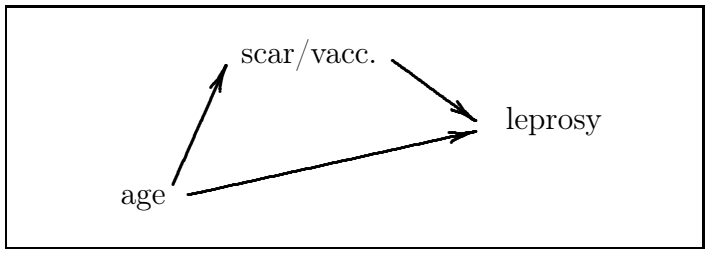
Biostats VHM 811 Course
Fall 2009, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to question on confounding/interaction on final exam

a) Survey data

The study could either be considered as a cohort study or a case-control study, depending on whether one considers the survey to have covered the entire population or to have constituted a sample from the population. This solution focuses on the cohort study interpretation, according to which cohorts of vaccinated and non-vaccinated persons were followed over time to detect new cases of leprosy. The information about vaccination status is only available retrospectively (through the survey), but that does not change the character of the study. The follow-up time for different age groups may have been different (as young children were included as well), and the follow-up periods referred to different physical ages where detection of leprosy have different incidences. Therefore it would have been desirable if person-time information was available, but this is not the case. The relation between the two variables age and vaccination derives from the fact that vaccinations may have had different coverages in different age groups. Thus `age` may affect `scar` (vaccination), whereas the converse is clearly impossible. This leads to the causal diagram:



The primary epidemiological question is whether age acts as a confounding variable, or possibly a moderating variable, for the relation between vaccination and leprosy.

The first Stata listing gives an analysis of the association between vaccination and leprosy; this is the relevant analysis. The shown measure of association is the odds-ratio (OR); other measures could have been chosen, but we will work with the OR. Our analysis has the following steps:

- The M-H test for homogeneity ($P = 0.82$) and the stratum-specific ORs show that there is no indication of an interaction between age and vaccination; thus, `age` is not a moderator variable.
- The M-H OR estimate differs somewhat from the crude OR; the measure of relative deviation can be computed in different ways: $(0.653-0.477)/0.477 = 37\%$, $(0.477^{-1}-0.653^{-1})/0.477^{-1} = 27\%$, or perhaps best as $(\ln(0.477) - \ln(0.653))/\ln(0.477) = 42\%$, but all of these indicate a substantial deviation.
- The association between age and leprosy in the non-exposed group would here most naturally be computed in the vaccinated group: $\text{OR} = (17 \cdot 43169)/(84 \cdot 2859) = 3.06$, so old age seems to a strong risk (the same conclusion is obtained from the non-vaccinated group: $\text{OR} = (83 \cdot 24993)/(76 \cdot 9601 = 2.84)$. With the large numbers involved, there is little doubt about the statistical significance. Alternatively, these estimates and 95% CIs can be read off the listing for the M-H analysis stratified by scar.
- The association between age and vaccination is evaluated across the entire population (as this is a cohort study): $\text{OR} = (2876 \cdot 25069)/(9684 \cdot 43253) = 0.17$, so older people have been vaccinated less (which is also obvious from the data). Again, the statistical significance is obvious.

- As all conditions for age to be a confounder for the relation between vaccination and leprosy have been met, we conclude that age is indeed a confounder. By the strength of all all associations explored, we could say that age is partial confounder.

- The relevant measure of association for the vaccination effect is the M-H adjusted odds-ratio, OR = 0.653 = 1/1.53. Vaccination is protective against leprosy, and reduces the risk (as measured by the OR) by an estimated 53%. The effect is strongly significant (different from the null).

b) Sample data

The study is now a case-control study because a sample has been drawn from the non-diseased population. It is not stated how the sample was drawn, but we will explore this as the first step of our analysis below. The causal diagram is unchanged from *a)*.

- If the sample from the population was a random sample, we would expect the distribution of counts on the four groups formed by age and vaccination to be similar to that from the survey. It is seen that the age distribution is very different, with relatively more older persons in the sample than in the population. On the other hand, the distribution for vaccination within each age group seems somewhat similar to that of the survey.

- If age was considered a potential confounder prior to analysis, one of the methods of dealing with the confounding of age would be to match cases and controls on age, that is, to select a sample of controls with the same age distribution as the cases. Among the cases, there are $76 + 84 = 160$ persons in age group 0, and $83 + 17 = 100$ persons in age group 1. In the sample, there are $214 + 426 = 640$ persons in age group 0, and $300 + 100 = 400$ persons in age group 1. That is, the sample is frequency-matched by age by taking four times as many controls as cases in each age group.

- A frequency-matched case-control design eliminates confounding (by `age`), but creates instead a selection bias because the distribution of the exposure (`scar`) is altered from the full population of controls. Therefore, a M-H analysis stratified by the matching variable (`age`) is still appropriate.

- The M-H analysis stratified by `age` still shows no indication of interaction ($P = 0.77$ for the homogeneity test).

- The M-H analysis shows that the crude estimate is biased towards the null, as expected theoretically. Thus, the M-H stratified estimate is preferred: OR = 0.572 = 1/1.75. Vaccination is protective, but now estimated to decrease risk by 75% (up from 53% in Part *a)*). The confidence intervals for the M-H estimates based on the survey and the sample indicate that the difference between could very well be due to chance.

- There is no pressing need to explore the association between age and vaccination among the controls (it should be similar to that in the population), or the association between age and leprosy in the non-exposed group (the age-matching destroys this distribution so that age effects can no longer be assessed).