8

Contribution of the case-control method to health program evaluation

Myriam Khlat
Institut National d'Etudes Démographiques (INED)
Paris, France

The general purpose of evaluation is the comparison of different groups of individuals with respect to the rate of occurrence of a particular outcome. The main difficulty in this undertaking is the non-comparability of the groups concerned as regards a number of factors likely to influence the outcome. Randomized trials are considered to be the ideal type of design in evaluation research, as they enable the avoidance of most types of biases in the comparison, but they have to be set up before the program is implemented and may prove too costly.

When a program has been in operation for a period of time, and when this program does not cover the entire population evenly, then the question is: is evaluation still feasible, and what are the alternatives to randomization? There are in fact alternative designs, and while they are certainly less conclusive they can nonetheless provide valuable clues. In particular, there is the case-control design which is rarely used by demographers, yet which is potentially very informative when properly planned and interpreted (Greenland, Watson and Neutra, 1981; Rhoads and Mills, 1984; Baker and Curbow, 1991; Selby, 1994).

In this chapter, the application of the case-control method to the evaluation of the demographic impact of health programs is presented and illustrated. The specific strengths and weaknesses of this type of study are highlighted, and the solutions to common problems of ordinary case-control research are discussed.

Background

Case-control studies have at times been called by other names, including: retrospective studies, case-referent studies, case-comparison studies, case-compeer studies, and trohoc studies. The term case-control did not come into general use until the 1950s, and case-control studies are now by far the type of epidemiological study

most frequently undertaken, as a less expensive and less intrusive approach than the cohort and community study techniques favored in the early days of the discipline.

The first case-control studies to be carried out examined the relationship between smoking and lung cancer, and for this purpose a series of lung cancer cases were compared with a series of non-cases of similar age, sex and socio-economic status. The main finding to emerge was that of a much higher proportion of smokers in cases than in non-cases; based on the observed difference, an estimate of the risk of lung cancer in smokers versus non smokers was derived.

More generally, in a case-control study, persons with a given disease or condition (designated as the " cases ") and similar persons without the disease or condition (designated as the " controls ") are selected. By comparing the members of the two groups on their present or past characteristics, behavior or experience, one can examine the association of these characteristics with the condition under study. The statistical expression of this comparison is an estimation of the risk of disease given exposure to such or such a factor, relative to the risk of disease given no exposure, known as the " relative risk ", that is obtained using appropriate statistical techniques.

How does this concept of relative risk fit in the framework of health program evaluation? If the condition under study is " death from a specific group of causes " (for example death from breast cancer in women aged 35 to 54), and if the characteristic of interest is " exposure to a health program " (for example mammographic screening), then the relative risk can be viewed as: the ratio of the cause-specific death rate in those individuals exposed to the program (women who have undergone regular mammographic screening) to the cause-specific death rate in those individuals not exposed to the program (women who have not had mammographic screening).

Program evaluation entails the statistical testing of the null hypothesis that the program does not affect the cause-specific death rate, against the alternative hypothesis that the program does reduce the cause-specific death rate. Under the alternative hypothesis, cases differ from the controls in their histories of exposure to the health program, with a greater proportion exposed in controls than in cases, resulting in a relative risk estimate significantly lower than 1. If this is so, then it can be concluded that the study results support the alternative hypothesis, and therefore that the program is successful in preventing deaths from the condition of interest, at the significance level adopted for the test. Based on the magnitude of the relative risk, the reduction in risk of death associated with exposure to the health program can then be estimated.

A step-by-step approach to implementing this design to the evaluation of the impact of health programs on mortality is presented below, using the example of a published case-control study (Horwitz and Feinstein, 1981) to illustrate and clarify each of the practical stages considered.

Definition and selection of cases

The proper constitution of the group of cases is essential when conducting a case-control study. It is particularly important that cases be homogeneous with respect to the condition under study, and for this purpose objective criteria of eligibility have to be specified for inclusion of cases in the study.

So as to ensure homogeneity of diagnostic criteria or certification rules for cause of death, it is usual to choose newly detected occurrences of the condition of interest for the series of case. This is the reason why cases or deaths generally enter the study as they are diagnosed or certified until the required sample size is reached, unless the study is dealing with very rare conditions, when it may be necessary to include conditions identified in the recent past. In addition, a sampling procedure is not usually needed, as case-control studies make use of all eligible cases appearing during a particular period of time. Potential cases may be identified from a number of different sources: hospital registries and other medical care institutions, disease registries, occupational sites, and in the community.

The particular case-control study considered here concerns the administration of anticoagulants to patients suffering from myocardial infarction. At the time the findings of this study were published, the benefits of anticoagulants in reducing mortality in the hospital management of these patients were controversial despite six randomized trials, of which one had shown a significant mortality reduction, and the other five had found negative results, possibly due to insufficient sample size. At that stage, a case-control approach was envisaged, and cases were selected from among the following group: patients who were hospitalized in the coronary or intensive care unit of the Yale-New Haven Hospital, and who died in the hospital with a diagnosis of acute myocardial infarction between January 1, 1974 and October 31, 1978.

Definition and selection of controls

Choosing the cases is to solve only half of the problem since by definition a case-control study is comparative, which poses the question of the selection of controls. As the number of eligible controls is typically greater than the number of eligible cases, sampling is more likely to be necessary for controls than for cases. Choice of the most appropriate control group is one of the most difficult and controversial aspects of study design, and it is probably far more difficult than the choice of the case group.

In experimental designs, the " control " group is the one which is not treated, and subjects from the same source population are assigned by random allocation methods to either the experimental group or to the control group. With a sufficient group size (at least 40 to 50 subjects per group), the random allocation ensures that the two groups are comparable in terms of age, sex, and the usual socio-demographic variables, and that the only difference between them is exposure to treatment.

In case-control studies, " controls " are actually persons without the disease, and comparability of the cases and the controls may be ensured either at the design stage or at the analysis stage, the former being more frequent in practice than the latter. At the design stage, the procedure which is used is referred to as "matching", which consists of selecting a comparison group with a parallel distribution on co-factors to the case group. Two types of matching may be used:

- individual matching: a control subject is chosen for each case subject with the same relevant attributes (age, sex, occupation, etc), which leads to the constitution of a series of matched pairs;

- frequency matching: the comparison group of subjects is chosen so as to parallel the case group in terms of overall distribution on matching factors. This type of matching involves the calculation of the observed number of cases within each level of the factors to be matched on, and then the selection of the appropriate number of controls from the pool of potential controls, in order to fill the quota for each category.

Frequency matching is more economical than individual matching, but in certain circumstances, such as the choice of neighborhood controls, individual matching is the only workable approach. At the analysis stage, an adjustment for the co-factors may be done, using post-stratification or regression analysis.

The four most commonly used control groups are: probability samples of the population from which the cases came; persons seeking medical care at the same institutions as the cases for conditions believed to be unrelated to the cases' diagnosis; visitors in hospital settings, or; neighbors of the cases.

The choice of which control group to use is generally dictated by the source of the cases, the relative costs of obtaining the various types of controls, and the facilities available to the investigator carrying out the study. The use of multiple control groups is often considered to be helpful in avoiding selection bias: one group is generally selected from the same source of care as the cases, and another group drawn randomly in the population or chosen from the same neighborhood to control for socio-economic differences (Ibrahim, 1985).

In the case-control study on anticoagulants and myocardial infarction, the controls were selected from among the patients who were hospitalized in the coronary or intensive care unit of the Yale-New Haven Hospital, and who were discharged between January 1, 1974 and October 31, 1978. From the 2,229 survivors, one patient was matched as a control to each fatality according to nearest date of hospitalization, age (within 4 years), gender and race.

Power considerations

The number of cases and controls to be included in the study is a statistical problem of sample size determination for the comparison of proportions. The larger the numbers included, the greater the power of the study in detecting differences between cases and controls in terms of their prior exposure to the health program being tested.

In order to calculate the minimum size required, the following parameters have to be considered: a rough estimate of the proportion of individuals in the population who have been exposed to the intervention under study; the minimum reduction in mortality risk which the investigators is interested in (20 %, 30 %), and the level of alpha error (risk of rejecting a null hypothesis which is true, usually set at 5 %) and the level of beta error (risk of accepting a null hypothesis which is really false, usually set at 10 to 20 %), and the number of controls per case. Once these parameters are fixed, the minimum sample size can be calculated using the appropriate formula (Schlesselmann, 1982).

In the study on anticoagulants and myocardial infarction, previously published data had suggested that anticoagulants would have been prescribed for 30 % of the fatalities and 50 % of the survivors. To get a power of 80 % at the 5 % significance level, it was found that at least 112 fatalities and an equal number of matched survivors would be required to demonstrate anticoagulant efficacy. To compensate for an expected loss of 25 % of the assembled patients who would be excluded by the clinical trial criteria for eligibility, the cases were 151 fatalities randomly selected as cases from the 234 patients who died.

Data collection

Part of the information needed in a case-control study will come from various types of records, such as: medical records, hospital charts, or death certificates. Other data will be obtained by interviewing subjects, or, in the case of deaths, from the relatives, through the mail or telephone or in person. The primary research instruments in a case-control study are therefore record abstract forms and questionnaires. In addition, approval of a number of committees and individuals will have to be obtained when trying to access the study subjects and their records, and research staff who will interview subjects or abstract records have to be trained.

In the study on anticoagulants and myocardial infarction, the following data were extracted from each patient's medical record by specially trained research assistants: contraindications (gastrointestinal bleeding, intracranial hemorrhage) which usually prohibit anticoagulants, and would have disqualified patients as possible candidates for a randomized trial; clinical conditions regarded as strong indications for anticoagulant use: thrombophlebitis, recent pulmonary embolism. In addition, information was obtained on clinical and paraclinical examinations made in the coronary care unit or intensive care unit, to allow a clinical stratification of the patients according to indexes of infarct severity.

Data analysis

Of basic concern here is the proportion of cases versus controls who have been exposed to or enrolled in the health program being evaluated. If the alternative hypothesis being tested by the study is met (" exposure to the program results in a mortality reduction "), then subjects exposed to the program would have a lower death rate than subjects not exposed. The ratio of the death rate in exposed subjects to the death rate in subjects not exposed is referred to as the relative risk (RR), and we expect this statistic to be significantly lower than 1 under the alternative hypothesis, that is if the program is successful in preventing the condition of interest.

Because subjects in a case control study are selected on the basis not of their exposure but of their health status, it is not possible to obtain a direct estimate of the relative risk. It is however possible to estimate the relative risk indirectly if cases and controls are assumed to be representative of persons with and without the disease in the basic population from which the cases derived, and if the disease is rare in the population.

Assuming that we have chosen study cases and controls so as to represent all cases and all noncases in the population being investigated, we can summarize the study results in tabular form:

Table 1
Distribution of cases and controls by exposure status to health program in the study sample.

Exposure to Health Program	Cases	Controls	Total
Yes	a	b	a+b
No	c	d	c+d
Total	a+c	b+d	a+b+c+d

It has to be pointed out that frequently b+d is chosen to equal c+d, that is the total number of controls equals the total number of cases, and for this reason the ratio a/a+b does not estimate the risk of disease for those with the risk factor. The key to understanding how the data in this table can be used to estimate the relative risk is in recognizing that a+c is a sample of total cases and b+d is a separate sample of non-cases. Therefore although the relationship of a to b or of a to a+b is not meaningful, the relationship of a to c provides an estimate of how all cases are divided into those with and without the risk factor. Similarly, the relationship of b to d provides an estimate of how all non-cases are divided into those with and without the risk factor.

The table below cross classifies the total population from which the cases and controls were selected according to their health status and exposure to the risk factor.

Table 2
Distribution of cases and non-cases by exposure status to health program in the population.

Exposure to Health Program	Cases	Non-cases	Total
Yes	A	B	A+B
No	C	D	C+D
Total	A+C	B+D	A+B+C+D

Based on this table, the relative risk of being sick for those with compared to those without the risk factor is:

(A/A+B)/(C/C+D)

For conditions which are uncommon, and fortunately most diseases and causes of death fit this requirement, A+B can be satisfactorily approximated by B. Similarly C+D can usually be approximated by D. By making these substitutions we obtain:

(A/A+B)/(C/C+D)=(A/B)/(C/D)=(A/C)/(B/D)

This is precisely the ratio of: A/C, that is between the probabilities of having and not having the factor for those who are cases, and; B/D, between the probabilities of having and not having the factor for those who are non cases; which reduces to AD/BC. This ratio is commonly called the odds ratio, which can reliably be derived from the study data as ad/bc, provided the assumptions of representativity and rarity are both satisfied.

How do we interpret the odds ratio value? In fact the odds ratio should not be taken at face value, but compared to the theoretical value of 1 using a significance test. An odds ratio value significantly lower than 1 is a strong argument in favor of a protective effect of the program. Various methods of statistical analysis are available for examining results of case-control studies. These include logistic regression models which allow: adjustments for confounding factors; assessment of individual and joint effects of two or more variables, and; tests for dose response, various matched analyses, etc.

In the study on anticoagulants and myocardial infarction, the patients with strong contraindications for anticoagulant use and those with strong indications for anticoagulant use, who would have been excluded in a therapeutic trial, were removed from the sample to avoid selection bias. Based on the smaller sample thus constituted, the odds ratio between use of anticoagulants on fatalities and their use on survivors was 0.6, indicating no significant difference in the history of anticoagulant use between fatalities and survivors.

The next step in the analysis consisted of removing the effects of any further selection bias which might occur if the use of anticoagulants is determined by the clinical severity of the infarction. Previous non-experimental studies had indeed been criticized for neglecting the selection bias which occurs when the decision to use or withhold anticoagulant treatment is determined by the clinical severity of the myocardial infarction. For this purpose the patients were in two groups: high-risk group, with clinical signs or symptoms of heart failure, and low-risk group, with no such signs. For the high-risk patients, the use of anticoagulants was associated with a protective odds ratio of 0.38, significant at the 5 % level.

Based on these figures, the authors concluded that anti-coagulants seem to be worthless in the hospital management of patients with a mild clinical presentation, and beneficial whenever severe infarcts have to be dealt with. In addition, they explained why previous randomized clinical trials had produced contradictory results: either because high-risk patients had tended to be excluded from the trials, or because low-risk patients predominated, which contributed to obscure the efficacy of anticoagulants in more severe infarctions. By applying in this observational study the same type of admission criteria that would have been imposed in a randomized clinical trial, the authors come up with a modified case-control method, which constitutes a valuable strategy for assessing medical therapies that cannot be tested with randomized trials.

Applications of the case-control design to program evaluation in the health field

There are several arguments in favor of the future use of the case control method in the evaluation of programs already in operation.

- it is a very efficient design for the study of rare conditions, and of uncommon outcomes such as death;

- it can be carried out over a much shorter period than a cohort study, which requires a large sample size to capture a rare outcome such as death; in case-control studies the number of people under investigation can be greatly reduced if all the fatalities are collected, while only a small proportion of the survivors are needed. As a result, in comparison with cohort studies, the case-control method allows a smaller sample size, economizes on subjects, time and on the costs associated with data collection, personnel and data processing;

- using a case-control study with a limited sample size, it is possible to collect a large amount of information on each subject, which is not feasible in large scale prospective cohort studies;

- it can be easily replicated. Not only are trials more expensive to duplicate, but ethical problems arising from preliminary results may prevent further use of this strategy. When a program or treatment is already in widespread use, or when its efficacy is strongly suspected, then a randomized trial may not be feasible.

Despite their usefulness and wide applicability, certain potential problems and limitations may be associated with case-control studies, and these have to be considered:

- the method is not applicable to the evaluation of new health program interventions, because the cases and controls would have had no opportunity for antecedent exposure to the program;

- one possible bias is the non-randomized assignment of a health program or preventive measure being evaluated: in screening for breast cancer, for example, women with a family history of breast cancer would be preferentially prescribed a mammography;

- to ensure homogeneity, cases and controls should only be compared if they are similar with respect to known risk factors. In order to classify subjects with respect to pertinent risk factors for the outcome under study, a substantial amount of information has to be collected to permit stratification. This is feasible in a case-control study generally based on a relatively small sample size, but quite impractical in a large-scale prospective cohort study;

- they are subject to recall bias, being dependent on patients' or informants' recall of a drug exposure that occurred long ago. Typically, controls may not have the same recall of past events as cases, since the subjects who have a condition may be more sensitive to the possible importance of past events. This bias ceases to be a problem when evidence for exposure is documented in medical records, and these are the situations for which the case-control technique is the most suited.

In recent years, the case-control method has become increasingly popular as an efficient and relatively inexpensive method for evaluating a number of different types of health interventions. These include:

- evaluation of immunization programs: for example, case-control studies of the BCG vaccine have been conducted in Columbia (Shapiro, Cook, Evans et al., 1985), in Malawi (Fine, Ponnighaus, Maine et al., 1986) and in Canada (Young and Hershfield, 1986);

- evaluation of screening programs: this is one of the areas where the case-control method has been most effectively used in program evaluation. For example, there have been case-control studies of the cervical smear as a screening device for invasive cervical cancer (Clarke and Anderson, 1979);

- evaluation of medical treatment programs: one famous example is the study of post-menopausal estrogen treatment which supported the hypothesis that therapy provides protection against osteoporotic fractures (Hutchinson, Polansky and Feinstein, 1979).

Conclusion

Although the case-control study clearly cannot be a substitute for experimental designs or for cohort surveys when a strong causal statement is imperative, it can be used as a very efficient tool when more controlled methods are not applicable, either for budgetary or ethical reasons. A further advantage is that it can easily be replicated, and an accumulation of concordant results across studies certainly contributes to strengthen the argument either in favor or against the effectiveness of a program.

References

Baker F. and Curbow B. (1991) The Case-control Study in Health Program Evaluation. *Evaluation and Program Planning,* 14: 263-272.

Clarke EA. and Anderson TW. (1979) Does Screening by " Pap " Smears Help Prevent Cervical Cancer? *Lancet* 2, 8132: 1-4.

Fine PEM., Ponnighaus JM., Maine N., Clarkson JA. and Bliss L. (1986) Protective Efficacy of BCG Against Leprosy in Northern Malawi. *Lancet,* 2, 8505: 499-503.

Greenland S., Watson E., Neutra RN. (1981) The Case-control Method in Medical Care Evaluation. *Medical Care*, XIX, 8 (August): 872-878.

Horwitz RI. and Feinstein AR. (1981) The Application of Therapeutic-trial Principles to Improve the Design of Epidemiologic Research: A Case-control Study Suggesting that Anticoagulants Reduce Mortality in Patients with Myocardial Infarction. *J Chron Dis*, 34: 575-583.

Hutchinson TA., Polansky SM. and Feinstein AR. (1979) Post-menopausal Oestrogens Protect Against Fractures of Hip and Distal Radius: A Case-control Study. *Lancet*, 2, 8145: 705-709.

Ibrahim MA. (1985) *Epidemiology and Health Policy*. Rockville, MD: Aspen Systems Corp..

Rhoads G. and Mills JL. (1984) The Role of the Case-control Study in Evaluating Health Interventions. Vitamin Supplementation and Neural Tube Defects. *American Journal of Epidemiology*, 120, 6 (December): 803- 808, 1984.

Schlesselmann JJ. (1982) *Case-control Studies. Design, Conduct, Analysis*. Oxford University Press, New York, Chapter 6. Sample size.

Selby J.(1994) Case-control Evaluations of Treatment and Program Efficacy. *Epidemiological Reviews*, 16, 1: 90-101.

Shapiro C., Cook N., Evans R., Willett W., Fajardo I. and Koch-Weser D. (1985) A Case-control Study of BCG and Child Tuberculosis in Cali, Columbia. International Journal of Epidemiology, 14: 441-446.

Young TK. and Hershfield ES. (1986) A case study to evaluate the effectiveness of mass neonatal BCG vaccination among Canadian Indians. American Journal of Public Health, 76: 783-786.