European Journal of Epidemiology **19**: 527–532, 2004.

Appending epidemiological studies to conventional case–control studies (hybride case–control studies)

Andreas Stang & Karl-Heinz Jöckel
Biometry and Epidemiology, Institute for Medical Informatics, University of Essen, Essen, Germany

Accepted in revised form 3 February 2004

Abstract. This paper summarizes several studies that can be appended to conventional case–control studies especially in the context of case–control studies that focus on etiologic questions. Appending studies to case–control studies may further add to the understanding of the epidemiology of diseases under investigation. We explain their uses, implications and limitations. One can append the following studies to a case–control study: (1) case-only study, (2) case–crossover study, (3) case cross-sectional study, (4) control cross-sectional study, (5) case follow-up study, and (6) control follow-up study. The choice of the additional studies that are appended to the conventional case–control study has implications for the set of data and biological material that has to be collected, the ethical review board and the informed consent. Due to several limitations, the attachment of additional studies to a case–control study should be carefully considered and limited to only few additional studies in order to avoid overburden of the study participants and study personnel.

Key words: Case–control studies, Cross-sectional studies, Epidemiological methods

Introduction

The case–control study is a powerful study design to investigate the association between a range of exposures and an outcome [1–5]. Famous examples include case–control studies investigating the association between smoking and lung cancer [6], oral contraceptives and thromboembolism [7], cryptorchidism and testicular cancer [8]. The application of the case–control methodology can go beyond etiologic research and may focus on the evaluation of the effectiveness of primary (e.g. vaccination) [9] and secondary prevention strategies (e.g. screening) [10], therapies and prognostic factors [11]. The case–control study has a number of advantages and disadvantages compared to cohort studies. For diseases that are sufficiently rare, cohort studies become too impractical to consider, and often case–control studies become the only useful alternative. However, for exposures that are extremely rare, case–control studies are not efficient, and cohort studies of special-exposure are necessary. Compared to cohort studies, case–control studies present more opportunities for bias including recall, selection and other biases [12, 13].

The understanding of the epidemiology of the studied disease within a case–control study can go beyond the primary study questions that led to the initiation of the case–control study, especially in the context of case–control studies that focus on etiologic questions. If planned well, case–control studies offer several design options that enable us to answer additional questions besides the primary study questions by appending additional studies to the case–control study.

In this paper, we want to give an overview and explain the indication, use, limitations and implications of studies that can be appended to conventional case–control studies. We discuss issues including ethical review, informed consent, data confidentiality and data collection that have to be considered during the planning phase of the case–control study to append studies on the case–control study.

In the classical case–control study that focuses on etiology, cases and controls are compared with regard to the retrospectively assessed exposure status. To enhance the efficacy of the control of confounding, the control group is often matched to the case group. The additional design options treat the groups of cases and controls in several ways to learn about the nature of the disease, the exposure or confounders, and their interdependence. These include retrospective, cross-sectional and prospective design options. Figure 1 gives a summary of a case–control studies and its additional design options.

Case-only study

The retrospective case-only study (Figure 1: no. 1) allows to investigate gene–gene, gene–environment and environment–environment interaction on a multiplicative scale [14, 15]. A crucial assumption for this study is the independence of the occurrence of the exposures of interest in the study base. Investigators

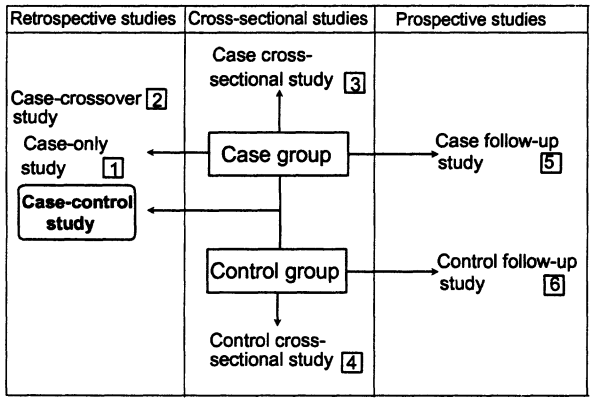

Figure 1. Case–control study design and additional design options.

who want to study gene–gene interaction should keep in mind that the genes under study must be in linkage equilibrium in the population being studied. The cross-product in a case-only study measures the ratio of OR_{11} to $OR_{01} \bullet OR_{10}$, a measure of departure from the multiplicative joint effects of two risk factors [15]. However, it has been suggested that for addressing public health concerns regarding disease frequency reduction, biological interaction, that is, assessing deviations from additivity is most relevant [13].

If genetic markers are considered, the investigators carefully have to plan which biological material (lymphocytes, leukocytes, serum, urine, liquor, etc.) from the cases has to be collected and how this material has to be processed and stored to enable the genetic analysis that is usually conducted at the end of the case recruitment. Even in the absence of a specific hypothesis concerning gene–gene or gene–environment interaction, the investigators should consider to collect and store biological material that enables a genetic analysis in future when a novel hypothesis on the genetic etiology of the disease comes up. However, with the gradual introduction of microarray techniques [16, 17], simultaneous genetic screening of several thousand markers may result in the problem of multiple testing. Regardless whether the genetic hypothesis existed before the study started or whether the hypothesis came up during or after the recruitment of cases, the investigator should inform the ethical review board and should make the informed consent as clear as possible. If the genetic markers that will be assessed are not known at the beginning or during the case recruitment the investigator should re-contact the ethical review board before he starts the genetic analysis. In addition, cases should be re-contacted to obtain the informed consent for the additional genetic studies.

Case-crossover study

The case-crossover study (Figure 1: no. 2) is a popular analytic tool for estimating the effects of triggers of acute outcomes by brief environmental exposures and involves only cases [18–21]. Each case serves as his or her own control. This approach controls for all measured and unmeasured time-invariant confounders by design, because these are constant within each individual's stratum. If intermittent exposure to factors with transient effects are of interest, a case-crossover study should be considered in addition to the conventional case–control study. For this purpose, the exposure assessment among cases has to be expanded. Each case contributes one case window and one or more control windows. The case window is defined as the 'at risk' period preceding the event. The control windows are periods of the same length as, and not overlapping with, the case window that provide an estimate of the expected frequency of exposure of each case. The case window and the control windows derive from the same person (case) at different times. The case-crossover odds ratio can be estimated by the ratio of the number of cases exposed only during the case window to the number of cases exposed only during the control window (i.e. ratio of discordant pairs) [22]. Obviously, the exact timing of the onset of the case disease in the case–control study has to be documented carefully.

While the case-crossover study avoids difficulties of selecting a valid control group, one still must define appropriate control windows that may be problematic because there may be induction periods and carry-over effects. A washout gap between the control and case window can avoid potential carry-over effects. In addition, a potential problem with restricting the set of eligible control windows in case-crossover studies is that time-selection bias can occur from time trends in the exposure of interest itself [23, 24]. A bidirectional sampling of control windows (i.e. before and after the case window) can eliminate the time-selection bias due to linear trend that occurs with unidirectional sampling. It is also possible to adjust case-crossover estimates for time selection biases through the use of longitudinal data from a truly nondiseased control group (case–time-controls) [25, 26]. Simulations by Bateson and Schwartz [27] show that the gross biases from seasonal variation can also be alleviated by choosing windows from a shorter period of time both before and after the case time. The study of environmental exposure effects has the advantage that exposure levels subsequent to the event are unaffected by the event occurrence [28]. However, if the outcome (disease of the case) affects subsequent exposure, then control time should precede the case window [23, 27].

Case cross-sectional study

The case cross-sectional study (Figure 1: no. 3) is an important source of the descriptive epidemiology

of the disease of interest, especially if the case recruitment is population-based. In study regions of population-based case–control studies, where population-based registries do not exist, the case cross-sectional study provides information on the occurrence of the disease of interest which is not available otherwise. That is, the population-based case–control study is, in essence, a temporary population-based disease registry. It allows to compute and compare the proportion of cases with different characteristics. In addition, for the calendar period of the case–control study, it enables the investigator to calculate and compare incidence rates by age, gender, etc. However, to estimate population-based incidence rates from the case–control study, the completeness of the case ascertainment is a crucial prerequisite and may be difficult to assess. In a recent study by Carpenter et al. [29] data from four large case–control studies of the etiology of sudden unexplained infant death (SIDS) were analyzed. Based on 745 cases, the authors calculated incidence rate estimates of autopsy-verified SIDS by study region and found that the rates ranged from 0.17 per 1.000 in Hungary to 1.3 per 1.000 in Northrhine–Westfalia. They also assessed the age-specific incidence of SIDS and found that the rates peak at 10 weeks of age.

Even when population-based disease registries (e.g. cancer registries) for the disease of interest are available in the study region, the cross-sectional characterization of the case group can go far beyond the characterization in population-based registries which routinely collect only a limited set of information. For example, the mandatory set of information of newly diagnosed cancer cases that is reported to population-based cancer registries in Germany does not include details of the diagnostic workup, treatment modalities or the comorbidity and detailed information on stage is often missing [30, 31]. Furthermore, the study-driven case ascertainment may be more complete than the routine registration procedure, especially for rare cancers like uveal melanoma [32]. This may occur, if the cancer registries are mainly based on routine reports from pathologists whereas a case–control study may additionally motivate treating clinicians to report eligible cases to the study.

To enable the case cross-sectional study, investigators carefully have to plan which clinical and other information beyond the necessary information to clarify the eligibility criteria for the cases has to be collected. For example, it may be important to collect detailed information of the diagnostic workup, exact anatomic location, stage and morphology of the disease, of the treatment modalities and the comorbidity of the cases. Often, these data are available from the treating clinicians so that the cases are not burdened with this part of the study. However, the informed consent of the case–control study should clearly state that the investigator is allowed to collect these additional data.

Control cross-sectional study

If the control group has been selected in a valid manner, it provides information on the prevalence of the exposure distribution in the study base from which the cases came and enables the investigator to calculate exposure prevalence estimates in the study population (control cross-sectional study, Figure 1: no. 4). If the case–control study is population based it enables the investigator to calculate population-based prevalence estimates of exposure. However, controls in case–control studies are often matched to the cases by age and gender (and sometimes further characteristics) and therefore prevalences can only be estimated within matching strata. Prevalence estimates within matching strata can become very instable if the sizes of the strata are small. Overall prevalence estimates can be derived by standardization to age and gender and can be compared to other studies whereas the crude prevalence may be misleading due to selection effects of matching controls to the cases.

In addition to the studies of exposure prevalence, one could conduct studies of the prevalence of diseases other than the case disease. For example, we did a control cross-sectional study [33] within a population-based case–control study [34] to estimate the prevalence of self-reported gallstones in the German population because current prevalence data on gallstones in the general population were not available. For this purpose, we estimated matching-factor specific, i.e. age- and gender-specific prevalence estimates of self-reported gallstones and thereafter calculated gender-specific age-standardized prevalences of gallstones in the study population. The age-standardized prevalence (European standard population) of self-reported gallstones among subjects aged 35–69 years was 4.2% among men and 14.5% among women [33]. One could also measure the prevalence of a wide range of diseases – many having nothing to do with the case disease – and compute prevalence odds ratios or prevalence ratios for these diseases and selected exposures as estimates of incidence rate ratios, given the necessary presumption of steady-state and no effect of exposure on disease duration [35].

Case follow-up study

Once the case group has been established, it also allows to start a case follow-up study (Figure 1: no. 5) to investigate potential determinants of the survival (overall, disease-specific, relapse-free, etc.) of the cases, as long as the case–control study included living cases (and not only dead cases). In addition, it

would allow to study risks of other diseases among the cases. For diseases where population-based disease registries are not available, this approach can offer population-based information on the prognosis of the disease, if the case ascertainment was population based. Even when the population-based disease registries are available, this approach might still offer some additional information. Compared to the relatively limited data set of routinely collected cases in population-based disease registries, the case follow-up study allows to investigate several prognostic factors, because the epidemiological information collected within the frame of the case–control studies usually goes far beyond the data collection in disease registries. It allows to simultaneously investigate the causal (or preventive) and prognostic effect of exposures of interest. For example, Sweeney and Farrow [36] found that smoking is not only a risk factor for renal cancer but also a prognostic factor for the disease. A case–control study on the etiology of renal carcinoma with a follow-up of the cases, therefore, might give further insights into the role of smoking regarding diseases risk and prognosis.

Owing to the observational nature of the case follow-up study, the investigator has to carefully consider the issue of confounding. Therefore, it is necessary – as within the case cross-sectional study – to characterize the case group with regard to all known prognostic factors of the disease (e.g. stage, location, morphology, comorbidity, treatment modalities, etc.) to be able to adjust for potential confounding. As has been mentioned in the context of a case cross-sectional study, the majority of prognostic factors may be available from the treating clinicians so that the cases are not burdened with this part of the study. However, when the follow-up is not only a mortality follow-up but also a morbidity follow-up, additional contacts with the cases after the inclusion into the case–control study may become necessary. The informed consent of the case–control study should clearly state that the investigator is allowed to collect additional prognostic data and may re-contact the cases in the future.

Control follow-up study

The control group within a case–control study can also be regarded as an epidemiologically well-described cohort of subjects who are free of the disease of interest at the inclusion in the study, as long as the case–control study included living controls (and not only dead controls). The follow-up of the control group (control follow-up study, Figure 1: no. 6) allows to prospectively study the association between exposures and diseases that may occur in the future. For this purpose, controls are divided into exposed and unexposed subjects so that this follow-up study includes an internal comparison cohort. The results of this cohort may also be compared with an external comparison cohort.

In addition, control subjects with diseases that do not belong to the diseases of interest within the case–control study can be followed over time and prognostic factors can be investigated. To be able to perform the case or control follow-up study, investigators have to consider the following issues: (1) additional data that are necessary for the follow-up study may have to be collected, (2) the ethical review board has to approve the study protocol, (3) study participants have to be appropriately informed about the follow-up, (4) data confidentiality issues have to be considered: personal data have to be safely stored over time in order to enable the re-contact of the cases.

Jöckel et al. [37] recently showed that a control cross-sectional and control follow-up study within a case–control study is feasible even under restrictive data protection laws in Germany. They collected not only interview data but also sampled blood and urine samples from a second population-based control group of their case–control study on the etiology of lung cancer and used this group as a cross-sectional population-based sample to investigate the incorporated concentrations of nickel, cadmium, chromium from the urine samples and lead from the blood sample [38]. In addition, they studied the association between these metals and the rate of oxidative DNA lesions in lymphocytes [39]. The informed consent included the possibility to re-contact the controls in future. It is planned to follow up this cross-sectional population-based sample to estimate the predictive value of the rate of oxidative DNA lesions on cancer.

There are several factors that may limit the suggested approaches. First, if the sample size calculation of the case–control study is solely based on the primary study questions within the context of the case–control study, the analyses of the appended studies may suffer from insufficient statistical power. This occurs especially often within case-only studies investigating gene–environment interaction. Second, for some approaches additional data of the patient or treating institution is necessary which may cost time, person input and money and may overburden participants of case–control studies. Third, appending further studies to the case–control study may increase the potential for data fishing if the study hypotheses are not clearly stated a priori. Fourth, the anonymization of the personal data has to be postponed especially in the context of case or control follow-up studies. Study participants have to be appropriately informed about the appended studies. Fifth, the plan to re-contact cases or controls in the future may reduce the participation in the case–control study and therefore may result in selection bias which is a threat to the validity of the case–control study. In view of the generally low response proportions of population controls, investigators may first get experience with

the response proportion within the sole case–control study before controls are also included in a control follow-up study.

Due to this list of limitations, the attachment of additional studies to case–control studies should be carefully considered and limited to only few additional studies in order to avoid overburden of the study participants and study personnel that may jeopardize the main study. Investigators should also trade off whether an independent study which may be more expensive but may have the benefit of providing better information should be preferred.

References

1. Rothman KJ. Modern epidemiology. Boston: Little, Brown and Company, 1986, pp. 62–69.
2. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case–control studies. I. Principles. Am J Epidemiol 1992; 135: 1019–1028.
3. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case–control studies. II. Types of controls. Am J Epidemiol 1992; 135: 1029–1041.
4. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case–control studies. III. Design options. Am J Epidemiol 1992; 135: 1042–1050.
5. Armenian HK, ed. Applications of the case–control method. Epidemiol Rev 1994; 16: 1–164.
6. Wynder EL, Graham EA. Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. A study of six hundred and eighty-four proved cases. JAMA 1950; 143: 329–336.
7. Royal College of General Practitioners. Oral contraception and thromboembolic disease. J Roy Coll Gen Pract 1967; 13: 267–279.
8. Henderson BE, Benton B, Jin J, Yu MC, Pike MC. Risk factors for cancer of the testis in young men. Int J Cancer 1979; 23: 598–602.
9. Comstock GW. Evaluating vaccination effectiveness and vaccine efficacy by means of case–control studies. Epidemiol Rev 1994; 16: 77–89.
10. Weiss NS. Application of the case–control method in the evaluation of screening. Epidemiol Rev 1994; 16: 102–108.
11. Weiss NS. Clinical epidemiology. 2nd edn. New York: Oxford University Press, 1996.
12. Austin H, Hill HA, Flanders WD, Greenberg RS. Limitations in the application of case–control methodology. Epidemiol Rev 1994; 116: 65–76.
13. Rothman KJ, Greendland S. Modern epidemiology. 2nd edn. Philadelphia: Lippincott-Raven, 1998.
14. Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene–environment interaction: Case–control studies with no controls! Am J Epidemiol 1996; 144: 207–213.
15. Yang Q, Khoury MJ, Sun F, Flanders WD. Case-only design to measure gene–gene interaction. Epidemiology 1999; 10: 167–170.
16. Gershon D. Microarray technology: An array of opportunities. Nature 2002; 416: 885–891.
17. Gottardo R, Pannucci JA, Kuske CR, Brettin T. Statistical analysis of microarray data: A Bayesian approach. Biostatistics 2003; 4: 597–620.
18. Maclure M. The case-crossover design: A method for studying transient effects on the risk of acute events. Am J Epidemiol 1991; 133: 144–153.
19. Mittleman MA, Maclure M, Sherwood JB, et al. Triggering acute myocardial infarction onset by episodes of anger. Circulation 1995; 92: 1720–1725.
20. Mittleman MA, Mintzer D, Maclure M, Tofler GH, Sherwood JB, Muller JE. Triggering of myocardial infarction by cocaine. Circulation 1999; 21: 2732–2741.
21. Petridou E, Mittleman MA, Trohanis D, Dessypris N, Karpathios T, Trichopoulos D. Transient exposures and the risk of childhood injury: A case-crossover study in Greece. Epidemiology 1998; 9: 622–625.
22. Hernandez-Diaz S, Hernan MA, Meyer K, Werler MM, Mitchell AA. Case-crossover and case-time-control designs in birth dectects epidemiology. Am J Epidemiol 2003; 158: 385–391.
23. Navidi W. Bidirectional case-crossover designs for exposures with time trends. Biometrics 1998; 54: 596–605.
24. Greenland S. Confounding and exposure trends in case-crossover and case-time-control designs. Epidemiology 1996; 7: 231–239.
25. Suissa S. The case-time-control design. Epidemiology 1995; 6: 248–253.
26. Suissa S. The case-time-control design: Further assumptions and conditions. Epidemiology 1998; 9: 441–445.
27. Bateson TF, Schwartz J. Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures. Epidemiology 1999; 10: 539–544.
28. Levy D, Lumley T, Sheppard L, Kuafman J, Checkoway H. Referent selection in case-crossover analyses of acute health effects of air pollution. Epidemiology 2001; 12: 186–192.
29. Carpenter RG, Irgens LM, Blair PS, et al. Sudden unexplained infant death in 20 regions in Europe: Case control study. Lancet 2004; 363: 185–191.
30. Krebsregister Saarland. Jahresbericht Krebsregister Saarland 1996/1997. Saarbrücken: Ministerium für Frauen, Arbeit, Gesundheit und Soziales, 2000.
31. Stang A, Stang K, Stegmaier C, Hakulinen T, Jöckel KH. Skin melanoma in Saarland: Incidence, survival and mortality 1970–96. Eur J Cancer Prev 2001; 10: 407–415.
32. Vajdic CM, Kricker A, Giblin M, et al. Incidence of ocular melanoma in Australia from 1990 to 1998. Int J Cancer 2003; 105: 117–122.
33. Timmer A, Ahrens W, Stegmaier C, et al. Risikofaktoren und Operationsraten des Gallensteinleidens. Ergebnisse einer bevölkerungsbezogenen Studie. Med Klinik 2000; 95: 672–677.
34. Stang A, Anastassiou G, Ahrens W, Bromen K, Bornfeld N, Jöckel KH. The possible role of radiofrequency radiation in the development of uveal melanoma. Epidemiology 2001; 12: 7–12.
35. Thompson ML, Myers JE, Kriebel D. Prevalence odds ratio or prevalence ratio in the analysis of cross sec-

tional data: What is to be done? Occup Environ Med 1998; 55: 272–277.
36. Sweeney C, Farrow DC. Differential survival related to smoking among patients with renal cell carcinoma. Epidemiology 2000; 11: 344–346.
37. Jöckel KH, Ahrens W, Jahn I, Pohlabeln H, Bolm-Audorff U. Occupational risk factors for lung cancer: A case–control study in West-Gemany. Int J Epidemiol 1998; 27: 549–560.
38. Jöckel KH, Ahrens W, Merzenich H, et al. Integrierter Ansatz zum Risikomonitoring auf der Basis eines hybriden Populationpanels. Abschlußbericht. Bremen, 1996.
39. Merzenich H, Hartwig A, Ahrens W, et al. Biomonitoring on carcinogenic metals and oxidative DNA damage in a cross-sectional study. Cancer Epidemiol Biomarkers Prev 2001; 10: 515–522.

Address for correspondence: Andreas Stang, Biometry and Epidemiology, Medical Faculty, Institute for Medical Informatics. University of Duisburg-Essen, Hufelandstr. 55, 45122 Essen, Germany
Phone: +49-201-723-4514; Fax: +49-201-723-5933
E-mail: andreas.stang@uni-essen.de