

# Case-Control Studies – An Efficient Observational Study Design

Karlijn J. van Stralen Friedo W. Dekker Carmine Zoccali Kitty J. Jager

Department of Medical Informatics J1b, Academic Medical Center, Amsterdam, The Netherlands

## Key Words

Case-control study · Epidemiology · Nephrology · Study design

## Abstract

Case-control studies are an efficient research method for investigating risk factors of a disease. The method involves the comparison of the odds of exposure in a patient group with that of the odds of exposure in a control group. As only a minority of the population is included in the study, less time can be devoted to those who remain free of the disease of interest. The design of a case-control study can be complex due to the selection of the appropriate cases and controls. Cases can be identified in a prospective and retrospective manner from various sources. Controls can be obtained via the patient, random digit dialing or in a hospital and all at different points in the time period of the study. All options have their own advantages and disadvantages. Furthermore, different forms of bias, such as recall bias and selection bias, can occur. When appropriately designed, case-control studies can provide the same information as in a cohort study, in a more rapid and efficient manner.

Copyright © 2009 S. Karger AG, Basel

## Introduction

In the previous paper, we discussed the cohort study design [1]. In cohort studies, much effort is devoted to the follow-up of subjects who remain free of the disease of interest, as usually only a minority of the study population actually develop the disease. Case-control studies aim to obtain the same information as cohort studies: the difference in exposure to risk factors in subjects in relation to the development of the disease. But, unlike in cohort studies, only a minority of the population is included in the study, namely all cases (patients) and a selected number of control subjects. Furthermore, as data on exposure are being collected in retrospect, case-control studies can be much more efficient. Nevertheless, case-control studies have, besides the 'general types of bias', their own specific sources of bias. Also, the selection of cases and controls may be quite complex. In the current paper, we will discuss the basic aspects of the design and the conduct of a case-control study and we will touch upon its major difficulties.

## The Case-Control Study Design

Like case reports and case series, case-control studies can be seen as the natural expansion of the daily practice of physicians [2]. Cases are selected on the basis of the

presence of the disease of interest. By asking questions to the patients or by obtaining data from medical records about the period prior to the occurrence of the disease, these cases are then classified into either exposed or unexposed to a particular risk factor. In a case series, one considers it sufficient to observe that the risk factor is present in many more cases than anticipated (some people call this 'expected' exposure to the risk factor a kind of 'mental control group'). However, one could easily imagine that this mental control group is dependent on a physician's experience and practice and that the responses of cases could depend on the way of questioning. Therefore, in a case-control study, a formal comparison with the exposure rate in controls is made. The controls are usually a representative sample of the population from which the cases originated ('source population'; fig. 1). As only a sample of the population under study is included, it is not possible to determine the incidence rates in the exposed and unexposed groups, and therefore, no relative risk can be determined. However, the odds of exposure can be assessed in both groups, resulting in an odds ratio. This odds ratio can be interpreted as a relative risk, under conditions that will be explained in detail in a future paper.

An example of a case-control study is the one by Forel et al. [3] on the relationship between aspirin use and chronic kidney disease (CKD). Cases were patients with newly diagnosed CKD, as determined by increased serum creatinine levels, who were identified in the Swedish population register. Out of the source population of 5.3 million individuals, 926 cases were identified who had developed CKD in the preceding 2 years. The 998 controls were a random sample from the same register. Cases more often used aspirin (37%) compared with controls (19%) resulting in a 2.5 times higher risk of CKD in aspirin users compared with nonusers. If one would study this same relation between an exposure and CKD in a cohort study, to determine such an effect, one would need to include over 150,000 subjects and follow them for 1 year, because CKD is such a rare outcome. Therefore a case-control study design seems, indeed, the most efficient study design to investigate this relationship.

### Sampling of Cases

After the formulation of a research question and the decision to make use of the case-control study design, one must select eligible cases and controls. As a first step, one

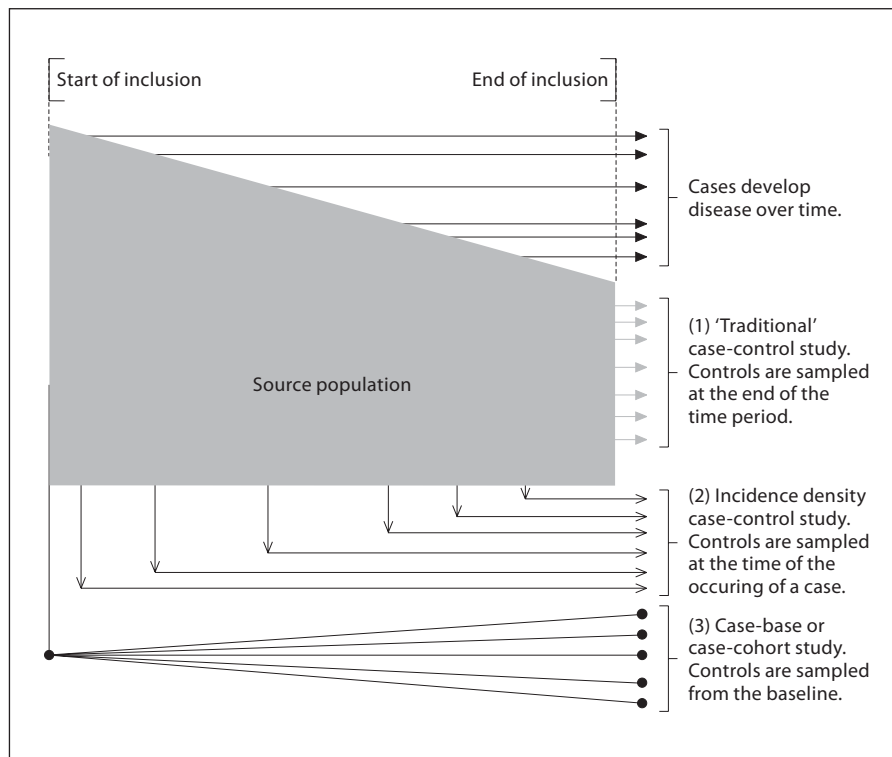
needs to define what a 'case' is. To avoid misclassification, objective criteria for the diagnosis of the disease under study are required. Moreover, a statement of eligibility criteria is essential in order to restrict the study to those who are potentially at risk for exposure. So, for example, in a study regarding the relation between oral contraceptives use and systemic lupus erythematosus, women who are postmenopausal or pregnant will need to be excluded, as they are not at risk of the use of oral contraceptives [4]. Naturally, the same eligibility criteria should be applied to the controls, as otherwise this could result in erroneous results [5].

Cases can be sampled from different sources, such as hospitals, medical records and disease or treatment registries and can be collected in 2 different ways. First, one could collect all prevalent cases, that is, all incident cases who developed the disease in the years before the start of the study, as in a retrospective cohort study. This is a very efficient and quick method. However, if the cases are patients with an unfavorable prognosis like end-stage renal disease for their study, it is likely that cases who would have been eligible died in the preceding years, resulting in a selection of only surviving cases. The second way of including cases is to collect all incident patients prospectively, i.e. to wait for a case to occur and include it into the study. This reduces the chances of including only a selection of cases (the survivors). The drawback is that this method is more time consuming and therefore less efficient.

### Sampling of Controls

In a case-control study, the decision which controls to select often proves to be more difficult than the selection of cases. If cases are difficult to sample one could increase the power of the study by collecting more controls than cases in the study. The first aspect needing careful attention is the choice of the population from which the controls will be derived. In principle, though not compulsory, controls should be derived from the same source population as the cases. Controls can be selected from the general population, by random-digit dialing (phone numbers), they can be obtained via the patient such as partners, friends or neighbors, or controls can be sampled from the same hospital as the case. Usually, there will be advantages for one type of control group that are missing in the other, and vice versa. For example, when studying risk factors associated with renal agenesis after a live birth [6], controls should be sampled from all live

**Fig. 1.** Both cases and controls are sampled from a (hypothetical) source population, free of disease at the start. While the cases are sampled over time, controls can be selected in 3 different fashions: (1) the traditional case-control study with controls sampled at the end of the time period; (2) the incidence density case-control study, in which controls are sampled each time a case occurs, and (3) the case-base study, in which the controls are sampled at the (hypothetical) start of the study. When using methods 2 and 3, controls can be included in the study multiple times, both as control and as a case. When using method 1, all controls are still free of disease at the end of the time period and can therefore never be included as case.



births. If patients with renal agenesis would be born more often in academic hospitals, it would not be adequate to sample controls only from babies born in the same hospital as the cases, as in this situation one would over-sample the number of babies born in academic hospitals. This may result in too high a number of babies and mothers with other diseases leading to an underestimation of the effect of potential exposures like diabetes.

A second point to address is the sampling scheme of controls. As in a cohort study, the source population must be free of the disease of interest at the start of inclusion in the study. A person is eligible to be selected as a control as long as they are free of this disease [7, 8]. Therefore, theoretically, they can be included in the study multiple times, both as control and as a case. One could sample controls in 3 different manners (fig. 1). First, one could sample controls at the end of inclusion in the study (traditional case-control study). A second option is to sample controls each time a case occurs (incidence density sampling), and the third option is to sample controls before the first diagnosis of a case, i.e. at the (hypothetical) start of inclusion in the study (case-base study). Each sampling method has different consequences for the interpretation of the odds ratio but this is beyond the scope of this paper.

In short, the case-base study and the incidence density sample case-control study will provide an accurate estimate of the relative risk. The traditional case-control study provides an accurate estimate if the disease is sufficiently rare, i.e. the 'rare disease assumption'.

### Bias

As in other types of study designs, in a case-control study, multiple types of bias can occur. If cases and controls are not derived from a similar population or if they have different chances of being exposed, selection bias may arise. Another important source of bias, which is unique for case-control studies, is recall bias. This occurs when a subject is interviewed to obtain exposure information after the disease has occurred. As the case is suffering from the disease, they may search their memory for any history of exposure. Controls do not have such a stimulus, which possibly results in less accurate or more socially acceptable answers. For example, in a study of risk factors for acute pyelonephritis, women were asked to give information about the number of urinary tract infections (UTI) in their mothers [9]. Women with pyelo-

nephritis may have asked their mothers on their history of UTI, resulting in accurate estimates of the percentage exposed. However, controls may be unaware of the history of UTI in their mothers, simply because they have not asked them. Therefore, even if there would be no association between the exposure and outcome, it is still possible to find an effect. This bias could be avoided in several ways. One could obtain information regarding exposure from sources that documented the information before the outcome was known such as medical records. However, such records may not be available for healthy individuals. It is also possible to send the questionnaire by mail so that cases and controls both have enough time to obtain accurate answers. Finally, it is important to use trained interviewers, preferably unaware ('blind') of the subject's disease status and use standardized questionnaires. Another solution is to obtain information from persons who have had a similar stimulus such as women with a different disease, resulting in an equal motivation to recall potentially relevant exposures. However, if both diseases would have 'joint' exposures, this would result in an underestimation of the effect of the exposure on the disease under study.

Although textbooks on epidemiology agree on the use of the terminology regarding case-control studies, authors of many papers have applied the term incorrectly. Recent reviews of articles in the fields of pediatrics and

surgery showed that in 25–65% of the articles describing themselves as case-control studies had in fact another study design [10, 11]. In these papers, the terms 'cases' and 'controls' were used to denote subjects who were affected or unaffected by a given risk factor, for example diabetic versus nondiabetic patients. Subjects were then followed forward in time to assess development of another outcome, for example CKD. However, although patients are compared with non-patients, this study design is a cohort study, as subjects are followed in time to assess the effect of exposure (in this case diabetes) to an outcome (in this case CKD). Although these studies may be of good quality, incorrect use of the terms of the used study design may lead to confusion with readers [12].

### Conclusion

In summary, case-control studies are an efficient design to study associations in a relatively cheap and rapid manner. When appropriately designed, they can provide the same information as obtained in a cohort study. Nevertheless, selection of both cases and controls is complex and also case-control studies are subject to multiple sources of bias. Therefore, before starting a case-control study, it is advisable to consult an epidemiologist or statistician.

### References

- 1 Euser AM, Jager KJ, Zoccali C, Dekker FW: Cohort studies: prospective versus retrospective. *Nephron Clin Pract*, in press.
- 2 Morabia A: *A History of Epidemiologic Methods and Concepts*. Basel, Birkhäuser, 2004.
- 3 Fored CM, Ejerblad E, Lindblad P, et al: Acetaminophen, aspirin, and chronic renal failure. *N Engl J Med* 2001;345:1801–1808.
- 4 Cooper GS, Dooley MA, Treadwell EL, et al: Hormonal and reproductive risk factors for development of systemic lupus erythematosus: results of a population-based, case-control study. *Arthritis Rheum* 2002;46:1830–1839.
- 5 Schlesselman JJ, Stolley PD: *Case-Control Studies, Design, Conduct, Analysis*. New York, Oxford University Press, 1982.
- 6 Parikh CR, McCall D, Engelman C, Schrier RW: Congenital renal agenesis: case-control analysis of birth characteristics. *Am J Kidney Dis* 2002;39:689–694.
- 7 Rothman KJ, Greenland S: *Modern Epidemiology*, ed 2. Philadelphia, Lippincott-Raven, 1998.
- 8 Greenland S, Thomas DC: On the need for the rare disease assumption in case-control studies. *Am J Epidemiol* 1982;116:547–553.
- 9 Scholes D, Hooton TM, Roberts PL, et al: Risk factors associated with acute pyelonephritis in healthy women. *Ann Intern Med* 2005;142:20–27.
- 10 Hellems MA, Kramer MS, Hayden GF: Case-control confusion. *Ambul Pediatr* 2006;6:96–99.
- 11 Mihailovic A, Bell CM, Urbach DR: Users' guide to the surgical literature. Case-control studies in surgical journals. *Can J Surg* 2005;48:148–151.
- 12 Marshall T: What is a case-control study? *Int J Epidemiol* 2004;33:612–613.