Lab 6 – Instructors guide

Analytic Study Designs

Goals: understand, differentiate, and identify relative strengths and weaknesses of the various types of epidemiologic study designs

1. Read the following passages. Identify the study design and determine what measures of association can be estimated. Please justify your choices.

   a. Questionnaires were mailed to every 10th person listed in the city telephone directory. Each person was asked to list age, sex, smoking habits, and respiratory symptoms during the preceding seven days. About 20% of the questionnaires were completed and returned. About 10% of respondents reported having upper respiratory symptoms.

      Cross-sectional or prevalence study; can estimate prevalence and prevalence odds ratio.

   b. 1,500 employees of a major aircraft company were initially examined in 1951 and were classified by diagnostic criteria for coronary artery disease (CAD). New cases of CAD have been identified by examinations every three years and through death certificates. Attack rates in different subgroups have been computed.

      Prospective cohort study; can estimate incidence proportion (CI), incidence rate (ID), and their ratios (CIR, IDR) and differences (CID,IDD), the latter serving as measures of impact.

   c. A random sample of middle-aged sedentary adults were selected from four census tracts, and each person was examined for coronary artery disease (CAD). All persons without disease were randomly assigned to either a two-year program of aerobic exercise or a two-year arthritis-prevention non-aerobic exercise program. Both groups were observed semi-annually for incidence of CAD.

      Randomized controlled trial/intervention trial; can estimate incidence (CI and ID), relative risk (CIR,IDR), and measures of impact.

   d. A 39-year old woman presents with a mild sore throat, fever, malaise and headache and is treated with penicillin, for presumed streptococcal infection. She returns in a week with hypertension, fever, rash and abdominal pain. She responds favorably to chloramphenicol, after a diagnosis of Rocky Mountain spotted fever is made.

      This is a case report. It is not usually regarded as an "epidemiologic" study, since there is neither "population" nor comparison group. Presentation of an interesting case can serve to alert other health care professionals of the possibility of misdiagnosing a potentially fatal disease.

e. 50 patients with thyroid cancer are identified and surveyed by patient interviews to identify previous radiation exposure.

This study is a case series. Since there is no comparison group of people without thyroid cancer, associations between thyroid cancer and various exposures cannot be examined except on the basis of information from outside of the study. Exposure prevalences among cases can be estimated and comparisons between different subtypes of the disease can be made. Case fatality can also be estimated by follow-up.

f. Patients admitted for carcinoma of the stomach and patients without a diagnosis of cancer are interviewed about their chewing tobacco history to assess the possible association of chewing tobacco and gastric cancer.

Case-control study. This study includes both diseased and nondiseased populations for comparisons of previous exposure to chewing tobacco. Incidence rates (either overall or by exposure status) of cancer cannot be determined without additional information.

g. Data on median income for households in census tracts within a large metropolitan county in the U.S. were obtained from the Census Bureau's Current Population Survey. Air pollution levels were measured in these same census tracts during a period of one-month. The data were analyzed using a geographic information system (GIS) to produce maps showing pollution and income levels by census tract.

Ecologic/community/group level/correlational/aggregate study; the census tract is the unit of analysis. The proportion of counties with high pollution levels can be estimated as a prevalence-type measure, but more often these data would be analyzed with cross-tabulation, correlation, regression, and graphical methods.

2. You have developed the hypothesis that automobile drivers who regularly sleep less than 6 hours/night have a higher incidence of fatal automobile accidents. Think about the design of a case-control study to test this hypothesis. Consider:

   a. How will you define a "case"? Are there any special considerations?

   b. Where would you find the cases?

   b. Some states (including NC) have a fatal accident reporting system, which provides a great deal of useful information. Medical examiner records are another possible source. If cases are identified through death certificates, police records will be needed to differentiate between drivers and passengers.

   c. Name a suitable population from which to choose the controls.

c. From the state's motor vehicle licensing records, choose a sample of drivers in the state who have not had fatal accidents. The controls should be drivers (or have been drivers during the period when the deaths occurred), since that is the population from which the cases arose (the study base). There is the question of what should be done about cases who were driving despite having their license revoked. Should different controls be chosen for them?

d. What major characteristic must you strive to measure similarly in each study participant?

d. Information about sleep, of course, as well as safety features of the automobile, passenger restraint systems in use, road conditions, weather, driver age, sex, and medical history, and any other factor that influences crash rates and probability of fatality.

e. What difficulties will be encountered in measuring this characteristic?

e. Information on sleep for deceased drivers will certainly be difficult to obtain with any degree of accuracy. Since such information will need to come from proxy respondents (e.g., spouse, other family member, co-worker), should similar sources be used to obtain data for controls?

3. You have developed the hypothesis that improving school lunches in elementary schools will decrease obesity among the students. You have randomly identified 6 elementary schools in the state of North Carolina to participate in this study. Think about the design of a group-level intervention trial to test this hypothesis. Consider:

   a. What is your unit of analysis?

The unit of analysis is the elementary school participating in the group-level intervention.

   b. What is your unit of measurement? (Think about what sort of measurements you need to take in order to evaluate the effectiveness of your intervention).

The unit of measurement is the individual students because you will need to measure the height and weight of each student.

   c. You decide to assign the elementary schools to control or intervention group by randomization. What sorts of problems could arise? What is another method of assignment that might work better?

Because you only have 6 schools, you do not have many units to randomize. Consequently, there is a high probability that the control and intervention groups won't be well balanced on important characteristics (for example, baseline level of obesity, average SES of the study body, size of the school, etc.) Another method of assignment would be to form three pairs by matching the schools on important characteristics. Then you could use randomization to

assign one member of each pair to the control group and the other member to the intervention group.

4. What is the importance of randomization in an intervention trial, and what does it accomplish?

   Randomization is important because it increases the likelihood that differences in outcome between groups can be attributed to the treatments applied. If randomization is not used, the differences in outcome may be due to differing characteristics in the groups being compared. Randomization of a large number of persons achieves equal baseline distributions of known (e.g. age, sex, severity of disease), unknown, and unmeasurable risk factors. Of course, equivalence at baseline does not ensure equivalence throughout the trial.

5. What is meant by the phrase "ecologic fallacy"?

   The "ecologic fallacy" is the inference that the individuals in a group share the characteristics of a larger population. Groups of individuals may differ greatly from the larger group. An association between two group-level measures does not imply an association between the two measures at the individual-level. If, however, the factors under consideration belong to the group, rather than to the individuals so that inference is not being made to the individual-level, then the potential for an ecologic fallacy does not arise.

6. Different study designs have particular advantages and disadvantages. Contrast the case-control and cohort designs with respect to the following factors, for a study collecting new data.

   a. Cost.

   b. Time required for completion of study.

   c. Efficiency (in terms of information per subject)

   d. Design issues

   e. Difficulty in obtaining information.

   f. Bias

   g. What can be estimated