ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

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

ARTICLE INFO

Article history:
Received 29 January 2016
Received in revised form 9 May 2016
Accepted 13 October 2016
Available online xxxx

Keywords:
Self-controlled case series
Competing risks
Multiple events

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.

* Corresponding author.
E-mail address: heather.whitaker@open.ac.uk (H.J. Whitaker).

http://dx.doi.org/10.1016/j.csda.2016.10.010
0167-9473/

Please cite this article in press as: Ghebremichael-Weldeselassie, Y., et al., Self-controlled case series with multiple event types. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.10.010

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 $k = 1, \ldots, m$. 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 $t_{ijk}, j = 1, \ldots, n_{ik}$, for $k = 1, \ldots, m$. Assuming a Poisson process, the likelihood contribution for individual i within the entire cohort including cases and non-cases, is then

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

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 $(n_{i1}, n_{i2}, \ldots, n_{im})$ (as well as the exposure history and the observation period). This gives the following product conditional likelihood contribution for each individual:

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

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 $\bigcup_{k=1}^{m} E_k = \{1, 2, \ldots, N\}$; note that the E_k may overlap. The conditional likelihood may be rewritten as

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

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 $(n_{i1}, \ldots, n_{im})$, an extra term would appear in the conditional likelihood, representing the relative marginal abundances of the different types. This term is of the form

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

Please cite this article in press as: Ghebremichael-Weldeselassie, Y., et al., Self-controlled case series with multiple event types. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.10.010