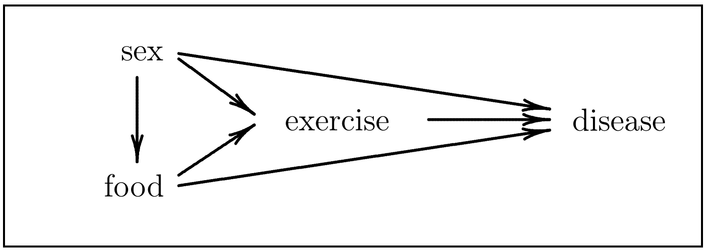
Biostats VHM 811 Course
Fall 2008, Atlantic Veterinary College, PEI
Henrik Stryhn

Brief solution to question on confounding/interaction on final exam

a) Study design and causal diagram

The study is a case-control study because groups of diseased and non-diseased cats were selected from the population of cats submitted to the clinic. The three variables under study here (sex, exercise and food) are all dichotomous variables measured as responses (in the questionnaires) but considered as explanatory variables for the disease in the causal diagram. The diagram was requested for the data for castrated cats, therefore the variable sex can be omitted from the diagram. Nevertheless, for completeness it is included in the diagram below:



The direction of the arrow between food and exercise is the one implied by the quote (alternatives could be argued). The direction of the arrow between sex and food cannot be determined from the information provided, but it seems more plausible that castration could affect food intake than conversely. As sex is not included in the analysis requested, there is no need to investigate further. The diagram shows that food could act as a confounder for the effect of exercise on the risk of disease. Conversely, exercise is an intervening (or intermediate) variable for the effect of food on the risk of disease. An opposite direction (or bi-direction) between exercise and food could be argued, but the quote given supports the interpretation of the diagram.

b) Epidemiological analysis

The first Stata listing gives an analysis of the association between food and disease stratified on exercise; with a causal diagram drawn as above, this analysis should be ignored (at first). The second listing gives an analysis for exercise stratified on food. The chosen measure of association is the odds-ratio; it is the only valid measure of association in a case-control study without additional information about the population. The listing shows that despite some differences in the odds-ratios between strata (of food) there is no evidence of an interaction between food and exercise ($P = 0.33$). This may be due to lack of power in the small dataset, but we have no choice but to proceed as if no interaction is present. The crude estimate and 95% CI for the effect of exercise is 0.25 (0.09,0.71), and the M-H stratified estimate and CI is 0.37 (0.11,1.22). This represents a major change: (0.366-0.252)/0.252 $\sim$ 45%. This large effect change when stratifying on food could indicate a confounding effect, but there are three further conditions to check.

As already discussed, the causal diagram enables a confounding effect. The association between the confounder and outcome should be evaluated in the exposure-negative group. A high level of exercise is protective and therefore constitutes the exposure-negative group; here the odds-ratio for food is 6.72 with a 95% CI that does not include 1. We conclude there is evidence of an association between the confounder and disease. Finally, the association between confounder and exposure should be evaluated in the control (non-diseased) group. No Stata listing is given for this analysis, but the four counts of the 2×2-table can be identified as 5, 2, 12, and 5 — giving an odds-ratio of $(5 \cdot 5)/(2 \cdot 12) = 1.04$. With a value so close to 1, there is no indication of an association between confounder and exposure. Thus, confounding does not exist. The effect change resulted from non-collapsibility of the odds-ratio.

The appropriate measure of association for the effect of exercise is the M-H stratified estimate: 0.37 (0.11,1.22), and we conclude there is no substantial evidence of this effect ($P = 0.10$). From the first listing we extract the total effect of food represented by the crude association: 12.75 (4.07,40.0), as well as the direct effect of food (if of interest) represented by the M-H stratified association: 9.80 (3.00,32.0).

c) Interaction

Recall that the only appropriate measure of association that can be computed from these data is the odds-ratio; that is, direct calculations of risk ratios or risk differences are not valid. By the assumption of a rare disease, the odds-ratio may be considered as an acceptable approximation of the (population) risk ratio.

A multiplicative interaction between sex and food is assessed directly by the stratified analysis of the effect of one variable by the other one (for which the roles of the two variables is irrelevant). According to the Stata listings, the test for homogeneity (no interaction) gives a test value of 0.10, and $P = 0.76$. There is absolutely no evidence of a multiplicative interaction between sex and food.

The presence of an additive interaction can be assessed by the given formulae, by substituting the relevant odds-ratios for the risk ratios. If we arbitrarily lay out the calculations with sex as the first factor and food as the second factor, and (low,low) as the reference cell, we get

$$\begin{aligned}
\mathrm{RR}_{10} \approx \mathrm{OR}_{10} &= (5 \cdot 12)/(1 \cdot 12) = 5.00, \\
\mathrm{RR}_{01} \approx \mathrm{OR}_{01} &= (4 \cdot 12)/(1 \cdot 11) = 4.36, \\
\mathrm{RR}_{11} \approx \mathrm{OR}_{11} &= (14 \cdot 12)/(1 \cdot 5) = 33.6.
\end{aligned}$$

The first two of these values are also found in the Stata listings. According to the additive formula (2), the risk ratio for cell (1,1) should be (approximately) computed as: $5.00 + 4.36 - 1 = 8.36$. This value is much less than the actual risk ratio (33.6); therefore, there is an indication of a positive interaction between sex and food. This reproduces the calculation and argument of Thrusfield (2007, p. 276). The statistical significance of the interaction cannot be determined from the information provided, and it is considered beyond the scope of the question to compute it manually or outline the relevant Stata command.