Joe Brew
Homework 5
PHC 6711

Read the attached article on maximization of response rates.

1. Suppose the cohort study of physical activity and colon cancer were to use a mailed questionnaire for follow-up data on vital status and colon cancer incidence. It is particularly important to have a high response proportion in follow-up data collection in cohort studies because non-response is likely to be related to outcome (and possibly to the exposures of interest). What techniques at baseline could help maximize response to follow-up information? What additional techniques would you use to maximize the response proportion to the follow-up questionnaire? Include techniques you would use to find those subjects who were "lost to follow-up" (e.g. questionnaire returned by post office due to person not at that address).

a. Techniques at baseline that could help maximize response to follow-up information
The authors found that several techniques could maximize response to follow-up information. These include financial incentives, "personalization," and the use of colors. They also included techniques such as recorded/registered delivery, the use of a previously stamped return envelope, contacting participants *before* sending questionnaires, and providing back-up questionnaires. Also, the source is of importance - if the questionnaire is sent from a university, it is more likely to be returned than if sent from a commercial entity.

b. Additional techniques to maximize the response proportion to the follow-up questionnaire
By making a questionnaire of more personal "interest" to the recipient, researchers can increase the likelihood of that questionnaire being returned. Inversely, by asking for senstive information, the likelihood of return lessens.

2. Heilbrun et al (1982) report data from the Honolulu Heart Program on smoking and lung cancer. 11,136 Japanese men were identified on the island of Oahu in 1965. 5,844 responded to a mailed questionnaire and were examined in person, 1,263 responded only to the mailed questionnaire and 1,259 did not respond at all. The remaining men were excluded due to death before Jan. 1, 1969, former smoking, or missing smoking information. Odds ratios of lung cancer for smoking among those who were and were not examined in person are compared in Table 2 in their paper (modified below). Ignore issues relating to the small numbers in some cells.

Table 2

Crude odds ratios of lung cancer in
examined and unexamined Japan-Hawaii
Cancer Study men according to smoking status*
based on the mailed questionnaire.

Cases identified on Oahu, Hawaii, 1969-1978.

Smoking Status	Total no. of men	cases	No. of ratio	95% Odds interval	confidence
Examined men					
Cigarette smoker	3,475	52	18.0	(5,75)	
Nonsmoker	2,369	2			
Unexamined men					
Cigarette smoker	818	12	6.6	(.8,50)	
Nonsmoker	445	1			
Total					
Cigarette smoker	4,293	64	14.2	(5,45)	
Nonsmoker	2,814	3			

*Excludes ex-cigarette smokers.
(Modified from Heilbrun et al to present crude odds ratios. These are very similar to the age-adjusted odds ratios.)
Reference: Heilbrun LK, Nomura A, Stemmerman GN: The effect of nonresponse in a prospective study of cancer. Am J Epidemiol 1982; 116:353-363.

a. Use the data in Table 2 to estimate examination participation rates in those who completed the questionnaire for each of the following groups:

(a) Cigarette smoker, developed lung cancer;
52/64 = 0.8125

(b) Cigarette smoker, did not develop lung cancer;
(3475-52) / (4293-64) = 0.809

(c) Nonsmoker, developed lung cancer;
2 / 3 = 0.667

(d) Nonsmoker, did not develop lung cancer.
(2369 - 2) / (2814-3) = 0.842

b. Show how the odds ratio for lung cancer in relation to smoking in all the men who completed the questionnaire can be derived from both the odds ratio in those who were examined and the response rates calculated above.

(64/(4293-64)) / (3/(2814-3)) = 14.18 is the “odds ratio for lung cancer in relation to smoking in all the men who completed the questionnaire”.

Since that answer, and those numbers, are already given in the table, I’m not sure why we would need to “derive” anything - the OR for all men who completed the questionnaire is apparent - perhaps the harder question is the OR for those that did NOT complete the questionnaire?

3. Often in case-control studies some potential cases will have died before an interview can take place. One can exclude those cases or include them by use of surrogate respondents. One should decide whether or not to include surrogate respondents, based on the effect of inclusion or exclusion of dead cases on selection bias and misclassification bias. Suppose in a case-control study of lung cancer in relation to an occupational exposure, 40% of cases die before they can be interviewed. Assume that it is known that surrogate respondents overestimate the occupational exposure, with the amount of error (sensitivity and specificity) independent of whether the surrogate is responding for a case or a control. Also assume that the occupational exposure is a risk factor for lung cancer (OR >1.0) and that among those with lung cancer, the exposure is a risk factor for death from lung cancer. Assume that there are no other sources of selection or misclassification bias. For each of the following situations, what is the primary factor that might bias the results of the study (differential misclassification, non-differential misclassification or selection bias) and in which direction would the bias in the odds ratio be (toward the null, away from the null, or toward the null with possible cross-over).

a. The cases in the study are limited to cases alive.
If only cases alive are allowed in the study, the odds ratio will go TOWARD the null, since it will underestimate the number of cases in the exposed group (artificially deflating the numerator in their odds).

b. The study uses surrogate respondents for dead cases.
In the case of the use of surrogate respondents, the odds ratio will go AWAY FROM the null, since it will overestimate the exposure without affecting the outcome

c. The study uses surrogate respondents for dead cases and for 40% of the controls (alive).
In this case it is hard to say in which direction the odds ratio will go. I suspect that it will affect the OR in the sense that exposure for controls will be overestimated (ie, OR towards the null), but exposure for the 100% of surrogate use in dead cases will offset and eventually overcompensate for this initial outcome (ie, OR away from the null).

d. Which approach would you take?
Since we know that surrogate behavior is independent of the group in question, we could safely use surrogates for ALL observations, thereby eliminating potential sources of bias.