

A Novel Weighting Method to Remove Bias from Within-subject Exposure Dependency in Case-crossover Studies

Kiyoshi Kubota (✉ kubota@dsrujp.org)

NPO Drug Safety Research Unit Japan

Lan Kelly

University of South Australia

Tsugumichi Sato

Tokyo University of Science

Nicole Pratt

University of South Australia

Elizabeth Roughead

University of South Australia

Takuhiro Yamaguchi

Tohoku University

Research Article

Keywords: case-crossover study, bias, autocorrelation

Posted Date: April 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-381953/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A novel weighting method to remove bias from within-subject exposure dependency in case-crossover studies

Kiyoshi Kubota ^{1,2}

Lan Kelly ³

Tsugumichi Sato ^{1,2}

Nicole Pratt ³

Elizabeth Roughead ³

Takuhiro Yamaguchi ⁴

1 NPO Drug Safety Research Unit Japan, Chiyoda-ku, Tokyo, Japan

2 Department of Pharmacy, Tokyo University of Science, Chiba, Japan

3 Quality Use of Medicines Pharmacy Research Centre, Clinical and Health Sciences, University of South Australia, Adelaide, Australia

4 Division of Biostatistics, Tohoku University Graduate School of Medicine, Miyagi, Japan

Correspondence: Kiyoshi Kubota, MD, PhD. NPO Drug Safety Research Unit Japan, 6-2-9-2F, Soto-Kanda, Chiyoda-ku, Tokyo, 101-0021, Japan.

Tel: +81-3-6284-4206 Fax: +81-3-6284-4207

e-mail: kubota@dsrujp.org.

Abstract

Background:

Case-crossover studies have been widely used in various fields including pharmacoepidemiology. Vines and Farrington indicated in 2001 that when within-subject exposure dependency exists, conditional logistic regression can be biased. However, this bias has not been well studied.

Methods:

We have extended findings by Vines and Farrington to develop a weighting method for the case-crossover study which removes bias from within-subject exposure dependency. Our method calculates the exposure probability at the case period in the case-crossover study which is used to weight the likelihood formulae presented by Greenland in 1999. We simulated data for the population with a disease where most patients receive a cyclic treatment pattern with within-subject exposure dependency but no time trends while some patients stop and start treatment. Finally, the method was applied to real-world data from Japan to study the association between celecoxib and peripheral edema and to study the association between selective serotonin reuptake inhibitor (SSRI) and hip fracture in Australia.

Results:

When the simulated rate ratio of the outcome was 4.0 in a case-crossover study with no time-varying confounder, the proposed weighting method and the Mantel-Haenszel odds ratio reproduced the true rate ratio. When a time-varying confounder existed, the Mantel-Haenszel method was biased but the weighting method was not. When more than one control period was used, standard conditional logistic regression was biased either with or without time-varying confounding and the bias increased (up to 9.4) when the study period was extended. In real-world analysis with a binary exposure variable in Japan and Australia, the point estimate of the odds ratio (around 2.5 for the association between celecoxib and peripheral edema and around 1.6 between SSRI and hip fracture) by our weighting method was equal to the Mantel-Haenszel odds ratio and stable compared with standard conditional logistic regression.

Conclusion:

Case-crossover studies may be biased from within-subject exposure, even without exposure time trends. This bias can be identified by comparing the odds ratio calculated by the Mantel-Haenszel method and that by standard conditional logistic regression. Our proposed method will remove bias from within-subject exposure dependency and can account for time-varying confounders.

Key words:

case-crossover study; bias; autocorrelation

Background

The case-crossover design has been widely used since it was proposed in 1991 [1]. The design has been used in various fields including pharmacoepidemiology [2], occupational epidemiology [3], studies on traffic safety [4] and air pollution health effects [5, 6]. In case-crossover studies, individuals who have experienced the outcome (cases) act as their own controls by including one or more periods before the onset of the outcome. The period including the outcome is the case period, while period(s) prior to the case period act as the controls. The number of control periods can be large: for example, in the original article [1], one analysis involved 8,766 control periods. The effect of the exposure should be brief; exposure in any period should affect the outcome in that period only, without any 'carryover effect' [1, 6, 7]. In addition, in case-crossover studies, the rate of outcome occurrence is usually assumed to be unchanged during exposed or unexposed periods, respectively. However, like other case-only studies, the case-crossover study has an advantage that the effect of time-invariant confounders is automatically controlled because the case period is compared with control period(s) of the same individual [7-9]. The case-crossover study also has unique characteristics. For example, the original unidirectional case-crossover study does not include periods after the outcome occurs. Thus, there is no bias due to the outcome influencing future exposures or future observation periods, unlike other case-only studies such as self-controlled case series [10-12].

However, case-crossover studies are susceptible to at least two types of major biases. The first is bias due to time trends in the exposure which can be removed using a variant of case-crossover studies, the case-time-control design [9, 13, 14]. The second is bias due to within-subject exposure dependency or autocorrelation in an individual's exposure history [6, 15, 16]. This bias is particularly important in pharmacoepidemiology because drug use on one day is rarely independent from use in the preceding days. However, the potential for this bias has had little attention in the pharmacoepidemiology literature, and standard conditional logistic regression has been used without assessing whether bias due to within-subject exposure dependency exists in many case-crossover studies [17-21].

In this paper, we describe how to remove bias from within-subject exposure dependency, when time trends in the exposure do not exist. We used these methods to analyze both simulated data and real-world data from Japan and Australia.

Methods

BIAS DUE TO CONDITIONAL LOGISTIC REGRESSION FOR CASE-CROSSOVER STUDIES

In 2001, Vines and Farrington proposed the likelihood for case-crossover studies as [15]:

$$L = \prod_{i=1}^N \frac{\exp(\beta x_{i0}) \sum_{\kappa^*} P\{X_{i0}=x_{i\kappa^*(0)}, \dots, X_{iM}=x_{i\kappa^*(M)}\}}{\sum_{\kappa} \exp(\beta x_{i\kappa(0)}) P\{X_{i0}=x_{i\kappa(0)}, \dots, X_{iM}=x_{i\kappa(M)}\}} \quad (1)$$

where X_{i0} is the exposure level at the case period ($m=0$) and X_{im} is the exposure level at the m -th control period at t ($t = -m$: $m=1, 2, \dots, M$), x_{i0} denotes the observed exposure level at the case period and x_{im} denotes the observed exposure level at the m -th control period, the sum in the denominator ranges over all permutations of κ of the integers $\{0, 1, \dots, M\}$ and the sum in the numerator ranges over the subset of the permutations for which $x_{i\kappa(0)} = x_{i0}$ of the individual i . In Equation (1), X_{i0} may be a binary exposure variable but can be a multi-level exposure variable or a vector of an exposure and time-varying confounders. When X_{im} denotes a binary exposure, the denominator in Equation (1) becomes $\exp(\beta) \sum_{\kappa_1} P(X_{i0} = 1, N_{exposed} = k) + \sum_{\kappa_0} P(X_{i0} = 0, N_{exposed} = k)$ where $N_{exposed}$ is the total number of exposed (case and control) periods, given by $N_{exposed} = \sum_{m=0}^M X_{im}$ and $\sum_{\kappa_1} P(X_{i0} = 1, N_{exposed} = k)$ is the sum of probabilities for all the permutations of k exposed and $(M+1-k)$ unexposed periods where $X_{i0} = 1$, and $\sum_{\kappa_0} P(X_{i0} = 0, N_{exposed} = k)$ is that where $X_{i0} = 0$. Data on $N_{exposed} = k$ is informative only when the positivity (non-zero probability) condition is satisfied or $\sum P(X_{i0} = 1, N_{exposed} = k) > 0$ and $\sum P(X_{i0} = 0, N_{exposed} = k) > 0$. Otherwise they do not contribute to the estimation of $\exp(\beta)$.

Let OR_{VF} be the estimate of $\exp(\beta)$ obtained from Equation (1). The likelihood in Equation (1) and OR_{VF} are in general different from the following likelihood for the standard conditional logistic (SCL) regression for individually matched case-control studies and its estimate, OR_{SCL} .

$$L = \prod_{i=1}^N \frac{\exp(\beta x_{i0})}{\sum_{j=0}^M \exp(\beta x_{ij})} \quad (2)$$

Vines and Farrington showed that the likelihoods in Equations (1) and (2) are equivalent if $P\{X_{i0} = x_{i0}, \dots, X_{iM} = x_{iM}\} = P\{X_{i0} = x_{i\kappa(0)}, \dots, X_{iM} = x_{i\kappa(M)}\}$ for all permutations κ of $\{0, 1, \dots, M\}$, that is, global exchangeability holds. For example, Equations (1) and (2) are

equivalent when the exposure status in one period is independent from the status in any other periods and the exposure probability is the same in all of case and control periods (Additional File 1, Appendix 1). If global exchangeability does not hold, OR_{SCL} can be biased.

Vines and Farrington did not show explicitly how to estimate $P\{X_{i0} = x_{ik(0)}, \dots, X_{iM} = x_{ik(M)}\}$ in Equation (1). These probabilities may be estimated by the proportion of each permutation in the population which contain cases, or in samples representing the population such as time-controls in the case-time-control design proposed by Suissa [14]. In the next section, however, we introduce a different approach to remove the bias from within-subject exposure dependency by assuming that pairwise exchangeability is satisfied but global exchangeability may not necessarily hold.

WEIGHTING METHOD TO REMOVE BIAS DUE TO WITHIN-EXPOSURE DEPENDENCY

Case-crossover studies with a binary exposure

When pairwise exchangeability, $P\{X_{i0} = 1, X_{im} = 0\} = P\{X_{i0} = 0, X_{im} = 1\}$ is satisfied for a binary exposure in all control periods m ($m=1, 2, \dots, M$), the estimate of $\exp(\beta)$ using the Mantel-Haenszel method (OR_{MH}) is unbiased whether within-subject exposure dependency exists or not [15]. In line with this finding, we will show that when pairwise exchangeability holds, the probabilities that the individual is unexposed (π_0) and exposed (π_1) at the case period, can be estimated from the cases in a case-crossover study (without requiring data from the population or time-controls). Once π_0 , π_1 , and the relative exposure probability $\pi_{10} = \pi_1/\pi_0$ are estimated, the following likelihood for case-crossover studies proposed by Greenland [22] can be used to obtain an unbiased estimate of $\exp(\beta)$, defined as OR_G :

$$L = \prod_{i=1}^N \frac{\exp(\beta x_{ic}) \pi_{ic}}{\sum_k \exp(\beta x_k) \pi_{ik}} = \prod_{i=1}^N \frac{\exp(\beta x_{ic}) \pi_c}{\sum_k \exp(\beta x_k) \pi_k} = \prod_{i=1}^N \frac{\exp(\beta x_{ic}) \pi_{c0}}{\sum_k \exp(\beta x_k) \pi_{k0}} \quad (3)$$

In the left-hand side of Equation (3), π_{ik} is the probability that individual i has the k -th exposure level at the case period ($k=0, 1$ for binary exposure), π_{ic} is π_{ik} observed, and x_{ic} is the exposure level observed when individual i has the outcome. In the middle of Equation (3), subscript i is omitted in the exposure probabilities as π_{ik} is replaced by the expected value in the population in the current study. In the right-hand side of Equation (3) $\pi_{k0} = \pi_k/\pi_0$. Equation (3) stands for the model where x_{ik} is a binary variable, multi-level exposure variable, or a vector of the observed exposure and time-varying confounders.

For a binary exposure, the right-hand side of Equation (3) can be rewritten as

$$L = \prod_{i=1}^N \frac{\exp(\beta x_{ic}) \pi_{c0}}{1 + \exp(\beta) \pi_{10}} \quad (4)$$

where $x_{ic} = 1$ and $\pi_{c0} = \pi_{10} = \pi_1/\pi_0$ when the individual i was exposed at the case period and $x_{ic} = 0$ and $\pi_{c0} = \pi_{00} = \pi_0/\pi_0 = 1$ when unexposed. As Vines and Farrington showed [15], Greenland did not estimate π_k for case-crossover studies in Equation (3) when within-subject exposure dependency exists.

We outline a novel weighting method using a modified version of Greenland's likelihood. We propose that π_k can be estimated, with or without within-subject exposure dependency, by assuming pairwise exchangeability. Let $P_{kl[m]}$ denote the joint probability that the subject

has the k -th exposure level at the case period and has the l -th exposure level at the m -th control period:

$$P_{kl[m]} = P(X_0 = x_k, X_m = x_l) \quad m=1, 2, \dots, M \quad (5)$$

In Equation (5), X_0 is the exposure status at the case period and X_m is the exposure status at the m -th control period. When the exposure variable is binary, both X_0 and X_m have two levels ($k=0, 1$ and $l=0,1$) and pairwise exchangeability is equivalent to stationary exposure (no time trends): when the exposure process is stationary, $P_{10[m]} + P_{11[m]} = P_{01[m]} + P_{11[m]}$ and this relationship leads to the pairwise exchangeability condition $P_{10[m]} = P_{01[m]}$ ($m=1, 2, \dots, M$).

Using conditional probabilities, pairwise exchangeability, $P\{X_{i0} = 1, X_{im} = 0\} = P\{X_{i0} = 0, X_{im} = 1\}$ can be rewritten as:

$$\pi_1 P(X_m = 0 | X_0 = 1) = \pi_0 P(X_m = 1 | X_0 = 0) \quad (6)$$

where $\pi_0 = P(X_0 = 0)$ and $\pi_1 = P(X_0 = 1)$. When both sides of Equation (6) are summed up over M control periods ($m=1, 2, \dots, M$), we obtain:

$$\pi_1 \sum_{m=1}^M P(X_m = 0 | X_0 = 1) = \pi_0 \sum_{m=1}^M P(X_m = 1 | X_0 = 0) \quad (7)$$

The quantity $\sum_{m=1}^M P(X_m = 0 | X_0 = 1)$ can be estimated by the average number of unexposed control periods (where $X_m = 0$) in those exposed at the case period ($X_0 = 1$). Similarly, the quantity $\sum_{m=1}^M P(X_m = 1 | X_0 = 0)$ can be estimated as the average number of exposed control periods ($X_m = 1$) in those unexposed at the case period ($X_0 = 0$). This average, defined as \overline{PT}_{10} and \overline{PT}_{01} , respectively, can be written as:

$$\overline{PT}_{10} = \sum_i PT_{10i}/a_1 \quad \text{and} \quad \overline{PT}_{01} = \sum_i PT_{01i}/a_0 \quad (8)$$

where PT_{10i} is the number of unexposed control periods (person-time) of case i who is exposed at the case period, PT_{01i} is the number of exposed control periods (person-time) of case i who is unexposed at the case period, and a_1 is the number of discordant exposed cases with at least one unexposed control period and a_0 is the number of discordant unexposed cases with at least one exposed control period. When $\sum_{m=1}^M P(X_m = 0 | X_0 = 1)$ and $\sum_{m=1}^M P(X_m = 1 | X_0 = 0)$ in Equation (7) are substituted by \overline{PT}_{10} and \overline{PT}_{01} , respectively, we obtain:

$$\pi_{10} = \pi_1/\pi_0 = \sum_{m=1}^M P(X_m = 1|X_0 = 0) / \sum_{m=1}^M P(X_m = 0|X_0 = 1) = \overline{PT_{01}}/\overline{PT_{10}} \quad (9)$$

When standard statistical software is used for conditional logistic regression analysis of case-crossover studies, we introduce weighting to ensure the denominator in Equation (2) equals that in Equation (4). Every exposed and unexposed period in case i should be weighted by w_{i1} and by w_{i0} , respectively, defined as follows:

$$w_{i1} = \pi_{10}/m_i^1 \quad \text{and} \quad w_{i0} = 1/m_i^0 \quad (10)$$

where m_i^1 is the number of exposed periods and m_i^0 is the number of the unexposed periods (including both case and control periods) in case i and π_{10} is estimated from Equation (9). In most statistical software, w_{ik} ($k=0,1$) may be specified as an offset variable in conditional logistic regression which uses the following likelihood:

$$L = \prod_{i=1}^N \frac{\exp(\beta x_{i0})}{\sum_{j=0}^M w_{ij} \exp(\beta x_{ij})} \quad (11)$$

where $w_{ij}=w_{i1}$ and $w_{ij}=w_{i0}$ when j-th period is exposed and unexposed, respectively, in case i.

Using a_1 and a_0 , Equation (4) can be rewritten as

$$L = \left(\frac{\exp(\beta)\pi_{10}}{1+\exp(\beta)\pi_{10}} \right)^{a_1} \left(\frac{1}{1+\exp(\beta)\pi_{10}} \right)^{a_0} \quad (12)$$

Equation (12) gives (see Additional File 1, Appendix 2) the following maximum likelihood estimate for OR_G :

$$OR_G = \exp(\beta) = \frac{a_1}{a_0} \frac{1}{\pi_{10}} \quad (13)$$

and the variance:

$$v(\beta) = \frac{1}{a_0} + \frac{1}{a_1} \quad (14)$$

From Equations (8), (9) and (13) we obtain;

$$OR_G = \exp(\beta) = \frac{a_1}{a_0} \frac{1}{\pi_{10}} = \frac{a_1}{a_0} \frac{\overline{PT_{10}}}{\overline{PT_{01}}} = \frac{\sum_i PT_{10i}}{\sum_i PT_{01i}} = OR_{MH} \quad (15)$$

Equation (15) shows that when the model involves only one binary exposure variable, the point estimate of OR_G is the same as OR_{MH} but the variance of $\log(OR_G)$ from Equation (14) can be greater or smaller than the variance of $\log(OR_{MH})$, though they can be the same (Additional File 1, Appendix 3).

Case-crossover studies with a binary exposure and a binary time-varying confounder

The Mantel-Haenszel estimator is unbiased for binary exposures when there is pairwise exchangeability but cannot be used when there is time varying confounding. On the other hand, our method can be extended to studies with time-varying confounder. For example, when there is a binary exposure (x) and a binary time-varying confounder variable (z), βx_{ij} in the likelihood in Equation (11) is specified as $(\beta, \gamma)(x_{ij}, z_{ij})^T$, which is equal to 0 when $(x_{ij}, z_{ij})=(0,0)$, β when $(x_{ij}, z_{ij})=(1,0)$, γ when $(x_{ij}, z_{ij})=(0,1)$, and $\beta + \gamma$ when $(x_{ij}, z_{ij})=(1,1)$ where $\exp(\gamma)$ is an estimate of the odds ratio of z . Similarly to the finding that OR_{SCL} in Equation (2) is unbiased when within-subject exposure dependency does not exist (Additional File1, Appendix 1), we may estimate an unbiased OR_G in the model involving the exposure and time-varying confounder by calculating the weight from the exposure variable x (Equation 10), if within-subject dependency does not exist for z during exposed periods (where the probability that the confounder is positive is f_1) and during unexposed periods (f_0) and the confounder is adjusted for as in the standard conditional logistic regression (see Appendix 4 in Additional File 1 for the detail; as to the relevant SAS codes, see 5-2d in Appendix 5 in Additional File1).

SIMULATION STUDIES

We simulated data relevant to drug therapy with no time trends and with autocorrelated exposure patterns (within-subject dependency) (Additional File 1, Appendix 5). The simulated data is created based on the following observations (i) drug treatment sometimes

has a cyclic pattern which often produces within-subject exposure dependency, (ii) some outcomes (e.g., acute adverse events) tend to occur soon after the treatment is initiated but they may also occur later during the drug therapy, and (iii) some patients stop treatment for various reasons while some patients start treatment. When the rate of stopping treatment is the same as that of starting treatment, the stationarity of the exposure may be maintained in the population as follows. We simulated scenarios with and without time varying confounding.

Assume that drug treatment involves a cyclic pattern where one cycle consists of 7 days and a patient has a drug on days 1 and 4 but no drug on days 2, 3, 5, 6, and 7. Figure 1 depicts three cycles of drug treatment with 8 subgroups consisting of 1 case period and 21 control periods, where 1 period is 1 day. Subgroup A represents stoppers who stop treatment at the case period while Subgroup H represents new users who start treatment at the case period. Subgroups B to G represent patients being treated with a different timing relative to the start of the treatment cycle. In Figure 1, the proportion of those exposed to a drug in the population is always $\frac{1}{4}$, indicating stationarity (no time trends).

Figure 2 shows 140 cases who had the outcome at the case period when the size of each of Subgroups A to H, $N=10,000$, the event rate in an unexposed period (r_0) is 0.001 per period and the rate ratio is 4. We assume that the expected number of cases is determined by exposure at the case period only, or Nr_0 when unexposed and $N RR r_0$ when exposed at the case period.

Figure 3 shows 184 cases who had the outcome at the case period when $N=10,000$, the event rate in the unexposed period without time-varying confounding is 0.001 per period and the rate ratio is 4 and 2 for the exposure and time-varying confounder, respectively. The status of the time-varying confounder in the unexposed or exposed period is related to the exposure status of that period only, and $f_0 = 0.2$ and $f_1 = 0.4$ where f_0 and f_1 are the expected values of the proportion of exposed periods and unexposed periods in the population, respectively, when the confounder is positive (See Appendix 4, Additional File 1). We assume that the number of cases is as expected, or $Nr_0(1 - f_0)$, $N RR_z r_0 f_0$, $N RR r_0(1 - f_1)$, and $N RR RR_z r_0 f_1$, when the combination of the exposure and time-varying confounder variables (x, z) at the case period is (0, 0), (0, 1), (1, 0), and (1, 1), respectively. The status of the time-varying confounder in the control periods was randomly generated and is therefore independent of the exposure or time-varying confounder at different periods.

To determine the effect of the length of each time period on the estimated odds ratio, we also analyzed the data by dividing 22 days in Figures 2 and 3 into 7, 5, 3 and 2 periods where 1 period included 3, 4, 7 and 11 days, respectively. The status of the exposure and

time varying confounder was defined by the last day of each period (Definition I in Table 1).

We analyzed the data by standard conditional regression, the Vines and Farrington method, our modified Greenland's method, and Mantel-Haenszel methods for the fixed study period of 22 days. We analyzed the data assuming 1 period = 1, 2, 3, 4, 7 and 11 days and M control periods, $M=1, \dots, 21$. Data simulation and analyses were performed using SAS 9.4.

CASE-CROSSOVER STUDIES OF REAL-WORLD DATA

The method was also applied to data from Japanese and Australian databases. The Japanese study on the association between celecoxib and peripheral edema from a previous study [23] was approved by the ethics committee of Tokyo University of Science (approval number: 18023) where obtaining the informed consent from study subjects was waived for the current study. The Japanese data came from 25 corporate-type health insurance plans provided by Medi-Scope® [24]. Claims data covering 60 months between May 2013 to April 2018 included 1,163,968 males (age (SD)=42.5 (13.2) years old) and 1,349,901 females (42.2 (13.1) years old) who were 20 years old or older (but younger than 75 years old). As detailed in Additional File 1 (Appendix 6), we examined 99,821 new users of celecoxib, who used celecoxib after at least 180 days of non-use. The occurrence of peripheral edema was defined by new use of furosemide after at least 180 days of non-use and the index date was the day when the outcome occurred. Daily exposure during an 84-day study period was determined using a 7-day grace period. We selected 311 cases who had both exposed and unexposed days during the study period. The 84-day study period was divided into $(M+1)$ periods where $M=1, 2, 5, 11, 27$ and 83 , resulting in periods of 42, 28, 14, 7, 3 and 1 days, respectively. Four different definitions were used to determine exposure in the case and control periods as shown in Table 1. Cases who started celecoxib on the index date were excluded from the analysis since furosemide could have been prescribed for prevention rather than treatment of edema.

The data was analyzed by standard conditional logistic regression (Equation (2)), the weighting method for a binary exposure (Equation (4)), and the Mantel-Haenszel method. We also extended the study period to 168 and 336 days. All the analyses were performed using SAS 9.4. Data and SAS codes to analyze the data are available upon request.

The Australian study investigated the association between hip fracture and psychoactive medicines, which has been described elsewhere [25,26]. The data were obtained from the Australian Department of Veterans' Affairs administrative claims database. The study was approved by Department of Defense and Veterans' Affairs Human Research Ethics (E016-007) and University of South Australia Human Research Ethics (P203/04) where obtaining the informed consent from study subjects was waived for the current study.

Psychoactive medicines included benzodiazepines, selective serotonin re-uptake inhibitors (SSRIs), opioids, antipsychotics and tricyclic antidepressants. The cases were 8,828 patients aged over 65 years who were hospitalized for hip fracture between 2009 and 2012. The index date for each case was the date of hospitalization.

A previous case-crossover study found an increased risk of hip fracture for opioids, SSRIs and antipsychotics [25]. A related case-control study found an association between hip fracture and SSRIs when used concurrently with other psychoactive medicines [26]. We have re-analyzed the case-control data as case-crossover study using the same methods as the Japanese study.

Both of the studies in Japan and Australia were carried out in accordance with the Declaration of Helsinki, and all methods were carried out in accordance with relevant guidelines and regulations in Japan and Australia.

Results

SIMULATION STUDIES: COMPARISON OF METHODS WITH OR WITHOUT TIME-VARING CONFOUNDING

Table 2 shows OR_{SCL} , OR_{VF} , OR_G , and OR_{MH} with their 95% confidence intervals (CIs) estimated for the scenario in Figure 2 for M control periods (1 period=1 day). The difference between OR_{SCL} and true RR (4.0) was more than 10 % of the true value when M=6 or >7 and increased when M increased. When control periods were extended to 10 and 20 cycles (70 and 140 control periods), the point estimate of OR_{SCL} was 7.85 and 9.37, respectively (not shown in Table 2). On the other hand, odds ratios from the remaining 3 methods (OR_{VF} , OR_G , and OR_{MH}) were unbiased irrespective of the value of M. The estimate of OR_{VF} cannot be estimated when M=7, 10, 14, 17 and 21 because the positivity condition was not satisfied for any data. For example, when M=7, $\sum P(X_{i0} = 0, N_{exposed} = 1) = 0$ because $X_{i0} = 1$ in Subgroup H which is only one subgroup where $N_{exposed} = 1$ and similarly for $N_{exposed} = 2$ or 3, X_{i0} was the same for all subgroups. When M=1 and 2, the CIs for OR_{VF} , OR_G , and OR_{MH} were the same. Otherwise, the CIs of OR_{VF} were wider than those of OR_G (except when M=5). On the other hand, the CIs of OR_{MH} were wider (when M=3 to 5) or narrower (when M>5) than those of OR_G being compatible with the finding that the variance of $\log(OR_G)$ can be larger or smaller than or equal to that of $\log(OR_{MH})$ (Additional File 1, Appendix 3),

Table 3 shows OR_{SCL} , OR_{VF} , OR_G , and OR_{MH} with their 95% CIs estimated for the time varying confounding scenario in Figure 3 for M control periods (1 period=1 day). The difference between OR_{SCL} for $\exp(\beta)$ and the true RR (4.0) was more than 10 % of the true value when M>5 and increased when M increased. The OR_{SCL} estimate for $\exp(\beta)$ was larger than the corresponding value in Table 2. When control periods were extended to 10 and 20 cycles, the point estimate of OR_{SCL} for $\exp(\beta)$ was 9.10 and 10.78, respectively. On the other hand, OR_{VF} and OR_G for both $\exp(\beta)$ and $\exp(\gamma)$ were in general unbiased, particularly when M>7 where the estimates of OR_{VF} and OR_G for $\exp(\beta)$ were within 3% of the true value. The estimate of OR_{MH} for $\exp(\beta)$ was stable with increasing M but overestimated as 4.67 (17% above the true value). When M=1, OR_{VF} was 4.27 for $\exp(\beta)$ (about 7% overestimated) and 1.86 for $\exp(\gamma)$ (7% underestimated). Similarly, when M=1, OR_G was 4.26 for $\exp(\beta)$ (7% overestimated) and 1.68 for $\exp(\gamma)$ (16% underestimated). When N was increased to 1,000,000, $\exp(\beta)$ and $\exp(\gamma)$ were 3.96 and 2.09 for OR_{VF} and 3.97 and 2.04 for OR_G when M=1 (not shown in Table 3).

Table 4 shows OR_{SCL} , OR_{VF} , OR_G , and OR_{MH} with their 95% CIs for the scenarios in Figures 2 and 3, where the length of the study period was fixed as 22 days but the length of

each time period varied between 1 and 11 days. For the scenario in Figure 2 with one binary exposure variable only, OR_{SCL} for $\exp(\beta)$ varied between 4.00 and 6.61 when the length of the time period varied, but OR_{VF} , OR_G , and OR_{MH} were stable and unbiased. For the scenario in Figure 3 with one binary exposure variable and one time varying confounder randomly generated in the control periods, the OR_{SCL} estimate for $\exp(\beta)$ varied between 4.34 and 6.79 when M and the length of the time period varied, while OR_{VF} and OR_G for $\exp(\beta)$ were stable and close to the true value, though the odds ratio was a little overestimated as 4.34 when the time period was 11 days (M=1). On the other hand, OR_{MH} for $\exp(\beta)$ was stable but overestimated as 4.67, as in Table 3. When M=1 (1 period=11 days), OR_{VF} was 4.31 (8% overestimated) for $\exp(\beta)$ and 1.94 (3% underestimated) for $\exp(\gamma)$. Similarly, when M=1, OR_G was 4.34 (9% overestimated) for $\exp(\beta)$ and 1.35 (32% underestimated) for $\exp(\gamma)$. When N was increased to 1,000,000, $\exp(\beta)$ and $\exp(\gamma)$ were 3.97 and 1.95 for both OR_{VF} and OR_G when M=1 (not shown in Table 4).

ANALYSES OF CASE-CROSSOVER STUDIES OF REAL-WORLD DATA

Tables 5 and 6 show the estimates using Exposure Definition I in Table 1 in the Japanese study. In general, OR_{SCL} was different from OR_{MH} and OR_G except when M=1, and OR_{SCL} estimated from Equation (2) increased when study period increased: OR_{SCL} was between 1.91 and 2.82 when study period = 84 days (Table 5), between 2.13 and 4.19 when study period = 168 days (Table 6), and between 3.73 and 7.16 when study period=336 days (Table 6). The point estimate of OR_G from Equation (4) was always the same as that of OR_{MH} , as expected (but the confidence interval differed between OR_G and OR_{MH}).

In Appendix Tables 6a and 6b in Additional File 1, the results corresponding to Table 5 and Table 6 for Exposure Definitions II, III and IV (Table 1) are presented. Those results indicated, as in Tables 5 and 6, that OR_{SCL} varied when M varied as well as when the study period varied, while OR_G and OR_{MH} were relatively stable. Detailed description for Definition II, III and IV is given in Appendix 6 in Additional File 1.

In the Australian study, after excluding the concordant cases, there were 1,316 discordant cases with daily exposed and unexposed periods to SSRIs in the 180 days before the index date. Exposure on the index date was excluded since hip fracture may have occurred the day before admission to hospital.

We used periods of 1, 5, 20, 30, 60 and 90 days with M=179, 35, 8, 5, 2 and 1, respectively. Exposure within each period was defined using Definition I (Table 1). Estimates of the OR_{SCL} were biased upwards (Table 7). When M=1 (exposure period = 90 days), estimates for OR_{SCL} was identical to OR_G and OR_{MH} as expected. As in the Japanese

study, OR_{SCL} from Equation (2) increased with M for $M > 1$.

Discussion

METHODS TO REMOVE BIAS FROM WITHIN-SUBJECT EXPOSURE DEPENDENCY WITH OR WITHOUT TIME-VARYING CONFONDERS

Using simulated data, we showed that OR_{SCL} can be biased when there is within-subject exposure dependency but no exposure time trend, except when only one control period is used. When only one control period is used ($M=1$), pairwise exchangeability is equivalent to global exchangeability and bias due to within-subject exposure dependency in standard conditional logistic regression does not occur (although bias due to time trends may occur). In Tables 2 and 3, OR_{SCL} increased with the increase of study period. This observation is similar to that in the previous case-crossover studies, where the odds ratio increased when a longer study period was employed [27-29]. In Table 2, OR_{VF} , OR_G and OR_{MH} were unbiased. Of those 3 unbiased estimates, OR_{VF} may be difficult to calculate when analyzing real-world data because it requires population data (or samples from the population) to estimate the exposure probabilities, unlike OR_G and OR_{MH} . In addition, probabilities for many exposure permutations must be reliably estimated which may be intractable. For example, in Figure 1 with 21 control periods, only 8 subgroups A to H were assumed to exist, but in real-world data, many more exposure patterns would occur. It is possible that the positivity condition is not satisfied for certain permutations and these do not contribute to the estimation of OR_{VF} . These limitations make OR_{VF} difficult to use in practical applications.

As Vines and Farrington showed, when no exposure time trend exists and there is one binary exposure variable which is pairwise exchangeable, OR_{MH} and OR_G are unbiased even when OR_{SCL} is biased. However, when the model involves a time-varying confounder variable in addition to the exposure variable, OR_{MH} may be biased while OR_G is unbiased.

In Tables 3 and 4, when $M=1$ (i.e., only 1 control period is used), OR_G and OR_{VF} for $\exp(\beta)$ were overestimated and those for $\exp(\gamma)$ were underestimated, but the estimates approached to the true values when N was increased, suggesting that the deviation from the true values observed when $M=1$ was due to random error. Conversely, it is likely that employing a larger number of control periods can produce more precise point estimates of $\exp(\beta)$ and $\exp(\gamma)$.

REAL-WORLD DATA ANALYSIS

In the Japanese and Australian studies, we found a discrepancy between OR_{SCL} and OR_{MH} when more than one control period was used. In the Japanese study, OR_{SCL} was

more than 2 times larger than OR_{MH} when the study period increased. We believe that this discrepancy occurred mainly due to within-subject exposure dependency, which increases as the study period increases. The estimates of OR_G were the same as OR_{MH} as expected.

In the Australian study, the discrepancy between OR_{SCL} and OR_{MH} was modest compared with the Japanese study. This was compatible with the finding by Vines and Farrington that the bias due to within-subject exposure dependency is minimal when $\exp(\beta) \approx 1$ [15].

LIMITATIONS OF THE CURRENT STUDY AND FUTURE DIRECTION

In the current study, our focus was on bias from within-subject exposure dependency. However, biases can occur from other sources as well. First, we have not applied the weighting method when a time-trend in the exposure exists [14]. Second, a 'washout period' has been used in some case-crossover studies [16, 17, 19] to allow for the uncertainty in the optimal length of one period and to reduce within subject exposure dependency. In the current study, when the length of the time period varied, OR_G (and OR_{MH}) was stable in both the simulation study and real-world data. Since exposure was defined by the last day of the period, any period with 2 or more days was equivalent to using a 'washout period' at the beginning of the period. However, much more analyses are needed to examine the need for a 'washout period' and to determine the optimal length of one period particularly when within-subject exposure dependency exists.

Another bias may occur when the event rate is not constant. For example, the event rate may be particularly high soon after the exposure is started, compared to later after treatment was started. One solution for this problem is to divide the exposed periods into high-risk and low-risk periods and considering this as different levels of exposure. When within-subject exposure dependency exists, this will require weighting for at least 3 exposure levels and the weighting method for binary exposures described in this study should be expanded for multi-level exposures.

Limitations of our study include that we assumed a time-varying confounder with no within-subject dependency. If within-subject dependency exists for a time-varying confounder, the weighting method may need to allow for both the exposure and time-varying confounders. Finally, when unmeasured time-varying confounders exist, the results still can be biased even when using our weighting method.

Conclusion

Despite autocorrelated exposures being common in pharmacoepidemiology, bias due to within-subject exposure dependency in the case-crossover study has had little attention in the pharmacoepidemiology literature and standard conditional logistic regression has been widely used without assessing whether this bias exists. Although using only one control period can avoid bias due to within-subject exposure dependency, this will reduce the accuracy of the estimate due to random error, as seen in our simulation study (the odds ratio when $M=1$ in Tables 3 and 4). To assess for the possibility of bias, we recommend comparing the Mantel-Haenszel odds ratio with the standard conditional logistic regression odds ratio before starting analysis of case-crossover data. If time-varying confounders exist in data set, they may be ignored when comparing two odds ratios to assess the possibility of bias due to within-subject exposure dependency. If a substantial discrepancy is found (e.g., more than a pre-specified threshold such as 5 to 10 % of the Mantel-Haenszel odds ratio), standard conditional regression should not be used. Either the Mantel-Haenszel method or our weighting method should be used instead. The weighting method has less bias than the Mantel-Haenszel method when a time-varying confounder exists. Furthermore, our weighting method is not subject to violations of the positivity assumption which make the Vines and Farrington odds ratio difficult to use in practice.

Future research will extend our weighting method to allow for time trends in the exposure, 'washout periods', and multi-exposure levels which are potentially important when the event rate changes during exposed periods.

Declarations

Ethics approval

The study using real-world data in Japan was approved by the ethics committee of Tokyo University of Science (reference no. 18023) where obtaining the informed consent from study subjects was waived for the current study. The study using real-world data in Australia was approved by Department of Defense and Veterans' Affairs Human Research Ethics (E016-007) and University of South Australia Human Research Ethics (P203/04) where obtaining the informed consent from study subjects was waived for the current study.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

We declare that the authors have no competing interests as defined by BMC, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Funding

This work was supported by an internal funding and but not by any external funding source.

Authors' contribution

The study was conceived by KK and the method was developed mainly by KK and LK with comments occasionally given from other authors. TS, TY and KK analyzed Japanese data and NP, ER and LK analyzed Australian data. The manuscript was drafted by KK and LK and reviewed by all authors critically. All authors approved the final version of the manuscript.

Acknowledgements

Not applicable.

REFERENCES

1. Maclure M. The case-crossover Design: A method for studying transient effects on the risk of acute events. *Am J Epidemiol* 1991; 133(2): 144-153.
2. Consiglio GP, Burden AM, Maclure M, et al. Case-crossover study design in pharmacoepidemiology: systematic review and recommendations. *Pharmacoepidemiol Drug Saf.* 2013 ;22(11):1146-53.
3. Checkoway H, Pearce N, Kriebel D. Selecting appropriate study designs to address specific research questions in occupational epidemiology. *Occup Environ Med.* 2007 ;64(9):633-8.
4. Kim JH, Mooney SJ. The epidemiologic principles underlying traffic safety study designs. *Int J Epidemiol.* 2016 ;45(5):1668-1675.
5. Carracedo-Martínez E, Taracido M, Tobias A, et al. Case-crossover analysis of air pollution health effects: a systematic review of methodology and application. *Environ Health Perspect.* 2010;118(8):1173-82.
6. Lumley T, Levy D. Bias in the case-crossover design: implications for studies of air pollution. *Environmetrics* 2000;11(6): 689–704.
7. Rothman KJ, Greenland S, Lash T. Case-control studies. In: Rothman KJ, Greenland S, Lash T, eds. *Modern Epidemiology*, 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008:111-128.
8. Mittleman MA, Mostofsky E. Exchangeability in the case-crossover design. *Int J Epidemiol.* 2014;43(5):1645-1655.
9. Maclure M, Fireman B, Nelson JC, et al. When should case-only designs be used for safety monitoring of medical products? *Pharmacoepidemiol Drug Saf.* 2012 ;21 Suppl 1:50-61.
10. Farrington CP, Whitaker HJ, Hocine MN. Case series analysis for censored, perturbed, or curtailed post-event exposures. *Biostatistics* 2009; 10(1):3-16.
11. Farrington CP, Anaya-Izquierdo K, Whitaker HJ, et al. Self-controlled case series analysis with event-dependent observation periods. *J Am Stat Assoc* 2011; 106(494): 417-426.
12. Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. *BMJ* 2016;354:i4515.
13. Hallas J, Pottegård A. Use of self-controlled designs in pharmacoepidemiology. *J Intern Med.* 2014;275(6):581-9.
14. Suissa S. The case-time-control design. *Epidemiology.* 1995;6(3):248-53.
15. Vines SK, Farrington CP. Within-subject exposure dependency in case-crossover studies. *Stat Med.* 2001 30;20(20):3039-49.

16. Jensen AK, Gerds TA, Weeke P, Torp-Pedersen C, Andersen PK. On the validity of the case-time-control design for autocorrelated exposure histories. *Epidemiology*. 2014;25(1):110-113.
17. Tubiana S, Blotière PO, Hoen B, Lesclous P, Millot S, Rudant J, et al. Dental procedures, antibiotic prophylaxis, and endocarditis among people with prosthetic heart valves: nationwide population based cohort and a case crossover study. *BMJ*. 2017;358:j3776.
18. Grimnes G, Isaksen T, Tichelaar YIGV, Brox J, Brækkan SK, Hansen JB. C-reactive protein and risk of venous thromboembolism: results from a population-based case-crossover study. *Haematologica*. 2018;103(7):1245-1250.
19. Sutcliffe S, Jemielita T, Lai HH, Andriole GL, Bradley CS, Clemens JQ, et al. A case-crossover study of urological chronic pelvic pain syndrome flare triggers in the MAPP research network. *J Urol*. 2018;199(5):1245-1251.
20. Ryu SY, Kim J, Kim DW, Chung EJ. Association between ranibizumab injections and risk of acute myocardial infarction in age-related macular degeneration: A case-crossover study. *Korean J Ophthalmol*. 2020;34(2):150-157.
21. Doubleday A, Schulte J, Sheppard L, Kadlec M, Dhammapala R, Fox J, et al. Mortality associated with wildfire smoke exposure in Washington state, 2006-2017: a case-crossover study. *Environ Health*. 2020;19(1):4.
22. Greenland S. A unified approach to the analysis of case-distribution (case-only) studies. *Stat Med*. 1999;18(1):1-15.
23. Wahab IA, Pratt NL, Wiese MD, et al. The validity of sequence symmetry analysis (SSA) for adverse drug reaction signal detection. *Pharmacoepidemiol Drug Saf*. 2013;22(5):496-502.
24. Medi-Scope® claims database system. https://www.jmiri.jp/medi_scope. Accessed December 22, 2020.
25. Leach MJ, Pratt NL, Roughead EE. Psychoactive medicine use and the risk of hip fracture in older people: a case-crossover study. *Pharmacoepidemiol Drug Saf*. 2015;24(6):576-82.
26. Leach MJ, Pratt NL, Roughead EE. Risk of hip fracture in older people using selective serotonin reuptake inhibitors and other psychoactive medicines concurrently: a matched case-control study in Australia. *Drugs Real World Outcomes*. 2017;4(2):87-96.
27. Hallas J, Pottegård A, Wang S, et al. Persistent User Bias in Case-Crossover Studies in Pharmacoepidemiology. *Am J Epidemiol*. 2016 Nov 15;184(10):761-769.
28. Wang P, Schneeweiss S, Glynn RB, et al. Use of the case-crossover design to study

prolonged drug exposures and insidious outcomes. *Ann Epidemiol.* 2004;14(4):296-303.

29. Orriols L, Wilchesky M, Lagarde E, et al. Prescription of antidepressants and the risk of road traffic crash in the elderly: a case-crossover study. *Br J Clin Pharmacol.* 2013;76(5):810-815.

Figure Legends

Figure 1

Exposure pattern in a hypothetical dynamic population of 80,000 patients. Patients receiving drug treatment with a cycle of 7 days are divided into 8 Subgroups A to H where 1 period is defined as 1 day. Subgroup A represents stoppers that stop treatment, Subgroup H represents starters that start treatment, and Subgroups B to G represents those being treated at the case period. The bold frame indicates that 21 control periods can be divided into 3 cycles with the same exposure patterns.

N: the size of each subgroup; c_0 : exposure status at the case period; c_m ($m=1, 2, \dots, 21$): exposure status at the m -th control period; Tx: treatment.

Figure 2

140 cases who had an outcome at the case period in a hypothetical population in Figure 1.

The incidence rate per period at unexposed period (r_0) is assumed to be 0.001 and the rate ratio of the exposure (RR) is assumed to be 4. Exposure status is shown in Figure 1.

n : the size of cases belonging to each ID_Subgroup in Figure 1; c_0 : exposure status at case period.

Figure 3

184 cases who had an outcome at the case period in a hypothetical population in Figure 1.

The incidence rate per period at unexposed period without confounder (r_0) is assumed to be 0.001 and the rate ratio of the exposure (RR) is assumed to be 4 and that of the time-varying confounder (RRz) is assumed to be 2. The proportion of the time-varying confounder at unexposed periods (f_0) and that at exposed periods (f_1) in the population are assumed as $f_0=0.2$ and $f_1=0.4$. Exposure status is shown in Figure 1. The status of the confounder at control periods are assigned randomly by the simulation.

n : the size of cases belonging to each ID_Subgroup in Figure 1; c_0 : exposure status at case period; z_0 : status of time-varying confounder at case period.

Table 1 Exposure Definitions

	Case period	Control periods
Definition I	Last day	Last day
Definition II	Last day	Half or more days
Definition III	Last day	Any 1 day
Definition IV	Any 1 day	Any 1 day

Last day= exposure status of the period was the exposure status on the last day; Half or more days=exposure status of the period is defined as 'exposed' if at least half of days during the period was exposed and 'unexposed' otherwise; Any 1 day=exposure status of the period is defined as 'exposed' if at least 1 day during the period is exposed and 'unexposed' otherwise.

Table 2 The estimates of, OR_{SCL} , OR_{VF} , OR_G , and OR_{MH} (95% confidence interval) from data in Figure 2 with a binary exposure only (study period=M+1 days)

M	OR_{SCL}	OR_{VF}	OR_G	OR_{MH}
1	4.00 (2.45 - 6.53)	4.00 (2.45 - 6.53)	4.00 (2.45 - 6.53)	4.00 (2.45 - 6.53)
2	4.00 (2.74 - 5.85)	4.00 (2.74 - 5.85)	4.00 (2.74 - 5.85)	4.00 (2.74 - 5.85)
3	3.55 (2.47 - 5.10)	4.00 (2.64 - 6.06)	4.00 (2.81 - 5.70)	4.00 (2.73 - 5.86)
4	3.84 (2.72 - 5.41)	4.00 (2.80 - 5.71)	4.00 (2.86 - 5.59)	4.00 (2.81 - 5.69)
5	4.28 (3.04 - 6.03)	4.00 (2.86 - 5.59)	4.00 (2.86 - 5.59)	4.00 (2.85 - 5.61)
6	4.60 (3.27 - 6.48)	4.00 (2.80 - 5.71)	4.00 (2.86 - 5.59)	4.00 (2.89 - 5.53)
7	4.38 (3.10 - 6.20)	-	4.00 (2.86 - 5.59)	4.00 (2.87 - 5.57)
8	4.73 (3.34 - 6.70)	4.00 (2.34 - 6.84)	4.00 (2.86 - 5.59)	4.00 (2.89 - 5.54)
9	5.01 (3.53 - 7.10)	4.00 (2.58 - 6.20)	4.00 (2.86 - 5.59)	4.00 (2.91 - 5.50)
10	4.86 (3.41 - 6.94)	-	4.00 (2.86 - 5.59)	4.00 (2.89 - 5.54)
11	5.09 (3.57 - 7.26)	4.00 (2.00 - 8.00)	4.00 (2.86 - 5.59)	4.00 (2.91 - 5.51)
12	5.33 (3.73 - 7.62)	4.00 (2.49 - 6.42)	4.00 (2.86 - 5.59)	4.00 (2.92 - 5.48)
13	5.54 (3.88 - 7.92)	4.00 (2.64 - 6.06)	4.00 (2.86 - 5.59)	4.00 (2.93 - 5.46)
14	5.39 (3.75 - 7.73)	-	4.00 (2.86 - 5.59)	4.00 (2.92 - 5.49)
15	5.60 (3.90 - 8.05)	4.00 (2.34 - 6.84)	4.00 (2.86 - 5.59)	4.00 (2.92 - 5.47)
16	5.79 (4.02 - 8.33)	4.00 (2.58 - 6.20)	4.00 (2.86 - 5.59)	4.00 (2.93 - 5.46)
17	5.66 (3.92 - 8.18)	-	4.00 (2.86 - 5.59)	4.00 (2.92 - 5.48)
18	5.83 (4.03 - 8.43)	4.00 (2.00 - 8.00)	4.00 (2.86 - 5.59)	4.00 (2.93 - 5.46)
19	6.00 (4.14 - 8.69)	4.00 (2.49 - 6.42)	4.00 (2.86 - 5.59)	4.00 (2.93 - 5.45)
20	6.16 (4.25 - 8.92)	4.00 (2.64 - 6.06)	4.00 (2.86 - 5.59)	4.00 (2.94 - 5.44)
21	6.02 (4.14 - 8.75)	-	4.00 (2.86 - 5.59)	4.00 (2.93 - 5.46)

Data shown in Figure 2 is analyzed by the standard conditional logistic regression (Equation (2)), the Vines and Farrington method (Equation (1)), the Greenland's method (Equation (11)) and the Mantel-Haenszel method. 1 period=1 day for all values of M. M is the number of control periods and study period=(M+1) days.

OR_{SCL} : odds ratio by the standard conditional logistic regression; OR_{VF} : odds ratio by the Vines and Farrington's method; OR_G : odds ratio by the Greenland's method; OR_{MH} : odds ratio by the Mantel-Haenszel method; M: the number of control periods.

Table 3 The estimates of, OR_{SCL} , OR_{VF} , OR_G , and OR_{MH} (95% confidence interval) from data in Figure 3 with a binary exposure and a binary time-varying confounder (study period=M+1 days)

M	variable	OR_{SCL}	OR_{VF}	OR_G	OR_{MH}
1	$\exp(\beta)$	4.26 (2.72 - 6.66)	4.27 (2.73 - 6.68)	4.26 (2.72 - 6.66)	4.67 (3.00 - 7.25)
	$\exp(\gamma)$	1.68 (1.02 - 2.78)	1.86 (1.11 - 3.11)	1.68 (1.02 - 2.78)	
2	$\exp(\beta)$	4.04 (2.85 - 5.75)	4.04 (2.85 - 5.73)	4.04 (2.85 - 5.75)	4.67 (3.33 - 6.54)
	$\exp(\gamma)$	1.84 (1.15 - 2.94)	2.06 (1.28 - 3.30)	1.84 (1.15 - 2.94)	
3	$\exp(\beta)$	3.62 (2.60 - 5.04)	4.11 (2.83 - 5.97)	4.10 (2.97 - 5.66)	4.67 (3.32 - 6.55)
	$\exp(\gamma)$	2.12 (1.43 - 3.12)	1.93 (1.31 - 2.86)	2.21 (1.49 - 3.28)	
4	$\exp(\beta)$	3.91 (2.86 - 5.34)	4.19 (3.03 - 5.79)	4.09 (3.02 - 5.55)	4.67 (3.41 - 6.38)
	$\exp(\gamma)$	2.14 (1.48 - 3.11)	2.43 (1.66 - 3.54)	2.20 (1.51 - 3.20)	
5	$\exp(\beta)$	4.36 (3.18 - 5.97)	4.07 (3.00 - 5.51)	4.02 (2.95 - 5.46)	4.67 (3.46 - 6.29)
	$\exp(\gamma)$	1.86 (1.29 - 2.66)	2.35 (1.62 - 3.41)	1.92 (1.33 - 2.76)	
6	$\exp(\beta)$	4.78 (3.51 - 6.52)	3.98 (2.88 - 5.50)	4.12 (3.04 - 5.59)	4.67 (3.50 - 6.22)
	$\exp(\gamma)$	1.92 (1.36 - 2.71)	2.00 (1.37 - 2.92)	1.87 (1.32 - 2.66)	
7	$\exp(\beta)$	4.48 (3.26 - 6.15)	-	4.04 (2.97 - 5.49)	4.67 (3.48 - 6.26)
	$\exp(\gamma)$	1.89 (1.34 - 2.67)	2.03 (1.33 - 3.08)	1.86 (1.31 - 2.64)	
8	$\exp(\beta)$	4.90 (3.57 - 6.74)	4.07 (2.50 - 6.63)	4.07 (2.99 - 5.52)	4.67 (3.50 - 6.22)
	$\exp(\gamma)$	1.81 (1.29 - 2.53)	1.76 (1.17 - 2.64)	1.79 (1.27 - 2.52)	
9	$\exp(\beta)$	5.24 (3.81 - 7.22)	4.10 (2.76 - 6.08)	4.08 (3.00 - 5.55)	4.67 (3.52 - 6.18)
	$\exp(\gamma)$	1.77 (1.27 - 2.48)	2.10 (1.42 - 3.10)	1.75 (1.24 - 2.46)	
10	$\exp(\beta)$	5.08 (3.67 - 7.03)	-	4.06 (2.99 - 5.52)	4.67 (3.50 - 6.22)
	$\exp(\gamma)$	1.77 (1.28 - 2.45)	1.62 (1.08 - 2.42)	1.76 (1.26 - 2.46)	
11	$\exp(\beta)$	5.33 (3.85 - 7.38)	4.07 (2.17 - 7.64)	4.07 (2.99 - 5.54)	4.67 (3.52 - 6.19)
	$\exp(\gamma)$	1.78 (1.29 - 2.46)	1.86 (1.26 - 2.76)	1.73 (1.24 - 2.42)	
12	$\exp(\beta)$	5.61 (4.06 - 7.75)	4.12 (2.69 - 6.32)	4.07 (3.00 - 5.52)	4.67 (3.53 - 6.17)
	$\exp(\gamma)$	1.92 (1.39 - 2.66)	1.81 (1.24 - 2.64)	1.86 (1.33 - 2.61)	
13	$\exp(\beta)$	5.81 (4.19 - 8.04)	4.06 (2.79 - 5.91)	4.02 (2.96 - 5.46)	4.67 (3.55 - 6.14)
	$\exp(\gamma)$	1.93 (1.39 - 2.68)	1.87 (1.28 - 2.73)	1.88 (1.34 - 2.64)	
14	$\exp(\beta)$	5.60 (4.02 - 7.79)	-	3.99 (2.94 - 5.43)	4.67 (3.53 - 6.17)
	$\exp(\gamma)$	1.93 (1.40 - 2.66)	2.10 (1.40 - 3.13)	1.87 (1.34 - 2.61)	
15	$\exp(\beta)$	5.83 (4.17 - 8.13)	3.94 (2.42 - 6.42)	3.97 (2.91 - 5.40)	4.67 (3.54 - 6.15)

exp(γ) 1.91 (1.38 - 2.63) 2.04 (1.39 - 3.00) 1.86 (1.33 - 2.61)

Table 3 -- continued

16	exp(β)	6.05 (4.35 - 8.43)	3.98 (2.68 - 5.92)	3.98 (2.93 - 5.40)	4.67 (3.55 - 6.14)
	exp(γ)	2.04 (1.47 - 2.82)	2.11 (1.45 - 3.07)	1.98 (1.41 - 2.77)	
17	exp(β)	5.94 (4.25 - 8.30)	-	3.97 (2.93 - 5.39)	4.67 (3.54 - 6.16)
	exp(γ)	2.06 (1.49 - 2.84)	2.19 (1.48 - 3.23)	2.02 (1.44 - 2.82)	
18	exp(β)	6.14 (4.39 - 8.59)	4.08 (2.18 - 7.64)	3.99 (2.94 - 5.42)	4.67 (3.54 - 6.14)
	exp(γ)	2.02 (1.47 - 2.77)	2.16 (1.48 - 3.15)	1.95 (1.41 - 2.72)	
19	exp(β)	6.32 (4.52 - 8.85)	4.00 (2.60 - 6.14)	3.98 (2.93 - 5.40)	4.67 (3.55 - 6.13)
	exp(γ)	2.07 (1.51 - 2.85)	1.98 (1.37 - 2.88)	2.01 (1.44 - 2.80)	
20	exp(β)	6.49 (4.64 - 9.08)	3.94 (2.71 - 5.74)	3.98 (2.93 - 5.40)	4.67 (3.56 - 6.12)
	exp(γ)	2.16 (1.57 - 2.98)	2.14 (1.47 - 3.11)	2.05 (1.47 - 2.87)	
21	exp(β)	6.35 (4.52 - 8.91)	-	3.96 (2.92 - 5.38)	4.67 (3.55 - 6.14)
	exp(γ)	2.20 (1.60 - 3.01)	2.20 (1.51 - 3.22)	2.10 (1.51 - 2.93)	

Data shown in Figure 3 is analyzed by the standard conditional logistic regression (Equation (2)), the Vines and Farrington method (Equation (1)), the Greenland's method (Equation (11)) and the Mantel-Haenszel method. 1 period=1 day for all values of M. M is the number of control periods and study period=(M+1) days.

OR_{SCL} : odds ratio by the standard conditional logistic regression; OR_{VF} : odds ratio by the Vines and Farrington's method; OR_G : odds ratio by the Greenland's method; OR_{MH} : odds ratio by the Mantel-Haenszel method; M: the number of control periods; exp(β): estimate for the exposure variable; exp(γ): estimate for the time-varying confounder.

Table 4 Estimates of Table 6 Estimates of OR_{SCL} , OR_G and OR_{MH} (95% confidence interval): Simulated data (study period=22 days)

Days in 1 period	1 day	2 days	3 days	4 days	7 days	11 days	
M	21	10	6	4	2	1	
One binary exposure variable only							
OR_{SCL}	$\exp(\beta)$	6.02 (4.14 - 8.75)	5.30 (3.73 - 7.53)	4.60 (3.27 - 6.48)	5.23 (3.65 - 7.50)	6.61 (3.27 - 13.33)	4.00 (2.45 - 6.53)
OR_{VF}	$\exp(\beta)$	-	4.00 (2.58 - 6.20)	4.00 (2.80 - 5.71)	4.00 (2.45 - 6.53)	-	4.00 (2.45 - 6.53)
OR_G	$\exp(\beta)$	4.00 (2.86 - 5.59)	4.00 (2.86 - 5.59)	4.00 (2.86 - 5.59)	4.00 (2.81 - 5.70)	4.00 (2.00 - 8.00)	4.00 (2.45 - 6.53)
OR_{MH}	$\exp(\beta)$	4.00 (2.93 - 5.46)	4.00 (2.92 - 5.48)	4.00 (2.89 - 5.53)	4.00 (2.87 - 5.57)	4.00 (2.19 - 7.29)	4.00 (2.45 - 6.53)
One binary exposure variable and one binary time-varying confounder							
OR_{SCL}	$\exp(\beta)$	6.35 (4.52 - 8.91)	5.57 (4.04 - 7.68)	4.76 (3.49 - 6.51)	5.29 (3.81 - 7.35)	6.79 (3.59 - 12.86)	4.34 (2.76 - 6.82)
	$\exp(\gamma)$	2.20 (1.60 - 3.01)	1.78 (1.28 - 2.48)	1.80 (1.27 - 2.54)	2.06 (1.39 - 3.06)	1.91 (1.25 - 2.91)	1.35 (0.83 - 2.20)
OR_{VF}	$\exp(\beta)$	-	4.08 (2.74 - 6.07)	4.02 (2.90 - 5.58)	4.02 (2.57 - 6.29)	-	4.31 (2.75 - 6.74)
	$\exp(\gamma)$	2.20 (1.51 - 3.22)	1.78 (1.20 - 2.63)	1.83 (1.25 - 2.67)	2.02 (1.24 - 3.31)	2.34 (1.43 - 3.85)	1.94 (1.16 - 3.24)
OR_G	$\exp(\beta)$	3.96 (2.92 - 5.38)	4.06 (2.98 - 5.52)	4.11 (3.03 - 5.58)	4.00 (2.89 - 5.53)	4.04 (2.15 - 7.60)	4.34 (2.76 - 6.82)
	$\exp(\gamma)$	2.10 (1.51 - 2.93)	1.77 (1.26 - 2.48)	1.76 (1.23 - 2.50)	2.01 (1.35 - 3.00)	1.90 (1.24 - 2.91)	1.35 (0.83 - 2.20)
OR_{MH}	$\exp(\beta)$	4.67 (3.55 - 6.14)	4.67 (3.54 - 6.16)	4.67 (3.50 - 6.22)	4.67 (3.48 - 6.26)	4.67 (2.72 - 8.01)	4.67 (3.00 - 7.25)

M is the number of control periods while study period is fixed as 22 days (precisely, study period = $\text{Int}(22/(M+1)) * (M+1)$ days).

OR_{SCL} : odds ratio by the standard conditional logistic regression; OR_{VF} : odds ratio by the Vines and Farrington's method; OR_G : odds ratio by the Greenland's method; OR_{MH} : odds ratio by the Mantel-Haenszel method; $\exp(\beta)$: the estimate for the rate ratio for the exposure; $\exp(\gamma)$: the estimate for the rate ratio for the time-varying confounder.

Table 5 Estimates of OR_{SCL} , OR_G and OR_{MH} (95% confidence interval): Japanese data on celecoxib-peripheral edema with study period=84 days and Exposure Definition I

Study period		84 days					
Days in 1 period	1 day	3 days	7 days	14 days	28 days	42 days	
M	83	27	11	5	2	1	
OR_{SCL}	2.82 (2.14-3.71)	2.73 (2.08-3.60)	2.53 (1.93-3.33)	2.28 (1.74-2.99)	2.05 (1.54-2.74)	1.91 (1.38-2.64)	
OR_G	1.99 (1.59-2.49)	1.98 (1.58-2.49)	1.97 (1.56-2.49)	1.94 (1.52-2.48)	1.93 (1.46-2.56)	1.91 (1.38-2.64)	
OR_{MH}	1.99 (1.58-2.51)	1.98 (1.57-2.50)	1.97 (1.55-2.50)	1.94 (1.51-2.48)	1.93 (1.46-2.56)	1.91 (1.38-2.64)	

OR_{SCL} : odds ratio by the standard conditional logistic regression; OR_{VF} : odds ratio by the Vines and Farrington's method; OR_G : odds ratio by the Greenland's method; OR_{MH} : odds ratio by the Mantel-Haenszel method.

Table 6 Estimates of OR_{SCL} , OR_G and OR_{MH} (95% confidence interval): Japanese data on celecoxib-peripheral edema with study period=168 and 336 days and Exposure Definition I

Study period	168 days			336 days			
	Days in 1 period	1 day	14 days	84 days	1 day	14 days	168 days
M		167	11	1	335	23	1
OR_{SCL}		4.19 (3.16-5.55)	3.39 (2.58-4.45)	2.13 (1.54-2.95)	7.16 (5.27-9.74)	5.97 (4.43-8.05)	3.73 (2.42-5.75)
OR_G		2.45 (1.96-3.07)	2.38 (1.88-3.02)	2.13 (1.54-2.95)	2.95 (2.33-3.74)	2.92 (2.27-3.74)	3.73 (2.42-5.75)
OR_{MH}		2.45 (1.94-3.09)	2.38 (1.87-3.03)	2.13 (1.54-2.95)	2.95 (2.28-3.82)	2.92 (2.24-3.79)	3.73 (2.42-5.75)

OR_{SCL} : odds ratio by the standard conditional logistic regression; OR_{VF} : odds ratio by the Vines and Farrington's method; OR_G : odds ratio by the Greenland's method; OR_{MH} : odds ratio by the Mantel-Haenszel method.

Table 7: Estimates of OR_{SCL} , OR_G and OR_{MH} (95% confidence interval): Australian data on SSRI-hip fracture (Study period=180 days, Exposure Definition III)

Study period		180 days				
Days in 1 period	1 day	5 days	20 days	30 days	60 days	90 days
M	179	35	8	5	2	1
OR_{SCL}	2.02 (1.71-2.37)	1.85 (1.58-2.17)	1.76 (1.49-2.06)	1.72 (1.46-2.02)	1.59 (1.33-1.91)	1.51 (1.24-1.83)
OR_G	1.67 (1.47-1.90)	1.60 (1.41-1.82)	1.59 (1.39-1.82)	1.61 (1.40-1.86)	1.55 (1.31-1.84)	1.51 (1.24-1.83)
OR_{MH}	1.67 (1.46-1.92)	1.60 (1.39-1.84)	1.59 (1.37-1.84)	1.61 (1.38-1.88)	1.55 (1.30-1.85)	1.51 (1.24-1.83)

OR_{SCL} : odds ratio by the standard conditional logistic regression; OR_{VF} : odds ratio by the Vines and Farrington's method; OR_G : odds ratio by the Greenland's method; OR_{MH} : odds ratio by the Mantel-Haenszel method.

ID_subgroup	N	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18	c19	c20	c21
A (stop Tx at c0)	10,000	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1
B (being treated)	10,000	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0
C (being treated)	10,000	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0
D (being treated)	10,000	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0
E (being treated)	10,000	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1
F (being treated)	10,000	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0
G (being treated)	10,000	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0
H (start Tx at c0)	10,000	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 1 Exposure pattern in a hypothetical dynamic population of 80,000 patients. Patients receiving drug treatment with a cycle of 7 days are divided into 8 Subgroups A to H where 1 period is defined as 1 day. Subgroup A represents stoppers that stop treatment, Subgroup H represents starters that start treatment, and Subgroups B to G represents those being treated at the case period. The bold frame indicates that 21 control periods can be divided into 3 cycles with the same exposure patters. N: the size of each subgroup; c0: exposure status at the case period; cm (m=1, 2, ---, 21): exposure status at the m-th control period; Tx: treatment.

ID_case	ID_subgroup	c0	n
1-10	A	0	10
11-20	B	0	10
21-30	C	0	10
31-40	D	0	10
41-80	E	1	40
81-90	F	0	10
91-100	G	0	10
101-140	H	1	40

Figure 2 140 cases who had an outcome at the case period in a hypothetical population in Figure 1.

The event rate per period at unexposed period (r_0) is assumed to be 0.001 and the rate ratio of the exposure (RR) is assumed to be 4. Exposure status is shown in Figure 1.

n: the number of cases belonging to each ID_Subgroup in Figure 1; c0: exposure status at case period.

ID_case	ID_subgroup	c0	z0	n
1-8	A	0	0	8
9-12	A	0	1	4
13-20	B	0	0	8
21-24	B	0	1	4
25-32	C	0	0	8
33-36	C	0	1	4
37-44	D	0	0	8
45-48	D	0	1	4
49-72	E	1	0	24
73-104	E	1	1	32
105-112	F	0	0	8
113-116	F	0	1	4
117-124	G	0	0	8
125-128	G	0	1	4
129-152	H	1	0	24
153-184	H	1	1	32

Figure 3 184 cases who had an outcome at the case period in a hypothetical population in Figure 1.

The event rate per period at unexposed period without confounder (r_0) is assumed to be 0.001 and the rate ratio of the exposure (RR) is assumed to be 4 and that of the time-varying confounder (RRz) is assumed to be 2. The proportion of the time-varying confounder at unexposed periods (f_0) and that at exposed periods (f_1) in the population are assumed as $f_0=0.2$ and $f_1=0.4$. Exposure status is shown in Figure 1. The status of the confounder at control periods are assigned randomly by the simulation.

n: the size of cases belonging to each ID_Subgroup in Figure 1; c0: exposure status at case period; z0: status of time-varying confounder at case period.

Figures

ID_subgroup	N	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18	c19	c20	c21
A (stop Tx at c0)	10,000	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1
B (being treated)	10,000	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0
C (being treated)	10,000	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0
D (being treated)	10,000	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0
E (being treated)	10,000	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1
F (being treated)	10,000	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0
G (being treated)	10,000	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0
H (start Tx at c0)	10,000	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 1

Exposure pattern in a hypothetical dynamic population of 80,000 patients. Patients receiving drug treatment with a cycle of 7 days are divided into 8 Subgroups A to H where 1 period is defined as 1 day. Subgroup A represents stoppers that stop treatment, Subgroup H represents starters that start treatment, and Subgroups B to G represents those being treated at the case period. The bold frame indicates that 21 control periods can be divided into 3 cycles with the same exposure patters. N: the size of each subgroup; c0: exposure status at the case period; cm (m=1, 2, ..., 21): exposure status at the m-th control period; Tx: treatment.

ID_subgroup	ID_case	n
A	1-10	10
B	11-20	10
C	21-30	10
D	31-40	10
E	41-80	40
F	81-90	10
G	91-100	10
H	101-140	40

Figure 2

140 cases who had an outcome at the case period in a hypothetical population in Figure 1. The event rate per period at unexposed period (r_0) is assumed to be 0.001 and the rate ratio of the exposure (RR) is assumed to be 4. Exposure status is shown in Figure 1. n: the number of cases belonging to each ID_Subgroup in Figure 1; c_0 : exposure status at case period.

ID_subgroup	ID_case	z0	n
A	1-8	0	8
A	9-12	1	4
B	13-20	0	8
B	21-24	1	4
C	25-32	0	8
C	33-36	1	4
D	37-44	0	8
D	45-48	1	4
E	49-72	0	24
E	73-104	1	32
F	105-112	0	8
F	113-116	1	4
G	117-124	0	8
G	125-128	1	4
H	129-152	0	24
H	153-184	1	32

Figure 3

184 cases who had an outcome at the case period in a hypothetical population in Figure 1. The event rate per period at unexposed period without confounder (r_0) is assumed to be 0.001 and the rate ratio of the exposure (RR) is assumed to be 4 and that of the time-varying confounder (RR_z) is assumed to be 2. The proportion of the time-varying confounder at unexposed periods (f_0) and that at exposed periods (f_1) in the population are assumed as $f_0=0.2$ and $f_1=0.4$. Exposure status is shown in Figure 1. The status of

the confounder at control periods are assigned randomly by the simulation. n: the size of cases belonging to each ID_Subgroup in Figure 1; c0: exposure status at case period; z0: status of time-varying confounder at case period.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFile1CaseCrossoverweighting0405.docx](#)