March 1994

Analytical study designs--II

M A. Lobo
Aga Khan University

A F. Qureshi
Aga Khan University, fozia.qureshi@aku.edu

Analytical Study Designs - II

Melvyn A. Lobo, Asma Fozia Qureshi (Department of Community Health Sciences, Aga Khan University, Karachi.)

"Is exposure to a specific risk factor related to development of disease?' is a question often asked by researchers. The case- control study design determines whether an association exists between disease and a suspected risk factor. The prospective cohort study design described below, is another way to examine such an association.

What is a Cohort Study?

Participants in cohort studies are categorized on the basis of presence or absence of exposure to a suspected risk factor for a disease. The investigator first selects exposed persons (persons exposed to the risk factor) and unexposed persons (persons without exposure to the risk factor). At the time exposure status is defined, subjects must be free from the outcome of interest such as lung cancer in smokers. The researcher next follows study subjects over time to assess the occurrence (incidence) of that outcome.

When Is it Appropriate to Select a Cohort Study Design?

The cohort study is best suited for analyzing the relationship between a risk factor and disease when: (i) the period between exposure and outcome is short, (ii) examining multiple effects of a single exposure and (iii) exposure is rare but outcome is common. The follow-up period in cohort studies depends on the hypothesized latent period between exposure and outcome. Latency could vary, depending on the dose of exposure and whether the subject has recently been exposed to the risk factor or been exposed to it over time. Follow-up provides empirical information about the latency period, i.e., a temporal telationship between exposure and disease, unlike case-control and cross-sectional studies where temporal sequence is a problem as it is often difficult to determine whether the risk factor preceded the outcome or vice- versa. In a cohort study where subjects are followed-up 'prospectively', researchers are limited by the time required for the outcome to occur. The 'historical' cohort study design overcomes this problem as relevant events have occurred at the start of the study. Exposed and unexposed groups are constructed on the basis of data from historical records and the outcome(s) measured currently. Though efficient for inyestigating disease with long latency periods, this design may be limited by records lacking the required information.

How Do I Conduct a Cohort Study?

Unlike the case-control study, the cohort study is a 4 step process.

Step 1. Eligibility of study subjects:

The researcher first determines the cohort of people to be studied. This could be, for example, a group of people born during the same time period (birth cohort) or persons working in a factory (occupational cohort). The cohort could consist of all those present or a sample of them. The study subjects should only include those that are free from the outcome of interest.

Step 2. Selection of the exposed population:

i. Definition of exposure: To prevent a distortion of results, strict inclusion criteria should be used to identify individuals to be included in the exposed group. For example, in studying the effect of cigarette smoking on development of lung cancer, the risk could be masked if occasional smokers were included with habitual smokers in the exposed group of individuals.

ii. Selection of exposed individuals: Exposed persons can be identified through interviews, medical records, medical examinations and laboratory results. A large number of exposed persons can be obtained for common exposures such as cigarette smoking. For rare exposures (occupational or environmental factors), it is more efficient to choose a group with a high probability of exposure (e.g., workers in a paint factory to study the effect of luminous paint on developing bladder cancer) to obtain

the required number of persons who will experience the outcome of interest.

Step 3. Selection of the unexposed comparison group:

i. Definition of unexposed: This group should be identified by strict criteria to prevent misclassification.

ii. Selection of unexposed individuals: These are persons unexposed to the risk factor, who can serveas a comparison group. The comparison group could be:

(a) internal comparison group: those determined to be unexposed from the initial cross-sectional/screening procedure applied. These individuals are likely to be similar to the exposed population in basic demographic and geographic characteristics except for exposure.

(b) external comparison group: the general population, with the assumption that the proportion of exposed individualsin this group is very small. The true risk will be under-estimated if a sizeable proportion of the population is in fact exposed to the risk factor under investigalion. For this reason, comparison of an employed group of exposed persons with the general population will under-estimate the true risk, as working persons are generally screened prior to employment and are on average healthier than the general population: the 'healthy worker effect'.

(c) other study groups: a single group or multiple groups of employed persons can serve as the comparison group especially to overcome the problem of using the general population for comparison. The researcher should try to ensure comparability between the exposed and unexposed.

Step 4. Ascertainment of outcome:

The goal is to obtain comparable, unbiased information on the subsequent health experience of every study subject. If the outcome of interest is death, outcome information for all members of a cohort may be obtained from death certificates or interviewing family members. For nonfatal outcomes, data can be obtained from medical/laboratory records or directly from the participants. Information from records helps to confirm the diagnosis and thus overcomes the possibility of bias due to a subject's awareness of the hypothesis under investigation. In the cohort study, ascertainment of outcome data involves the follow-up of all study participants from exposure into the future, to determine whether they develop the outcome. Failure to obtain such data on every subject, or collecting it for a greater proportion of individuals in either the exposed or unexposed group, is the major source of cohort bias. For example, if the proportion of those lost to follow-up due to death or migration is greater in smokers than non-smokers, the occurrence of lung cancer cases may artificially be equal in both groups. The researcher could thus erroneously conclude that there is no added risk of lung cancer in smokers.

How Do I Analyze Data from the Cohort Study?

The exposed and unexposed populations should first be compared to ensure similarity with respect to known confounding factors. The measure of risk calculated from a cohort study is the Relative Risk which describes the increased risk of developingthe disease in the exposed compared with the unexposed. As initially disease free individuals are followed over time, incidence (new cases) rates can be calculated for both the exposed and unexposed groups, for those exposed to different doses of the factor or to a combination of factors. Incidence rates generally cannot be calculated with case-control studies, as the case might have had the disease before exposure to the risk factor.

How Do I Calculate the Relative Risk?

First create a two-by-two table as shown above. Second, calculate incidence of disease in the exposed and unexposed groups. Third, divide the incidence in the exposed (le) by the incidence in the unexposed (Io).

How Do I Interpret the Relative Risk?

	Diseased	Not Diseased
Exposed	**Cell A** Exposed persons who developed the disease	**Cell B** Exposed persons who did not develop the disease
Unexposed	**Cell C** Unexposed persons who developed the disease	**Cell D** Unexposed persons who did not develop the disease

There are three possibilities when calculating the relative risk:
1. A relative risk of one: the risk of developing the disease is equal for the exposed and unexposed subjects, i.e., as the risk in both groups is the same and the risk factor is not associated with the outcome.

$$\text{Relative Risk (RR)} = \frac{I_e}{I_o} = \frac{A/(A+B)}{C(C+D)}$$

2. A relative risk of more than one: the risk of developing the disease is that many times more in the exposed than in the unexposed, e.g., a RR= 4 signifies a fourfold higher risk in the exposed.
3. A relative risk of less than one: the risk of developing the disease in the exposed is that many times less than in the unexposed.
Issues In Interpretation
As with other methods of epidemiologic study, the researcher needs to evaluate the validity of study results. Three factors commony responsible for distorting results in cohort studies and need to be considered are:
i. Misclassification, i.e., inaccuracy in classifying
study subjects with respect to exposure and disease status.
ii. Loss to follow-up could result from death, migration or infrequent follow-up. Validity of the study is called into question if this proportion is large i.e., above 30%, or if the loss was greater in one group.
iii. Non-participation could give rise to selection bias if subjects who agree to participate differ from non-participants with respect to exposure and other risk factors that affect the outcome of the study.

References

1. Hennekena, CH. and Buring J.E. Edited by S.L Myrent. Epidemiology in medicine. 1st S. Boston, Little Brown and Company, 1987.

2. Mausner, J.S. and Kramer, S. Epidemiology, an introductory text, 2nd ed. Philadelphia, W.B. Saunders Company, 1985.

3. Health Research Methodology: a guide for training in research methods, education in athion. Western Pacific, WHO Regional Publications, Series No.5.