		88/173	24/72	
CIR	1.51			1.53		

Some confounding is apparently present, since the crude CIR is stronger (farther from the null value) than the stratum-specific CIR's. However, an association between low birth weight and delayed development remains even after controlling for elevated blood lead.

4. Is there evidence of an association between the confounder (blood lead level) and the primary exposure (low birth weight)? To determine this association would you use (a) the entire cohort of children, (b) only those children with delayed development, or (c) only those children with normal development? Why? If there is an association, are low birth weight children more or less likely to have elevated blood lead levels?

	Birth weight		
Lead level	**Low**	**Normal**	**Total**
High	**173**	**72**	**245**
Low	**187**	**288**	**475**
Total	**360**	**360**	**720**

Prevalence ratio = (173*360) / (72*360) = 2.4

Confounding results when the exposed and unexposed groups in the study base are different in respect to a determinant of the outcome. Since the cases in a cohort study arise from the

cohort (i.e., the cohort is the study base), in order for confounding to occur, there must be an association between confounder and exposure in the cohort as a whole. From the table it is clear that there is an association between blood lead level (the confounder) and low birth weight (the primary exposure); low birth weight children are substantially more likely (prevalence ratio 2.4) to have elevated blood lead levels. (Note that the OR = (173 x 288) / (72 x 187) = 3.7 means that the odds of an elevated blood level are 3.7 times as high for low birth weight babies, which is not equivalent to "3.7 times as likely". In this case the outcome (elevated blood lead level) is not "rare", so odds are proportions are different from each other, as are their respective ratios.)

5. Is there an association between the confounder (blood lead level) and the outcome (delayed development)? To determine this association would you use (a) the entire cohort of children, (b) only those children with low birth weight, or (c) only those children with normal birth weight? Why?

   In this case, we want to know if there is an association between the confounder and the outcome independent of the exposure. Therefore we look in the UNEXPOSED (normal birth weight) group.

	Lead level		
Development	**High**	**Low**	**Total**
Delayed	**24**	**53**	**77**
Normal	**48**	**235**	**283**
Total	**72**	**288**	**360**

   CIR = (24/72) / (53/288) = 1.81

   There is an association between the confounder (blood lead level) and the outcome (developmental status).

6. Using all of the above information, do you think blood lead level is a confounder of the association between low birth weight and delayed development in this study population?

   If our conceptual model is that neither blood lead level nor low birth weight is an intervening variable on the causal pathway between delayed development and the other, then the crude CIR between low birth weight and delayed development is confounded by the different distribution of blood lead levels by birth weight. Blood lead levels are associated with the exposure (low birth weight) and are a known determinant of delayed development.

However, although confounding is present, elevated blood lead accounts for only a small amount of the crude association between low birth weight and delayed development. With or without adjustment, the latter association is of only modest strength.

7. What changes in the study design would have avoided the potential confounding effects blood lead level? What are the advantages and disadvantages of these alternatives?

   Restriction: The investigators could have restricted their study population to children without elevated blood lead levels. However, since lead exposure was not assessed until well after the cohort was selected, many children would probably have to be later dropped from the cohort when the investigators determined that their blood lead levels were high. (This is the same as carrying out a stratified analysis and discarding the stratum for elevated blood levels.) If baseline blood lead level was an adequate predictor of blood lead level during the relevant part of the follow-up, then babies with elevated levels at baseline could have been excluded. However, information would not then have been available on an important subset of children.

   Matching: If blood lead levels had been measured at baseline, then the study could have matched normal weight babies to low birth weight babies based on blood lead level. Since matching would eliminate an association between the potential confounder and the exposure variable in the study base, this procedure would have avoided confounding by blood lead level IF baseline lead levels tracked perfectly with lead levels during the three year period. (This strategy might increase statistical efficiency [better statistical precision per subject] since it would have avoided having relatively few normal weight babies with elevated blood lead levels to compare to the low weight babies with blood lead levels.)