Analytic study designs - Assignment solutions

Part I

1. A case control study design was employed. Subjects were selected on the basis of their having or not having the disease of interest (i.e., selected cancers). A strength of this particular case control study, however, is that exposure information had been obtained before the diagnosis was known.

2. The case-control design has a number of subtleties, and a full understanding of the central issue of control selection came only after several decades of case-control studies. The principle for selection of the control group is that it should permit estimation of the exposure distribution in the study base, i.e., the population from which the cases arose. However, previously it was thought that, on analogy with an experimental comparison, the cases and controls should be "comparable" or "as similar as possible except for their disease status".

   Belief in the importance of comparability between cases and controls has probably been the rationale for control selection procedures in many published studies and perhaps also for the widespread use of matched controls. There is a role for matching in selection of controls, primarily where a known strong risk factor is prominent in the cases but much less frequent in the study base. In this situation, most cases will have the known risk factor (e.g., smoking in a study of lung cancer) but most controls in an unmatched control group will not. That difference will make it difficult or at least inefficient to compare the exposure distributions of cases and controls while taking account of the known risk factor. Selecting the control group in a way that forces it to have the same distribution of the established risk factor as will be found in the case group avoids this problem. However, if the matching criterion is not a strong risk factor, then there is the danger that a matched control group will have an exposure distribution that is more like that of the case group than is the exposure distribution in the study base, a situation referred to as "over-matching".

   There are other considerations that enter into the selection procedure for controls, however, including feasibility, cost, anticipated response rate, and availability of exposure information. For example, people selected from the general population may not be able to provide their exposure status (e.g., if measuring it involves a costly or invasive procedure, such as coronary angiography, or access to records they will not have, such as workplace exposure to various chemicals). Also, if the condition under study is not rare but requires a costly or invasive procedure to detect, a control group randomly selected from the general population could include within it a meaningful proportion of people with the condition. So the actual selection procedure often represents a compromise between the objective of constructing a window into the study base and constructing a window that is clear enough to see through, even if one knows that the view through it may not quite be the study base. Thus, even though they are not fully representative of the population from which cases arose, hospital controls may be more motivated and available than community controls and may provide opportunities for detection of exposure more similar to those for cases.

In this study the use of hospitalized controls produces an overestimate of the prevalence of smoking in the population from which cases arose. The reason is that many of the controls are symptomatic for diseases associated with smoking. Thus, this control group has a greater proportion of smokers than would a general population sample and the contrast with the lung cancer case group is therefore diminished, resulting in a negative bias in the measure of association, similar to Berkson's or referral bias. (Note: this is *not* what was meant by "Briefly"!)

3. E. Selective Recall - Cases generally have an increased motivation or tendency to recollect exposure to a putative causal agent. To avoid bias, it is helpful if controls have a similar stimulus to remember and report.

4. This statement is imprecise in the present context. In a case control study, such as this, one directly estimates and compares the proportion of exposed (smokers) among the cases to the proportion of exposed among the noncases, not the proportion of diseased among the exposed and unexposed as implied in this statement. The author's statement would be more appropriate to a cohort design or to most cross-sectional studies.

5. "E" - none of the statements is correct: Statement "C" comes the closest to describing the data presented in Table 4, though its usefulness is doubtful in a case control study. Interpretation of findings would be more appropriate in terms of the proportion exposed among the cases and controls. [Statement "C" is not actually correct because Table 4 does not deal with all smokers and nonsmokers admitted to Roswell Park, but only with those with certain cancers or symptoms referable to those same sites (see first paragraph of second column.]

Part II

1. A. case control

   - Exposure variable can be measured more reliably, with no danger of its measurement being influenced by the disease outcome;
   - Selection bias in the choice of a control group is avoided, while selection bias from loss to follow-up can often be controlled or at least its likelihood guaged;
   - The temporal sequence (i.e., that exposure precedes disease) is clearly demonstrated;
   - Risk and rates can be estimated directly.

2. B. Cohort -- The investigators measured initial exposure status (by survey) before death (and ideally before disease onset) and subsequent mortality experience by surveillance of death records.

3. a.

   - Better informed than general population
   - Motivated to respond to medical research

- Possible motivation to change behavior
- Busy schedule may lead to poor response rate from causes other than lung cancer
- Better medical care may decrease mortality from causes other than lung cancer.

Techniques to minimize problems of survey:

- Keep questionnaire brief
- Do not reveal research hypothesis
- Follow-up interviews with non-responders

b.

- Large sample required and lengthy follow-up, with attendant expense, effort, and delay in obtaining results
- Difficult to assess prior exposure history
- Poor response rate from initial survey
- Loss to follow-up.

4. a. Since both involve questionnaire responses, the differences are in: the timing of the survey and the quantification of exposure. Doll and Hill might be favored for less selective recall, whereas Levin et al. might be selected for a more thorough inquiry with more precise quantification of exposure. The "real" exposure factor is presumably some constituent or combination of constituents of tobacco smoke. The "true" agents could vary with type of cigarette, smoking practices, cigarette brand contents, and so on. (Cigarettes smoke contains thousands of substances, including many that are known carcinogens.)

   b. The assumption is that current smoking practice, which is more reliably measured, is a good indicator of previous smoking practice. A prior case-control study showed current smoking history was almost as great a risk factor for lung cancer as total smoking history.

5. Advantages:

- Routine notification is available;
- Easy access to records;
- Necropsy confirmation available in many cases.

Disadvantages:

- Differential survival may influence the findings, though this should be less of a problem with a rapidly fatal disease like lung cancer;
- Cause of death may be inaccurately recorded, particularly where multiple conditions are present;

- Follow-up must be longer for any given number of cases.

6. The distribution of smoking habits varies considerably with age, as does cancer incidence. Therefore, age adjustment or age-specific comparisons are needed.

7. To demonstrate that disease diagnosis was not based not only on clinical judgment but documented by examination of tissue specimens. Particularly with a rare disease, like lung cancer, false positives can readily distort the observed association, leading to information (misclassification) bias.

9. Yes: Prevalent cases may have failed to respond initially, and therefore been deleted from the cohort; short period of follow-up of disease with long induction; doctors may receive better medical care, thereby reducing mortality.

12. Ultimately, it is not possible to exclude all possible alternate explanations except by conducting an experiment in which smoking (or stopping smoking), can be randomly allocated in sufficiently large groups with sufficient adherence to the experimental regimen. For ethical and practical reasons, such an experiment cannot be conducted with smoking onset as the experimental variable. Smoking cessation programs have not been sufficiently effective to serve as an alternate intervention. If they were, it is doubtful that they could be used in a lung cancer trial, since the evidence for the harmful effects of smoking is so abundant that ethical considerations would probably preclude such a trial.

    Though "absolute" proof requires human experiments, the weight of observational and animal evidence is convincing. Any particular factor proposed as an alternate explanation (e.g., exercise, diet) can be examined simultaneously with smoking, to see which factor is associated with lung cancer when the other factor is controlled. There are no serious contenders. Traditional criteria for causal inference are well satisfied:

    a. strong association (RR about 8 for smokers overall);
    b. dose-response effect exists;
    c. replication in many studies;
    d. cohort studies demonstrate that exposure precedes disease;
    e. biologic explanation (animal and tissue culture models, autopsy studies);
    f. experimental confirmation available in animal models;
    g. analogy to other carcinogen-cancer associations.

    Smoking and lung cancer is one of the strongest of epidemiologically-established relationships. Controversy remains, but that should not paralyze policy planning (see M.A. Ibrahim, The cigarette smoking/ lung cancer hypothesis. Editorial. *Am J Public Health* 1976; 66:131-132). Nevertheless, the debate has continued even into recent years (see references in Bibliography).