SEMINAR

Introduction to Self-controlled Study Design

Masao Iwagami^{1,2}, Yoshinori Takeuchi³

¹ Department of Health Services Research, Faculty of Medicine, University of Tsukuba

² Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine

³ Department of Biostatistics, School of Public Health, Graduate School of Medicine, The University of Tokyo

ABSTRACT

Self-controlled study designs, also known as case-only designs or Self-controlled Crossover Observational PharmacoEpidemiologic (SCOPE) studies, include case-crossover (CCO) and self-controlled case series (SCCS). These designs compare different time windows (i.e., lengths of time) within the same person. An SCCS compares the occurrence of an outcome (event) during periods with and without exposure in the same person, whereas a CCO compares periods with and without the outcome for exposure. The main strength of self-controlled study designs is that they can ignore confounding factors that do not change over time (e.g., sex, genetics, habitual healthy or unhealthy behaviors). The effect of these factors are canceled out through statistical analyses, even if they are unknown or unmeasured. However, self-controlled study designs cannot be used for all research questions. Assumptions specific to each study design are needed. In CCO, there should be no substantial changes in exposure trends during the study period, the exposure should be transient (intermittent), and the outcome should be abrupt (sudden). In SCCS, event rates should be constant within each defined period and events must be independently recurrent or rare. In addition, the occurrence of an event should not affect subsequent exposures. Self-controlled study designs may be particularly useful in studies using electronic health records, in which some (time-invariant) confounding factors may not have been recorded, provided that the research question meets the assumptions required for each study design.

KEY WORDS

observational study, epidemiological method, case-crossover study, self-controlled case series, sequence symmetry analysis

1. INTRODUCTION

First, we will briefly review traditional epidemiologic designs and then introduce self-controlled study designs to aid a better understanding of them.

1.1. Traditional Epidemiologic Designs (Cohort and Case–Control Studies)

Traditional epidemiologic designs include cohort and case–control studies. Cohort studies compare people with and without exposure for the incidence of an outcome, whereas case–control studies compare people with and without an outcome for previous exposure. Let us take an example, whether benzodiazepine prescription increases the risk of hip fracture, possibly due to dizzi-

ness and falls, as a side-effect of benzodiazepine. **Fig. 1** is a graphical representation of cohort study and case–control study designs for an association between benzodiazepine (exposure) and hip fracture (outcome) [1].

An important objective of clinical or epidemiological research is to assess the causality between the exposure and outcome of interest. However, in observational studies, confounding factors can distort the effect estimates (e.g., odds ratio, hazard ratio) unless their influences are appropriately controlled. Some statistical methods, such as multivariable regression models and propensity score analysis, are frequently used to adjust for measured confounding factors. However, these methods are not useful for unknown or unmeasured confounders. In **Fig. 1**, people with and without benzodiazepine may be

systematically different for the measured and unknown/unmeasured confounders.

1.2. Self-Controlled Study Designs (Self-Controlled Case Series and Case-Crossover)

Instead of comparing different people, a comparison can be made between different time windows (i.e., lengths of

time) within a case who experienced an outcome of interest. Such study designs are known as self-controlled study designs, which include self-controlled case series (SCCS) and case-crossover (CCO) designs. An SCCS compares the incidence of outcome(s) during periods with [“risk period(s)”] and without exposure [“baseline period(s)”] in the same case. A CCO compares different periods with (“case period”) and without the outcome [“control period(s)”] for exposure. **Fig. 2** is a graphical representation of SCCS and CCO designs, using the same research question as that in **Fig. 1** [2].

As shown in **Table 1**, an SCCS is similar to a cohort study in that both designs are “exposure-anchored” [3]. That is, a comparison is made between people/periods with and without exposure for the incidence of an outcome. In contrast, a CCO is similar to a case-control study in that they are “outcome-anchored” [3]. That is, a comparison is made between people/periods with and without an outcome for exposure.

Self-controlled study designs do not require information on people who did not experience an outcome during the study period because they do not contribute to the statistical analysis, even if they are included in the analytical dataset. Therefore, it is sufficient to collect information for cases only (i.e., people with the outcome of interest) during the study period. For this reason,

Table 1 Traditional observational study designs and self-controlled study designs

	Exposure-anchored	Outcome-anchored
Traditional epidemiological design	Cohort study (comparing people with and without exposure)	Case-control study (comparing people with and without an outcome)
Self-controlled study design	Self-controlled case series (comparing periods with and without exposure)	Case-crossover design (comparing periods with and without an outcome)

self-controlled study designs are also called case-only designs [4]. A working group of the International Society for Pharmacoepidemiology recently published guidance for the application of self-controlled study designs in pharmacoepidemiology. The guidance recommends calling the study designs “Self-controlled Crossover Observational PharmacoEpidemiologic (SCOPE) studies” [3].

An advantage of self-controlled study designs is that they can ignore factors that do not change throughout the study period (e.g., sex, risk genes, healthy or unhealthy behaviors within a short follow-up length) in the statistical analyses. Their effects are canceled out, even if they are unknown or unmeasured. However, statistical adjustments are still necessary for factors that change over time within an individual (e.g., age, some medications).

Despite the value of self-controlled study designs, caution is needed because they cannot be used for all research questions. CCO and SCCS designs require several assumptions specific to each study design (as explained later) to estimate the effect of an exposure on an outcome (e.g., odds ratio, rate ratio). When these assumptions are violated, self-control study designs result in biased estimates.

2. CASE-CROSSOVER DESIGN

2.1. History

The CCO design was proposed by Maclure et al. in 1991 to examine the potential risk of acute exposures, such as coffee consumption and sexual activity, on the incidence of myocardial infarction [5]. Since then, the CCO design has been used to examine the risk of air pollution on cardiovascular and respiratory outcomes [6], and the risk of prescription drugs on specific adverse events [7].

2.2. Details of the Study Design

First, people with the outcome of interest should be identified. The index date is defined as the timing (e.g., day) when the first outcome occurred. Then, the time window (i.e., lengths of time) of the “case period” before the outcome occurred should be determined, assuming that the

exposure during that period could cause the outcome. Then, one or several “control period(s)”, which are typically prior to the “case period” for the same person, should be defined. In a CCO, researchers can freely define the “control period(s)”, such as the number and length of “control period(s)” and length of interval between “case period” and “control period(s)”. Thus, showing how the results change using several different definitions of “control periods” in the sensitivity analyses is recommended. Finally, the “case period” and “control period(s)” should be compared for the odds of (presence of) exposure to estimate the odds ratio.

2.3. Statistical Analyses

Similar to a matched case-control study design, a CCO requires the use of the Mantel-Haenszel method, a conditional logistic regression model (most frequently used), or conditional Poisson regression model to obtain the odds ratio between the “case period” and “control period(s)” in the same case [8]. Only discordant sets (where the case and at least one control period have a different exposure status) are used in the analysis, whereas concordant sets (where the case and all control periods have the same exposure status) do not contribute to the analysis. Therefore, people with constant exposure, as well as people with no exposure, throughout the entire study period are excluded from the analysis. Time-varying confounders, such as drug prescriptions, should be explicitly adjusted in the statistical analysis.

2.4. Assumptions Required for Case-Crossover Design

A constant exposure trend is assumed for an unbiased estimation via a CCO; that is, there should be no substantial change in exposure trends during the study period. For example, the assumption is violated if a new drug is coming to market and prevailing rapidly. In the case of a violation, some variants of a CCO have been proposed to deal with such an exposure time trend; a case-time-control design (which adjusts the population-level exposure time trend using data from people without the outcome) [9] or a case-case-time-control design

(where non-cases are sampled exclusively from future cases) [10].

In addition, the exposure should be transient (intermittent), and the outcome onset should be abrupt (sudden) [3]. The appropriateness of these conditions is judged from the viewpoint of the biological mechanism and/or common sense. For example, Maclure et al. expected that coffee consumption and sexual activity were intermittent and that myocardial infarction occurred abruptly [5]. Their expectations seem to be true and acceptable. In contrast, for a research question of the potential risk of an anti-hypertensive drug on cancer incidence, the prescription of anti-hypertensive drugs is mostly continuous once they are started, whereas cancer often develops gradually. Therefore, these situations do not meet the transient exposure and abrupt outcome conditions.

2.5. An Example of Case-Crossover Design

Miyamoto et al. used a CCO to examine the potential risk of pregabalin on injury, using Japanese medical claims data [11]. During the study period (from January 2014 to December 2016), they extracted 304 patients with records of injury and pregabalin prescription within 180 days prior to the date of injury (i.e., index date). The authors compared the odds of pregabalin prescription during a 30-day “case period” immediately before the index date and five 30-day “control periods” as the main analysis (Fig. 3), and during a 15-day “case period” and eleven 15-day “control periods” as a sensitivity analysis. The authors considered that people were using pregabalin if they had a ≥ 1 -day supply within the “case period” or “control periods”. Pregabalin prescription was significantly associated with an increased risk of injury, with an adjusted odds ratio (95% confidence interval) of 1.48

(1.10–2.00) in the main analysis and 1.92 (1.43–2.59) in the sensitivity analysis. The authors concluded that a prescription of pregabalin may be causally associated with an increased risk of injury.

3. SELF-CONTROLLED CASE SERIES

3.1. History

The SCCS was proposed by Farrington et al. in 1995 to examine the potential risk of the Measles Mumps Rubella vaccine on the incidence of aseptic meningitis [12]. Since then, it has been used primarily for vaccine safety assessments [13] as well as different types of exposures, including infectious episodes [14] and surgeries [15].

3.2. Details of the Study Design

First, people with the outcome of interest should be identified. Then, the study periods for cases should be defined arbitrarily (e.g., from January 2018 to December 2020, from 1 year before the event date to 1 year after the event date). The study periods of individuals are split into a “risk period(s)” (during which the risk of outcome is assumed to be increased because of the exposure) and a “baseline period(s)” for the remaining study periods. A person can have two or more “risk periods” if they are exposed several times during the study period. Finally, the “risk period(s)” are compared with “baseline period(s)” to determine the incidence of the outcome to estimate an incidence rate ratio. In an SCCS, there may be two or more occurrences of outcome in the same person.

3.3. Statistical Analyses

The conditional Poisson regression model is used for the SCCS analysis to obtain an incidence rate ratio for the

outcome between “risk period(s)” and “baseline period(s)”. Time-varying confounding factors, such as age and season, should be explicitly adjusted in the statistical model. If they are not adjusted, people with the outcome but without any exposure during the study period are excluded from the analysis because they have only a “baseline period”, which is canceled out. For this reason, some studies using an SCCS only recruit patients with an outcome *and* exposure during the study period [14]. If data are adjusted for confounding factors, the information of people with an outcome but without exposure can be used in the analysis. Commands and practice datasets for STATA, R, and SAS are available on a website (<http://sccs-studies.info/index.html>) maintained by Farrington et al.

3.4. Assumptions Required for Self-Controlled Case Series

SCCS studies require the following assumptions [16]. First, event rates should be constant within each defined period (i.e., each “risk period” and “baseline period”). Second, events should be independently recurrent or rare. Therefore, if two or more outcomes can occur in the same case, the prior outcome should not change the probability of subsequent outcomes. For example, the first incidence of stroke is expected to increase the probability of subsequent stroke. In such a case, this assumption is violated. If the event is rare, this violation of assumption can be overcome by focusing only on the first occurrence of the outcome. Third, the occurrence of an event should not affect subsequent exposures. More specifically, an event should not temporarily decrease or increase the probability of exposure. For example, to assess the potential risk of influenza vaccination on seizure, vaccination is probably delayed when a person has a

seizure event. In such a case, this assumption is violated. One way to minimize (correct for) this violation of assumption is to define a “pre-exposure period” just before exposure, and estimate rates separately for the “risk period”, “baseline period”, and “pre-exposure period”. Furthermore, an event should not be fatal because no exposure can occur after death. One way to deal with this violation of assumption is to conduct a sensitivity analysis excluding cases who died just after the event.

3.5. An Example of Self-Controlled Case Series

Ohbe et al. used an SCCS to examine the potential risk of a traumatic skin wound on the first incidence of infective endocarditis, using two Japanese medical claims databases [17]. The researchers extracted 159 patients with an outpatient diagnosis of traumatic skin wounds and hospitalization for infective endocarditis during the study period (from April 2012 to August 2018) in one database for younger people, and 290 patients between March 2012 and February 2017 in another database for older people. The authors defined the “risk period” as within 1–16 weeks after exposure to the traumatic skin wound, and split it into four “risk periods”, 1–4, 5–8, 9–12, and 13–16 weeks after exposure to the traumatic skin wounds. Other observational periods were regarded as the “baseline period” (Fig. 4). Compared with the “baseline period”, the incidence rate ratios (95% confidence intervals) for young people were 3.78 (2.07–6.92), 1.58 (0.64–3.89), 1.60 (0.65–3.94), and 1.29 (0.47–3.53) at 1–4, 5–8, 9–12, and 13–16 weeks after exposure to the traumatic skin wounds, respectively. The corresponding figures for older people were 2.61 (1.67–4.09), 1.73 (1.01–2.94), 1.19 (0.63–2.27), and 1.52 (0.82–2.74), respectively.

Fig. 5 Graphical representation of sequence symmetry analysis for the association between benzodiazepine and hip fracture

The authors concluded that traumatic skin wounds may be causally associated with an increased risk of infective endocarditis.

4. SEQUENCE SYMMETRY ANALYSIS

In addition to SCCS and CCO designs, sequence symmetry analysis (SSA) is a method related to the self-controlled study design. SSA was developed to examine symmetry in the distribution of an event before and after an exposure of interest [18]. SSA generally reports a sequence ratio, whereby the ratio of the exposure-event sequence orders approximates the incidence ratio in exposed and non-exposed person-time within a case [19]. In SSA, as with other self-controlled methods, the effect of time-invariant confounders is adjusted by statistical analysis without explicit modeling. SSA has been applied for pharmacoepidemiological studies [20] and signal detection in pharmacovigilance [21].

Fig. 5 shows a typical setting of analytical periods in SSA for an association between benzodiazepine and hip fracture for example. The sequence ratio is calculated by dividing the number of subjects who experienced hip fracture during a defined number of days after benzodiazepine exposure (indicated by the black arrow, i.e., from exposure to outcome) by the number of subjects who experienced hip fracture during the same number of days before benzodiazepine exposure (indicated by the white arrow, i.e., from event to exposure). If a sequence ratio is larger than 1, this indicates that exposure may increase the risk of outcome.

5. CONCLUSION

We provided an overview and some examples of self-

controlled study designs, including CCO and SCCS designs. The main strength of self-controlled study designs, compared with traditional epidemiologic designs (cohort and case-control studies), is that time-invariant confounding factors can be canceled out, even if they are unknown or unmeasured. However, self-controlled study designs should be used carefully. If the aforementioned assumptions are not met, the results can be biased. Self-controlled study designs may be particularly useful in studies using electronic health records, in which some (time-invariant) confounding factors may not be recorded, provided that the research question meets assumptions required for each study design.

ACKNOWLEDGMENTS

We would like to thank Dr. Yoshihisa Miyamoto in the Epidemiology and Prevention Group, Center for Public Health Sciences, National Cancer Centre, and Dr. Hiroyuki Ohbe in the Department of Clinical Epidemiology and Health Economics, School of Public Health, The University of Tokyo, for their critical reading of the manuscript and feedback. The English editing of the current paper was supported by a grant-in-aid from the Ministry of Health, Health, Labour and Welfare Policy Research Grants, Japan; (19AA2007).

CONFLICT OF INTERESTS

Y.T. has received consultant fees from Pharmaceuticals and Medical Devices Agency and EPARK, Inc. Additionally, Y.T. has conducted a collaborative study with Pfizer inc., which is not associated with this paper. No other potential competing interests relevant to this paper are reported.

REFERENCES

1. Requena G, Huerta C, Gardarsdottir H, Logie J, González-González R, Abbing-Karahagopian V, et al. Hip/Femur Fractures Associated With The Use Of Benzodiazepines (Anxiolytics, Hypnotics And Related Drugs): A Methodological Approach To Assess Consistencies Across Databases From The PROTECT-EU Project. *Pharmacoepidemiol Drug Saf* 2016;25:66–78.
2. Requena G, Logie J, Martin E, Boudiaf N, González González R, Huerta C, et al. Do Case-Only Designs Yield Consistent Results Across Design And Different Databases? A Case Study Of Hip Fractures And Benzodiazepines. *Pharmacoepidemiol Drug Saf* 2016;25:79–87.
3. Cadarette SM, Maclure M, Delaney JAC, Whitaker HJ, Hayes KN, Wang SV, et al. Control Yourself: ISPE-Endorsed Guidance In The Application Of Self-Controlled Study Designs In Pharmacoepidemiology. *Pharmacoepidemiol Drug Saf* 2021;30:671–84.
4. Nordmann S, Biard L, Ravaud P, Esposito-Farèse M, Tubach F. Case-Only Designs In Pharmacoepidemiology: A Systematic Review. *PLoS One* 2012;7:e49444.
5. 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.
6. Carracedo-Martínez E, Taracido M, Tobias A, Saez M, Figueiras A. Case-Crossover Analysis Of Air Pollution Health Effects: A Systematic Review Of Methodology And Application. *Environ Health Perspect* 2010;118:1173–82.
7. Consiglio GP, Burden AM, Maclure M, McCarthy L, Cadarette SM. Case-Crossover Study Design In Pharmacoepidemiology: Systematic Review And Recommendations. *Pharmacoepidemiol Drug Saf* 2013;22:1146–53.
8. Mostofsky E, Coull BA, Mittleman MA. Analysis of Observational Self-matched Data to Examine Acute Triggers of Outcome Events with Abrupt Onset. *Epidemiology* 2018;29:804–16.
9. Suissa S. The Case-Time-Control Design. *Epidemiology* 1995;6:248–53.
10. Wang S, Linkletter C, Maclure M, Dore D, Mor V, Buka S, et al. Future Cases As Present Controls To Adjust For Exposure Trend Bias In Case-Only Studies. *Epidemiology* 2011;22:568–74.
11. Miyamoto Y, Iwagami M, Aso S, Matsui H, Doi K, Yasunaga H. Pregabalin And Injury: A Nested Case-Control And Case-Crossover Study. *Pharmacoepidemiol Drug Saf* 2020;29:558–64.
12. Farrington CP. Relative Incidence Estimation From Case Series For Vaccine Safety Evaluation. *Biometrics* 1995;51:228–35.
13. Gault N, Castañeda-Sanabria J, De Rycke Y, Guillo S, Foulon S, Tubach F. Self-Controlled Designs In Pharmacoepidemiology Involving Electronic Healthcare Databases: A Systematic Review. *BMC Med Res Methodol* 2017;17:25.
14. Kwong JC, Schwartz KL, Campitelli MA, Chung H, Crowcroft NS, Karnauchow T, et al. Acute Myocardial Infarction After Laboratory-Confirmed Influenza Infection. *N Engl J Med* 2018;378:345–53.
15. Shimada YJ, Tsugawa Y, Brown DFM, Hasegawa K. Bariatric Surgery And Emergency Department Visits And Hospitalizations For Heart Failure Exacerbation: Population-Based, Self-Controlled Series. *J Am Coll Cardiol* 2016;67:895–903.
16. Petersen I, Douglas I, Whitaker H. Self Controlled Case Series Methods: An Alternative To Standard Epidemiological Study Designs. *BMJ* 2016;354:i4515.
17. Ohbe H, Iwagami M, Sasabuchi Y, Yasunaga H. Increased Risk Of Infective Endocarditis After Traumatic Skin Wound. *Heart* 2021;heartjnl-2020-318632.
18. Hallas J. Evidence Of Depression Provoked By Cardiovascular Medication: A Prescription Sequence Symmetry Analysis. *Epidemiology* 1996;7:478–84.
19. Takeuchi Y, Shinozaki T, Matsuyama Y. A Comparison Of Estimators From Self-Controlled Case Series, Case-Crossover Design, And Sequence Symmetry Analysis For Pharmacoepidemiological Studies. *BMC Med Res Methodol* 2018;18:4.
20. Takeuchi Y, Kajiyama K, Ishiguro C, Uyama Y. Atypical Antipsychotics And The Risk Of Hyperlipidemia: A Sequence Symmetry Analysis. *Drug Saf* 2015;38:641–50.
21. Pratt N, Chan EW, Choi NK, Kimura M, Kimura T, Kubota K, et al. Prescription Sequence Symmetry Analysis: Assessing Risk, Temporality, And Consistency For Adverse Drug Reactions Across Datasets In Five Countries. *Pharmacoepidemiol Drug Saf* 2015;24:858–64.