

Open Research Online

The Open University's repository of research publications and other research outputs

Self-controlled case series with multiple event types

Journal Item

How to cite:

Ghebremichael-Weldeselassie, Yonas; Whitaker, Heather J.; Douglas, Ian J.; Smeeth, Liam and Farrington, C. Paddy (2017). Self-controlled case series with multiple event types. *Computational Statistics & Data Analysis*, 113 pp. 64–72.

For guidance on citations see [FAQs](#).

© 2016 Elsevier B.V.

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1016/j.csda.2016.10.010>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Self-controlled case series with multiple event types

Yonas Ghebremichael-Weldeselassie^a, Heather J. Whitaker^{a,*}, Ian J. Douglas^b,

Liam Smeeth^b, C. Paddy Farrington^a

^a *Department of Mathematics and Statistics, The Open University, Milton Keynes, MK7 6AA, United Kingdom*

^b *Non Communicable Disease Unit, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom*

ABSTRACT

Self-controlled case series methods for events that may be classified as one of several types are described. When the event is non-recurrent, the different types correspond to competing risks. It is shown that, under circumstances that are likely to arise in practical applications, the SCCS multi-type likelihood reduces to the product of the type-specific likelihoods. For recurrent events, this applies whether or not the marginal type-specific counts are dependent. As for the standard SCCS method, a rare disease assumption is required for non-recurrent events. Several forms of this assumption are investigated by simulation. The methods are applied to data on MMR vaccine and convulsions (febrile and non-febrile), and to data on thiazolidinediones and fractures (at different sites).

1. Introduction

The self-controlled case series (or SCCS) method is an alternative to other study designs, such as cohort or case-control methods, to investigate the potential temporal association between a transient exposure and an event of interest (Farrington, 1995; Whitaker et al., 2006). Estimation is conditional on the total number of events observed for each individual over a pre-defined observation period (time over which a full event and exposure history is available). As a result, only cases (that is, individuals who have experienced at least one event) are included in the analysis, and all time-invariant confounders that act multiplicatively on the baseline incidence are automatically controlled for. It yields estimates of relative incidence; that is, the incidence in pre-defined exposure risk periods relative to all other time within the observation period.

The SCCS method is used with a single event type—which may be independently recurrent, or rare and non-recurrent. In this paper, extending the SCCS method to events which may be classified as one of several types is investigated. These events may be recurrent or non-recurrent. For example, when investigating possible associations between vaccination and convulsions, one might distinguish between febrile and non-febrile convulsions: in this case the event of interest is convulsion, and the two types are febrile and non-febrile. The analysis of non-recurrent events of multiple types is usually referred to as ‘competing risks’ (Aalen et al., 2008, pages 114–117). The most common instance of competing risks relates to death, the event of interest, the different event types corresponding to different and mutually exclusive causes of death. The SCCS method is not best suited for analysing deaths, or other events that affect subsequent exposures (Farrington et al., 2009). However, one can also envisage non-terminal competing risks, for example the first occurrence of a potentially recurrent multi-type event. In this case, event types are mutually exclusive.

In this paper, the analysis of events of multiple types is investigated. The questions of interest are whether the SCCS method requires any modification or additional assumptions to handle multiple event types; whether it is best to analyse different event types separately or together; and what degree of confounder control the SCCS method provides.

Section 2 discusses the SCCS method for multiple event types for independently recurrent events. Considered in Section 3 are competing risks—that is, terminal events of multiple types, or unique events of which only one type of event can occur. Section 4 presents simulations to investigate the scenarios discussed in Sections 2 and 3. Section 5 presents an application of the method to data on febrile convulsions and fractures.

2. Independently recurrent multi-type events

Suppose that the event of interest can take one of several, mutually exclusive types labelled $k = 1, \dots, m$. The event is potentially recurrent and, for each event type, recurrences are independent and arise over time as non-homogeneous Poisson processes. In this context, an individual can experience more than one event type, but cannot experience different event types simultaneously. All event types are of interest. Two cases were considered, according to whether the marginal counts of each type of event are independent or not.

2.1. Marginally independent event types

In this setting, the type-specific event processes are independent non-homogeneous Poisson processes. Let $\lambda_k(t|x^t, y)$ denote the incidence of events of type k , which may depend on time-varying exposures $x^t = \{x(s) : s \leq t\}$ and time-constant variables y . The risk periods may or may not be the same for all event types. Without loss of generality, take the covariates to be the same for all types. The incidence rate of the event, irrespective of event type, is

$$\lambda_+(t|x^t, y) = \sum_{k=1}^m \lambda_k(t|x^t, y).$$

Suppose an individual i experiences n_{ik} events of type k within the observation period $(a_i, b_i]$. The total number of events experienced by individual i is $n_i = \sum_{k=1}^m n_{ik}$. If $n_{ik} > 0$, denote the event times by $t_{ijk}, j = 1, \dots, n_{ik}$, for $k = 1, \dots, m$. Assuming a Poisson process, the likelihood contribution for individual i within the entire cohort including cases and non-cases, is then

$$L_i = \prod_{k=1}^m \prod_{j=1}^{n_{ik}} \lambda_k(t_{ijk}|x_i^{t_{ijk}}, y_i) \exp\left(-\int_{a_i}^{b_i} \lambda_+(t|x_i^t, y_i) dt\right), \quad (1)$$

with the convention that if for any event type k , $n_{ik} = 0$, then the corresponding term in the product is replaced by 1. Now condition on the total numbers of events of each type experienced by individual i , that is, on the vector $(n_{i1}, n_{i2}, \dots, n_{im})$ (as well as the exposure history and the observation period). This gives the following product conditional likelihood contribution for each individual:

$$L_{ci}^m = \prod_{k=1}^m \frac{\prod_{j=1}^{n_{ik}} \lambda_k(t_{ijk}|x_i^{t_{ijk}}, y_i)}{\left(\int_{a_i}^{b_i} \lambda_k(t|x_i^t, y_i) dt\right)^{n_{ik}}}.$$

Note that if $n_i = 0$, then $L_{ci}^m = 1$ and hence only cases (that is, individuals with $n_i \geq 1$ irrespective of event type) need be sampled. From now on, it is assumed that $n_i \geq 1$. Let $E_k = \{i : n_{ik} > 0\}$, the subset of individuals who have experienced events of type k . Thus $\bigcup_{k=1}^m E_k = \{1, 2, \dots, N\}$; note that the E_k may overlap. The conditional likelihood may be rewritten as

$$L_c^m = \prod_{k=1}^m \prod_{i \in E_k} \frac{\prod_{j=1}^{n_{ik}} \lambda_k(t_{ijk}|x_i^{t_{ijk}}, y_i)}{\left(\int_{a_i}^{b_i} \lambda_k(t|x_i^t, y_i) dt\right)^{n_{ik}}}.$$

The SCCS likelihood is just the product of the m type-specific SCCS likelihoods. As in the standard case, only cases (that is, individuals with $n_i > 0$) need be sampled, and time-invariant confounders y acting multiplicatively on any of the m type-specific incidence rates $\lambda_k(t|x^t, y)$ are automatically adjusted.

Note that if n_i were conditioned on, rather than (n_{i1}, \dots, n_{im}) , an extra term would appear in the conditional likelihood, representing the relative marginal abundances of the different types. This term is of the form

$$\prod_{i=1}^N \prod_{k=1}^m \left(\frac{\Lambda_{ik}}{\Lambda_{i+}}\right)^{n_{ik}},$$

where

$$\Lambda_{ik} = \int_{a_i}^{b_i} \lambda_k(t|x_i^t, y_i) dt, \quad \Lambda_{i+} = \int_{a_i}^{b_i} \lambda_+(t|x_i^t, y_i) dt.$$

Multiplicative confounders for specific event types do not generally factor out of this term, whence the preference for the analysis conditional on individual type-specific event counts.

The likelihood L_c^m offers the opportunity to test whether age and exposure effects (provided the risk periods are the same for all event types) differ between event types. This may be achieved by including interactions between a factor coding the event types, and the exposure and age effects. If they do not differ significantly, it may be reasonable to conclude that

$$\lambda_k(t|x^t, y) = \alpha_k \lambda_+(t|x^t, y)$$

for some constants $\alpha_k, k = 1, \dots, m$. In this case, event types can be combined. The multi-type SCCS likelihood L_c^m is then equivalent to the standard SCCS likelihood for the combined event types, namely

$$L_c = \prod_{i=1}^N \frac{\prod_{j=1}^{n_i} \lambda_+(t_{ij}|x_i^{t_{ij}}, y_i)}{\left(\int_{a_i}^{b_i} \lambda_+(t|x_i^t, y_i) dt \right)^{n_i}},$$

where the $t_{ij}, j = 1, \dots, n_i$ are the event times for individual i .

2.2. Marginally dependent type-specific counts

In this setting, the marginal counts n_{ik} of the different event types are correlated, but the type-specific event times remain independent within individuals. Consider the following framework for such a scenario, which allows for a wide range of correlation structures, both positive and negative. Suppose that, for individual i , events of type k arise according to a non-homogeneous Poisson process with rate

$$\lambda_{ik}(t|x_i^t, y_i) = U_{ik} \mu_k(t|x_i^t, y_i)$$

for $k = 1, \dots, m$, where U_{ik} is a positive random variable, $k = 1, \dots, m$, and (U_{i1}, \dots, U_{im}) has some m -variate density. Conditioning on (U_{i1}, \dots, U_{im}) , the analysis in the previous subsection is retrieved. The multi-type likelihood contribution for individual i is

$$\begin{aligned} L_{ci}^m &= \prod_{k=1}^m \frac{\prod_{j=1}^{n_{ik}} \lambda_{ik}(t_{ijk}|x_i^{t_{ijk}}, y_i)}{\left(\int_{a_i}^{b_i} \lambda_{ik}(t|x_i^t, y_i) dt \right)^{n_{ik}}} \\ &= \prod_{k=1}^m \frac{\prod_{j=1}^{n_{ik}} \mu_k(t_{ijk}|x_i^{t_{ijk}}, y_i)}{\left(\int_{a_i}^{b_i} \mu_k(t|x_i^t, y_i) dt \right)^{n_{ik}}}, \end{aligned}$$

which is of the same form as in the previous subsection: the random effects U_{ik} are time-constant and act multiplicatively, and hence factor out.

3. Competing risks

The analyses described in the previous section apply when recurrences of event types within an individual are independent. Typically, the SCCS method is only used for rare events, and recurrences tend to be uncommon. However, in some situations the independent recurrences assumption is clearly invalid. This is the case for myocardial infarction, for example. An SCCS model has been proposed to allow the event rate to depend on the previous number of events (Simpson, 2013). However, a simple alternative is to analyse just the first events. More generally, the SCCS method can be used for non-recurrent events, but then requires a rare disease assumption.

In the present section, we assume that the event is unique, but may still arise as m distinct, mutually exclusive, types: in this setting the m event type hazards may be regarded as competing risks.

The cohort likelihood contribution of an individual i with n_{ik} events of type $k, k = 1, \dots, m$, with $n_i = \sum_{k=1}^m n_{ik} = 0$ or 1 , arising at time t_i , in that individual's observation period $(a_i, b_i]$, is

$$L_i = \prod_{k=1}^m \lambda_k(t_i|x_i^{t_i}, y_i)^{n_{ik}} \exp\left(- \int_{a_i}^{t_i} \lambda_+(t|x_i^t, y_i) dt\right).$$

Conditioning on (n_{i1}, \dots, n_{im}) , yields the conditional likelihood contribution

$$L_{ci} = \frac{\prod_{k=1}^m \lambda_k(t_i | x_i^{t_i}, y_i)^{n_{ik}} \exp\left(-\int_{a_i}^{t_i} \lambda_+(t | x_i^t, y_i) dt\right)}{\int_{a_i}^{b_i} \prod_{k=1}^m \lambda_k(s | x_i^s, y_i)^{n_{ik}} \exp\left(-\int_{a_i}^s \lambda_+(t | x_i^t, y_i) dt\right) ds}$$

In the context of SCCS analyses, where interest primarily focuses on rare events, there are two main cases of interest to consider, corresponding to different rare disease assumptions.

3.1. All event types are rare

In this setting, write $\lambda_+(t | x^t, y) = \phi \nu_+(t | x^t, y)$ and consider the limit as $\phi \rightarrow 0$. The exponential terms in L_{ci} tend to 1. Thus, for a sample of N cases, the multi-type conditional likelihood is retrieved:

$$L_c^m = \prod_{i=1}^N \prod_{k=1}^m \left(\frac{\lambda_k(t_i | x_i^{t_i}, y_i)}{\int_{a_i}^{b_i} \lambda_k(t | x_i^t, y_i) dt} \right)^{n_{ik}},$$

with $n_{ik} = 0$ or 1 for unique events. If E_k is the subset of $\{1, 2, \dots, N\}$ corresponding to individuals experiencing event type k , this likelihood may be rewritten as

$$L_c^m = \prod_{k=1}^m \prod_{i \in E_k} \frac{\lambda_k(t_i | x_i^{t_i}, y_i)}{\int_{a_i}^{b_i} \lambda_k(t | x_i^t, y_i) dt}.$$

Thus, the analysis can proceed for independently recurrent multi-type events, as described in Section 2.

3.2. Only rare event types are of interest

In this setting, primary interest focuses on the association with the exposure x^t for $m_0 < m$ rare event types, labelled $k = 1, \dots, m_0$. The remaining event types $k = m_0 + 1, \dots, m$ are not associated with the exposure, and are not of primary interest; for this reason only events of type $k \leq m_0$ are sampled. If the ‘nuisance’ event types $k = m_0 + 1, \dots, m$ are rare, they can be ignored, and the same analysis as described in the previous subsection can be used with m_0 in place of m . If they are not, then the analysis has to be modified, as indicated by Andersen (2006).

Situations in which this scenario applies are not common. It may be relevant for studies of mortality due to one specific uncommon cause, in elderly populations in which mortality due to other causes cannot be ignored. For example, Hubbard et al. (2005) analysed sudden deaths after Bupropion. Sudden deaths are uncommon, and deaths due to other causes are competing events which may or may not also be uncommon. Note that because death is a terminal event, observation periods in this SCCS study all started at the first Bupropion dose.

In order to accommodate event types that are not of primary interest, the notation used hitherto is adapted slightly. Write $\lambda_k(t | x^t, y) = \phi \nu_k(t | x^t, y)$ for $k \leq m_0$, and consider the limit as $\phi \rightarrow 0$. The ‘nuisance’ hazards no longer depend on x^t and so are now written $\lambda_k(t | y)$, $k > m_0$. Let

$$\lambda_{++}(t | y) = \sum_{k=m_0+1}^m \lambda_k(t | y).$$

Finally, only individuals for whom $n_{ik} > 0$ for $k = 1, \dots, m_0$ are sampled. Recall that events are unique, an individual cannot experience more than one event type. Individuals for whom $n_{ik} > 0$ for $k > m_0$ are ignored.

In the limit as $\phi \rightarrow 0$,

$$\exp\left(-\int_{a_i}^{t_i} \lambda_+(t | x_i^t, y) dt\right) \rightarrow \exp\left(-\int_{a_i}^{t_i} \lambda_{++}(t | y) dt\right).$$

Thus, the conditional likelihood contribution for individual i becomes

$$L_{ci} = \frac{\prod_{k=1}^{m_0} \lambda_k(t_i | x_i^{t_i}, y_i)^{n_{ik}} \exp\left(-\int_{a_i}^{t_i} \lambda_{++}(t | y_i) dt\right)}{\int_{a_i}^{b_i} \prod_{k=1}^{m_0} \lambda_k(s | x_i^s, y_i)^{n_{ik}} \exp\left(-\int_{a_i}^s \lambda_{++}(t | y_i) dt\right) ds}$$

The integrals involving $\lambda_{++}(t | y)$ do not disappear in general, because the nuisance events may not be rare. Suppose now that

$$\lambda_k(t | x^t, y) = \mu(t | y) \exp(\beta^T x^t).$$

Since $\lambda_{++}(t|y)$ does not involve the exposure of interest, redefine

$$\lambda_k^*(t|x^t, y) = \mu^*(t|y) \exp(\beta^T x^t)$$

with

$$\mu^*(t|y) = \mu(t|y) \exp\left(-\int_0^t \lambda_{++}(s|y) ds\right).$$

With this substitution, the likelihood now takes the standard multi-type event form:

$$\begin{aligned} L_c^m &= \prod_{i=1}^N \prod_{k=1}^{m_0} \left(\frac{\lambda_k^*(t_i|x_i^{t_i}, y_i)}{\int_{a_i}^{b_i} \lambda_k^*(t|x_i^t, y_i) dt} \right)^{n_{ik}} \\ &= \prod_{k=1}^{m_0} \prod_{i \in E_k} \frac{\lambda_k^*(t_i|x_i^{t_i}, y_i)}{\int_{a_i}^{b_i} \lambda_k^*(t|x_i^t, y_i) dt}. \end{aligned}$$

Note that if the combined nuisance hazard $\lambda_{++}(t|y)$ is multiplicative in the fixed covariates y , the baseline event rate $\mu^*(t|y)$ may no longer be so. As a result, the effects of the fixed covariates y may not factor out fully. In this case, it may be necessary to stratify the analysis by the levels of the covariates y to avoid residual confounding.

4. Simulations

We conducted a simulation study related to the rare nuisance disease assumptions and residual confounding explained in Section 3.

4.1. Rare nuisance disease assumption

In this subsection, the effect of a nuisance event, with varying degrees of rarity, on the age specific relative incidence estimates related to the event of interest is investigated using simulations. We first simulated a population of 20,000 individuals with start and end of observation periods of 0 and 730 days, respectively. Ages at start of a single exposure were generated from a uniform distribution between 0 and 730 days for the entire population. Individuals who experienced the event of interest were randomly selected with an incidence rate of $p_1 = 0.01$, equivalent to 1% of the population. To identify individuals that experienced the nuisance event, we used incidence rates (p_2) of 0.01, 0.05, 0.08, 0.10, 0.15, 0.2, 0.3, 0.4 and 0.5.

The true age-related relative incidence parameters for both the event of interest and the nuisance event were taken to be 1 (baseline), 2, 5, 3 and 2, respectively over the following age groups (0, 150], (150, 300], (300, 450], (450, 600] and (600, 730]. The true exposure-related relative incidence for the event of interest was taken to be 4, over a risk period of 20 days. The nuisance event was not assumed to be associated with the exposure and hence its exposure-related relative incidence was taken to be 1. Based on these age and exposure groups, the observation periods of all cases were divided into intervals. A multinomial distribution was used to simulate the interval in which an event occurred for each case, given the event type and specified age- and exposure-related relative incidences. Age at event was simulated using a uniform distribution within the given interval. If, according to the simulation, an individual happened to experience both event types, the event that occurred second was removed, in line with the theory based on unique terminal events. Then the analysis suggested in Section 3.2, that is, a standard SCCS analysis that considers only event types of interest and ignores the nuisance event types, was applied to the generated data. Each scenario was replicated 500 times.

The bias and Mean Square Error (MSE) were used to evaluate the effects of the nuisance event on the age-related relative incidence estimates. The bias was computed as the difference between the mean of the 500 log relative incidence estimates and the true log relative incidence value. The MSE was computed as the sum of the variance of the estimated log relative incidences and the square of the bias. Results are presented in Table 1. The mean number of events of interest remaining in the sample after those who experienced the nuisance event first were removed is also given.

From Table 1, it can be seen that the bias in estimating the age-related relative incidence increases as the nuisance event becomes more common, while the exposure-related relative incidence estimates remain unbiased, as expected. When both event types are rare, for example when $p_1 = p_2 = 0.01$, the absolute value of bias in estimating the age effects is negligible; all are less than 0.005 except for the fifth age group which is 0.013. Similarly the MSE values increase as the nuisance event becomes less rare.

4.2. Residual confounding

In this subsection, we explore via simulation the effect of ignoring fixed covariates associated with non-rare competing events on the association between exposure and the events of interest. We considered one event of interest and one nuisance event with a confounding covariate: gender. All individuals in the population were randomly assigned a gender, 50% male and 50% female. The data sets were largely simulated in the same way as in Section 4.1, with the following differences. We

Table 1

Results of the rare disease assumption simulation study: Bias and Mean Square Error (MSE) of exposure- and age-related relative incidence. p_1 and p_2 are the rarity of the event of interest and nuisance event in the population, respectively.

Rarity		No. cases	Exposure		Age							
p_1	p_2		Bias	MSE	(150, 300]		(300, 450]		(450, 600]		(600, 730]	
					Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
0.01	0.01	200	-0.04	0.06	0.004	0.12	0.01	0.09	0.003	0.10	-0.01	0.11
0.01	0.05	195.0	-0.01	0.06	0.002	0.10	-0.003	0.08	-0.02	0.09	-0.05	0.11
0.01	0.08	192.0	-0.01	0.06	-0.01	0.10	-0.01	0.09	-0.04	0.10	-0.08	0.11
0.01	0.10	190.4	-0.03	0.06	0.02	0.12	-0.01	0.09	-0.05	0.12	-0.08	0.13
0.01	0.15	185.2	-0.03	0.06	-0.02	0.11	-0.04	0.09	-0.09	0.12	-0.15	0.13
0.01	0.2	180.3	-0.03	0.07	-0.01	0.10	-0.06	0.08	-0.15	0.11	-0.19	0.15
0.01	0.3	170.1	-0.03	0.07	-0.003	0.11	-0.10	0.1	-0.22	0.16	-0.30	0.21
0.01	0.4	159.9	-0.03	0.08	-0.03	0.11	-0.14	0.12	-0.30	0.19	-0.42	0.31
0.01	0.5	149.8	-0.01	0.08	-0.04	0.12	-0.20	0.13	-0.45	0.32	-0.60	0.5

Table 2

Simulation results. Bias and mean square error (MSE) of the exposure related relative incidence, Analysis with and with out a confounder gender.

Rarity		No. cases		No confounder		Confounder included	
p_1	p_2	Female	Male	Bias	MSE	Bias	MSE
0.01	0.1	233.334	244.096	-0.0145	0.0251	-0.0136	0.0252
0.01	0.15	225.592	240.298	-0.0124	0.023	-0.0115	0.023
0.01	0.2	218.978	236.138	-0.0185	0.022	-0.0176	0.0226
0.01	0.3	203.296	229.060	-0.0142	0.0322	-0.0128	0.0321
0.01	0.4	186.756	224.004	-0.0039	0.0272	-0.0031	0.0273
0.01	0.5	171.540	216.216	-0.0124	0.0296	-0.0108	0.0295

explored only the higher incidence rates for the nuisance event, $p_2 = 0.1, 0.15, 0.2, 0.3, 0.4$ and 0.5 . Females were assumed to have a greater chance of experiencing the nuisance event than males, of those experiencing the nuisance event 70% are female. The true exposure and age related relative incidence parameters were same as for Section 4.1.

For each simulated data set a standard SCCS model was fitted with and without the confounder gender. The confounding covariate was taken into account by including it as an interaction with age ($age * gender$). The bias and MSE in estimating the relative incidences related to the association between exposure and the event of interest are presented in Table 2.

The reduction in the number of female cases of interest relative to males is apparent. The results show that the bias when the confounder gender is ignored is slightly higher than the bias for when the confounder is included. This indicates the presence of a residual confounder effect, though the effect is very small.

5. Examples

As shown in Sections 2 and 3, the multi-type SCCS likelihood is simply the product of the type-specific SCCS likelihoods. The standard parametric method of fitting an SCCS model is to subdivide the observation period of each case into age and exposure categories, and fit a log-linear Poisson model with piecewise constant effects for age (Age) and exposure (Exposure), and a (nuisance) individual effect to mimic the product multinomial likelihood (Indiv). The standard model formula is thus $Age + Exposure + Indiv$. In applications involving a large number of cases, it is desirable to absorb the nuisance factor $Indiv$ which has a large number of levels, depending on study size; see Whitaker et al. (2006).

Suppose that the age and exposure categories are the same for all event types. A simple way to analyse multi-type data using the SCCS method is to concatenate the data sets prepared for a standard SCCS analysis of each of the m types, and define a new factor Type taking levels 1 to m for each of these data sets. The factors Age and Exposure are defined in the same way, and concatenated. The individual factor, however, should not be concatenated, but replaced by a new factor $Indiv2$. Let $N_k = |E_k|$ denote the number of individuals with event type k ; note that $\sum_{k=1}^m N_k \geq N$: the inequality is strict if some individuals experience events of more than one type. The cases of type k will correspond to levels of $Indiv2$ ranging from $N_1 + \dots + N_{k-1} + 1$ to $N_1 + \dots + N_k$, with multiplicities n_{ik} . The new individual factor $Indiv2$ thus has levels 1 to $\sum_{k=1}^m N_k$. This nuisance factor simply ensures that marginal totals are correctly constrained for each event type.

A variety of models can then be fitted. The model formula $Age + Exposure + Indiv2$ corresponds to the model for all event types combined; the model $Age + Exposure * Type + Indiv2$ allows for different exposure effects for the different types, but a common age effect. The model $Age * Type + Exposure * Type + Indiv2$ allows for different exposure and age effects. As these models are nested, tests to assess the need for type-specific parameters may be based on the difference in deviances.

We present two examples of analyses based on these models, using data on convulsions and measles, mumps and rubella vaccination, and data on fractures and thiazolidinediones.

Table 3

Exposure-related relative incidence (RI), with 95% confidence interval (CI) with $-2 * \log \text{Likelihood}$ and degrees of freedom (df) from various analyses of MMR vaccine and different types of convulsion.

Exposure	Recurrences included			First convulsions only		
	RI	CI	$-2 \log L$ (df)	RI	CI	$-2 \log L$ (df)
Common exposure and age effects						
All convulsions	1.76	(1.44, 2.15)	1295.8 (5029)	1.90	(1.55, 2.33)	1225.5 (4529)
Common age effects						
Febrile	1.84	(1.47, 2.30)		2.00	(1.59, 2.51)	
Non-febrile	1.51	(0.98, 2.32)		1.58	(1.004, 2.49)	
			1295.1 (5028)			1224.7 (4528)
Different exposure and age effects						
Febrile	1.84	(1.47, 2.31)		2.00	(1.59, 2.51)	
Non-febrile	1.50	(0.98, 2.32)		1.58	(1.003, 2.49)	
			1295.0 (5022)			1224.6 (4522)
Type specific separate analyses						
Febrile	1.84	(1.47, 2.31)		2.00	(1.59, 2.51)	
Non-febrile	1.50	(0.98, 2.32)		1.58	(1.003, 2.49)	

5.1. Convulsions and MMR vaccine

Here, we investigate the association between convulsions and measles/mumps/rubella (MMR) vaccine. The adverse event, convulsions, is of two types, febrile and non-febrile. The data set comprises of 894 cases, of age 1–2344 days, who experienced 988 convulsion events, of which 755 are febrile and 233 are non-febrile.

The data were analysed using the multiple event type SCCS method proposed in Section 2. The risk periods for both febrile and non-febrile events were defined as 0–25 days after MMR vaccination. Seven age groups with end dates at 365, 730, 1096, 1461, 1827 and 2192 days of age were used. Analyses both including the recurrences and using only first convulsions were performed. Three models were fitted for each analysis: (1) a model including both an `exposure*event` type interaction and an `age*event` type interaction, that is allowing different exposure and age effects for the different event types, (2) a model including an `exposure*event` type interaction but with the same age effect and (3) a model with no interactions, with same age and exposure effects for both event types. Separate models for febrile and non-febrile event types were also fitted. Results are presented in Table 3.

With a difference in deviance of 0.1 on 6 degrees of freedom from the two models including different exposure and age effects to the model including different exposure effects only, the results indicate that there is no need to allow different age effects in either model. Likewise, the difference in deviances (0.7 and 0.8 on 1 degree of freedom) from the models including different exposure effects and the model with the same exposures effect indicate that the exposure effects for febrile and non-febrile convulsions can be combined. Relative incidences estimates when only first event is considered are all higher than the corresponding estimates when recurrences are allowed. The combined analysis results in slightly narrower confidence intervals as compared to type-specific separate analyses.

5.2. Fractures and thiazolidinediones

In this subsection, similar analyses to 5.1 are performed but there are more than two event types. The aim is to investigate whether there is an increased risk of fracture associated with the use of thiazolidinediones, a class of medicines used to treat type 2 diabetes. The data are primary care computerized clinical records from the United Kingdom-based General Practice Research Database (GPRD). 1819 patients aged 40 years or older prescribed at least one thiazolidinedione and with at least one fracture were included in the analysis. The event types are classified based on the site where the fracture occurred. There are three fracture sites; hip, spine and 'FAWH' (foot, arm, wrist or hand). In the data set there are 158, 70, and 1025 events (fractures) on hip, spine and FAWH, respectively. However, there were 11 occurrences of individuals experiencing two events on the same day, we reduced these to one event in the following order of precedence: spine, hip, FAWH.

Douglas et al. (2009) investigated the effect of thiazolidinediones on fracture by site using separate SCCS models for each site. We analyse the same data using both separate models and a combined analysis. Three combined analyses, as in Section 5.1, were performed with common exposure and age effects, common age effect but different exposure effects, and different age and exposure effects. The exposure risk period began with the start of exposure (first thiazolidinedione prescription) and ended at the end of the observation period, so the baseline/control period only included time before the start of exposure. 42 age groups were defined: the first age group was less than 14 610 days (40 years) of age, followed by 5 age groups of length two years, 28 groups of length one year, 7 groups of length two years and the last age group with age greater than 33 603 days. The 42nd age group was taken as a reference group in the analysis. Results are given in Table 4.

Allowing different age effects was found to be important with a difference in deviance of 122.1 on 82 degrees of freedom, $p = 0.002$ (and 109 on 82 d.f. for the first fractures analysis $p = 0.025$), there was a very apparent difference in exposure risk

Table 4

Relative incidence (RI), 95% confidence interval (CI) of FAWH (Foot, arm, wrist, or hand) Hip, and Spine fractures, with $-2 * \log\text{Likelihood}$ and degrees of freedom (df).

Exposure	Recurrences included			First fractures only		
	RI	CI	$-2\log L$ (df)	RI	CI	$-2\log L$ (df)
Common exposure and age effects						
Any fracture	1.36	(1.13, 1.62)	5529.8 (13 482)	1.44	(1.19, 1.75)	4785.9 (11 591)
Common age effects						
FAWH	1.12	(0.93, 1.36)		1.23	(1.00, 1.51)	
Hip	3.50	(2.35, 5.21)		3.40	(2.18, 5.30)	
Spine	2.41	(1.36, 4.29)		2.52	(1.37, 4.63)	
			5496.6 (13 480)			4763.6 (11 589)
Different exposure and age effects						
FAWH	1.26	(1.03, 1.54)		1.36	(1.10, 1.69)	
Hip	1.79	(1.07, 3.00)		1.77	(1.00, 3.14)	
Spine	2.19	(0.99, 4.85)		2.17	(0.95, 4.98)	
			5374.5 (13 398)			4656.0 (11 507)
Site specific separate analyses						
FAWH	1.26	(1.03, 1.54)		1.36	(1.10, 1.69)	
Hip	1.79	(1.07, 3.00)		1.77	(1.00, 3.14)	
Spine	2.19	(0.99, 4.85)		2.17	(0.95, 4.98)	

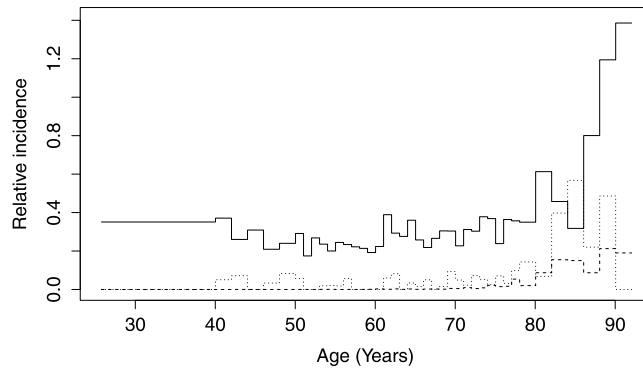


Fig. 1. Age related relative incidence functions for Foot, arm, wrist or hand (FAWH) (solid line), Hip (dashed line) and Spine (dotted line) fractures.

effects with and without common age effects. Allowing different exposure effects was also found to be important $p < 0.001$ for first and all fractures. In this case, separate site specific analyses coincide with the combined analysis. There was a significant increased risk of hip and foot, arm, wrist or hand fracture during the exposure risk period. Spine fracture had the highest, but non significant, exposure-related relative incidence estimate of 2.19(0.99, 4.85); the large standard error is due to the small number of spine events. The relative incidence estimates for FAWH fractures increased when only first fractures per individual are used in the analysis. Hip related relative incidences are larger when recurrences are allowed.

Age related site specific relative incidence estimates are presented in Fig. 1. Generally, the age related relative incidence of all types of fractures increases with age. The baseline relative incidence of FAWH fracture is higher than hip and spine fractures. However, hip and spine fractures seem to have similar baseline relative incidences.

6. Discussion

The SCCS method has been extended to situations in which events may be classified as different types. It has been shown that, quite generally, the multi-type likelihood is just the product of the type-specific likelihoods. Hitherto, analyses of different event types were done separately. While this may often remain a reasonable approach, this paper offers the further opportunity formally to investigate several types in the same analysis, with potential gains in parsimony and efficiency, for example in the analysis of MMR convulsions we have a slight gain in efficiency when using the new method.

The equivalence of the multi-type likelihood with the product of type-specific likelihoods holds under a broad range of scenarios which are likely to arise in practice. Thus, it holds for recurrent event types, both when the marginal type-specific counts are independent, or when they are dependent through multiplicative random type-specific individual effects. It also holds for rare non-recurrent events—a special case of competing risks. We investigated rare non-recurrent events, when competing events that are not of primary interest may not be rare. The general model also holds in this case, though the interpretation of the age effect differs, provided that the competing events not of primary interest are not associated

with the time-varying exposure under study. Bias in estimates of the exposure-related relative incidence was unaffected by the presence of a nuisance event, even when the nuisance event affected half of the population. We struggle to think of circumstances in which we would be interested in common competing events that did depend on the exposure of interest: in such circumstances, designs other than the SCCS should probably be used.

The original SCCS method requires that event times (of a single event type) are independent within individuals. The only context in which the SCCS method has been extended to handle dependence between events is that in which the event rate depends on the previous number of events (Simpson, 2013). A common solution is to analyse only first events. This paper does not explore possible dependence in the *timing* of multi-event types, only dependence in the marginal counts of event types. Using fractures and thiazolidinediones as an example, it is possible that all three fracture types: hip, spine and FAWH, could happen simultaneously, or notionally, as a direct result of one fracture type occurring or in its healing, another fracture type could be either more or less likely to occur in the short term. The first scenario of simultaneous fractures can be handled easily in the framework presented, either by defining a new event type 'multiple fractures', or by allocating a multiple fracture to the type of its most serious component. The second scenario of short-term dependence in timing of events is not straight forward, but an analysis of first events only would be valid. Dependence of an event of interest on a measured nuisance events could be allowed for by including 'nuisance risk periods' as covariates in the model, similar to the exposure risk periods, but any such dependence could only be one-way: the event of interest must not influence the occurrence of nuisance events.

The rare nuisance disease assumption was investigated in more detail than hitherto by simulation, in order to provide guidance as to when the SCCS method may be used with non-recurrent events. Only bias in the age-related relative incidence was apparent.

Acknowledgements

This research was supported by an MRC Methodology Grant (MR/L009005/1). We thank Professor Elizabeth Miller and Dr Nick Andrews, of Public Health England, for sharing the convulsions data with us.

References

- Aalen, O., Borgan, O., Gjessing, H., 2008. *Survival and Event History Analysis, A Process Point of View*. Springer-Verlag.
- Andersen, P.K., 2006. Discussion on the paper by Farrington and Whitaker "semiparametric analysis of case series data". *Appl. Stat.* 55 (5), 587–589.
- Douglas, I.J., Evans, S.J., Pocock, S., Smeeth, L., 2009. The risk of fractures associated with thiazolidinediones: a self-controlled case-series study. *PLoS Med.* 6 (9), e1000154.
- Farrington, C.P., 1995. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* 51, 228–235.
- Farrington, C.P., Whitaker, H.J., Hocine, M.N., 2009. Case series analysis for censored, perturbed or curtailed post-event exposures. *Biostatistics* 10 (1), 3–16.
- Hubbard, R., Lewis, S., West, J., Smith, C., Godfrey, C., Smeeth, L., Farrington, P., Britton, J., 2005. Bupropion and the risk of sudden death: a self-controlled case-series analysis using the health improvement network. *Thorax* 60 (10), 848–850.
- Simpson, S.E., 2013. A positive event dependence model for self-controlled case series with applications in postmarketing surveillance. *Biometrics* 69 (1), 128–136.
- Whitaker, H.J., Farrington, C.P., Spiessens, B., Musonda, P., 2006. Tutorial in biostatistics: the self-controlled case series method. *Stat. Med.* 25 (10), 1768–1797.