



# Self controlled case series methods: an alternative to standard epidemiological study designs

Irene Petersen,<sup>1,2</sup> Ian Douglas,<sup>3</sup> Heather Whitaker<sup>4</sup>

<sup>1</sup>Department of Primary Care and Population Health, University College London, London NW3 2PF, UK

<sup>2</sup>Department of Clinical Epidemiology, Aarhus University, Aarhus N, Denmark

<sup>3</sup>London School of Hygiene and Tropical Medicine, London, UK

<sup>4</sup>Department of Mathematics and Statistics, Open University, Milton Keynes, UK

Correspondence to:

I Petersen [i.petersen@ucl.ac.uk](mailto:i.petersen@ucl.ac.uk)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2016;354:i4515 <http://dx.doi.org/10.1136/bmj.i4515>

Accepted: 24 July 2016

The self controlled case series (SCCS) method is an epidemiological study design for which individuals act as their own control—ie, 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 exact information on the size of the population at risk is lacking or identification of an appropriate comparison group is difficult—eg, for studies of adverse effects of drug treatments. We provide an overview of the SCCS method, with examples of its use, discuss limitations, assumptions, and potential biases that can arise where assumptions are not met, and provide solutions and examples of good practice.

In 1951, Richard Doll and Austin Bradford Hill wrote to all registered doctors 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.<sup>1</sup> Since then, epidemiological study designs,

such as cohort and case-control studies, have been widely applied in medical research (see supplementary web table 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 (eg, if most of the population receives the vaccine). Likewise, studies on hospital data may not have information on the exact catchment areas and hence it is a 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.<sup>2,3</sup>

In this paper we provide an overview of SCCS methodology and some examples of how the method has been applied, in order to give an idea 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.

## SCCS method

The SCCS method aims to estimate a relative incidence, which compares the incidence of adverse events within periods of hypothesised excess risk due to exposure with incidence during all other times. 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.

Having identified cases, the next 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 third step is to define the periods when exposure may have had an impact within the observation period, eg, 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. Studies using SCCS might 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 the same or different lengths). Observation periods can be further divided according to age groups or seasons, or any other relevant time

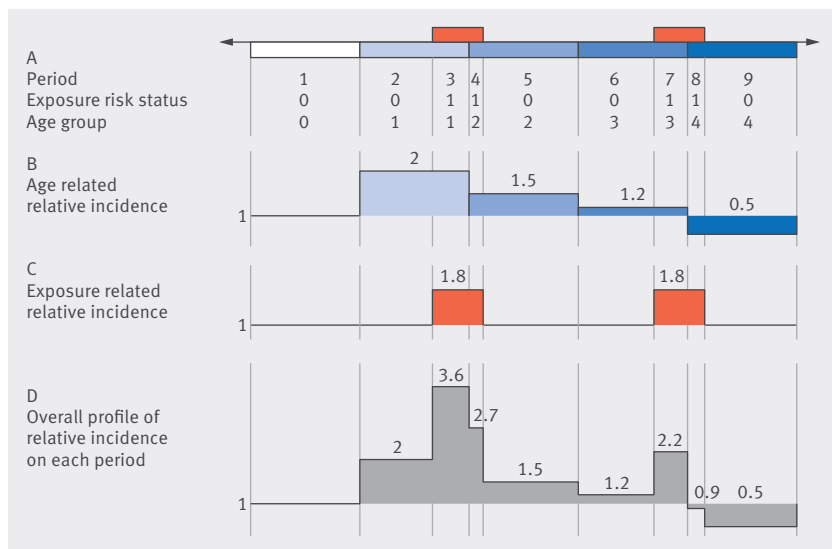
## SUMMARY POINTS

The self controlled case series (SCCS) 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 minimise bias



**Fig 1 | Illustration of SCCS model output using a hypothetical 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). 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

varying factors so they can be accounted for in the analysis. The fourth step is to “map” events in relation to the different periods identified. Once the data are set up, the final step is to estimate the relative incidence of events in prespecified exposure risk periods compared with baseline periods, while taking into account the effect of any time varying confounders such as age groups (see supplementary file).

In figure 1 we illustrate the SCCS set up for someone 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 to D).

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 (see supplementary file). 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 that illustrates how to fit the SCCS model is included in the supplementary materials, and further details of the theory and model fitting are given in Whitaker et al.<sup>3</sup> 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**

Douglas et al used both a cohort and an SCCS approach to study a drug interaction between clopidogrel and proton pump inhibitors on the risk of myocardial infarction.<sup>4</sup> The cohort analysis found an increased risk of myocardial infarction when clopidogrel was combined with a proton pump inhibitor, supporting the notion that an interaction between the two drugs would reduce the effectiveness of clopidogrel (relative incidence 1.30, 95% confidence interval 1.12 to 1.50). However, the effect was non-specific, as an increased risk of death from non-vascular causes was also seen, suggesting it was difficult to account for underlying confounding. The SCCS analysis examined whether the risk of myocardial infarction was increased in the period when the proton pump inhibitor was added to clopidogrel treatment (fig 2). Thus the observation period was the duration of clopidogrel treatment, and the exposure risk periods were periods of proton pump inhibitor treatment, which varied in length and number among the study participants. SCCS found no increased risk for myocardial infarction (0.75, 0.55 to 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 aimed to evaluate the association between inflammation and myocardial infarction or stroke.<sup>5</sup> The exposures included inflammatory stimuli: vaccinations and acute infection. SCCS was used because of concerns that people receiving vaccinations may differ from those who were 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 after 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 (fig 2). An increased risk of both myocardial infarction and stroke was seen over the risk period after either a systemic respiratory tract infection or a urinary tract infection, with the strongest effect in the first seven days. The risk was still increased up to 91 days, but appeared to be returning towards baseline. No increased risk was seen after influenza, tetanus, or pneumococcal vaccinations.<sup>5</sup> Figure 2 provides the incidence rate ratios for myocardial infarction after respiratory tract infections.

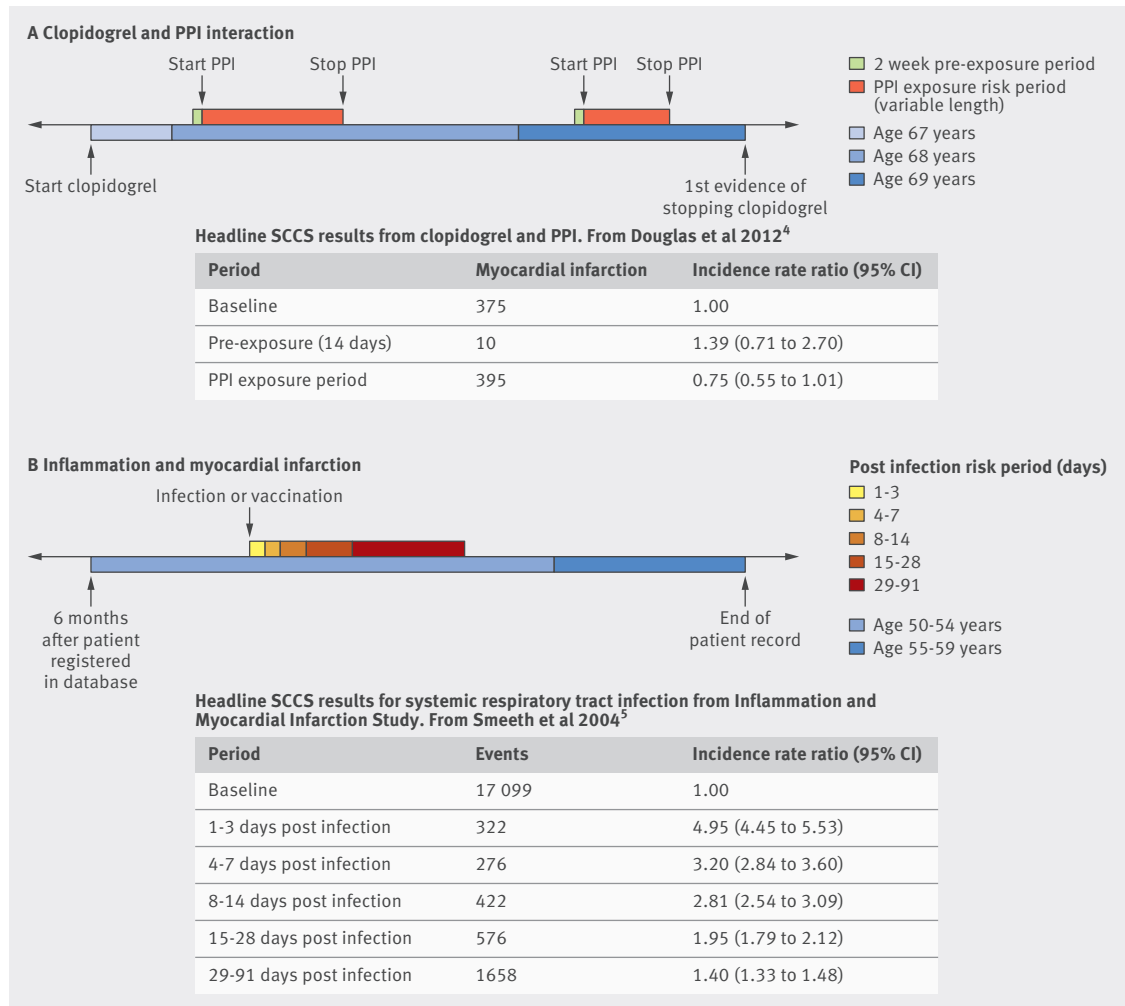


Fig 2 | Typical observation periods, risk periods, and headline results for two examples of applied SCCS. PPI=proton pump inhibitor

### 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, and the solutions and examples of good practice against each.

#### 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 contraindication 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 or moderations to the SCCS method that can mitigate

potential biases (table 1). Note that independent causes of death or 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) just before an exposure.<sup>3</sup> 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), which would otherwise bias estimates towards the null. A pre-exposure period was included in the clopidogrel and proton pump inhibitor interaction example, anticipating that the chance of starting treatment with a proton pump inhibitor could be temporarily altered by having a myocardial infarction (fig 2).<sup>4</sup> The pre-exposure period “trick” only works for short term delay (in relation to the length of observation). If there is a long

**Table 1 | Violation of assumptions, impact, and solutions with examples of good practice**

Problems	Solutions	Examples
Event temporarily decreases probability of exposure	Include a pre-exposure period	Stowe et al studied the risk of infections after mumps, measles, and rubella vaccination. Vaccination is delayed when a child has an infection; a 14 day pre-exposure period was included to allow for this <sup>6</sup>
Event temporarily increases probability of exposure	Include a pre-exposure period	Gibson et al studied the association between prescription drugs and motor vehicle crashes. A four week pre-exposure period was included as some drugs may be used to treat anxiety or pain caused by the crash <sup>7</sup>
No exposure can occur (or is observed) after event	For single exposures that cannot be repeated, begin observation period at exposure, so only exposed cases are included. End observation at planned end	Hubbard et al studied the association between first bupropion prescription and sudden death. The observation period began with date of first prescription and ended (beyond death) with date of last data collection for the cohort. At the time, bupropion could only be prescribed as a single course of treatment <sup>8</sup>
	For single or multiple exposures, use method outlined in Farrington et al <sup>9</sup>	Dodd et al studied the association between influenza vaccination and Guillain Barré syndrome (GBS). Some practitioners will not vaccinate patients with previous GBS. <sup>10</sup> Various analyses were performed: standard SCCS, vaccinated cases only, and the method outlined in Farrington et al <sup>9</sup>
Event increases probability of death	Undertake suitable tests or sensitivity analyses excluding cases who died as a result of event	Langan et al studied the risk of stroke after herpes zoster infection. A sensitivity analysis was performed excluding cases who died within 90 days of stroke. Findings were not modified <sup>11</sup>
	If necessary, use methods outlined in Farrington et al <sup>12</sup> to adjust for bias	Brauer et al studied the association between antipsychotic drugs and myocardial infarction using both methods outlined in Farrington et al <sup>12</sup> and a standard SCCS analysis. There was little difference in results <sup>13</sup>
Recurrences of an event are not independent	Study 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 et al began the observation period 12 months into follow-up time to ensure first stroke events had been correctly identified <sup>11</sup>

term delay in exposure after events, the methods in the next paragraph may be applied.

#### *No exposure can occur after the event, eg, if the outcome is death*

If the outcome is death, then 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 contraindication to exposure (table 1). 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 or censoring not occurred.<sup>3,8</sup> 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 or censoring. Alternatively, there is an extension of the SCCS method that produces unbiased estimates given these scenarios.<sup>9</sup>

If the outcome is death, but the exposure is external to the case and fully observable after death—eg, 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). If the event mortality is low, bias will be neg-

ligible 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.<sup>12</sup>

#### **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.<sup>14,15</sup>

#### **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.<sup>12</sup> A test for independence has been developed.<sup>16</sup> If events are dependent, a simple solution is to study just first events (table 1). Alternatively, an extension that allows a first event to increase the future event risk may be used<sup>17</sup>

#### **Final remarks**

With this paper we have sought to demonstrate how the SCCS 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 characteristic that remains constant over the 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.

#### **Linked information**

The Open University hosts a website for the SCCS methodology, with example datasets and code for analysis in Stata, SAS, R, Genstat, and GLIM (<http://statistics.open.ac.uk/sccs>) A link to the tutorial by Whitaker et al<sup>3</sup> can also be found at <http://statistics.open.ac.uk/sccs/sccs.pdf>

**Contributors:** IP, ID, and HW jointly wrote this paper. HW is a leading expert 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 encourage colleagues and students to consider the application of SCCS in (pharmaco)-epidemiological studies. HW is the guarantor.

**Funding:** HW was supported by an Medical Research Council methodology grant (MR/L009005/1).

**Competing interests:** We have read and understood the BMJ policy on declaration of interests and declare the following: none.

**Provenance and peer review:** Not commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 3.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: <http://creativecommons.org/licenses/by/3.0/>.

- 1 Doll R, Hill AB. The mortality of doctors in relation to their smoking habits; a preliminary report. *BMJ* 1954;1:1451-5. doi:10.1136/bmj.1.4877.1451.
- 2 Farrington CP. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* 1995;51:228-35. doi:10.2307/2533328.
- 3 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.
- 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;345: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;351:2611-8. doi:10.1056/NEJMoa041747.
- 6 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;27:1422-5. doi:10.1016/j.vaccine.2008.12.038.
- 7 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;169:761-8. doi:10.1093/aje/kwn364.
- 8 Hubbard R, Lewis S, West J, et al. Bupropion and the risk of sudden death: a self-controlled case-series analysis using The Health Improvement Network. *Thorax* 2005;60:848-50. doi:10.1136/thx.2005.041798.
- 9 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.
- 10 Dodd CN, Romio SA, Black S, et al. Global H1N1 GBS Consortium. International collaboration to assess the risk of Guillain Barré Syndrome following Influenza A (H1N1) 2009 monovalent vaccines. *Vaccine* 2013;31:4448-58. doi:10.1016/j.vaccine.2013.06.032.
- 11 Langan SM, Minassian C, Smeeth L, Thomas SL. Risk of stroke following herpes zoster: a self-controlled case-series study. *Clin Infect Dis* 2014;58:1497-503. doi:10.1093/cid/ciu098.
- 12 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;106:417-26doi:10.1198/jasa.2011.ap10108.
- 13 Brauer R, Smeeth L, Anaya-Izquierdo K, et al. Antipsychotic drugs and risks of myocardial infarction: a self-controlled case series study. *Eur Heart J* 2015;36:984-92. doi:10.1093/eurheartj/ehu263.
- 14 Ghebremichael-Weldeselassie Y, Whitaker HJ, Farrington CP. Self-controlled case series method with smooth age effect. *Stat Med* 2014;33:639-49. doi:10.1002/sim.5949.
- 15 Farrington CP, Whitaker HJ. Semiparametric analysis of case series data. *J R Stat Soc Ser C Appl Stat* 2006;55:553-94doi:10.1111/j.1467-9876.2006.00554.x.
- 16 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;59:457-75doi:10.1111/j.1467-9876.2009.00703.x.
- 17 Simpson SE. A positive event dependence model for self-controlled case series with applications in postmarketing surveillance. *Biometrics* 2013;69:128-36. doi:10.1111/j.1541-0420.2012.01795.x.

© BMJ Publishing Group Ltd 2016

**Supplementary table:** Standard epidemiological study designs

**Supplementary material:** Example of how to fit the SCCS methods in practice based on a study of the association between a specific vaccine and risk of viral meningitis