The Use of active Comparators in self-controlled Designs

Jesper Hallas, Heather Whitaker, Joseph A Delaney, Suzanne M Cadarette, Nicole Pratt, and Malcolm Maclure

Correspondence to Dr. Jesper Hallas, Clinical Pharmacology and Pharmacy, University of Southern Denmark, JB Winsløws Vej 19,2, 5000 Odense C, Denmark. Phone +45 65503010
(Email: jhallas@health.sdu.dk)

Author affiliations:
Clinical Pharmacology and Pharmacy, IST, University of Southern Denmark, Odense, Denmark (Jesper Hallas)
Dept of Clinical Pharmacology and Biochemistry, Odense University Hospital, Odense, Denmark (Jesper Hallas), Department of statistics, modelling and economics, Public Health England, London, UK (Heather Whitaker), College of Pharmacy, University of Manitoba, Winnipeg, MB, Canada (Joseph A Delaney), Leslie Dan Faculty of Pharmacy and Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada (Suzanne M Cadarette), Quality Use of Medicines and Pharmacy Research Centre, Clinical and Health Sciences, University of South Australia, Australia (Nicole Pratt), and Dept Anesthesiology, Pharmacology and Therapeutics, University of British Columbia, Vancouver, BC, Canada (Malcolm Maclure).

Funding
This paper emerged from discussions in the “Self-controlled study design Working Group”, SCWiG, established after funding from International Society of Pharmacoepidemiology’s strategic manuscript initiative.

Conflicts of interest
None of the authors has any conflicts of interest to declare.

According to Danish law, the raw data underlying this paper cannot be made available to third parties.

Running title: Active comparators in self-controlled designs

© The Author(s) 2021. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
Abstract

For self-controlled studies of medication-related effects, time-varying confounding by indication may occur if the indication varies over time. We describe how active comparators might mitigate such bias, using an empirical example.

Approaches to using active comparators are described for case-crossover design, case-time-control design, self-controlled case-series and sequence symmetry analyses. In the empirical example, we used Danish data 1996-2018 to study the association between penicillin and venous thromboembolism (VTE), using roxithromycin, a macrolide antibiotic, as comparator. Upper respiratory infection is a transient risk factor for VTE, thus representing time-dependent confounding by indication.

Odds ratios (OR) for case-crossover analysis were 3.35 (95% confidence interval (CI) 3.23-3.49) for penicillin and 3.56 (CI 3.30 – 3.83) for roxithromycin. We used a Wald-based method or an interaction term to estimate the OR for penicillin with roxithromycin as comparator. These two estimates were 0.94 (CI 0.87 – 1.03) and 1.03 (CI 0.95 – 1.13). Results were similar for the case-time-control analysis, but both the self-controlled case-series and sequence symmetry analysis suggested a weak protective effect of penicillin, seemingly explained by VTE affecting future exposure exclusively for penicillin.

The strong association of antibiotics with VTE suggests presence of confounding by indication. Such confounding can be mitigated by using an active comparator.

Key words: Case-only designs, active comparators, confounding-by-indication

Abbreviations:
CCO : Case-crossover design
SCCS: Self-controlled case series design
CTC: Case-time-control design
SSA: Sequence symmetry analysis
VTE: Venous thromboembolism
Background

One of the most pervasive and challenging methodological issues within pharmacoepidemiology is confounding by indication (1). This occurs when the indication for prescribing the drug is an important risk factor for the studied outcome. A widely used way to limit this bias is to apply an active comparator design. The researcher identifies a comparator drug with a similar indication but an entirely different mode of pharmacological action that would not, theoretically, cause the outcome in question. Thereby, a substantial part of confounding by indication is eliminated by using the comparator as the reference (2).

Active comparators are also commonly used to address research questions that are comparative by nature, for example which drug, A or B, to prefer in a given clinical scenario. Typically, conventional adjustments for other potential confounders are implemented in parallel with the active comparator approach (2).

Self-controlled study designs use the individual’s own experience as reference rather than other individuals (fig 1), and they are thereby immune towards confounders that are stable over time (3). None of the self-controlled designs is inherently robust towards confounders that vary over time. Unfortunately, the indication for prescribing is often time dependent; antibiotics are prescribed when the patient is infected, antidepressants when the patient is depressed, proton pump inhibitors when the patient has dyspepsia etc. If any of these indications are also risk factors for the outcome of interest, they will be time-dependent confounders, and thereby not inherently adjusted by the self-controlled design. In other words, the self-controlled designs can be as vulnerable towards confounding by indication as a conventional cohort design, and the arguments for employing an active comparator can be as strong in a self-controlled study as it is in a conventional cohort study. Yet, to our knowledge, active comparators have never been described or recommended as a practice in self-controlled designs.

The aim of this paper is to demonstrate an approach to using active comparators to reduce the potential for within-person confounding in self-controlled designs. To illustrate the approach, we present an empirical example. We present approaches applicable to the case-crossover (CCO) (4), the self-controlled case series (SCCS) (5), the case-time-control design (CTC) (6); and for the sequence symmetry analysis (SSA) (7).

Methods

We describe two active comparator approaches in the self-controlled designs; a simple ratio approach and an effect modifier approach. In addition, we analyze an empirical case: oral narrow-spectrum penicillin
(phenoxymethyl-penicillin) as a cause of venous thromboembolism (VTE), an assumed null association. Penicillin is primarily used for respiratory infections, which by itself is a risk factor for VTE (8). Since respiratory infections are transient, this is an example of time-dependent confounding by indication which is not inherently adjusted by any self-controlled design. A simple self-controlled analysis would thus be expected to show a spurious, confounded association, whereas a self-controlled analysis using an active comparator would not. As the comparator, we chose roxithromycin, a macrolide antibiotic used for the same indication and recommended as second-line drug, for example in patients having penicillin allergy. Our assumption was that roxithromycin would show a similar same degree of time-dependent confounding by indication. Owing to a very restrictive prescribing policy, narrow-spectrum penicillin is highly dominant on the Danish market and is used over seven times more often than roxithromycin (9). A literature review conducted by us revealed no evidence of any intrinsic venous thrombogenic effect for any of these two antibiotics.

As with other applications of active comparators, the appropriateness of our comparator choice relies on the two drugs being used for similar indications. If, for example, roxithromycin was consistently used for more severe infections (with a stronger association with VTE), our comparator-adjusted estimates could be biased.

**Design**

Two overarching analytic approaches were used.

1) The simple ratio approach. The effects are estimated for each drug (penicillin and roxithromycin) separately, without considering the other. The comparator adjusted estimate then emerges as a simple ratio of the estimate for the drug of interest and the comparator. Confidence intervals for the ratio can be calculated by the Wald-test based method (10).

2) The effect modifier approach. The analysis is first conducted using a composite exposure of the drug of interest and the comparator. Then an interaction term that identifies the drug of interest is introduced, to quantify the change in effect that emerges by going from the comparator to the drug of interest. The estimand is this interaction term, and its confidence interval is provided directly by the regression.

**Population**

We retrieved data from nation-wide Danish data sources, i.e., the Danish National Patient Registry (11), the Danish Prescription Registry (12) and the Central Person Registry (13). To avoid survival bias in the application of bidirectional designs, we required that subjects were alive for at least one year after their first...
VTE. All patients aged 18 or more from Jan 1996 were eligible.

Outcomes
Outcomes were defined by the occurrence of a venous thromboembolism (VTE), i.e., a composite of deep venous thrombosis and pulmonary embolism (International classification of disease version 10: I802, I803 and I26). For all analyses except SCCS, only the individual’s first occurrence of a VTE diagnosis was considered. A first-time VTE has a positive predictive value about 88% in The Danish National Patient Registry (14).

Setup
Data were analyzed using Stata v 15.2 (StataCorp, College Station, Texas, US). We used the analytical setup provided by The Danish Health Data Board, a governmental institution that maintains nation-wide data resources for statistical or scientific purposes (15). According to Danish law, pure registry studies are exempt from review by an Ethics Committee (15).

Analytic specification

The case-crossover design, CCO
For each individual, only the first occurrence of the outcome was analyzed. We used a focal window of four weeks, a washout window of four weeks and four referent windows each lasting four weeks before the washout, to improve statistical precision (16). A given window was considered exposed if a prescription occurred in it. Data were analyzed by conditional logistic regression.

The case-time-control design, CTC
The definition of windows in the case-time-control study was the same as in the CCO. In addition, risk-set sampling by outcome date was used to select four non-cases per case, and their exposure histories were charted similarly as for the cases. Each matched non-case was assigned an index date identical to that of the corresponding case. In addition, we required non-cases to match the index case with respect to sex and birthyear, to have no history of VTE at the time of their index date and to have at least one penicillin or roxithromycin prescription in either the focal or referent windows.

Data were analyzed using conditional logistic regression. The estimand for a traditional CTC without active comparator is an interaction term between exposure and being a case. For the effect modifier approach, the estimand for an active comparator adjusted CTC is thus a second order interaction.
Logit(R) = \beta_0 + \beta_1 * E + \beta_2 * (E \times G) + \beta_3 * (E \times G \times A) + \epsilon

where R is the probability of the outcome event, \beta_0 - \beta_3 regression coefficients, E the combined exposure with either the drug of interest or comparator, G the subject’s group designation as either case or non-case, A that the exposure is the drug of interest and not the comparator and \epsilon an error term. The odds ratio of interest, i.e., the one that measures the risk for the drug of interest, above and beyond the comparator is given by \text{OR} = \exp(\beta_3)

The CTC was devised to adjust for population-level exposure trend bias in CCO (6). The use of a second order interaction allows the two drugs in question to have differing trends, even trends in opposite directions. To evaluate possible failure of parametric assumptions underlying the confidence intervals provided by the regression or by the simple ratio method, we computed confidence intervals by bootstrapping as well.

The self-controlled case series, SCCS

For all subjects, we split their follow-up in eleven two-year segments, starting with a segment from Jan 1997 through December 1998 until a last segment from Jan 2017 through December 2018. We then ascertained whether each segment contained at least one VTE outcome and at least one antibiotic prescription of interest. Those segments that did not were excluded. This segmented analysis was introduced to avoid analyzing the entire 22-year interval as one study period, which would render the analysis vulnerable to both time-dependent confounding, ageing and exposure trends.

If more than one VTE occurred during a two-year segment, they were all included as outcomes unless occurring within four weeks after a prior VTE admission, in keeping with the rationale for SCCS that independent recurrent events can be included. This four-week quarantine was introduced to avoid inclusion of admissions that were possibly readmissions associated with a previous VTE rather than a new VTE event. We mapped use of penicillin or roxithromycin throughout the entire segment, assigning an exposure length of four weeks following each antibiotic prescription. As described in the original SCCS methodology, the effect estimate was estimated using Poisson regression, conditional on the given individual (17).

The sequence symmetry analysis, SSA

In the sequence symmetry analysis, two symmetrical time intervals before and after a first prescription for a
drug are analyzed. Only subjects who have an incident outcome in either of these intervals are included. If there is an association between the drug and the outcome, then it is more likely that the drug will be prescribed before the outcome than vice versa, and the simple ratio between sequences estimates the incidence rate ratio, possibly with a conservative bias (18). We used a six-month window before and after the first antibiotic prescription as our observation window.

Effect modification in the SSA can be analyzed as predictors of one sequence of events vs the opposite order in subjects who have both the exposure and outcome event (7). The rationale is that if two strata have the same sequence ratio, then being in a given stratum has no predictive value for the estimated sequence ratio.

For simplicity and because we were unaware of other time-varying confounders, we did not include any covariates in the analyses other than the two antibiotic exposures of interest. A schematic presentation of the four analytic principles is shown in figure 1.

Sensitivity Analyses
A number of post-hoc exploratory analyses were performed in response to the findings, with the purpose of aiding interpretation. First, to clarify whether our findings could be affected by protopathic bias (i.e., initial diagnosis misinterpreted and prompting treatment), we conducted an analysis with the outcome restricted to deep venous thrombosis without pulmonary embolism. The primary manifestation of mild/moderate pulmonary embolisms is sudden dyspnea, occasionally with cough and low-grade fever. It is conceivable that some are misinterpreted as pneumonias and treated with antibiotics before a diagnosis of pulmonary embolism is established. Deep venous thrombosis is usually confined to the legs and is hardly ever confused with pneumonia. Second, in order to interpret the findings for the SCCS, we conducted a post-hoc analysis, in which the 4-week referent window preceding the exposure focal window was disregarded in the analysis. Such analysis is a standard approach to deal with short-term effects of an outcome on the chance of being exposed and was prompted by the observation of an elevated rate of penicillin prescriptions after VTEs (see below). Third, to interpret the differences in estimates from the two analytic approaches, simple ratio and effect modification, we conducted an analysis restricted to patients who only used one of the two antibiotics during the included observation windows. Fourth, as an aid in interpreting the bidirectional results, we graphed the density of prescribing for penicillin and roxithromycin in one-week intervals, beginning 30 weeks before the first VTE and ending 30 weeks after. Finally, to interpret the post-outcome usage pattern of roxithromycin, we charted the density of roxithromycin prescribing separately for subjects who had their VTE treated with new oral anticoagulants.
Our rationale was that fear of an interaction between roxithromycin and vitamin K antagonists might affect post-outcome roxithromycin prescribing selectively in patients with recent VTE (see below).

Results
We identified 69,751 eligible individuals with VTE. 47,669 (68.3%) of their first episodes were coded with a deep venous thrombosis diagnosis and 22,082 (31.7%) as pulmonary embolism. 34,067 (49%) were men, and their median age was 67 years with an interquartile range of 55-77.

The results of the four main analyses are shown in table 1. As expected, all analyses showed an association between penicillin use and VTE, ranging from 1.56 (SSA) to 3.35 (CCO). For CCO and CTC, the estimates for roxithromycin were similar to those for penicillin, resulting in ratio-based estimates close to the null value, 0.94 and 0.95 for the CCO and CTC, respectively. The estimates based on interaction terms tended to be slightly higher than the ratio-based estimates, (1.03 and 1.06 for the CCO and CTC, respectively). The active comparator adjusted CCO or CTC estimates for penicillin had confidence intervals that comfortably included the null value, whether they were based on ratios or interaction terms. In addition, the width of the comparator adjusted CIs were similar, whether they were based on simple ratio or interaction terms. The bootstrapped CIs for the CTC analysis were similar as well, whether they were carried out for the interaction term or for the simple ratio (data not shown).

For SCCS and SSA, all individual estimates for penicillin and roxithromycin were clearly above the null value, as expected (range 1.56-3.37). However, the associations were moderately stronger for roxithromycin than for penicillin, resulting in comparator adjusted estimates for penicillin below the null value, suggesting a protective effect towards VTE (Table 1).

A number of post-hoc analyses were conducted to aid the interpretation of the findings for SCCS and SSA. To elucidate whether our result could be explained by reverse causation by pulmonary embolism being misinterpreted as pneumonia, we conducted a post-hoc analysis including only cases that were defined by deep vein thrombosis without pulmonary embolism (Web Table 1). Except for wider confidence intervals, the findings of this analysis were similar to the main analysis, including the apparent protective effect of penicillin in the SSA and SCCS analyses.

To aid in interpreting the slightly differing values in the interaction-based estimates and the ratio-based estimates, we conducted a post-hoc analysis including subjects who had only used one antibiotic during the analyzed study periods (Web Table 2). In this analysis, estimates based on ratios or interaction terms were similar, in no instances differing by more than 7%.

Finally, we charted the density of penicillin and roxithromycin prescriptions during a 30-week interval.
before and after the first VTE episode in each patient (figure 2). Both graphs showed a rising density before the VTE outcome and in addition, the graph for penicillin showed a downward slope after the VTE. No such downward slope was seen for roxithromycin. A set of graphs was generated for persons starting new oral anticoagulants instead of warfarin after their VTE, essentially showing the same pattern as in the main analysis (Web Figure 1).

As a response to the density graphs, we conducted an SCCS analysis, excluding the four weeks preceding each exposure focal window (antibiotic prescriptions). These analyses showed a ratio-based estimate of 0.93 (CI: 0.87 - 1.00) and an interaction-based estimate of 1.04 (CI: 0.96 – 1.12).

**Discussion**

For the CCO and CTC, all results were consistent with our preconceptions; we found clearly elevated estimates for both penicillin and roxithromycin viewed separately, demonstrating confounding by indication. Active comparator adjusted estimates for penicillin were close to the null value, whether they were based on ratios or interaction terms, and their width of confidence intervals were similar, as well.

For the SSA and SCCS, however, associations were moderately stronger for roxithromycin than for penicillin, resulting in a comparator adjusted estimate below unity. We interpret this apparent protective effect of penicillin as a spurious finding. SSA and SCCS are, in contrast to CCO and CTC, both exposure-anchored and both bi-directional. The latter implies that they observe exposure patterns after the outcome, using this follow-up as part of the reference. One possible explanation for the findings is reverse causation, caused by confusing pulmonary embolism with pneumonia. Something presenting as an atypical pneumonia might trigger second-line choice of antibiotics. This is, however, largely refuted by the analysis restricted to deep venous thrombosis of the legs, showing the same results as in the main analysis. Another possibility is that roxithromycin is selectively avoided after a VTE diagnosis as patients are often treated with warfarin. Some macrolides exert a substantial and clinically relevant inhibition of warfarin metabolism, thereby increasing the risk of bleeding (19). This potential interaction is much less relevant for roxithromycin (20), but the mere concern of an interaction could potentially lower use of roxithromycin after a VTE diagnosis, causing the estimates for roxithromycin in bi-directional designs to rise. We performed a supplementary analysis, generating density graphs for persons being prescribed direct oral anticoagulants (DOACs) instead of warfarin after their VTE. No interaction is known or suspected for DOACs, but the figures were very similar to those of the overall population (web figure 1).

In the density graph for antibiotic use after VTE, we identified an increased use of penicillin but not
macrolides in the four-week period after diagnosis. After removing this period from the reference period in the SCCS, the active comparator estimate was similar to CCO and CTO. We interpret all of these observations as a clear indication that the bias in the SSA and SCCS arises in the follow-up after the VTE outcome, but we are currently unable to pinpoint the exact mechanism.

In principle, the estimates based on interaction terms are derived from one population whereas the estimates based on ratios are derived from two populations, albeit with some overlap. The interaction approach fits all exposures and potential covariates into one model. Thereby, it is assumed that such one-model-fits-all is valid. However, it is conceivable, although we do not have data supporting it, that the effect of covariates in the population of penicillin users is not the same as in the roxithromycin users. The ratio approach makes no such assumption, and it is even possible to incorporate different covariates or different classification of exposures for the two drugs in question in the ratio approach. We currently have no strong preference for either approach and instead recommend using both approaches and discuss the clinical appropriateness of their different assumptions for the drugs under investigation.

The approaches demonstrated here can be used to incorporate any negative control in a self-controlled study, without this necessarily being an active comparator in a strict sense. However, there is a spectrum of time-dependency for the indication, and for the least time-dependent, an active comparator might not be warranted in a self-controlled study. For example, the first applications of SCCS was within vaccine safety in children (5). Childhood vaccines are given at times that are unrelated to the child's clinical status. Another example is the use of CTC in a study of cardiovascular outcomes among users of a weight-loss product (21). These patients would most likely be overweight during the entire study period, not just when they took the drug. However, self-controlled designs are typically applied in studies of short-term exposures, i.e., where the indication is most likely to be time dependent.

In conclusion, we believe there is a good rationale for considering the use of active comparators in self-controlled designs to mitigate indication bias. It can be implemented in all self-controlled designs either by simple ratios between individual estimates for the drug of interest and the active comparator or by use of an interaction term identifying the drug of interest. In our empiric example, results were very close to our expectations for the unidirectional designs, whereas there is reason to suspect a moderate bias for the bidirectional designs. It illustrates the importance of understanding both the clinical scenario and the particular design features very well.

Density of penicillin and roxithromycin prescriptions in a 30-week interval before and after the first
episode of venous thromboembolism. 69,751 individuals identified in Danish registers during the period Jan 1996 - Dec 2018.

References
9. MedStat.dk. https://medstat.dk/en/viewDataTables/medicineAndMedicalGroups/%7B%22year%22:[%222019%22],%22region%22:[%20%22],%22gender%22:[%22A%22],%22ageGroup%22:[%22A%22],%22searchVariable%22:[%22people_count%22],%22errorMessages%22:[],%22atcCode%22:[%22J01CE02%22, %22J01FA06%22],%22sector%22:[%220%22]]D Accessed 16-3-2021
Table 1. Self-controlled analyses of the association between penicillin or roxithromycin and venous thromboembolism. The simple ratio estimate and effect modifier estimate are two approaches to estimate the effect of penicillin, using roxithromycin as reference. Based on nation-wide data from Denmark 1996-2018. See text for analytic specifications.

<table>
<thead>
<tr>
<th>Method</th>
<th>Penicillin</th>
<th>Roxithromycin</th>
<th>Simple Ratio Estimate</th>
<th>Effect Modifier Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>95% CI</td>
<td>Exposed outcomes</td>
<td>Estimate</td>
</tr>
<tr>
<td>Case-crossover</td>
<td>3.35</td>
<td>3.32, 3.49</td>
<td>5087</td>
<td>3.56</td>
</tr>
<tr>
<td>Case-time-control</td>
<td>2.98</td>
<td>2.83, 3.14</td>
<td>5087</td>
<td>3.14</td>
</tr>
<tr>
<td>Sequence symmetry</td>
<td>1.56</td>
<td>1.45, 1.60</td>
<td>1841</td>
<td>2.23</td>
</tr>
<tr>
<td>analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-controlled case</td>
<td>2.93</td>
<td>2.85, 3.02</td>
<td>6180</td>
<td>3.37</td>
</tr>
<tr>
<td>series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CI: 95% confidence interval
Figure 1. Schematic presentation of the self-controlled designs.

Figure 2