Sources of error - Assignment

1. The continuing controversy over the health effects of low frequency magnetic fields began with an epidemiologic study of electrical wiring configurations and childhood cancer (Wertheimer N. *Am J Epidemiol* 1979; 109:273-284).

   Cases consisted of persons dying of cancer in Colorado before age 19 in the years 1950-1973 who also had a Colorado birth certificate and whose birth or "death" address was in the greater Denver area and had been occupied from 1946-1973. Controls consisted of next (non-sibling) Denver-area birth certificates. Residential status for cases and controls is shown below.

   For all study participants, birth and death addresses were visited and a map of the wires and transformers of the electric power distribution was drawn. Homes were categorized as having "high-current configurations" (HCC), "low-current configurations" (LCC), or "very low current configurations" (VLCC), according to their proximity to high current distribution lines. Table 2 shows the distribution of expected current of residence addresses for cases and controls:

Table 1
Residential status of cases and controls

Residential status	Cases	Controls
Stable, same birth & death address	109	128
Moved, birth & death addresses available	145	128
Only one address available, either birth or death	88	88

Table 2
Case-control distribution for the amount of current expected from different wiring configurations, based on all known addresses for study participants

Wiring configuration	Expected current	Cases	Controls	% Cases
HCC	High-very high	182	103	64
LCC	Low	289	324	47
VLCC	Very low	20	45	31

   a. Calculate the odds ratios comparing HCC to LCC houses, and LCC to VLCC houses.

   b. What is the meaning of the percentages in the right-most column? Are they incidences? Do they indicate a dose-response relationship between electric current and cancer occurrence?

   c. Identify likely sources of selection and information bias in this study, with particular attention to those sources that are common in case-control studies. Can you suggest methods to minimize these sources of bias? (Use your imagination—you are not expected to consult the article, though if you wish to do so, try answering this question first).

   d. What was the purpose of selecting controls by taking the next birth certificate?

e. OPTIONAL: If the reported cumulative incidence (CI) of childhood cancer (by age 19) is 10 cases per 10,000 children in the general population, what would you estimate the cumulative incidence to be among children living next to HCC's assuming that one out of five children live next to HCC's? What proportion of childhood cancer might be attributable to HCC's?

2. This question is based on the attached extract from Rosenberg et al., "Oral contraceptive use in relation to nonfatal myocardial infarction". (*Am J Epidemiol* 1980; 111:59-66).

   a. What type of study design has been used?

      A. case-control with prevalent cases

      B. case-control with incident cases

      C. prospective cohort

      D. retrospective (historical) cohort

      E. ecologic

      F. case-control nested in a cohort

   b. Give 2 possible sources of selection bias that might interfere with using these data to obtain estimates of the relative risk of MI in women taking oral contraceptives. For each source, give an example of how it might cause the true relative risk to be overstated in the data.

   c. Briefly assess the likelihood of misclassification of the outcome measure, reported MI.

   d. Briefly (2-3 sentences) assess the likelihood of misclassification of the exposure measure, reported oral contraceptive use.

3. The issue of "detection bias" sparked a vigorous controversy in the investigation of the relationship between endometrial cancer and the use of exogenous estrogen preparations. The case for the importance of "detection bias" was presented by Horwitz and Feinstein (Alternative analytic methods for case-control studies of estrogens and endometrial cancer. *N Engl J Med* 1978; 299:1089-1094). The following questions refer to that article and Horwitz and Feinstein's detection bias argument.

   a. Give a definition of the term "detection bias" as it is applied by Horwitz and Feinstein to studies of the endometrial cancer-exogenous estrogen relationship.

   b. Would "detection bias" tend to overstate or understate a truly positive association between estrogen use and endometrial cancer? Explain briefly.

   c. Is "detection bias" (in the above sense) a form of selection bias or information (misclassification) bias? Justify your position.

   Consider "detection bias" to be a form of selection bias, and let alpha, beta, gamma, and delta be probabilities by which individuals in the target population are included in a study, according to the following scheme:

Alpha (α) is the probability by which individuals with endometrial cancer and a history of estrogen use are included in the study;

Beta (β) is the probability by which individuals with endometrial cancer but without a history of estrogen use are included in the study;

Gamma (γ) is the probability by which individuals without endometrial cancer but with a history of estrogen use are included in the study;

Delta (δ) is the probability by which individuals without endometrial cancer and without a history of estrogen use are included in the study.

Selection probabilities of inclusion in the actual population, by disease and exposure characteristics of individuals in the target population

	E	$\overline{E}$
D	α	β
$\overline{D}$	γ	δ

d. Assuming that no other source of selection bias is present, what relationship among or between selection probabilities most closely characterizes the detection bias situation described by Horwitz and Feinstein? Justify your answer.

e. Characterize, in terms of the above selection probabilities, the impact of the approach adopted by Horwitz and Feinstein to correct for the effect of "detection bias" (i.e., their use of "alternative controls"). Justify your answer.

f. What is an alternate, and presumably theoretically preferable, approach to avoiding detection bias and how would it be characterized in terms of the above selection probabilities? Comment on the practicality of the preferable approach.

4. The major community studies of CVD, such as the Framingham study, began before the availability of exercise ECGs, echocardiography, and other sophisticated methods of detecting CHD. For example, in the Evans County Study (Cassel JC *et al.*, *Arch Intern Med* 1971 (December); 128 [entire issue]), CHD case detection in living subjects was accomplished using clinical judgment based on history, symptoms, physical examination, resting ECG, and chest X-ray. Opportunities for bias from misclassification include both an incorrect exclusion decision at enrollment (so that a subject later diagnosed as having CHD may have been a prevalent case at the outset and therefore not a new case) and misclassification of CHD status at follow-up.

   The following questions are based on data from the Evans County Cardiovascular Disease Study (Cornoni JC, Waller LE, Cassel JC, Tyroler HA, Hames CG. The incidence study—study design and methods. *Arch Intern Med* 1971 [December];128:896-900):

Results from the Evans County Cardiovascular Study

Persons examined in 1960-62	3102
CHD cases at initial examination	93
Died before 1967-1969 reexamination (includes 36 deaths among those with CHD in 1960-62 and 56 CHD deaths among those without CHD in 1960-62)	320
Moved, vital status could not be determined	40
Known alive, but not re-examined (migrated or refused)	212
Re-examined in 1967-69 (includes 57 subjects who had CHD in 1960-62)	2530
CHD cases detected at reexamination in 1967-69 among survivors initially free of CHD	87

a. Diagram the above data.

b. Calculate the observed (87-month) cumulative incidence of CHD in the Evans County incidence study (exclude subjects who migrated, refused re-examination, were lost to follow-up, or died of non-CHD causes).

c. Estimate the "true" incidence of CHD between 1960-62 and 1967-69 that would be expected if the sensitivity and specificity of the diagnostic procedure in the Evans County study were, respectively, 70% and 98%. Assume that the 93 CHD cases detected in 1960-62 had in fact constituted all those and only those with CHD at the time and that there was no misclassification of cause of death for subjects who did not survive until reexamination.

d. *Optional*: It is reasonable to suppose that the sensitivity and specificity in 1960-62 would have been worse than in 1967-69. What is the lowest that specificity could have been given the data (you may set sensitivity at any level you like).