

Self-controlled case series methods – an alternative to standard epidemiological study designs.

Irene Petersen (1,2), Reader in Statistics and Epidemiology, Ian Douglas, Senior lecturer in Pharmacoepidemiology (3), Heather Whitaker (4), Senior Lecturer in Statistics

- 1) University College London, Department of Primary Care and Population Health, Rowland Hill St, London NW3 2PF, UK
- 2) Aarhus University, Department of Clinical Epidemiology, Olof Palmes Alle 43 – 45, Aarhus N, Denmark
- 3) London School of Hygiene and Tropical Medicine, Keppel St, London, WC1E 7HT, UK
- 4) Open University, Department of Mathematics and Statistics, Walton Hall, Milton Keynes, MK7 6AA, UK

Word count: 1983

Stand first

The self-controlled case series (SCCS) method is an epidemiological study design for which individuals act as their own control i.e. comparisons are made within-individuals. Hence, only individuals who have experienced an event are included and all time invariant confounding is eliminated. The temporal association between a transient exposure and an event is estimated. SCCS was originally developed for evaluation of vaccine safety, but has since been applied in a range of settings where we may not have exact information on the size of the population at risk or it can be difficult to identify an appropriate comparison group, for example for studies of adverse effects of drug treatments. We provide an overview of the SCCS method, examples of its use, we discuss limitations, assumptions and potential biases that can arise where assumptions are not met along with solutions and examples of good practice.

Introduction

In 1951, Richard Doll and Austin Bradford Hill, wrote to all registered physicians in the United Kingdom to obtain information about their smoking habits. This research was groundbreaking as they were one of the first teams to demonstrate a link between smoking and various diseases (1). Since then, epidemiological study designs, such as cohort and case-control studies, have been widely applied in medical research (w1). There are several situations, however, where standard epidemiological study designs fall short. For example, in the research of adverse effects of vaccines, it can be difficult to identify suitable comparison groups (e.g. if most of the population receives the vaccine). Likewise, for studies on hospital data we may not have information on the exact catchment areas and hence we struggle to find suitable controls for cases of a particular event. In these situations the self-controlled case series (SCCS) method provides an alternative epidemiological study design to investigate the association between a transient exposure and an outcome event. The SCCS method is a case-only method; it has the advantages that no separate controls are required and any fixed confounder is automatically controlled for.(2,3)

In this paper we provide an overview of SCCS methodology and some examples of how the method has been applied in order to give a flavour of the potentials of SCCS. As for any epidemiological study design the SCCS method has some assumptions. We discuss these and identify the key limitations and potential biases.

Self-controlled case series (SCCS) methodology

The SCCS method aims to estimate a relative incidence, which compares the incidence of adverse events within periods of hypothesized excess risk due to exposure with incidence during all other time. Asking 'when?' rather than 'who?' becomes the key question. Precise timings are needed, so the SCCS method is best suited to acute events and transient exposures for which periods of exposure risk can be clearly defined. Only those who have experienced an event (cases), can contribute any information on *when* the event occurred, so only data on these individuals need to be collected.

After having identified cases, the first step in setting up the data for SCCS is to define observation periods, these are the study periods for each individual, over which a full history on the timing of events and exposures are available. The next step is to define the periods when exposure may have had an impact within the observation period, for example, a fixed number of days after receipt of a first drug

prescription or a vaccine; these are known as exposure risk periods. All remaining time within the observation period constitutes baseline exposure periods, to which the exposure risk periods will be compared. SCCS studies may include multiple exposure risk periods as it may be necessary to capture either multiple doses or varying exposure-related risk using several risk periods (which may be of same or different lengths). Observation periods can be further divided according to age groups, seasons and/or any other relevant time-varying factors so they can be accounted for in the analysis. The third step is to 'map' events in relation to the different periods identified. Once the data is set up, the final step is to estimate the relative incidence (RI) of events in pre-specified exposure risk periods compared to baseline periods, whilst taking into account the effect of any time-varying confounders such as age groups (w2).

In Figure 1 we illustrate the SCCS setup for an individual with two exposure risk periods and five age groups (panel A) and use this individual to illustrate the output of an SCCS model (panels B, C and D).

Figure 1 here

Comparisons are not made between-individuals as in a cohort or case-control study; estimation is within-individuals. Any factor or characteristic that remains constant over observation periods cancels out of the SCCS model, this includes individual-specific underlying risk and factors such as sex, ethnicity and deprivation. It is in this sense that the SCCS model is *self-controlled*.

Conceptually, the SCCS method builds on the principles of a cohort study (w1). Individuals are followed through time, the exposure history is fixed and events are random. However, the SCCS method differs in that the total *number* of events occurring within an individual's observation period is fixed and follow-up is *not* censored at an event. Hence all exposures occurring within the observation period - both *before* and *after* individuals have experienced the event - are included in the analysis.

A worked example which illustrates how to fit the SCCS model is included in the web materials (w2) and further details of the theory and model fitting are given in Whitaker *et al.*(3). Below we present two examples of studies that applied the SCCS method. Figure 2 provides a schematic overview of the SCCS study design for each of the studies and the key results.

Example applications of the SCCS method in medical research

Example 1: A possible drug interaction between clopidogrel and Proton Pump Inhibitors (PPI)

Douglas *et al.* (4) used both a cohort and SCCS approach to study a drug interaction between clopidogrel and proton pump inhibitors (PPI) on the risk of myocardial infarction (MI). The cohort analysis found an increased risk of MI when clopidogrel was combined with a PPI, supporting the notion that an interaction between the two drugs would reduce the effectiveness of clopidogrel (RI 1.30 (1.12-1.50)). However, the effect was non-specific as an increased risk of non-vascular death was also seen, suggesting underlying confounding was difficult to account for. The SCCS analysis examined whether the risk of MI was increased in the period when PPI was added to clopidogrel treatment (Figure 2). Thus, the observation period was the duration of clopidogrel treatment, and the exposure risk periods were periods of PPI treatment, which varied in length and number amongst the study participants. SCCS found no increased risk for MI (RI 0.75 (0.55-1.01)), suggesting that the underlying confounding had been dealt with through this design.

Example 2: Inflammations and Myocardial Infarction or Stroke

Smeeth *et al.* (5) aimed to evaluate the association between inflammation and myocardial infarction or stroke. The exposures included inflammatory stimuli: vaccinations and acute infection. SCCS was used because of concerns that people receiving vaccinations may differ from those unvaccinated and likewise people acquiring infections may differ from uninfected people in ways that are difficult to measure and account for. Exposure risk periods were defined up to 91 days following the recording of either an infection or a vaccination, and were further subdivided to allow the relative incidence of events to vary over this period (Figure 2) An increased risk of both MI and stroke was seen over the risk period following either a systemic respiratory tract infection or a urinary tract infection, with the strongest effect in the first 7 days. The risk was still elevated up to 91 days, but appeared to be returning towards baseline. No increased risk was seen following influenza, tetanus or pneumococcal vaccinations.(5) Figure 2 provides the incidence rate ratios for MI following respiratory tract infections.

SCCS assumptions

As for other epidemiological study designs the SCCS model makes certain assumptions that should be met to provide valid and unbiased estimates. Below we outline these assumptions. Table 1 provides a summary of assumption violations that can arise, the solutions and examples of good practice against each.

A. Occurrence of an event should *not* (appreciably) affect subsequent exposures

A key assumption is that subsequent exposures should not appreciably be affected by previous events. However, this may often be the case. For example, occurrence of an event may delay exposure, the event may be a contra-indication for treatment or the event may result in, or is death. This assumption also means that the event itself should not determine the timing of the end of the observation period.

Ignoring this assumption may potentially produce biased estimates, but there are various extensions/moderations to the SCCS method that can mitigate potential biases (Table 1). Note that independent causes of death/end of observation do not cause bias.

Temporary delay or increase in exposure after an event

If the event only temporarily delays exposure this will result in a deficit of events in the period just before exposure that reduces the overall incidence in the baseline period. This results in relative incidence estimates that are biased upwards. One way to correct for this bias is to include a 'pre-exposure period' (Table 1.1) just before an exposure.(3) A pre-exposure period can similarly be applied if there is a short-term increase in the probability of exposure after an event (Table 1.2), which would otherwise bias estimates toward the null. A pre-exposure period was included in the clopidogrel and PPI interaction example anticipating that the chance of being started on a PPI could be temporarily altered by having an MI, this is illustrated in figure 2.(4) The pre-exposure period 'trick' only works for short-term delay (in relation to the length of observation). If there is long-term delay in exposure after events the methods in the next paragraph may be applied.

No exposure can occur after the event e.g. if the outcome is death

If the outcome is death exposures that might have otherwise occurred after the event will never be known. This is also true if the exposure history is only collected up until the time of the event or if the event is a contra-indication to exposure (Table 1.3). If a fixed length exposure can only occur once, a

simple solution is to define the observation to begin with the start of exposure and finish with the end of study that would have applied had death/censoring not occurred.(3,6) This a priori definition of the observation period means that only exposed cases are included, and if the exposure risk period is of a fixed length then the full exposure history is always known even beyond the time of death/censoring. Alternatively, there is an extension of the SCCS method which produces unbiased estimates given these scenarios.(7)

If the outcome is death, but the exposure is external to the case and fully observable after death e.g. a weather phenomenon, the standard SCCS method can be applied using the full planned observation period, had the case not died.

Event increases the probability of death

If the event carries high mortality, such as myocardial infarction or stroke, then there is a chance that observation periods could be cut short as a direct result of the event. Resulting bias can be in either direction (Table 1.4). If the event mortality is low, bias will be negligible and the situation ignorable. Comparison of results from fitting SCCS models to all cases and excluding those who died can be made; major differences would suggest bias. Bias may be corrected by fitting an extension that involves modelling post-event survival times.(8)

B. Event rates are constant within intervals

Event rates are assumed to be constant within each defined period. While such an assumption is often unrealistic, it is convenient, makes relative incidence estimates easy to interpret and is commonly used in other study designs. Control of strong age or season effects is important and SCCS models with greater flexibility have been developed.(9,10)

C. Events must be independently recurrent or rare

The SCCS method was developed for *independent* recurrent events, but it has been demonstrated that the method is valid for non-recurrent events when the risk of occurrence over the study period in the entire cohort is 10% or less.(8) A test for independence has been developed. (11) If events are dependent a simple solution is to study just first events (Table 1.5). An extension for which events may depend on the prior event history exists.(12)

Final remarks

With this paper we have sought to demonstrate how the self-controlled case series method provides an alternative to standard epidemiological designs when investigating associations between a transient exposure and an outcome event. A major strength of the SCCS is that it is self-controlled and accounts for any factor or characteristics that remains constant over observation period. Thus, where uncertainty over the control of fixed confounders exists in a cohort or case-control study, SCCS may provide a superior design, given careful thought is made to applying the methodology correctly.

Summary

- The self-controlled case series method provides an alternative to established epidemiological designs.
- SCCS is best suited to acute recurrent or non-recurrent events and transient exposures for which precise timings are available.
- Estimation is within-individuals and no separate controls are required, hence the method is self-controlled and time invariant factors are cancelled out.
- Follow up is not censored at the event, so when events can impact on subsequent exposure, care must be taken to ensure analyses are carried out that eliminate or minimize bias.

Linked information

The Open University host a website for the Self-controlled Case series methodology with example datasets and code for analysis in Stata, SAS, R, Genstat and GLIM.

A link to the tutorial by Whitaker *et al.*(3) can also be found on this website.

<http://statistics.open.ac.uk/sccs/index.htm>

Statements

This paper was written jointly by Irene Petersen (IP), Ian Douglas (ID) and Heather Whitaker (HW). HW is one of the leading experts in SCCS and has been involved in the development of the SCCS methodology from 2004 and onwards. ID and IP have applied the SCCS in a range of settings and keep encouraging colleagues and students to consider the application of SCCS in (pharmaco)-epidemiological

studies. HW was supported by an MRC methodology grant (MR/L009005/1). None of the authors have any competing interests. Guarantor HW.

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide license

(<http://www.bmj.com/sites/default/files/BMJ%20Author%20Licence%20March%202013.doc>) to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution and convert or allow conversion into any format including without limitation audio, iii) create any other derivative work(s) based in whole or part on the on the Contribution, iv) to exploit all subsidiary rights to exploit all subsidiary rights that currently exist or as may exist in the future in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above. All research articles will be made available on an Open Access basis (with authors being asked to pay an open access fee—see <http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/copyright-open-access-and-permission-reuse>). The terms of such Open Access shall be governed by a [Creative Commons](#) license—details as to which Creative Commons license will apply to the research article are set out in our worldwide license referred to above.

Table 1: Violation of assumptions, impact and solutions with examples of good practice.

Problem	Solution	Example
1) Event temporarily decreases the probability of exposure	Include a pre-exposure period	Stowe <i>et al</i> studied the risk of infections after mumps, measles and rubella (MMR) vaccination. Vaccination is delayed when a child has an infection; a 14-day pre-exposure period was included to allow for this. (13)
2) Event temporarily increases the probability of exposure	Include a pre-exposure period	Gibson <i>et al</i> studied the association between prescription medications and motor vehicle crashes. A 4-week pre-exposure period was included since some medications may be used to treat anxiety or pain caused by the crash. (14)
3) No exposure can occur (or is observed) after the event	For single exposures that cannot be repeated, begin the observation period at exposure, so only exposed cases are included. End observation at the planned end. For single or multiple exposures, use the method outlined in Farrington <i>et al.</i> (7)	Hubbard <i>et al</i> studied the association between first bupropion prescription and sudden death. The observation period began with date of first prescription and ended (beyond death) with the date of last data collection for the cohort. At the time, bupropion could only be prescribed as a single course of treatment.(6) Dodd <i>et al</i> studied the association between influenza vaccination and Guillain Barré Syndrome (GBS). Some practitioners will not vaccinate previous GBS patients.(15) Various analyses were performed: standard SCCS, vaccinated cases only, and the method outlined in Farrington <i>et al.</i> (7)
4) Event increases the probability of death	Undertake suitable tests or sensitivity analyses excluding cases who died as a result of the event. If necessary, use the methods outlined in Farrington <i>et al.</i> (8) to adjust for bias.	Langan <i>et al</i> studied the risk of stroke following herpes zoster infection. A sensitivity analysis was performed excluding cases who died within 90 days of stroke, findings were not modified.(16) Brauer <i>et al.</i> studied the association between antipsychotic drugs and myocardial infarction using both the methods outlined in Farrington <i>et al.</i> (8) and a standard SCCS analysis. There was little difference in results.(17)
5) Recurrences of an event are not independent	Study the first event only	Recurrence of stroke is not independent of the first occurrence, so only a first stroke event can be studied. In a study on the risk of stroke, Langan <i>et al.</i> began the observation period 12 months into follow up time to ensure first stroke events had been correctly identified.(16)

References

1. Doll R, Hill AB. The Mortality of Doctors in Relation to Their Smoking Habits. *Br Med J*. 1954 Jun 26;1(4877):1451–5.
2. Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics*. 1995 Mar;51(1):228–35.
3. Whitaker HJ, Farrington CP, Spiessens B, Musonda P. Tutorial in biostatistics: the self-controlled case series method. *Stat Med*. 2006 May 30;25(10):1768–97.
4. Douglas IJ, Evans SJW, Hingorani AD, Grosso AM, Timmis A, Hemingway H, et al. Clopidogrel and interaction with proton pump inhibitors: comparison between cohort and within person study designs. *BMJ*. 2012 Jul 10;345(jul10 1):e4388–e4388.
5. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. *N Engl J Med*. 2004 Dec 16;351(25):2611–8.
6. Hubbard R, Lewis S, West J, Smith C, Godfrey C, Smeeth L, et al. Bupropion and the risk of sudden death: a self-controlled case-series analysis using The Health Improvement Network. *Thorax*. 2005 Oct;60(10):848–50.
7. Farrington CP, Whitaker HJ, Hocine MN. Case series analysis for censored, perturbed, or curtailed post-event exposures. *Biostat Oxf Engl*. 2009 Jan;10(1):3–16.
8. Farrington CP, Anaya-Izquierdo K, Whitaker HJ, Hocine MN, Douglas I, Smeeth L. Self-Controlled Case Series Analysis With Event-Dependent Observation Periods. *J Am Stat Assoc*. 2011 Jun;106(494):417–26.
9. Ghebremichael-Weldeselassie Y, Whitaker HJ, Paddy Farrington C. Self-controlled case series method with smooth age effect. *Stat Med*. 2014 Feb 20;33(4):639–49.
10. Farrington CP, Whitaker HJ. Semiparametric analysis of case series data. *J R Stat Soc Ser C Appl Stat*. 2006 Nov 1;55(5):553–94.
11. Farrington CP, Hocine MN. Within-individual dependence in self-controlled case series models for recurrent events. *J R Stat Soc Ser C Appl Stat*. 2010 May;59(3):457–75.
12. Simpson SE. A positive event dependence model for self-controlled case series with applications in postmarketing surveillance. *Biometrics*. 2013 Mar;69(1):128–36.
13. Stowe J, Andrews N, Taylor B, Miller E. No evidence of an increase of bacterial and viral infections following Measles, Mumps and Rubella vaccine. *Vaccine*. 2009 Feb;27(9):1422–5.
14. Gibson JE, Hubbard RB, Smith CJP, Tata LJ, Britton JR, Fogarty AW. Use of Self-controlled Analytical Techniques to Assess the Association Between Use of Prescription Medications and the Risk of Motor Vehicle Crashes. *Am J Epidemiol*. 2009 Jan 6;169(6):761–8.

15. Dodd CN, Romio SA, Black S, Vellozzi C, Andrews N, Sturkenboom M, et al. International collaboration to assess the risk of Guillain Barré Syndrome following Influenza A (H1N1) 2009 monovalent vaccines. *Vaccine*. 2013 Sep;31(40):4448–58.
16. Langan SM, Minassian C, Smeeth L, Thomas SL. Risk of Stroke Following Herpes Zoster: A Self-Controlled Case-Series Study. *Clin Infect Dis*. 2014 Jun 1;58(11):1497–503.
17. Brauer R, Smeeth L, Anaya-Izquierdo K, Timmis A, Denaxas SC, Farrington CP, et al. Antipsychotic drugs and risks of myocardial infarction: a self-controlled case series study. *Eur Heart J*. 2015 Apr 2;36(16):984–92.

Figure 1 Graphic illustration of the SCCS model output using a fictive individual with two exposure risk periods and five age groups.

Panel A: An individual observation period segmented by two exposure risk periods (red boxes, labelled exposure risk status 1) and five age groups (blue boxes, labelled 0 to 4). The baseline categories for age and exposure are labelled age group 0 and exposure risk status 0 respectively.

Panel B: Exposure-related relative incidences. Note for the baseline category the relative incidence is 1 and the exposure relative incidence is arbitrarily set to 1.8.

Panel C: Age-related relative incidence on age groups 1-4 compared with age group 0. The age relative incidences for age groups 1-4 have been set to 2, 1.5, 1.2 and 0.5 respectively.

Panel D: Overall profile of relative incidence on each of the nine intervals, this is the age-related relative incidence multiplied by the exposure-related relative incidence.

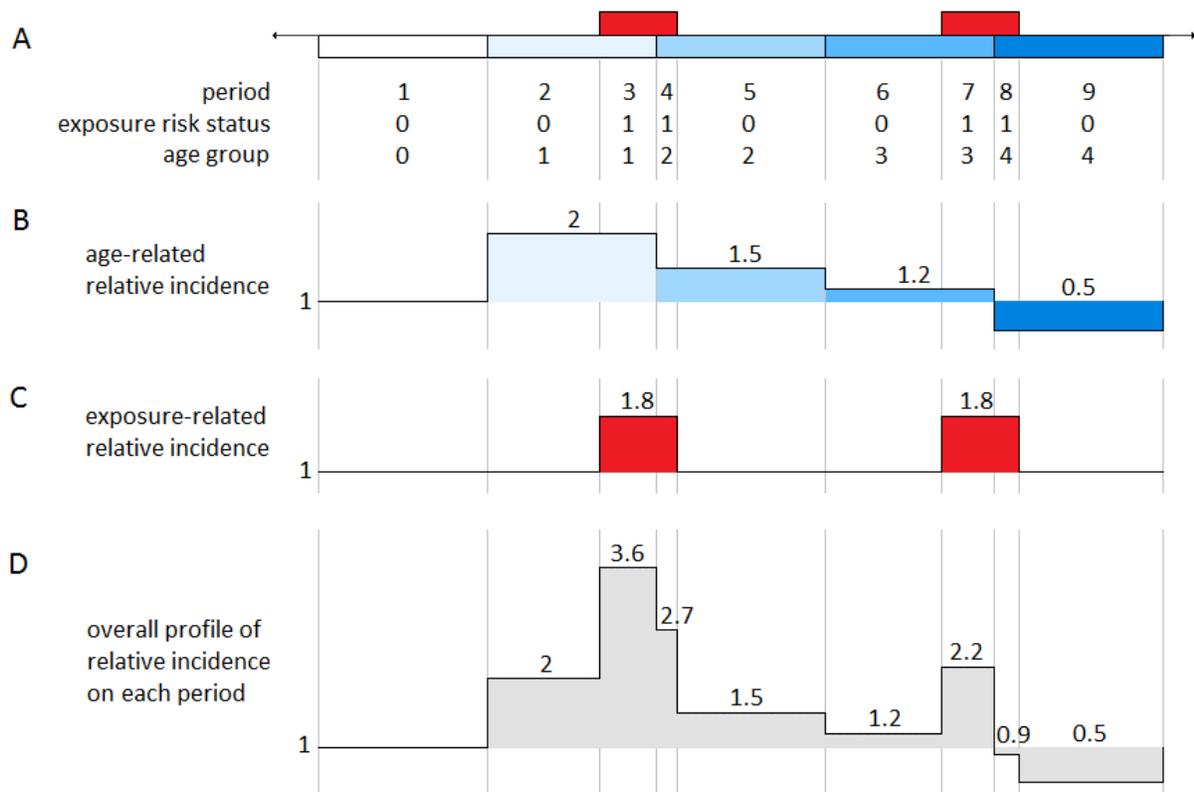
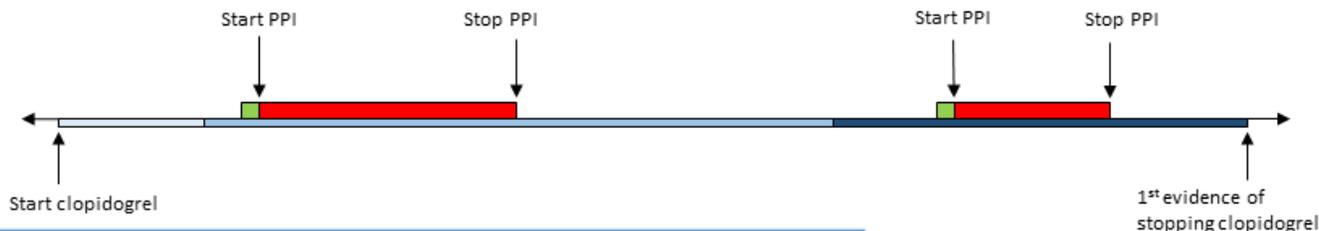


Figure 2: Typical Observation periods, risk periods and headline results for two examples of applied SCCS

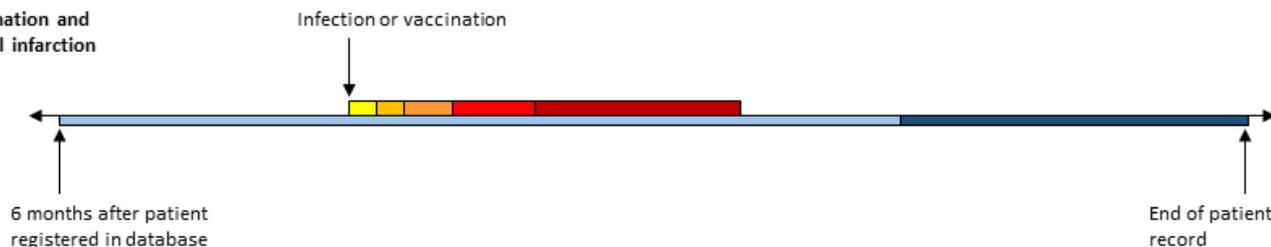
A. Clopidogrel and PPI interaction



- 2 week pre-exposure period
- PPI exposure risk period (variable length)
- Age 67 years
- Age 68 years
- Age 69 years

Headline SCCS results from Clopidogrel and PPI Interaction study N=444		
Period	n myocardial infarction	Incidence rate ratio (95% confidence interval)
Baseline	375	1.00
Pre-exposure (14 d)	10	1.39 (0.71-2.70)
PPI exposure period	395	0.75 (0.55-1.01)

B. Inflammation and myocardial infarction



- 1-3 days post infection risk period
- 4-7 days post infection risk period
- 8-14 days post infection risk period
- 15-28 days post infection risk period
- 29-91 days post infection risk period
- Age 50-54 years
- Age 55-59 years

Headline SCCS results from Inflammation and Myocardial Infarction Study Systemic Respiratory Tract Infection N=20,921		
Period	n events	Incidence rate ratio (95% confidence interval)
Baseline	17,099	1.00
1-3 d post infection	322	4.95 (4.45-5.53)
4-7 d post infection	276	3.20 (2.84-3.60)
8-14 d post infection	422	2.81 (2.54-3.09)
15-28 d post infection	576	1.95 (1.79-2.12)
29-91 d post infection	1,658	1.40 (1.33-1.48)