



School of Mathematics and Statistics,
The Open University, Milton Keynes MK7
6AA, UK

paddy.farrington@open.ac.uk

Cite this as: *BMJ* 2022;377:o625

<http://dx.doi.org/10.1136/bmj.o625>

Published: 06 April 2022

FAST FACTS

The self-controlled case series method and covid-19

The self-controlled case series method is one of two approaches used to estimate the association between covid-19 and venous thromboembolism or bleeding. This article briefly describes the method, its assumptions, and how it was implemented in the linked study, and offers some pointers to guide the interpretation of the results.

Paddy Farrington *professor emeritus of statistics*

The self-controlled case series method

The self-controlled case series method is an epidemiological design for estimating the association between an exposure and a health outcome.^{1,2} In the linked study (doi:10.1136/bmj-2021-069590), the exposure is covid-19 and the outcome is deep vein thrombosis, pulmonary embolism, or bleeding.³

Key features

The self-controlled case series method automatically adjusts for all multiplicative confounders that do not vary over the duration of the study—automatically meaning that such confounders need not be adjusted for explicitly measured, or even known. This is because estimation is within individuals: individuals act as their own control (hence the term self-controlled). Time varying confounders (such as time, age, or other exposures), however, must be adjusted for explicitly. Also, cases (people who have experienced the outcome) only need be sampled as they contribute to the estimation (hence the term case series).

For these reasons, the method is well suited to the analysis of uncommon outcomes, using data from pre-existing databases with possibly incomplete information on potential confounders.

The method proceeds by specifying risk periods during which each individual is considered to be—potentially at least—at higher (or lower) risk of the outcome owing to the exposure of interest. In the linked study, we chose the period up to 180 days after the covid-19 date (the earliest recorded date of covid-19), subdivided into shorter segments. A peculiarity of the self-controlled case series, compared with other epidemiological techniques, is that time after the event is used. This is because the method derives from a conditional argument based on the question “Given that the outcome event occurred, how likely is it that it arose during a risk period?”; the answer to which involves all observation times at which the event could have occurred, including those after it actually did occur.

The self-controlled case series method requires two conditions stemming from this feature. The first is that the outcome event should not affect subsequent exposures, and the second is that the event should not censor subsequent observation.

Application to covid-19 data

Are these two conditions met in our study? Strictly speaking, probably not (indeed, rare are the situations in which conditions required by any statistical method are strictly fulfilled). But simple work arounds exist. For the first condition, outcomes might affect subsequent exposures—for example, owing to nosocomial acquisition of SARS-CoV-2 after admission to hospital. But such an effect is time limited and may be circumvented by the inclusion of a dummy pre-exposure risk period. To take care of this, we chose a 30 day interval. Another mechanism resulting in inverse causality is the delay between SARS-CoV-2 infection and its identification, which is dealt with similarly by including the covid-19 date in the pre-risk period; this was the subject of a separate investigation.⁴

For the second condition, some events—notably pulmonary embolism, may result in the patient’s death, at which point observation is censored. But a simple sensitivity analysis (repeating the analysis without the cases who died) can be used to determine whether this contravention actually affects the results in meaningful ways—and as it turns out, it does not.

Thus, in our study, departures from assumptions are not so serious as to invalidate the results, and the standard self-controlled case series model can be used. Had this not been the case, other (more complicated) self-controlled case series models could have been deployed that do not require these conditions to be met.^{2,5,6}

Issues of interpretation

In our study, we used both the self-controlled case series method and a matched cohort method to estimate the incidence rate ratio associated with covid-19. In both cases, the incidence rate ratios represent the relative incidence of the event in a defined post-covid period, compared with the incidence in the absence of infection.

In the self-controlled case series method, all fixed confounders are adjusted for automatically, but time varying confounders must be adjusted for explicitly. We only adjusted for period effects, owing to difficulties in documenting other time varying confounders throughout the study period, such as cancer treatment. In the matched cohort study analysis, however, these concurrent treatments could be included. Nonetheless, fixed confounders also

needed to be adjusted for explicitly, and only limited information on them was available. Thus, the two methods are to some extent complementary with respect to control of confounders. Obtaining similar results from similar analyses using different methods provides reassurance about the validity of the two approaches.

A shortcoming of the self-controlled case series method is that it only yields incidence rate ratios and not absolute measures of risk. An additional benefit of using both methods is that the matched cohort study yields absolute risks as well as incidence rate ratios. These estimates of absolute risk are essential to contextualise the associations. For example, in our study, the incidence of a first deep vein thrombosis, pulmonary embolism, and bleeding event in the population in the absence of covid-19 is low. Large incidence rate ratios such as those we obtained might represent low incidences, divided by very low incidences.

Key features of self-controlled case series

- Uses only cases
- Automatically controls for fixed multiplicative confounders
- Time varying confounders must be adjusted for explicitly
- Easy to check sensitivity to failure of key assumptions
- Provides relative and not absolute measures of risk

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/doi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review: Commissioned; not externally peer reviewed.

The SCCS (self-controlled case series) website is available at www.sccs-studies.info

- 1 Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. *Stat Med* 2006;25:1768-97. doi: 10.1002/sim.2302. pmid: 16220518
- 2 Farrington P, Whitaker H, Ghebremichael Weldeselassie Y. *Self-controlled case series studies: A modelling guide with R*. Chapman & Hall/CRC Press, 2018;doi: 10.1201/9780429491313.
- 3 Katsoularis I, Fonseca-Rodríguez O, Farrington P, et al. Risks of deep vein thrombosis, pulmonary embolism, and bleeding after covid-19: nationwide self-controlled case series and matched cohort study. *BMJ* 2022;377:e069590.
- 4 Fonseca-Rodríguez O, Fors Connolly AM, Katsoularis I, Lindmark K, Farrington P. Avoiding bias in self-controlled case series studies of coronavirus disease 2019. *Stat Med* 2021;40:6197-208. doi: 10.1002/sim.9179. pmid: 34470078
- 5 Farrington CP, Whitaker HJ, Hocine MN. Case series analysis for censored, perturbed, or curtailed post-event exposures. *Biostatistics* 2009;10:3-16. doi: 10.1093/biostatistics/kxn013. pmid: 18499654
- 6 Farrington CP, Anaya-Izquierdo K, Whitaker HJ, et al. Self-controlled case series analysis with event-dependent observation periods. *J Am Stat Assoc* 2011;106:417-26. doi: 10.1198/jasa.2011.ap10108.