Curr Epidemiol Rep (2016) 3:262–264
DOI 10.1007/s40471-016-0091-7

INVITED COMMENTARY

Case-Control Studies & the Hierarchy of Study Design

Henrik Toft Sørensen [1]

Published online: 7 October 2016

Keywords Case-control design · Data collection · Selection bias · Information bias · Cohort study

The case-control design is a cost-efficient approach that has been used in numerous studies, providing important knowledge about risk factors for many diseases. A classic example is Doll's 1950 study, which revealed an association between smoking and lung cancer [1]. Over time more advanced subtypes of case-control designs have been developed. Examples are the case-crossover design [2] and the case-time-control design [3].

In the hierarchy of study designs, the case-control design has a low ranking [4], mainly because of inherent methodological issues. Among these, selection bias (such as in the selection of controls) and information bias (such as recall bias), are key challenges [5]. Thus, the case-control design is considered weaker than both the randomized controlled trial and the cohort study [6]. Decades ago, Feinstein described a case-control study based on retrospective data collection as a trohoc study (cohort spelled backwards) [7].

Within the field of epidemiology, there have been two major trends in data collection and use. The first is the establishment of large cohorts based on collection of prospective data, often in conjunction with biobanked material such as blood and tissue samples. The second involves use of existing medical registries and databases, allowing studies to make use of already collected secondary data. As both data collection methods have limitations, reconsideration of the case-control design as a valid and efficient study design is needed. It must be stressed that case-control studies can be based on primary and secondary data, collected prospectively or retrospectively [8].

Six arguments can be made for the value and utility of the case-control design.

1. Data collection for many existing large cohorts [8–11] is limited to baseline characteristics and exposures, supplemented with follow-up of health events. Important examples are the Nurses' Health Study and the European Prospective Investigation into Cancer and Nutrition (EPIC) study [12, 13]. Such cohort studies rarely allow examination of the effect of acute exposures or exposures whose effects vary over short periods of time or are diphasic. In contrast, a case-control study can capture such effects. For instance, a case-control study conducted during 1979–1989 in Washington State [14] examined the risk of cardiac arrest during vigorous exercise. The study included 133 men who experienced primary cardiac arrest and a control group of healthy men. Information was obtained through interviews with the men's wives. The study found that the risk of cardiac arrest during exercise was higher than at other times, especially among men with low levels of habitual activity. Among habitually vigorous men, the overall risk of cardiac arrest (both during vigorous activity and at other times) was 40 % lower than that among sedentary men. This association could not have been revealed in a classic cohort study in which there is no possibility of collecting exposure information right before the effect. Thus, case-control studies allow for collection of very detailed information at the time of a health event or diagnosis of a disease. The later case-crossover design represents a fundamental step forward in the study of acute effects.

✉ Henrik Toft Sørensen
hts@clin.au.dk

[1] Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark

2. Even large cohorts with hundreds of thousands of participants are too small to study rare health events or diseases. They are appropriate only for investigating the etiology of relatively common conditions. In contrast, the case-control design is very cost-effective in the context of rare diseases. As an example, in a cohort drawn from Danish databases (1977–2009), a nested nationwide population-based case-control study was conducted among individuals who had undergone colonoscopies (n = 272,342). It identified 2045 CRC cases and 8105 CRC-free individuals (controls). The aim of the case-control study was to examine CRC risks associated with serrated polyps (SSA/Ps). For each case and control, tissue blocks were obtained for the first hyperplastic colorectal polyp(s) biopsied or excised during or after the initial colonoscopy. Four expert pathologists reviewed these lesions using current terminology for serrated polyps. Seventy-nine cases and 142 controls were diagnosed with SSA/Ps (odds ratio (OR), 3.07; 95 % confidence interval [CI], 2.30–4.10). SSA/Ps with cytology markers for dysplasia were associated with a particularly high OR (4.76; 95 % CI, 2.59–8.73) [15] for a subsequent diagnosis of CRC. This case-control study showed that the same information could be obtained by examining 10,015 individuals instead of the entire cohort of approximately 270,000 individuals [15].
3. A case-control design using interviews may provide more valid and complete exposure information than cohort studies relying on medical databases. For instance, interviews may provide more accurate information on actual drug exposure than prescription databases, although information bias is a risk [16]. While prescription databases contain information on when a prescription was issued or reimbursed by a pharmacy, they lack information on the time of actual drug intake. Such misclassification will most likely bias the risk estimates towards the null [5]. As well, many large health care databases have limited information on lifestyle factors. Even those with information on body mass index, alcohol intake, and smoking, such as the general practice databases in the UK, have a substantial amount of missing data, which can lead to problems with residual confounding [17]. In contrast, case-control studies based on interviews can provide much better and more detailed data on exposures and potential confounding factors.
4. A case-control study also can be used to examine novel hypotheses and even multiple exposures not known at the time when a large cohort study is designed. Moreover, information collected using a case-control design can be combined readily with biobank data. An example is the Leiden case-control study of venous thromboembolism (VTE) risk factors in interaction with biomarkers. This study revealed the now well-known interaction between oral contraceptives and VTE risk in carriers of Leiden factor 5 [18]. The nested case-control design also can be used to validate outcomes in database studies. For example, Jick et al. have applied the case-control design to validate VTE diagnoses within a large cohort, in order to confirm the specificity of the diagnosis [19].
5. A particular problem in pharmacoepidemiological studies of long-term drug effects is to determine present drug intake and the cumulative amount of drug consumption. Although lack of information on non-compliance and exact drug intake is a well-known limitation of registry studies, many pharmacoepidemiological studies assume that the dose and cumulative amount described in the medical record or reimbursed at the pharmacy is the same as the exposed dose and cumulative amount. Actual drug intake and the cumulative dose might be difficult to estimate using a cohort design since exposures represent dynamic variables throughout the follow-up period. In a case-control design, it is easier to estimate the real cumulative dose [20].
6. The case-control design may be intuitively more attractive for physicians taking care of specific patients. As well, it is a concern that many students do not have a personal relation to data collection and classification of patients, since the majority of dissertations are based on medical databases and existing cohorts. A case-control design facilitates students' participation in data collection and gives them a better understanding of data quality. Finally, it is possible to use a case-control study of risk factors as a prospective prognostic study based on the same cases. A good example is provided by a Swedish study in which cases from a case-control study of risk factors for hip fracture were used for a later cohort study focusing on survival following hip fracture [21, 22].

Currently, the case-control design is relatively low in the hierarchy of study designs. However, there are no strong universal arguments for keeping this ranking. Choice of design must depend on the topic in question. In a number of contexts, the case-control design is more valuable than other designs in providing valid information on an exposure/effect relation.

Compliance with Ethical Standards

Conflict of Interest Henrik Toft Sørensen declares that he has no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

1. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. Br Med J. 1950;2:739–48.
2. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol. 1991;133:144–53.
3. Suissa S. The case-time-control design. Epidemiology. 1995;6:248–53.
4. Fletcher RH, Fletcher SW, Fletcher GS. Clinical epidemiology, page 206. The essentials. 5th ed. Philadelphia: Lippincott Williams & Wilkins; 2014.
5. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
6. Feinstein AR. Clinical biostatistics. XX. The epidemiologic trohoc, the ablative risk ratio, and "retrospective" research. Clin Pharmacol Ther. 1973;14:291–307.
7. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. JAMA. 1995;274:1800–4.
8. Acute myocardial infarction and combined oral contraceptives: results of an international multicentre case-control study. WHO Collaborative Study of Cardiovascular Disease and Steroid Hormone Contraception. Lancet. 1997;349:1202–9.
9. Lukic M, Licaj I, Lund E, Skeie G, Weiderpass E, Braaten T. Coffee consumption and the risk of cancer in the Norwegian Women and Cancer (NOWAC) Study. Eur J Epidemiol. 2016;31:905–16.
10. Dewi NU, Boshuizen HC, Johansson M, Vineis P, Kampman E, Steffen A, Tjønneland A, Halkjær J, Overvad K, Severi G, Fagherazzi G, Boutron-Ruault MC, Kaaks R, Li K, Boeing H, Trichopoulou A, Bamia C, Klinaki E, Tumino R, Palli D, Mattiello A, Tagliabue G, Peeters PH, Vermeulen R, Weiderpass E, Torhild Gram I, Huerta JM, Agudo A, Sánchez MJ, Ardanaz E, Dorronsoro M, Quirós JR, Sonestedt E, Johansson M, Grankvist K, Key T, Khaw KT, Wareham N, Cross AJ, Norat T, Riboli E, Fanidi A, Muller D, Bueno-de-Mesquita HB. Anthropometry and the Risk of Lung Cancer in EPIC. Am J Epidemiol. 2016;184:129–39.
11. Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. Lancet. 2015;386:533–40.
12. Tjønneland A, Olsen A, Boll K, Stripp C, Christensen J, Engholm G, Overvad K. Study design, exposure variables, and socioeconomic determinants of participation in diet, cancer and health: a population-based prospective cohort study of 57,053 men and women in Denmark. Scand J Public Health. 2007;35:432–41.
13. Ardisson Korat AV, Willett WC, Hu FB. Diet, lifestyle, and genetic risk factors for type 2 diabetes: a review from the nurses' health study, nurses' health study 2, and health professionals' follow-up study. Curr Nutr Rep. 2014;3:345–54.
14. Siscovick DS, Weiss NS, Fletcher RH, Lasky T. The incidence of primary cardiac arrest during vigorous exercise. N Engl J Med. 1984;311:874–7.
15. Erichsen R, Baron JA, Hamilton-Dutoit SJ, Snover DC, Torlakovic EE, Pedersen L, Frøslev T, Vyberg M, Hamilton SR, Sørensen HT. Increased risk of colorectal cancer development among patients with serrated polyps. Gastroenterology. 2016;150:895–902.
16. Rockenbauer M, Olsen J, Czeizel AE, Pedersen L, Sørensen HT, EuroMAP Group. Recall bias in a case-control surveillance system on the use of medicine during pregnancy. Epidemiology. 2001;12:461–6.
17. McGlynn KA, Hagberg K, Chen J, Graubard BI, London WT, Jick S, Sahasrabuddhe VV. Statin use and risk of primary liver cancer in the Clinical Practice Research Datalink. J Natl Cancer Inst. 2015;107(4).
18. Bloemenkamp KW, Rosendaal FR, Helmerhorst FM, Büller HR, Vandenbroucke JP. Enhancement by factor V Leiden mutation of risk of deep-vein thrombosis associated with oral contraceptives containing a third-generation progestagen. Lancet. 1995;346:1593–6.
19. Jick H, Derby LE, Myers MW, Vasilakis C, Newton KM. Risk of hospital admission for idiopathic venous thromboembolism among users of postmenopausal oestrogens. Lancet. 1996;348:981–3.
20. Robertson DJ, Larsson H, Friis S, Pedersen L, Baron JA, Sørensen HT. Proton pump inhibitor use and risk of colorectal cancer: a population-based, case-control study. Gastroenterology. 2007;133:755–60.
21. Michaëlsson K, Baron JA, Farahmand BY, Persson I, Ljunghall S. Oral-contraceptive use and risk of hip fracture: a case-control study. Lancet. 1999;353:1481–4.
22. Farahmand BY, Michaëlsson K, Ahlbom A, Ljunghall S, Baron JA, Swedish Hip Fracture Study Group. Survival after hip fracture. Osteoporos Int. 2005;16:1583–90.