

Semi-parametric analysis of case series data

C. P. Farrington and H. J. Whitaker

Department of Statistics

The Open University

Milton Keynes MK7 6AA, UK

March 5, 2006

Abstract

The case series model for estimating the association between an age-dependent exposure and an outcome event requires information only on cases and implicitly adjusts for all age-independent multiplicative confounders, while allowing for an age-dependent baseline incidence. In this paper the model is presented in greater generality than hitherto, including more general discussion of its derivation, underlying assumptions, applicability, limitations and efficiency. A semi-parametric version of the model is developed, in which the age-specific relative incidence is left unspecified. Modelling covariate effects and testing assumptions are discussed. The small sample performance of this model is studied in simulations. The methods are illustrated with several examples from epidemiology.

1 Introduction

The case series model was originally developed to estimate the relative incidence of acute events following transient exposures (Farrington 1995). The model requires data only on cases, that is to say, individuals who experience the event of interest. It is self-matched and hence controls implicitly for all age-independent confounders that act multiplicatively on the baseline incidence. In certain circumstances it has been found to have good efficiency relative to a cohort model. The method has been used in vaccine safety studies, see for example Andrews (2002), in non-vaccine pharmaco-epidemiology, for example Hubbard, Farrington, Smith, Smeeth & Tattersfield (2003) and Hocine, Guillemot, Tubert-Bitter & Moreau (2005), and, independently, in other

areas of epidemiology (Becker, Li & Kelman 2004). The implementation and applications of the method are described in Whitaker, Farrington, Spiessens & Musonda (2006).

The case series model is one of several that have been proposed for analysing data on cases only. Prentice, Vollmer & Kalbfleisch (1984), developing an idea of Aalen, Borgan, Keiding & Thorman (1980), suggested that a Cox proportional hazards model could be used to analyse case series. This gives a valid test for no association, but does not yield readily interpretable effect estimates. The case-crossover model of Maclure (1991), which is a case-control method with referents selected from the case's own history, is self-matched, but only yields consistent estimates under rather strong exchangeability conditions (Vines & Farrington 2001). The method of Feldmann (1993), on the other hand, assumes a constant baseline incidence but gives consistent estimates, though it is only approximately self-matched for rare events. The case series method incorporates aspects of these earlier approaches. In particular, it coincides with Feldmann's method when the disease is rare and the baseline incidence is constant. A further close connection is to the work of Cox (1972b) on modulated Poisson processes. Cox's conditional likelihood for a modulated Poisson process is identical to a case series likelihood for a single individual.

The case series model, in its original form, assumed categorical exposures and piecewise constant baseline incidence. In Section 2 we present the case series model in greater generality, including arbitrary exposure functions. This more general model encompasses the full-stratum and time-stratified case-crossover models proposed by Navidi (1998) and Lumley & Levy (2000). We discuss the derivation of the model, the assumptions required, and efficiency relative to the underlying cohort model. This provides the basis for the development of a semi-parametric case series model in Section 3, in which the age-specific relative risk function is left unspecified. This semi-parametric model is motivated by experience with the parametric model, which has been found on occasion to lack robustness to mis-specification of the baseline age-specific relative incidence, particularly in the analysis of non-transient exposures. We discuss which covariate effects are estimable, confounded, and implicitly adjusted. We indicate how some of the assumptions can be tested. In Section 4 we discuss inference, and report on the performance of the model in small and medium samples. In Section 5 we describe applications in four contrasting settings, which serve to illustrate some of the features described earlier.

2 The case series model

We consider events occurring within a specified observation period $(a, b]$, usually determined by age and calendar time boundaries. This determines an individual-specific observation period $(a_i, b_i] \subseteq (a, b]$ for individual i , $i = 1, \dots, N$. We also consider age-dependent exposures over $(a, b]$: each individual's exposure status is documented over their observation period. Our purpose is to draw inferences about the association, if any, between exposures and events. For simplicity, the underlying time line used throughout is age. In some applications the relevant time line might be calendar time. In others, both age and time may be relevant.

2.1 The case series likelihood

In the course of their observation period, individuals experience exposures from one or more risk factors which may alter their probability of experiencing an event. Let $x_i(t)$ denote the vector of exposures experienced by individual i at age t , and let $x_i^t = \{x_i(s) : s \leq t\}$ be the exposure history to age t . Suppose that events for individual i arise in a counting process with intensity process $\lambda_i(t|x_i^t)$. The likelihood that individual i experiences n_i events at times t_{i1}, \dots, t_{in_i} in $(a_i, b_i]$ is

$$L_i^u = \prod_{j=1}^{n_i} \lambda_i(t_{ij}|x_i^{t_{ij}}) \times \exp\left(-\int_{a_i}^{b_i} \lambda_i(t|x_i^t) dt\right). \quad (1)$$

This is a cohort likelihood, in the sense that, at each age t and for each individual i , the exposure history x_i^t is regarded as fixed. Note that the covariate x_i^t can depend on prior events, in which case (1) is a partial likelihood (Kalbfleisch & Prentice 2002). Now let $x_i = x_i^{b_i}$ denote the entire history of exposure to age b_i . Suppose that, for all $t \in (a_i, b_i]$, $\lambda_i(t|x_i^t) = \lambda_i(t|x_i)$. This is equivalent to requiring that $X_i(t)$ is a vector of external covariates, and is further discussed in Subsection 2.3. Thus, conditionally on x_i , the counting process for individual i is a non-homogeneous Poisson process with rate $\lambda_i(t|x_i)$. Conditioning on the number of events n_i observed in $(a_i, b_i]$ gives the following conditional likelihood for individual i :

$$L_i^c = \frac{\prod_{j=1}^{n_i} \lambda_i(t_{ij}|x_i)}{\left(\int_{a_i}^{b_i} \lambda_i(t|x_i) dt\right)^{n_i}}. \quad (2)$$

The case series method bases inference on this conditional likelihood. Note that individuals with no events in $(a_i, b_i]$ contribute $L_i^c = 1$ and hence need not be sampled: inference is based solely on a random sample of cases, that is to say a case series, whence the name of the method. Note also that the case series method is derived from the cohort model (1). It retains key features of a cohort design, namely conditioning on exposure with outcome events treated as random. In this sense it differs conceptually from the case-crossover method, also based on cases only, but which is derived from case-control logic and does not allow for variability in baseline incidence (Maclure 1991, Vines & Farrington 2001).

2.2 Self-control and the proportional incidence model

The incidence $\lambda_i(t|x_i)$ is most conveniently parameterised according to the proportional incidence model

$$\lambda_i(t|x_i) = \varphi\psi(t) \exp\{\gamma_i + x_i(t)^T \beta\}$$

where φ is the underlying incidence at age a , $\psi(t)$ is the age-specific relative incidence, assumed to be common to all individuals, with $\psi(a) = 1$, say, γ_i is a sum of fixed and random individual effects, which might depend on covariates that do not vary over the period $(a_i, b_i]$, and $x_i(t)$ is the vector of age-dependent exposures previously introduced. The focus of inference is the parameter β . Let $\Psi(t) = \int_a^t \psi(s)ds$ denote the cumulative relative incidence to age t . Then, given a random sample of N cases, the case series likelihood is

$$L = \prod_{i=1}^N \prod_{j=1}^{n_i} \left(\frac{\psi(t_{ij}) \exp\{x_i(t_{ij})^T \beta\}}{\int_{a_i}^{b_i} \exp\{x_i(t)^T \beta\} d\Psi(t)} \right). \quad (3)$$

Note that the terms φ and $\exp(\gamma_i)$ in (2), which represent age-independent fixed and random individual effects, cancel out. In this sense, the case series method parameterised in this way is self-controlled: estimation is ‘within individuals’. The likelihood (3) generalises that of Farrington (1995), in which $\psi(t)$ and $x_i(t)$ were chosen to be step functions. It also generalises the full-stratum bidirectional case-crossover model of Navidi (1998) and the time-stratified case-crossover model of Lumley & Levy (2000).

2.3 Assumptions and scope of the method

The condition $\lambda_i(t|x_i^t) = \lambda_i(t|x_i)$, introduced in Subsection 2.1, is critical to the validity of the case series method. This condition states that the event rate at age t , given

the exposure history to age t , is equal to the event rate at age t , given the exposure history over the entire observation period. Equivalently, exposure histories X_i^s for all $s \geq t$ must be independent of the occurrence of an event at t . Thus $X_i(t)$ must be an external time-varying covariate (Kalbfleisch & Prentice 2002). In practice, this means that exposures must be independent of prior events. This condition can be circumvented in some special cases. For example, if $X_i(t)$ is wholly determined at some age c_i then the method may be applied with observation periods $(c_i, b_i]$ since the subsequent exposure history is known. Nevertheless, the requirement that $X_i(t)$ should be an external variable is potentially restrictive. The validity of this assumption can be tested in certain circumstances, discussed in Subsection 3.4.

The case series method was derived from a Poisson cohort model, conditionally on the exposures, and so is applicable to potentially recurrent events. In fact, it is also applicable to rare non-recurrent events, as is now demonstrated; see also Farrington (1995). Suppose that $\lambda_i(t|x_i^t) = \lambda_i(t|x_i)$ now denotes a hazard function, where x_i^t and x_i now are exposure histories in $(0, t]$ and $(0, b_i]$, respectively, and let $S_i(t|x_i)$ denote the survivor function. The conditional likelihood for individual i , who experiences the event at t_i , given x_i and that an event occurred in $(a_i, b_i]$, is

$$L_i^c = \frac{S_i(t_i|x_i)\lambda_i(t_i|x_i)}{S(a_i|x_i) - S(b_i|x_i)}. \quad (4)$$

Suppose that $\lambda_i(t|x_i) = \varphi\nu_i(t|x_i)$ where φ is the baseline hazard at age a , say, and $\nu_i(t|x_i)$ is a relative hazard function, bounded on $(0, b_i]$. Then as $\varphi \rightarrow 0$, $S_i(t|x_i) \rightarrow 1$ and $S(a_i|x_i) - S(b_i|x_i) \rightarrow \int_{a_i}^{b_i} \lambda_i(t|x_i)dt$. Hence the conditional likelihood (4) tends to (2), the ratio of the two being $1 + O(\varphi)$. It follows that the case series method may be used with rare non-recurrent events. It may also be used with clustered recurrent events, that is, events for which the incidence is increased following a first event, provided that this initiating event is rare, by restricting the analysis to the first event. Nevertheless, the case series method is not applicable to clustered data more generally, or to frequent non-recurrent events. In practice, however, the method is most commonly used for rare events, so the Poisson property required in deriving the model is seldom seriously restrictive.

A third assumption of the case series method is that covariates act multiplicatively on the baseline incidence. This is critical to the ‘within-individual’ character of the model: time-independent covariates that are not multiplicative will not cancel out. The assumption must also hold for the time-varying exposures of interest. The proportional incidence assumption can be tested using methods similar to those used for

the proportional hazards model, and will be discussed briefly in Subsection 3.4.

A fourth assumption of the method is that the observation periods $(a_i, b_i]$ are independent of event times. This is a stronger requirement than the independent censoring assumption in cohort models, in which censoring after time t can depend on the event history prior to t (Kalbfleisch & Prentice 2002). In particular, in case series analyses, individuals must generally remain observable after an event occurs. In some special circumstances, deaths can be analysed using the case series method, for example when the entire exposure history is known at some age c_i . Provided that it is reasonable to ascribe a notional study end b_i , a case series analysis can be undertaken with observation periods $(c_i, b_i]$. Nevertheless, the independence assumption is potentially restrictive: for example it may be violated if the events of interest increase short-term mortality. In practice, the case series method is often robust to failure of this assumption. An example is described in Subsection 5.4.

As described above, the case series model requires stronger assumptions than cohort or case-control models. Taken together, these four conditions undoubtedly restrict the applicability of the case series method. In one fundamental respect, however, its requirements are notably less stringent than those of cohort and case-control methods. Time-independent confounders and individual random effects need not be modelled, or even identified, provided that they act multiplicatively on the baseline incidence. Thus confounding due to the multiplicative effects of genetic factors, location, socio-economic background, sex, underlying health state, to list a few examples, are automatically controlled for in the analysis. It is seldom possible to identify all relevant confounders, let alone control adequately for them, in case-control or non-randomized cohort studies, the validity of which is predicated on the correctness of the model.

2.4 Relative efficiency

An attractive feature of the case series method is its limited data requirements compared to other commonly used methods. However, conditioning on the number of events for each individual generally results in some loss of information about β . The information lost is, roughly, that contained in the marginal distribution of numbers of events, given the exposure histories x_i . This marginal distribution does not involve the timing of events relative to exposures.

In this subsection we explore the relative asymptotic efficiency of $\widehat{\beta}$ estimated using the case series model (3) and using the cohort model (1). To simplify the calculations we assume that $(a_i, b_i] = (a, b]$ and $\gamma_i + \log(\varphi) = \gamma$ for all i , and that the age-specific

relative incidence $\psi(t)$ may be written $\psi(t) = \exp\{g(t; \alpha)\}$ where $g(t; \alpha)$ is linear in the parameters α (as is the case with polynomial or step function representations). Thus the cohort model involves the unknown parameters α , β and γ and the case series model involves just α and β .

At its heart, the relative efficiency of the case series and cohort methods derives from the interplay of within and between individual variability in exposure and age effects. We begin by defining appropriate measures of within-individual variability for each individual. For an individual with exposure history $x = \{x(t) : a \leq t \leq b\}$ define the density

$$f_x(t) = \frac{\exp\{g(t; \alpha) + x(t)^T \beta\}}{\int_a^b \exp\{g(s; \alpha) + x(s)^T \beta\} ds}.$$

Let $\mu_x = \int x(t) f_x(t) dt$, and $\nu_x = \int x(t) x(t)^T f_x(t) dt - \mu_x \mu_x^T$ denote the within-individual mean and variance of x with respect to this density. Let $g'(t)$ denote the vector function $\partial g(t; \alpha) / \partial \alpha$, and define $\theta_x = \int g'(t) f_x(t) dt$, $\delta_x = \int g'(t) g'(t)^T f_x(t) dt - \theta_x \theta_x^T$ and $\kappa_x = \int x(t) g'(t)^T f_x(t) dt - \mu_x \theta_x^T$.

Define the weights

$$\omega_x = \frac{\int_a^b \exp\{g(t; \alpha) + x(t)^T \beta\} dt}{E \int_a^b \exp\{g(t; \alpha) + X(t)^T \beta\} dt}$$

where the expectation in the denominator is over X , the exposure history of a randomly selected individual. If f denotes the population density of X then $\omega_x f(x)$ is also a density, which we refer to as the weighted density.

Now define the average within-individual variances and covariance as follows: $Var_w(X) = E(\omega_X \nu_X)$, $Var_w(g') = E(\omega_X \delta_X)$ and $Cov_w(X, g') = E(\omega_X \kappa_X)$. The subscripts w indicate that these refer to within-individual quantities, as distinct from between-individual variances and covariance, subscripted by the letter b . These are defined as follows: $Var_b(X) = E(\omega_X \mu_X \mu_X^T) - E(\omega_X \mu_X) E(\omega_X \mu_X^T)$, $Var_b(g') = E(\omega_X \theta_X \theta_X^T) - E(\omega_X \theta_X) E(\omega_X \theta_X^T)$ and $Cov_b(X, g') = E(\omega_X \mu_X \theta_X^T) - E(\omega_X \mu_X) E(\omega_X \theta_X^T)$. Note that these are indeed variances and covariance relative to the weighted density of X . The following proposition then holds.

Proposition 1 Suppose that, for all exposure histories x the within and between individual moments defined above are finite. Then, as the cohort size and hence the

number of events n increase, the variances of $\widehat{\beta}$ under the case series and cohort models are asymptotically equivalent to the following expressions:

$$\begin{aligned} \text{Case series} & : \frac{1}{n} [Var_w(X) - Cov_w(X, g')Var_w(g')^{-1}Cov_w(X, g')^T]^{-1} \\ \text{Cohort} & : \frac{1}{n} [Var_{w+b}(X) - Cov_{w+b}(X, g')Var_{w+b}(g')^{-1}Cov_{w+b}(X, g')^T]^{-1} \end{aligned} \quad (5)$$

where $Var_{w+b} \equiv Var_w + Var_b$ and $Cov_{w+b} \equiv Cov_w + Cov_b$.

Outline proof The matrix of second derivatives of the case series log likelihood, times -1 , may be written

$$\sum_{i=1}^{N_c} n_i \begin{pmatrix} \nu_{x_i} & \kappa_{x_i} \\ \kappa_{x_i}^T & \delta_{x_i} \end{pmatrix}$$

where N_c is the cohort size (non-cases have $n_i = 0$). As N_c and hence n increase, this is asymptotically equivalent to

$$I_{cs} = n \begin{pmatrix} Var_w(X) & Cov_w(X, g') \\ Cov_w(X, g')^T & Var_w(g') \end{pmatrix}.$$

For the cohort model a similar argument applies. The calculations are greatly simplified if $X(t)$ is replaced by $X(t) - E(\omega_X \mu_X)$, $g(t)$ is replaced by $g(t) - \alpha^T E(\omega_X \theta_X)$ and the model is reparameterised in γ . Then the matrix of second derivatives of the cohort log likelihood, times -1 , is asymptotically equivalent to

$$I_c = n \begin{pmatrix} Var_{w+b}(X) & Cov_{w+b}(X, g') & 0 \\ Cov_{w+b}(X, g')^T & Var_{w+b}(g') & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and the expressions (5) follow by inverting the matrices I_{cs} and I_c . ■

The expression $Var_w(X) - Cov_w(X, g')Var_w(g')^{-1}Cov_w(X, g')^T$ is zero if $X(t)$ is of the form $Ag'(t) + c_X$, where A is a constant matrix and c_X does not vary with t , but depends on X and hence may vary between individuals. In this case, $X(t) - Ag'(t)$ does not vary with t : as functions of t , $X(t)$ and $Ag'(t)$ have the same shape, though the constant difference between them, c_X , may vary between individuals.

Thus, if the $X(t)$ have the same shape as a linear combination of the components of $g(t; \alpha)$ then, from (5), exposure and age effects are wholly confounded and $\widehat{\beta}$ cannot be estimated. Note that such non-identifiability also arises with the cohort method when $X(t)$ is of the form $Ag'(t) + c$, c now representing a constant vector that does not vary between individuals.

In practice, if $X(t) = h(t) + c_X$, so that the exposure functions for different individuals have the same shape, age and exposure effects will not be readily identifiable in case series analyses without strong parametric assumptions about $\psi(t)$ (and might not be identifiable even with such assumptions). If $\psi(t)$ is to be estimated semi-parametrically, it is therefore necessary to insist that the individual exposure functions $X(t)$ vary in shape between individuals. Also, in designing case series analyses it is usually important to ensure that there is adequate between-individual variation in the exposure histories of the cases. An exception, in which all individuals have the same exposure history, is discussed in Subsection 5.3.

Suppose now that β has dimension 1. The asymptotic relative efficiency of $\widehat{\beta}$ in the case series design compared to the cohort design is the ratio of within-individual to within plus between-individual precisions. Rearranging equations (5), this may be expressed compactly as

$$ARE = \frac{Var_w(X)}{Var_{w+b}(X)} \left\{ \frac{1 - R_w^2}{1 - R_{w+b}^2} \right\}$$

where R_w^2 and R_{w+b}^2 are multiple correlation-like quantities between $X(t)$ and $g'(t)$. Aside from issues of identifiability, the asymptotic relative efficiency of the case series method will tend to be high when the ratio of average within-individual to between-individual variances in exposure is high. (We emphasize that these are weighted variances, where the weights depend both on age and on exposures.)

To gain further insight we examine in more detail an important special case in which, for simplicity, we assume $\psi(t)$ known to be identically 1, so that $ARE = Var_w(X)/Var_{w+b}(X)$. Let $I_S(\cdot)$ denote the indicator function for a set S . Suppose first that

$$X(t) = \begin{cases} I_{(a, a+d]}(t) & \text{with probability } p \\ 0 & \text{with probability } 1 - p, \end{cases}$$

corresponding to exposure of duration d in a proportion p of the population, and no exposure in the remainder. Then

$$ARE = \frac{1 + pr e^\beta / (1 - pr)}{1 + re^\beta / (1 - r)} \quad (6)$$

where $r = d/(b-a)$ is the proportion of the observation period at risk. This expression was previously derived in Farrington (1995). The ARE increases as r and β decline and as p increases.

This discussion of relative efficiency has assumed that there are no age-independent covariates. Estimating the effects associated with such covariates may reduce the efficiency of the cohort method, but not of the case series method. If relevant fixed or random age-independent effects are ignored, it is possible that the efficiency achieved with the case series model may be better than with the cohort model. This has been observed in a direct comparison of the two methods on the same data (Farrington, Nash & Miller 1996).

3 A semi-parametric model

The primary interest is usually inference about β , the relative incidence associated with exposure; the age-specific relative incidence $\psi(t)$ is a nuisance parameter. Farrington (1995) used a step function to represent $\psi(t)$. This is at best an approximate representation. Furthermore, mis-specification of the age groups can produce biased estimates of β . These considerations motivate the development of a semi-parametric model in which the age effect is left unspecified.

3.1 Semi-parametric likelihood

Suppose that $\psi(t)$ is left completely unspecified, other than it being non-negative and bounded. Let \mathfrak{S} denote the set of all distinct event ages t_{is} for all N individuals combined. Suppose that there are M distinct event ages, listed in increasing order as s_1, \dots, s_M . The non-parametric maximum likelihood estimator of $\Psi(t)$ is selected within the family of non-decreasing step functions constant outside \mathfrak{S} and with jumps of height $\Delta\Psi(t)$ for $t \in \mathfrak{S}$. The maximum likelihood estimator $(\hat{\beta}, \hat{\Psi}(t))$ maximises the semi-parametric likelihood

$$L_{\mathfrak{S}} = \prod_{i=1}^N \prod_{j=1}^{n_i} \left(\frac{\Delta\Psi(t_{ij}) \exp\{x_i(t_{ij})^T \beta\}}{\int_{a_i}^{b_i} \exp\{x_i(t)^T \beta\} d\Psi(t)} \right).$$

To derive a form of the likelihood suitable for modelling purposes, write $\Delta\Psi(s_r) = \exp(\alpha_r)$, and without loss of generality let $\alpha_1 = 0$. For each individual i and for each $s_r \in \mathfrak{S}$ define the weight $w_{ir} = I_{(a_i, b_i]}(s_r)$. Let $\alpha_{ij} = \sum_{r=1}^M I_{\{t_{ij}\}}(s_r) \alpha_r$, the value of α_r corresponding to t_{ij} . The likelihood (3) can then be written

$$L_{\mathfrak{S}} = \prod_{i=1}^N \prod_{j=1}^{n_i} \left(\frac{\exp\{\alpha_{ij} + x_i(t_{ij})^T \beta\}}{\sum_{r=1}^M w_{ir} \exp\{\alpha_r + x_i(s_r)^T \beta\}} \right). \quad (7)$$

Note that the likelihood depends only on the event ages s_r through their ranks and through the covariate values $x_i(s_r)$.

3.2 Fitting the model

The likelihood (7) is product multinomial, and is readily fitted via an associated Poisson model (McCullagh & Nelder 1989). Let n_{ir} be the number of events experienced by individual i at age s_r . The associated Poisson model has weights w_{ir} , responses n_{ir} with mean λ_{ir} , and link function

$$\log(\lambda_{ir}) = \gamma_i + \alpha_r + x_i(s_r)^T \beta. \quad (8)$$

Note that the individual effect γ_i must be fitted in order to constrain the marginal totals to their observed values. This parameter is usually of large dimension, which may cause fitting problems if the model is fitted directly. In log-linear modelling software with an absorption facility it is possible to sidestep this problem entirely as the model is fitted without the parameters γ_i being estimated (Thompson 1977). However, the parameter α is also of high dimension and must be estimated. In consequence, fitting the semi-parametric model can be computationally demanding for large datasets.

In some applications, events are grouped in time units, for example by day of age. In this case the semi-parametric model (7) coincides with the parametric step function model with a separate step for each time unit. However, fitting such a parametric model can be awkward as many age groups have zero counts, which may cause numerical problems. This is avoided in the semi-parametric formulation.

The fitting method is similar in spirit to that used for parametric models: see Whitaker et al. (2006) for further details. Software to fit the semi-parametric model is available from the website <http://statistics.open.ac.uk/sccs>.

3.3 Covariates, interactions and temporal effects

As previously remarked, the main effects of all age-independent covariates are implicitly adjusted for in the case series model. The main effects of such variables on the incidence are not estimable. However, their interactions with the age-dependent exposure variables $x_i(t)$ are estimable. Let z_i be a univariate age-independent covariate. The incidence model with interaction is

$$\lambda_i(t|x_i, z_i) = \varphi\psi(t) \exp\{\gamma_i + x_i(t)^T \beta + z_i \cdot x_i(t)^T \delta\}. \quad (9)$$

For a case series analysis, the likelihood (7) is extended to:

$$L_{\mathfrak{S}} = \prod_{i=1}^N \prod_{j=1}^{n_i} \left(\frac{\exp\{\alpha_{ij} + x_i(t_{ij})^T \beta + z_i \cdot x_i(t_{ij})^T \delta\}}{\sum_{r=1}^M w_{ir} \exp\{\alpha_r + x_i(s_r)^T \beta + z_i \cdot x_i(s_r)^T \delta\}} \right)$$

where δ is the interaction parameter. The model (9) is fitted via the log-linear model

$$\log(\lambda_{ir}|x_i, z_i) = \gamma_i + \alpha_r + x_i(s_r)^T \beta + z_i \cdot x_i(s_r)^T \delta. \quad (10)$$

In some situations, the baseline incidence may depend on calendar time u as well as the individual's age t . A simple incidence model with calendar time effects might be written:

$$\lambda_i(t, u|x_i) = \varphi\psi(t)\eta(u) \exp\{\gamma_i + x_i(t)^T \beta\}. \quad (11)$$

Suppose now that $\eta(u) = \exp\{u\delta\}$, and that individual i is of age 0 at calendar time u_{0i} . Thus $u = t + u_{0i}$, and (11) may be re-written as a model with birth cohort effect u_{0i} :

$$\lambda_i(t|x_i) = \varphi\psi(t) \exp(t\delta) \exp\{\gamma_i + u_{0i}\delta + x_i(t)^T \beta\}$$

Now set $\psi^*(t) = \psi(t) \exp(t\delta)$ and $\gamma_i^* = \gamma_i + u_{0i}\delta$. In the semi-parametric case series analysis, the term γ_i^* drops out and the $\exp(t\delta)$ term is absorbed into the baseline incidence function, with $\log\{\psi^*(s_r)\} = \alpha_r^* = \alpha_r + s_r\delta$. Thus the time dependence disappears and, in effect, we fit the log-linear model

$$\log(\lambda_{ir}) = \gamma_i^* + \alpha_r^* + x_i(s_r)^T \beta \quad (12)$$

which is of the same form as (8). In other words, the semi-parametric case series model adjusts implicitly for exponential calendar time trends in the baseline incidence. The estimated baseline incidence function, however, now reflects both age effects and exponential calendar time trends. More generally, the case series model is robust to monotone calendar time trends in the baseline incidence. Departures from exponential time trends, and non-monotone variation such as seasonality, can be modelled explicitly by incorporating further parametric terms in the model.

The exposure covariate $x_i(t)$ can include external age-dependent categorical or quantitative variables. The values of any such variables are only required at the event times $s_r, r = 1, \dots, M$. The case series method is best suited to the investigation of age-dependent exposures that are transient (if categorical) or non-monotone (if quantitative), since long-lasting or monotone effects may be confounded to a greater

or lesser extent with age. If it is required to estimate the effect of long-lasting or monotone exposures, then unexposed cases may be required, that is, cases for which $x_i(t) = 0$ for all $t \in (a_i, b_i]$, in order to separate out age and exposure effects.

3.4 Checking assumptions

The validity of the assumptions can sometimes be assessed by tests based on fitting expanded models. For example, the case series model requires, rather fundamentally, that covariates act multiplicatively on the baseline incidence. This applies both to age-independent covariates Z and age-dependent exposures $X(t)$. Some indication of the validity of the multiplicativity assumption may be obtained by fitting an interaction between the covariate and a parametric function of age, typically age itself or a step function of age. Thus, the model is expanded to include terms $Z.t$ or $X(t).t$ and the likelihood ratio test is used to assess their significance, as originally suggested for the proportional hazards model (Cox 1972a).

A further assumption of the method is that exposures should not depend on prior events. We present a simple test of this assumption, applicable to a binary exposure. The issue of interest is whether occurrence of an event delays or brings forward subsequent exposures. In some circumstances it might be reasonable to assume that, were such an effect to occur, it would be restricted to some time period τ after the event. A simple way of testing this is by including in the model an additional age-dependent covariate $z(t)$ that takes the value 1 in pre-exposure periods of duration τ and 0 elsewhere: if occurrence of an event delays or brings forward subsequent exposures, then the relative incidence in the pre-exposure period will be < 1 or > 1 , respectively. If there is no such effect then the relative incidence in the pre-exposure period should be 1. Proposition 2 formalises this idea in the case of univariate point exposures and rare adverse events.

Proposition 2 Suppose that, for each individual i , adverse events and exposures arise in counting processes with intensity processes $\varphi_i \psi(t) \exp\{x_i(t)\beta\}$ for the events and $\mu(t) \exp\{y_i(t)\xi\}$ for the exposures, where $y_i(t) = 1$ if an event occurred in $(t - \tau, t)$ and $y_i(t) = 0$ otherwise. Suppose furthermore that the probability of occurrence of two or more adverse events in an age interval of length τ is negligible. Then the likelihood contribution of an individual i with n_i adverse events in an observation period $(a_i, b_i]$,

conditionally on n_i and the realized exposures, is approximately

$$L_i^{c*} \simeq \prod_{j=1}^{n_i} \frac{\psi^*(t_{ij}) \exp\{x_i(t_{ij})\beta + z_i(t_{ij})\xi\}}{\int_{a_i}^{b_i} \psi^*(t) \exp\{x_i(t)\beta + z_i(t)\xi\} dt} \quad (13)$$

where $z_i(t)$ is the number of exposures experienced by individual i in $(t, t + \tau)$ and $\psi^*(t) = \psi(t) \exp\left\{-\int_{a_i}^{b_i} \exp\{\xi I_{(t, t+\tau)}(s)\} \mu(s) ds\right\}$.

Outline proof Suppress individual markers i for clarity. The joint density of event times $t \equiv t_1, \dots, t_n$ and exposure times $v \equiv v_1, \dots, v_m$ is

$$\begin{aligned} f(n, m, t, v) &= \prod_{j=1}^n \left\{ \varphi \psi(t_j) e^{x(t_j)\beta} \right\} \exp\left\{ -\int_a^b \varphi \psi(u) e^{x(u)\beta} du \right\} \\ &\quad \times \prod_{k=1}^m \left\{ \mu(v_k) e^{y(v_k)\xi} \right\} \exp\left\{ -\int_a^b \mu(u) e^{y(u)\xi} du \right\}. \end{aligned} \quad (14)$$

Let $z(t) = \sum_{k=1}^m I_{(v_k - \tau, v_k)}(t)$. Note that $\sum_{j=1}^n z(t_j) = \sum_{k=1}^m y(v_k)$. Also, if $t_j - t_{j-1} \geq \tau$ for $j = 2, \dots, n$ then

$$\int_a^b \mu(u) e^{y(u)\xi} du = \sum_{j=1}^n \int_a^b \mu(u) \exp\{\xi I_{(t_j, t_j + \tau)}(u)\} du - (n-1) \int_a^b \mu(u) du. \quad (15)$$

Substituting these expressions in (14) and integrating out t_1, \dots, t_n yields

$$\begin{aligned} f(n, m, v) &\simeq \frac{\varphi^n}{n!} \left\{ \int_a^b \psi^*(u) e^{\beta x(u) + \xi z(u)} du \right\}^n \exp\left\{ \int_a^b \mu(u) du \right\}^{n-1} \\ &\quad \times \prod_{k=1}^m \mu(v_k) \exp\left\{ -\int_a^b \varphi \psi(u) e^{x(u)\beta} du \right\}. \end{aligned} \quad (16)$$

The approximation results from the fact that (15) is assumed to hold even if $t_j - t_{j-1} < \tau$ for some j . By assumption such realizations arise with negligible probability. Combining (14) and (16) and substituting (15) gives the approximate conditional density of t upon which the likelihood (13) is based. ■

The likelihood (13) is a case series likelihood. The case series method may thus be used to estimate the parameters β and ξ , the latter parameter relating to the short-term inhibitory ($\xi < 0$) or excitatory ($\xi > 0$) effect of events on exposures. If $\xi = 0$ then the exposure process is a non-homogeneous Poisson process with rate $\mu(t)$. The additional covariate $z(t)$ counts the number of exposures between t and $t + \tau$. Usually, τ is small so there is seldom more than one exposure in $(t, t + \tau)$, and so including a pre-exposure period of duration τ is the same as fitting $z(t)$. Note that $\psi^*(t)$ differs from the age-specific relative incidence $\psi(t)$.

4 Inference

The hypothesis $\beta = 0$ may be tested using the standard likelihood ratio test. Approximate confidence intervals and standard errors for β may be obtained from the profile log likelihood, or from the Fisher information derived from (7) for the model treated as finite-dimensional. The profile likelihood for β is particularly simple to calculate from model (8). These methods are justified below.

The semi-parametric model also provides an empirical estimate of the baseline cumulative relative incidence function $\Psi(t)$, namely $\Sigma \exp(\hat{\alpha}_r) I_{(-\infty, s_r]}(t)$, for $t \in (a, b]$. The $\hat{\alpha}_r$ can also be used to estimate the age-specific relative incidence function $\psi(t)$. However, in view of (12), the parameters α_r represent the combined effects of age and exponential time trends, and hence should be interpreted with some caution. As noted previously, however, the primary focus of inference is usually β , not $\Psi(t)$.

4.1 Asymptotics

In the semi-parametric model, the parameter $\Psi(t)$ is infinite dimensional. Let $\|\cdot\|_\infty$ denote the supremum norm and $\|\cdot\|_2$ the L_2 norm, and let β_0 and Ψ_0 be the true parameter values, with $\beta_0 \in B$ a compact subset of R^p and Ψ_0 absolutely continuous and non-decreasing on $[a, b]$ with $\Psi(a) = 0$. The parameter n denotes the total number of events, Σn_i . For each case, $X : (a, b]^p \rightarrow C$ is the exposure function of uniformly bounded variation (with C a compact subset of R^p), and W is the observation period $W \equiv (a_W, b_W] \subseteq (a, b]$. We write $W(t) = I_{(a_W, b_W]}(t)$. Consistency and asymptotic normality and efficiency of the semi-parametric maximum likelihood estimators are established in the theorem below, which is proved using the following assumptions:

Assumptions

- (a1) With probability $q > 0$, $W = (a, b]$.
- (a2) There is a positive c such that for all W , $\int W(t) d\Psi_0(t) \geq c$.
- (b1) For any constant vector c_1 and function $f(t_1, t_2)$ not depending on X , if $P\{\forall t_1, t_2 \in (a, b], c_1^T [X(t_1) - X(t_2)] = f(t_1, t_2)\} = 1$ then f is identically zero and $c_1 = 0$.

Conditions (a1) and (a2) ensure sufficient overlap and width of the observation periods for the limiting processes in the proof to work, while condition (b1) ensures that there is sufficient within and between individual variation in exposures to identify the parameters and invert the information operator. In particular, condition (b1) is equivalent to the following:

(b2) For any constant vector c_1 and function $f(t)$ not depending on X , if $P\{\forall t \in (a, b], f(t) + c_1^T X(t) \text{ does not depend on } t\} = 1$ then f is constant and $c_1 = 0$.

Note that (b2) rules out exposure functions of the form $X(t) = h(t) + c_X$ as discussed in Subsection 2.4.

Theorem. Under the stated assumptions, the following hold asymptotically as the number of cases N and hence the number of events $n \rightarrow \infty$.

(a) The maximum likelihood estimators of β and Ψ are consistent: $\|\hat{\beta} - \beta_0\|_2$ and $\|\hat{\Psi} - \Psi_0\|_\infty$ converge almost surely to zero.

(b) If β_0 is an interior point of B , then $\sqrt{n}\|\hat{\Psi} - \Psi_0\|_\infty$ is bounded in probability and $\sqrt{n}(\hat{\beta} - \beta_0)$ converges in distribution to a normal distribution with mean zero and efficient variance.

(c) Let $\hat{\Psi}_\beta$ denote the value of Ψ that maximises $\log(L_{\mathfrak{S}})$ with β fixed. Under the null hypothesis $\beta = \beta_0$ the profile loglikelihood $2\{\log L_{\mathfrak{S}}(\hat{\beta}, \hat{\Psi}) - \log L_{\mathfrak{S}}(\beta_0, \hat{\Psi}_{\beta_0})\}$ is distributed χ_p^2 . ■

The proof is given in the Appendix. From part (c) of the theorem, inference may be based on the profile log likelihood, at least provided that it is approximately quadratic. Since the profile log likelihood for β is the same whether the model is regarded as finite or infinite dimensional, the variances of the parameter estimates may also be approximated by treating the model as if it were finite dimensional, as suggested in a different context by Huang & Rossini (1997). The validity of these approximations is investigated in the next subsection.

As noted above, if event times are measured in discrete units then, for all practical purposes, the semi-parametric model is equivalent to an associated model with $\psi(t)$ modelled as a step function with jumps at unit intervals. Since the maximum observation period $(a, b]$ is fixed, the model is finite dimensional and standard asymptotic theory applies, though its relevance in small samples requires confirmation.

Note also that the theorem can be generalised to situations in which the interval $(a, b]$ is replaced by a union of disjoint intervals.

4.2 Simulations

The performance of the semi-parametric model in samples of small to moderate size was investigated by simulation. In these simulations, the observation period ranges from 0 to 500 arbitrary time units. The exposure variables $x_i(t)$ are categorical, taking

the value 1 in $(c_i, c_i + d]$ and 0 elsewhere, where d is the post-exposure risk period. Every case is assumed to have been exposed at some point in the observation period. The c_i were sampled from a unimodal beta distribution on $(0, 500]$ with mean 250 and standard deviation 100.

Two contrasting scenarios were investigated for the age-specific relative incidence, represented in each case by a step function. Scenario A assumes symmetric age effects of mild amplitude, the age-specific relative incidence taking the values 1 (0.25) 2, 2(0.25)1 on successive 50-unit intervals. Scenario B assumes strong monotonic increasing age effects, the age-specific relative incidence taking the values 1 (0.5) 5.5 on successive 50-unit intervals.

For each scenario two post-exposure risk periods were considered: short ($d = 50$ units, 10% of total observation time) and long ($d = 250$ units, 50% of total observation time). Exposure-related relative incidences of 1, 2 and 5 over these risk periods were investigated, with sample sizes of 25, 50 and 100 cases.

For each of the 36 combinations of scenario, choice of parameters and sample size, 10000 independent datasets of cases were generated. In line with the conditioning upon which the case series method is based, we kept the number of events per individual fixed. We assumed that the underlying rate was very low, so that each case involves a single event. (Performance improves with more frequent events.) We then randomly allocated event times using the cumulative distribution function $F_i(t) = \int_0^t \lambda_i(s|x_i)ds / \int_0^{500} \lambda_i(s|x_i)ds$. Event times were not grouped, and hence are all distinct (similar results were obtained when events were grouped into discrete time units). Each simulated dataset was analysed by the semi-parametric model. We randomly re-allocated event times 10000 times for each run, keeping the exposure distribution fixed in line with the conditioning underlying the case series model. (The exposure distribution thus varied randomly between runs, but not within runs.) For each run of 10000, the median of the relative incidence estimate $\exp(\hat{\beta})$ was calculated. We used the median rather than the arithmetic or geometric mean as there is necessarily a non-zero probability that $\hat{\beta} = \pm\infty$. We also calculated the coverage probability of the approximate 95% confidence interval $\exp(\hat{\beta} \pm 1.96\hat{s})$ where \hat{s} is the estimated standard error obtained from the Fisher information based on (7), for the model treated as if it were finite-dimensional. The results are presented in Tables 1 and 2. Similar results were obtained for the two scenarios, and for the two risk periods. For sample size 25 there is evidence of bias of about 10% at relative incidence 5. By sample size 50 the bias is of no material importance. The coverage probability of the

confidence intervals is close to 95%, though generally slightly conservative.

Table 1 Simulations under scenario A

Relative incidence		Sample size					
		Risk period 50			Risk period 250		
		25	50	100	25	50	100
1	M	0.947	0.979	0.983	1.024	1.000	0.997
	P95	0.969	0.965	0.956	0.955	0.949	0.952
2	M	1.993	1.984	1.982	2.122	2.027	2.007
	P95	0.963	0.956	0.952	0.956	0.951	0.951
5	M	5.410	5.094	4.968	5.549	5.098	5.004
	P95	0.947	0.950	0.948	0.964	0.955	0.945

M: median; P95: coverage of 95% CI

Table 2 Simulations under scenario B

Relative incidence		Sample size					
		Risk period 50			Risk period 250		
		25	50	100	25	50	100
1	M	0.934	0.967	0.995	1.012	1.000	1.000
	P95	0.970	0.965	0.955	0.954	0.952	0.950
2	M	1.964	1.986	1.979	2.099	2.030	2.004
	P95	0.962	0.953	0.953	0.963	0.954	0.953
5	M	5.365	5.066	4.983	5.459	5.187	5.079
	P95	0.950	0.947	0.948	0.966	0.965	0.950

M: median. P95: coverage of 95% CI

The bias in the log relative incidence (calculated as $\text{median}(\hat{\beta}) - \beta$) was typically an order of magnitude smaller than the median of the standard errors of $\hat{\beta}$ (not shown). These results suggest that standard asymptotic theory may validly be used even in studies of moderate size.

5 Applications

The first two applications relate to adverse events putatively associated with the measles, mumps and rubella (MMR) vaccine, which was introduced in the UK in 1988. The first relates to recurrent events and transient exposures, the second to non-recurrent events and long-lasting exposures. In the third application, to asthma and

air pollution, we propose a model that extends the time-stratified case-crossover model of Lumley & Levy (2000). Finally, in the fourth application, to respiratory infections and myocardial infarction, we investigate robustness to failure of some of the modelling assumptions and illustrate methods for checking these assumptions.

5.1 Bleeding disorders

Idiopathic thrombocytopenic purpura (ITP) is an uncommon, potentially recurrent bleeding disorder. In rare instances, MMR vaccine is thought to cause ITP. The putative effect, if present, is caused by the live vaccine virus and is transient. ITP is not a contra-indication to vaccination, hence vaccination rates should be unaffected by prior ITP.

The data relate to 35 children admitted to hospital with ITP when aged between 1 and 2 years (366 to 730 days of age, to be precise). Twenty nine children suffered a single event, five suffered 2 events and one suffered 5 events. This distribution suggests perhaps that children vary in their propensity to develop ITP; the case-series method allows for such variation as it is self-controlled. Of the 44 events, 13 occurred within 6 weeks of receipt of MMR vaccine. In this analysis, the variable $x_i(t)$ is a 2-level factor, taking level 2 during the period 0-42 days after MMR vaccination (day 0 is the day of MMR) and 1 (the reference level) at all other times. All 35 children received MMR vaccine, but only 31 received it on days 324 to 730 and hence were exposed during the period 366 to 730 days. The histogram of age at vaccination for these 31 is shown in Figure 1(a).

Table 3 Relative incidence of ITP after MMR vaccine

<i>Risk period (days)</i>	<i>Events</i>	<i>Relative incidence (95% CI)</i>	
Unexposed	31	1	(-)
0-42 post MMR	13	3.01	(1.38, 6.54)

The semi-parametric model gave the results shown in Table 3. The likelihood ratio test statistic for the vaccine effect is 7.09 on 1 degree of freedom, $p = 0.008$. The log-likelihood profile for the parameter corresponding to the 0-42 day risk period is shown in Figure 1(b). This is roughly quadratic; the profile likelihood 95% confidence interval is (1.36, 6.48), similar to that given in Table 3. Figure 1(c) shows the estimated cumulative baseline incidence, relative to that at age 374 days.

The occurrence of a child with 5 episodes might suggest that the risk of subsequent episodes of ITP increases after the first. The analysis was therefore repeated using

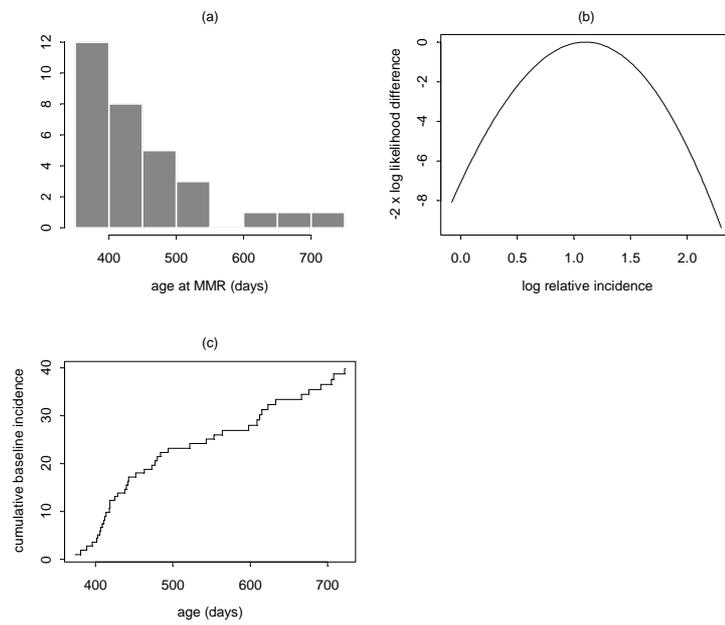


Figure 1: ITP and MMR vaccination. (a) Histogram of age at MMR vaccination in exposed cases; (b) profile log likelihood for log relative incidence; (c) estimated cumulative relative baseline incidence.

only the 35 first episodes. The results were similar. They are shown in Table 4, along with those of an independent nested case-control study of first ITP (Black, Kaye & Jick 2003).

Table 4 First ITP and MMR vaccination within 42 days

<i>Study</i>	<i>Sample size</i>	$\exp(\hat{\beta})$	<i>95% CI</i>
Case series	35 cases	3.73	(1.60, 8.71)
Case-control	23 cases, 116 controls	6.3	(1.3, 30.1)

The case series estimate of the log relative risk is more precise than that from the nested case-control analysis, even allowing for the different numbers of cases. As suggested by equation (6), the case series design has high efficiency relative to the other designs since the proportion of individuals exposed is high and the risk period is short relative to the observation period.

5.2 Autism

In 1998 it was suggested that MMR vaccine may cause autism, a claim that provoked a long and damaging controversy. In this example a reanalysis is presented of a dataset on autism diagnosis and MMR vaccination, originally analysed with the parametric case series model (Taylor, Miller, Farrington, Petropoulos, Favot-Mayaud, Li & Waight 1999, Farrington, Miller & Taylor 2001). The dataset includes 357 cases of autism diagnosed in children aged up to 16 years of age born between 1979 and 1998. As the observation period substantially predates concerns over a potential role for MMR in autism, receipt of MMR vaccine may reasonably be regarded as unaffected by a prior diagnosis. Ages were recorded by month, and the maximum observation period stretches from month 1 to month 192 of age.

Autism is a non-recurrent event, but is rare, and hence the case series method may be applied. In response to claims that the vaccine may cause autism long after vaccination, an indefinite post-MMR risk period is used: thus an individual is considered to be potentially at increased risk at all times after receipt of the vaccine. The presence of 64 unvaccinated cases makes it possible separately to identify the age and vaccine effects in older age groups. In the original analysis, the age effect was parameterised as piecewise constant in 17 age groups, and the estimated relative incidence of autism after MMR was 1.24, 95% CI (0.67, 2.27).

In view of possible confounding between age and exposure effects, it is desirable to avoid parametric assumptions. Accordingly, the analysis was repeated with the semi-

parametric model. The exposure variable $x_i(t)$ was a time-varying factor with 5 levels as follows: level 1 prior to MMR vaccination, level 2 over the period 0-24 months after MMR, level 3 over the period 25-48 months post-MMR, level 4 over the period 49-72 months post MMR, and level 5 over the period 72+ months post MMR. The results are shown in Table 5. Clearly, it is sensible to combine all post-MMR risk periods. This yields a relative incidence of 0.882, 95% CI (0.399, 1.95). Thus the parametric model with 17 age groups may have slightly exaggerated the relative incidence, though the substantive conclusions are unaffected. The profile log likelihood for the log relative incidence post-MMR is shown in Figure 2(a), and appears roughly quadratic. The profile likelihood-based 95% confidence interval is (0.394, 1.94).

Table 5 Relative incidence of autism after first MMR vaccine

<i>Risk period (months)</i>	<i>Events</i>	<i>Relative incidence (95% CI)</i>	
Unexposed	94	1	(-)
0-24 post MMR	131	0.892	(0.400, 1.99)
25-48 post MMR	109	0.755	(0.310, 1.84)
49-72 post MMR	17	0.849	(0.275, 2.63)
72+ post MMR	6	0.903	(0.236, 3.45)

Autism is gender related, being more prevalent in boys. It is therefore of interest to investigate whether the effect of MMR vaccine is gender-related. There were 291 males and 66 females. The likelihood ratio statistic for the interaction between gender and vaccine effect was 0.0030 on 1 degree of freedom, not suggestive of any difference in the effect of MMR between males and females.

Figure 2(b) shows the estimated cumulative baseline incidence, relative to its value at age 18 months. Figure 2(c) shows the estimated baseline relative incidence. There is a peak around 40 months, and thereafter a steep linear trend. This trend not an age-related effect but a temporal one, as shown in Figure 2(d). Here we have plotted the logarithms of the counts of cases diagnosed before age 60 months, by year of birth, 1979 to 1992. We truncated the data in this way, and used data only to 1992, so as to avoid bias in the later years. The full line is the regression line obtained from a Poisson log-linear fit to these counts (slope 0.22). The dashed line is that obtained from the upper tail of the display in panel (c) (slope 0.25): the two nearly coincide. Panel (c) thus illustrates the combined effect of age and linear time trends. As discussed in Subsection 3.3, the case series model allows for such effects, the presence of which does not bias the estimate of β .

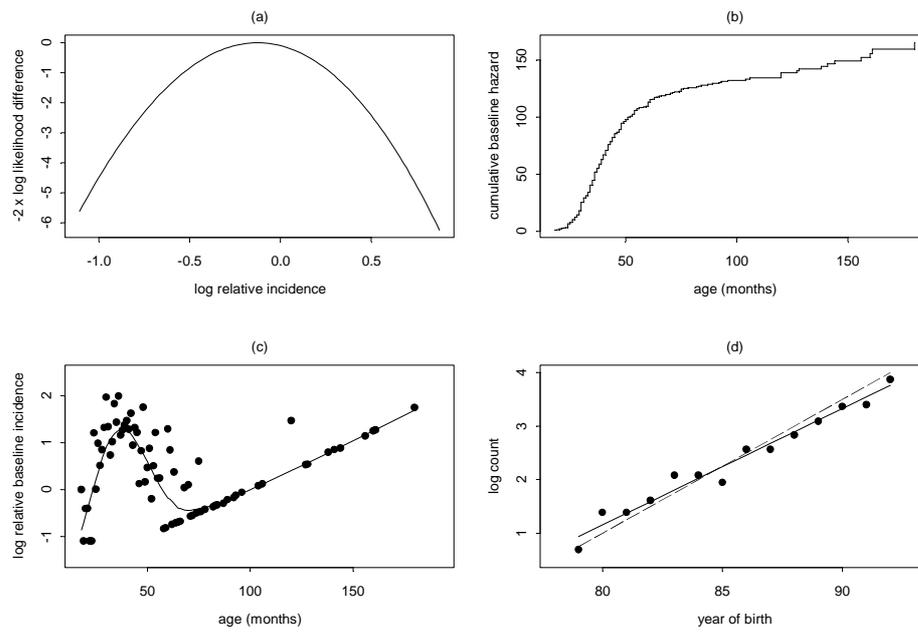


Figure 2: Autism diagnosis and MMR. (a): profile log likelihood for the log relative incidence. (b): estimated cumulative baseline relative hazard. (c): estimated baseline relative hazard with lowess curve. (d): logarithms of counts of autism cases diagnosed before 60 months of age by year of birth, with two regression lines (see text).

Several independent studies of autism at any time post-MMR have been conducted, including a cohort study (Madsen, Hviid, Vestergaard, Schendel, Wolfahrt, Thorsen, Olsen & Mellbye 2002) and a nested case-control study (Smeeth, Cook, Fombonne, Heavey, Rodrigues, Smith & Hall 2004). The three studies are summarized in Table 6.

Table 6 Three studies of autism and MMR

<i>Study</i>	<i>Sample size</i>	$\exp(\hat{\beta})$	<i>95% CI</i>
Case series	357 cases	0.88	(0.40, 1.95)
Cohort	537 303 children, 316 cases	0.92	(0.68, 1.24)
Case-control	1294 cases, 4469 controls	0.86	(0.68, 1.09)

The three studies give similar effect estimates. Even allowing for different numbers of cases, the case series estimate is less precise than those obtained by the cohort and case-control methods. In this instance, the case series method is appreciably less efficient than the cohort method because the risk period (which comprises all time after MMR vaccination) is very long relative to the observation period. Nevertheless, the case series study has very much smaller overall size, a factor which contributed to its rapid completion.

5.3 Air pollution and asthma

In this application we propose a new model for investigating the association between air pollution and disease. A feature of studies of air pollution undertaken at a single location is that all individuals have the same, or highly correlated, exposures, so that the effects of exposure are confounded with changes over time. A standard method of analysis of such data is the case-crossover method. However, logistic regression analysis of the case-crossover model may not yield consistent estimates (Lumley & Levy 2000, Vines & Farrington 2001). Several alternatives have been suggested. The full-stratum bi-directional case-crossover method of Navidi (1998) is a case series model with a constant baseline relative incidence of disease, that is $\psi(t) = 1$ over an observation period that may span several years. Since this is usually inappropriate, Lumley & Levy (2000) suggested that the model be used with short observation periods (for example, months), over which the baseline incidence could reasonably be assumed roughly constant. This is the time-stratified case-crossover method, which is also a case series model with $\psi(t) = 1$.

In the presence of strong seasonal effects, however, the assumption $\psi(t) = 1$ may be questionable even with short observation periods. We propose a modification that allows for testing for residual seasonal effects, and may reduce the sensitivity of the results to the choice of observation period.

Suppose that data on cases are available over a period which, for simplicity, we assume exactly spans K years, with $K \geq 2$. Each year is divided into L disjoint observation periods. For definiteness, we shall use months, so $L = 12$. The time line is time of the month. Let $x_{kl}(t)$ denote the vector of exposures at time t of month l in year k . Suppose that there are M_l distinct event times within month l over the K years, listed s_{l1}, \dots, s_{lM_l} in increasing order with $\mathfrak{S}_l = \{s_{l1}, \dots, s_{lM_l}\}$, and suppose that at time s_{lj} , n_{klj} events occurred in month l of year k . Then the case series likelihood (7) for month l is

$$L_{\mathfrak{S}_l} = \prod_{k=1}^K \prod_{j=1}^{M_l} \left(\frac{\exp\{\alpha_{lj} + x_{kl}(s_{lj})^T \beta\}}{\sum_{r=1}^{M_l} \exp\{\alpha_{lr} + x_{kl}(s_{lr})^T \beta\}} \right)^{n_{klj}} \quad (17)$$

where the α_{lj} are now log relative incidences within month l , $j = 1, \dots, M_l$. As shown in section 3.3, the model also implicitly adjusts for an exponential trend within month l common to each of the K years. Thus the model should be robust to monotone secular trends. The full likelihood for all months $l = 1, \dots, L$ is $L_{\mathfrak{S}} = \prod_{l=1}^L L_{\mathfrak{S}_l}$. This model thus adjusts for both seasonal and exponential secular variation in the incidence of the disease of interest. The model relies on replication over several years to identify the seasonal effects, whence the requirement $K \geq 2$.

We illustrate the method with data on air pollution and asthma from Nottingham. Particulate matter (PM₁₀) counts for central Nottingham were obtained for the period 11.09.1996 to 10.09.2004, together with daily admissions for asthma to one Nottingham hospital over the same period. We investigate whether PM₁₀ counts are associated with hospital admissions for asthma. The PM₁₀ data are shown in Figure 3, and the asthma data in Figure 4.

There were 3264 admissions for asthma over this 8-year period (0 on 36% of days, daily maximum 14). The exposure variable for a given day is the average of the PM₁₀ level on that day and the day before. We fitted the case series model (17) with and without seasonal adjustment. The model without seasonal adjustment is obtained by setting $\alpha_{jk} = 0$; this is the time-stratified case-crossover model of Lumley & Levy (2000). The results are shown in Table 7. The parameter $e^{10\beta}$ represents the relative incidence of asthma associated with an increase in average PM₁₀ count of ten μgm^{-3} .

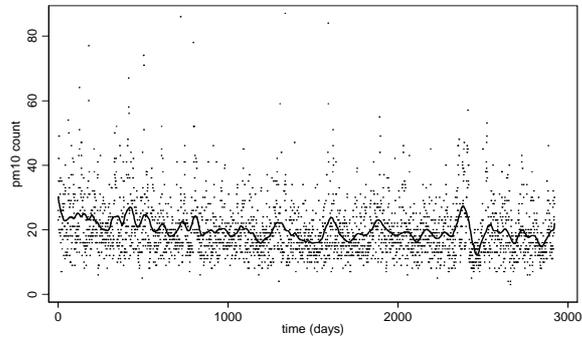


Figure 3: Daily PM₁₀ count (μgm^{-3}) in Central Nottingham, 11.09.96 - 10.09.04, and lowess smooth with 3 month bandwidth.

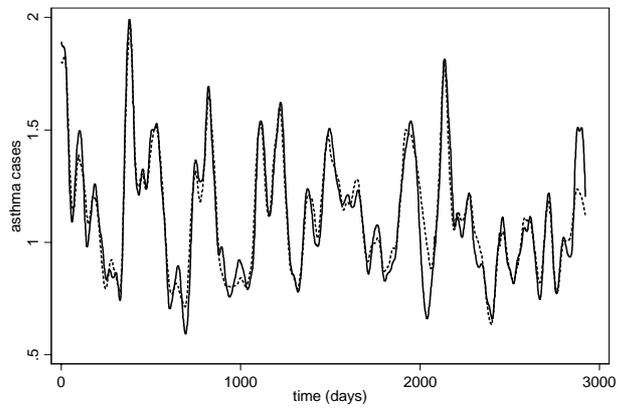


Figure 4: Solid line: smoothed daily hospital admissions for asthma. Dotted line: smoothed fitted values from the model with seasonality. Both smooths use a lowess with 3-month bandwidth.

Table 7 Asthma and PM₁₀ counts

Seasonal adjustment	No	Yes
$\exp(10\beta)$	1.007	0.997
95% CI	(0.958, 1.057)	(0.944, 1.051)

In these data, residual confounding by season is clearly not a problem. Allowing for residual seasonal variation does not alter the conclusion that there is little evidence that PM₁₀ levels are associated with asthma. In line with other publications on the topic, we also undertook an analysis using day of the week within months as the observation periods, in which case $L = 7 \times 12$. The results were similar.

In the environmental epidemiology literature, case-crossover methods are often contrasted with loglinear time series models in which temporal effects are modelled by smoothing techniques. We have argued elsewhere that this distinction is unhelpful (Whitaker & Farrington 2005). The case series model presented above is in fact equivalent to a loglinear model with the $K \times L$ month effects modelled as a step function, and non-parametric seasonal adjustment for time of year within months. The case series formulation really comes into its own when applied to cases with different exposures.

5.4 Respiratory infections and myocardial infarction

Chronic inflammation is believed to promote certain types of cardiovascular disease. This raises the question as to whether acute inflammation caused by infection also has a detrimental impact on the cardiovascular system. In this example, the question of interest is whether a respiratory infection increases the risk of myocardial infarction (MI). A large study of the impact of infections and vaccination on cardiovascular disease has recently been undertaken (Smeeth, Thomas, Hall, Hubbard, Farrington & Vallance 2004). We use a random subsample from this study to investigate assumptions for the case series model.

The data consist of 602 cases of first MI experienced by individuals aged between 65 and 75 years of age, together with the ages at which these individuals contracted acute respiratory tract infections. The data are taken from the General Practice Research Database (GPRD). An individual's observation period is the time registered with the GPRD between ages 65 and 75 years. Suppose that individual i is registered with the GPRD between ages A_i and B_i . Then the observation period is $(a_i, b_i]$ where $a_i = \max\{65, A_i\}$ and $b_i = \min\{75, B_i\}$, expressed in days.

The first issue we examine is the possible dependence of the observation period

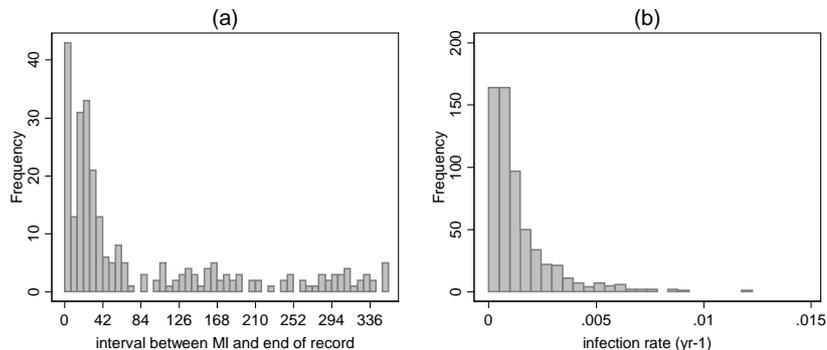


Figure 5: Myocardial infarction (MI) and respiratory tract infections. (a) Interval from MI to end of record (truncated at 365 days), (b) Individual infection rates

$(a_i, b_i]$ on the event of interest (first MI). Information on deaths was not available, but it is reasonable to surmise that a proportion of cases will have died immediately or shortly after their MI, and hence that the b_i are not independent of the event times t_i . Figure 5(a) shows the histogram of $B_i - t_i$ (truncated at 365 days). There is an excess of short intervals. In particular, 156 had $B_i - t_i \leq 42$ days. We shall assume that all of these cases died from their MI. This assumption is conservative: some will have died of other causes, including infection. Figure 5(b) shows the distribution of individual annual infection rates in the cases.

We investigate the robustness of the case series method to failure of the assumption that observation periods are independent of events by imputing data for the 156 cases with presumed MI-related deaths. The imputed data represent the post-MI infection history these cases might have experienced, had they not died from their MI. For each of the 156 cases, we first extended their GPRD registration up to 3 years after their MI, by randomly sampling an interval $u_i \sim U(B_i - t_i, 3 \text{ years})$ and replacing b_i by $b_i^* = \min\{B_i + u_i, 75\}$. We generated a Poisson number of post-MI infections, with rate $(b_i^* - b_i) \times \hat{\mu}_i$ where $\hat{\mu}_i$ is the observed infection rate of case i during their actual GPRD registration period $(A_i, B_i]$. The infection times were then independently sampled from the uniform densities $U(b_i, b_i^*)$ to simulate a Poisson process. Finally, we combined our 156 resurrected cases with the remainder of the data set and analysed the data using the case series method with risk periods 1-14 and 15-91 days after each infection. We repeated the imputation 5000 times. (In view of the computational burden involved,

we used a parametric model with two 5-year age groups; the estimates from the original data are similar to those obtained with the semi-parametric model.) The results are shown in Table 8.

Table 8 Impact of dependent censoring

<i>Exposure period</i>	<i>Actual data</i>		<i>Imputed data</i>			
	$\hat{\beta}$	$SE(\hat{\beta})$	Mean $\hat{\beta}$	SD $\hat{\beta}$	Mean $SE(\hat{\beta})$	SD $SE(\hat{\beta})$
1-14 days	1.3180	0.1761	1.3518	0.0155	0.1766	0.0007
15-91 days	0.1928	0.1445	0.2034	0.0122	0.1443	0.0003

The estimates in the *Actual data* columns are the values obtained by applying the case series method to the 602 cases, ignoring the possible effect of MI-dependent censoring. Those in the *Imputed data* columns relate to the 5000 imputed datasets without MI-dependent censoring. The imputed values of $\hat{\beta}$ were approximately normally distributed, with small standard deviation (SD $\hat{\beta}$ in Table 8). With these data, dependent censoring results in a small negative bias, which reduces the relative incidence for the 1-14 day period by a factor of 0.97, and the relative incidence in the 15-91 day period by a factor of 0.99. Bias of this magnitude is of no practical importance. Note also that the standard errors are little affected. Overall, these results suggest that, for these data, the case series method is robust to failure of the assumption that observation periods are independent of event times.

Next, we investigate the assumption that infections occur independently of prior MIs. Arguments, plausible *a priori*, can be made that occurrence of an MI could affect the short-term infection rate, either by increasing it (through reduced resistance) or decreasing it (through reducing the chance of infectious contacts). We use Proposition 2 to estimate the relative rate of infection occurring within $\tau = 14, 31$ and 91 days of an MI, and to investigate whether allowing for this affects the relative incidence estimates in the post-infection period. The results for $\tau = 91$, obtained using the semi-parametric model, are shown in Table 9. The covariate $z(t)$ counts the number of infections in $(t, t + \tau)$. There is little evidence that the infection rate is altered in the 91 days following an MI (non-significant effects were also observed for $\tau = 14$ and $\tau = 31$). Furthermore, allowing for such an effect has only a marginal impact on the estimated effect of infections on MI. We conclude that there is little evidence to doubt the validity of the assumption that exposures are independent of prior events, and in any case the results are robust to violation of this assumption.

Table 9 Impact of event-dependent exposures

<i>Exposure variable or period</i>	<i>Relative incidence (95% CI)</i>	
	<i>Without $z(t)$</i>	<i>With $z(t)$</i>
$z(t)$	-	0.868 (0.663, 1.136)
1-14 days	3.810 (2.675, 5.427)	3.748 (2.627, 5.347)
15-91 days	1.227 (0.920, 1.637)	1.215 (0.910, 1.621)

Finally, we investigate the validity of the proportional incidence assumption. We augment the model with an exposure \times age interaction, where age is a continuous covariate. Specifically we fit the model

$$\lambda_i(t|x_i) = \varphi\psi(t) \exp\{\gamma_i + x_i(t)^T\beta + (t - \bar{t}).x_i(t)^T\delta\}$$

where \bar{t} is the average age over the observation period. The augmented model includes two interaction terms, one for each exposure period. The estimates of the interaction parameters δ suggest a reduction in the 1-14 day relative incidence by the factor 0.977 (95% CI 0.859 to 1.111) per year of age, and an increase in the 15-91 day relative incidence by the factor 1.081 (95% CI 0.971 to 1.202) per year of age. The likelihood ratio test statistic was 2.438 on 2 degrees of freedom. We conclude that there is no compelling evidence for failure of the proportional incidence assumption.

Overall, the assumptions either hold or results are robust to their failure. From Table 9, the substantive conclusion from this analysis is that the risk of MI is substantially increased in the fortnight following a respiratory tract infection.

6 Discussion

In this paper the case series model has been extended to cope with arbitrary exposures, in a semi-parametric framework that avoids the need explicitly to specify the age-specific baseline incidence.

The case series model has two main advantages: it involves only cases, and controls for age-independent confounders that act multiplicatively on the baseline incidence. Requiring only cases makes the method attractive for use with certain types of data such as registry data and hospital admission databases. It is also convenient when the population from which the cases are generated is not clearly defined. Controlling confounding is an advantage when known confounders are approachable only via proxy variables, as is the case with socio-economic factors and underlying state of health, or when confounders are unknown or unmeasured. In addition, the semi-parametric

model also implicitly controls for exponential time trends in the incidence of the events of interest.

Nevertheless, the method requires some assumptions which are stronger than those needed for cohort studies. Perhaps the most restrictive assumptions in practice are that exposures must be external time-varying covariates, and that observation periods are independent of event times. If these or other assumptions do not hold, other designs, including case-control, nested case-control, case-cohort and other risk-set sampling methods (Prentice 1986, Borgan, Goldstein & Langholz 1995) can be used even if a full cohort study is too onerous to undertake. In some instances it may also be possible to undertake a ‘nested’ case series analysis based on the cases from these designs, thus helping to throw further light on the relative merits of the different approaches, and on the extent of unmeasured confounding.

7 Acknowledgements

We thank Liz Miller and Nick Andrews (HPA Centre for Infections) for the ITP and autism data, Richard Hubbard and Joe West (University of Nottingham) for the asthma data, and Liam Smeeth and Sara Thomas (London School of Hygiene and Tropical Medicine) for the MI data. We also thank the referees for their helpful suggestions. This work was supported by Wellcome Trust project grant 070346.

Appendix

The proofs of parts (a) and (b) of the theorem follow broadly the same strategy as those in Murphy, Rossini & Van der Vaart (1997) and Parner (1998). We therefore emphasize points that are specific to the case series model. We use the notation $Pf = \int fdP$ and let P_n denote the empirical distribution of the data and P_0 the true distribution with β_0, Ψ_0 the true parameter values. To reduce clutter we suppress individual labels throughout. Let T denote the event time, X the exposure function, and W the indicator function for the observation period (see Subsection 4.1). Note that, conditionally on (X, W) multiple event times for one individual are independent.

Consistency

Let $Y = (X, T, W)$ for one event. The log likelihood comprises terms of the form

$$l(\beta, \Psi; Y) = \log(\Delta\Psi(T)) + X(T)^T\beta - \log \int_a^b W(t) \exp(X(t)^T\beta) d\Psi(t). \quad (18)$$

Because $\Psi(t)$ represents a relative effect, it is not identifiable without some further constraint. We shall require $\Psi(b) = 1$ (in applications a different constraint is used, namely $\Delta\Psi(T_{(1)}) = 1$). This constraint applies to the estimates as well: thus $\widehat{\Psi}(b) = 1$.

$\widehat{\Psi}(t)$ is a sequence of non-decreasing functions uniformly bounded above and below and hence, by Helly's theorem, there exists a subsequence converging weakly to a non-decreasing function $\Psi^*(t)$. Since the $\widehat{\Psi}(t)$ are step functions on a compact interval, convergence is uniform in the supremum norm. By the compactness of B , there is a further subsequence such that $\widehat{\Psi}(t) \rightarrow \Psi^*(t)$ and $\widehat{\beta} \rightarrow \beta^*$ for some value β^* .

The following representation for $\widehat{\Psi}(t)$ may be derived by differentiating the log likelihood with respect to the increments $\Delta\Psi(T)$:

$$\widehat{\Psi}(t) = \int_a^t \frac{dF_n(u)}{V_n(u; \widehat{\Psi}, \widehat{\beta})}$$

where

$$V_n(u; \Psi, \beta) = P_n \left\{ \frac{W(u) \exp(X(u)^T \beta)}{\int_a^b W(t) \exp(X(t)^T \beta) d\Psi(t)} \right\}, \quad F_n(u) = P_n \{I_{(a, u]}(T)\}.$$

Also define the step function

$$\widetilde{\Psi}(t) = \int_a^t \frac{dF_n(u)}{V_n(u; \Psi_0, \beta_0)}.$$

This is well defined for sufficiently large n : $V_n(u; \Psi_0, \beta_0)$ is bounded away from 0 and ∞ as $n \rightarrow \infty$; formally we should normalise $\widetilde{\Psi}(t)$ so that $\widetilde{\Psi}(b) = 1$, but the normalising constant cancels out in all expressions.

The functions $W(u) e^{X(u)^T \beta_0} \left(\int W(s) e^{X(s)^T \beta_0} d\Psi_0(s) \right)^{-1}$ form a Donsker class and hence $V_n(u; \Psi_0, \beta_0) \rightarrow V_0(u; \Psi_0, \beta_0)$ uniformly in u , and, using the dominated convergence theorem, $\|\widehat{\Psi} - \Psi_0\|_\infty \rightarrow 0$.

We now show that $V_n(u; \widehat{\Psi}, \widehat{\beta})$ is uniformly bounded away from 0 and ∞ for sufficiently large n . Clearly, by the compactness of B and C , there is a positive constant m such that $W(u) e^{X(u)^T \widehat{\beta}} \left(\int W(s) e^{X(s)^T \widehat{\beta}} d\widehat{\Psi}(s) \right)^{-1} \geq mW(u)$. The functions $W(u)$ form a Glivenko-Cantelli class and hence, as $n \rightarrow \infty$, $V_n(u; \widehat{\Psi}, \widehat{\beta})$ tends uniformly to a limit $\geq mP\{W(u) = 1\} \geq mq$.

Suppose now that $V_n(u; \widehat{\Psi}, \widehat{\beta})$ is not uniformly bounded above. Then, with positive probability, there is a sequence t_n such that $V_n(t_n; \widehat{\Psi}, \widehat{\beta}) \rightarrow \infty$, and hence a sequence W_n such that $\int W_n(s) d\widehat{\Psi}(s) \rightarrow 0$. By the compactness of $[a, b]$ there is a subsequence

W_{n_k} whose endpoints tend to limits defining an interval $W_0 = (a_0, b_0] \subset (a, b]$ such that $\int W_0(s)d\widehat{\Psi}(s) \rightarrow 0$. Since $\int W_{n_k}(s)d\Psi_0(s) \geq c$ and Ψ_0 is absolutely continuous, $\int W_0(s)d\Psi_0(s) \geq c$ and hence $\int W_0(s)dP_0(s) = \int W_0(s)V_0(s; \Psi_0, \beta_0)^{-1}d\Psi_0(s) > 0$ since $V_0(s; \Psi_0, \beta_0)$ is bounded above. Let A be the union of all intervals W_0 that are the limits of sequences W_n such that $\int W_n(s)d\widehat{\Psi}(s) \rightarrow 0$. Since $\int_a^b d\widehat{\Psi}(s) = 1$, $\overline{A} = (a, b] - A \neq \emptyset$.

We show that existence of such a set A leads to a contradiction. Consider the difference of the mean log likelihoods evaluated at $(\widehat{\beta}, \widehat{\Psi})$ and at $(\beta_0, \widetilde{\Psi})$, which is non-negative, and the contribution of observation i , $T = t_i$, $W = W_i$, $X = X_i$. Suppose that, in the sample of n events, a total m_i event times $s_1, \dots, s_{m_i} \in W_i$ (including t_i) and, without loss of generality, assume that $\widehat{\Psi}(s_1) \geq \widehat{\Psi}(s_j)$, $j = 1, \dots, m_i$. The contribution of observation i to the mean log likelihood difference may be written

$$\begin{aligned} \widehat{l}_i - \widetilde{l}_i &= n^{-1} \left\{ \log \left(\frac{\Delta \widehat{\Psi}(t_i)/\Delta \widehat{\Psi}(s_1)}{\Delta \widetilde{\Psi}(t_i)/\Delta \widetilde{\Psi}(s_1)} \right) + X_i(t_i)^T(\widehat{\beta} - \beta_0) \right. \\ &\quad \left. - \log \left(\frac{\int W_i(s)e^{X_i(s)^T \widehat{\beta}} d\widehat{\Psi}(s)/\Delta \widehat{\Psi}(s_1)}{\int W_i(s)e^{X_i(s)^T \widetilde{\beta}} d\widetilde{\Psi}(s)/\Delta \widetilde{\Psi}(s_1)} \right) \right\} \end{aligned}$$

The first term on the right may be written

$$\log\{V_n(s_1; \widehat{\Psi}, \widehat{\beta})/V_n(t_i; \widehat{\Psi}, \widehat{\beta})\} - \log\{V_n(s_1; \Psi_0, \beta_0)/V_n(t_i; \Psi_0, \beta_0)\}.$$

The third term on the right is dominated by

$$-X_i(s_1)^T \widehat{\beta} + \log \sum_j \exp(X_i(s_j)^T \beta_0) V_n(s_1; \Psi_0, \beta_0)/V_n(s_j; \Psi_0, \beta_0).$$

By the compactness of B and C , and since $V_n(t; \Psi_0, \beta_0)$ is uniformly bounded away from 0 and ∞ , we have, for all i and some constant R ,

$$\begin{aligned} \widehat{l}_i - \widetilde{l}_i &\leq n^{-1} \left\{ \log \left(V_n(s_1; \widehat{\Psi}, \widehat{\beta})/V_n(t_i; \widehat{\Psi}, \widehat{\beta}) \right) + R \right\} \\ &\leq n^{-1} R. \end{aligned}$$

The ratio $V_n(s_1; \widehat{\Psi}, \widehat{\beta})/V_n(t_i; \widehat{\Psi}, \widehat{\beta})$ can be made arbitrarily small by selecting n sufficiently large, $W_i = (a, b]$, $t_i \in A$ and $s_1 \notin A$. This combination arises with positive probability by the construction of A , the assumption that $W = (a, b]$ with positive probability, and the independence of T and W . Thus, as $n \rightarrow \infty$, the mean log likelihood difference eventually becomes negative, almost surely. This is impossible, hence $V_n(u; \widehat{\Psi}, \widehat{\beta})$ is uniformly bounded away from ∞ .

Now return to the mean difference of the log likelihoods. We have

$$P_n\{l(\widehat{\beta}, \widehat{\Psi}; Y) - l(\beta_0, \widetilde{\Psi}; Y)\} \geq 0.$$

Since $V_n(u; \widehat{\Psi}, \widehat{\beta})$ is uniformly bounded away from 0 and ∞ for sufficiently large n , we can apply a uniform law of large numbers (for example, Proposition 3 of Parner (1998), or the argument of Murphy et al. (1997)) to show that the mean log likelihood difference converges to $P_0\{l(\beta^*, \Psi^*; Y) - l(\beta_0, \Psi_0; Y)\} \geq 0$ a.s. This is minus the Kullback-Leibler divergence, and therefore must also be non-positive. Hence $P_0\{l(\beta^*, \Psi^*; Y) - l(\beta_0, \Psi_0; Y)\} = 0$. Note also that

$$\Psi^*(t) = \int_a^t \frac{dF(u)}{V^*(u)}, \quad \Psi_0(t) = \int_a^t \frac{dF(u)}{V_0(u; \Psi_0, \beta_0)}$$

where $V^*(u)$ is the limit of $V_n(u, \widehat{\Psi}, \widehat{\beta})$ and $F(u) = P_0\{I_{(a, u]}(T)\}$. Hence Ψ^* is absolutely continuous.

Finally we prove that the parameters are identifiable, and hence $\beta^* = \beta_0$ and $\Psi^* = \Psi_0$.

Suppose that Ψ and β are such that $l(\beta, \Psi; Y) = l(\beta_0, \Psi_0; Y)$ a.s., with Ψ absolutely continuous. Let $D(\Psi, \Psi_0)(t) = d\Psi(t)/d\Psi_0(t)$. Then from (18),

$$\log \left\{ \frac{D(\Psi, \Psi_0)(t_1)}{D(\Psi, \Psi_0)(t_2)} \right\} = [X(t_1) - X(t_2)]^T (\beta - \beta_0)$$

for all t_1, t_2 . Hence, since the left hand side does not depend on X , it follows by assumption (b1) that $\beta = \beta_0$. Rearranging (18) again and using $\beta = \beta_0$ gives

$$\frac{d\Psi(t)}{\int_a^b W(s) e^{X(s)^T \beta_0} d\Psi(s)} = \frac{d\Psi_0(t)}{\int_a^b W(s) e^{X(s)^T \beta_0} d\Psi_0(s)}.$$

Integrating both sides from a to t

$$\frac{\Psi(t)}{\Psi_0(t)} = \frac{\int_a^b W(s) e^{X(s)^T \beta_0} d\Psi(s)}{\int_a^b W(s) e^{X(s)^T \beta_0} d\Psi_0(s)}$$

for all t . Since $\Psi(b)/\Psi_0(b) = 1$ it follows that $\Psi(t) = \Psi_0(t)$.

Since every subsequence of $(\widehat{\beta}, \widehat{\Psi})$ contains a subsequence converging uniformly to (β_0, Ψ_0) , the original sequence also converges uniformly, and so $(\widehat{\beta}, \widehat{\Psi})$ is consistent.

Asymptotic normality

Suppose that β has dimension p . Let h_1 be a bounded non-negative function on $[a, b]$. For ϵ in a fixed neighbourhood of zero, define

$$d\Psi_\epsilon = \frac{(1 + \epsilon h_1)d\Psi}{\int_a^b (1 + \epsilon h_1(u))d\Psi(u)}.$$

The integral in the denominator is there simply to ensure that $\Psi_\epsilon(b) = 1$, but plays no other substantive role. Since h_1 is bounded, Ψ_ϵ is a valid cumulative relative incidence function for sufficiently small ϵ . The score operator $l_{1\beta\Psi}$ for Ψ in the directions h_1 is obtained by differentiating the log likelihood with respect to ϵ , yielding

$$\begin{aligned} l_{1\beta\Psi}(Y)[h_1] &= h_1(T) - \frac{\int_a^b W(s)e^{X(s)^\top \beta} h_1(s) d\Psi(s)}{\int_a^b W(s)e^{X(s)^\top \beta} d\Psi(s)} \\ &= h_1(T) - E[h_1(T)|W, X; \beta, \Psi] \end{aligned}$$

where the expectation is taken with respect to T , conditionally on W, X . Similarly for $h_2 \in R^p$ the score operator for β in the direction h_2 is

$$\begin{aligned} h_2^T l_{2\beta\Psi}(Y) &= h_2^T \left\{ X(T) - \frac{\int_a^b W(s)e^{X(s)^\top \beta} X(s) d\Psi(s)}{\int_a^b W(s)e^{X(s)^\top \beta} d\Psi(s)} \right\} \\ &= h_2^T \{X(T) - E[X(T)|W, X; \beta, \Psi]\}. \end{aligned}$$

The score operators for Ψ in the directions h_1 and $h_1 + c$ are identical for any constant c . Thus we regard $l_{1\beta\Psi}$ as an operator on the quotient space BV/K where BV is the space of functions of bounded variation on $[a, b]$, and K is the subspace of constant functions. We equip BV/K with the quotient variation norm $\|\cdot\|_{BV}$ and R^p with the Euclidean norm, and the product space $BV/K \times R^p$ with the supremum of these two norms.

Let $l_n(\beta, \Psi) = P_n(l_{1\beta\Psi}, l_{2\beta\Psi})$, $l(\beta, \Psi) = P_0(l_{1\beta\Psi}, l_{2\beta\Psi})$. We verify the conditions of Theorem 3.3.1 of Van der Waart & Wellner (1996) for $\sqrt{n}(l_n - l)$. Clearly, when β_0 is interior to B , $l(\beta_0, \Psi_0) = 0$ and by consistency, $l_n(\hat{\beta}, \hat{\Psi}) = 0$ asymptotically. The compactness of B and C and the fact that $\int_a^b W(s)d\Psi_0(s)$ is bounded away from 0 and ∞ ensure that the score function is Fréchet differentiable at (β_0, Ψ_0) (Bickel, Klaasen, Ritov & Wellner 1993).

Let $S_{W,X}(T) = h_1(T) - E[h_1(T)|W, X; \beta, \Psi] + h_2^T \{X(T) - E[X(T)|W, X; \beta, \Psi]\}$. The information operator $\sigma = (\sigma_1, \sigma_2) : BV/K \times R^p \rightarrow BV/K \times R^p$ may be succinctly

written

$$\begin{aligned}\sigma_1[h](u) &= P_0 \left\{ \frac{W(u)e^{X(u)^\top \beta}}{\int_a^b W(s)e^{X(s)^\top \beta} d\Psi(s)} S_{W,X}(u) \right\} \\ \sigma_2[h] &= P_0 \{ \text{Cov}[X(T), S_{W,X}(T) | W, X; \beta, \Psi] \}\end{aligned}$$

where the covariance is taken with respect to T conditionally on W, X .

To prove that σ is invertible at β_0, Ψ_0 , we prove that it is compact and 1-1. $\sigma[h]$ may be written in the form $A + K_1 + K_2$ where $A[h](u) = (h_1(u)V_0(u; \Psi_0, \beta_0), h_2)$, $\|K_1[h_1]\| \leq D \int |h_1| d\Psi_0$, and K_2 maps into R^p . Specifically,

$$K_1[h_1](u) = P_0 \left\{ -W(u)e^{X(u)^\top \beta_0} \frac{\int_a^b W(s)e^{X(s)^\top \beta_0} h_1(s) d\Psi_0(s)}{\left(\int_a^b W(s)e^{X(s)^\top \beta_0} d\Psi_0(s)\right)^2} \right\}.$$

Since $V_0(u; \Psi_0, \beta_0) > d' > 0$, A is invertible and the compactness of σ follows as in Murphy et al. (1997).

To prove that σ is 1-1, we show that $\|\sigma[h]\| = 0$ implies that $h \equiv 0$, a.s. Suppose that $\|\sigma[h]\| = 0$. Then $\|\sigma_1[h]\|_{BV} = 0$ and $\|\sigma_2[h]\|_2 = 0$. Hence, by completeness, $\sigma_1[h] \equiv 0$ and $\sigma_2[h] = 0$. Now

$$\begin{aligned}E_{T,W,X}\{S_{W,X}(T)^2\} &= P_0\{(l_{1\beta\Psi}(Y)[h_1] + h_2^T l_{2\beta\Psi}(Y))^2\} \\ &= \int \sigma_1[h](u)h_1(u)d\Psi_0 + h_2^T \sigma_2[h] \\ &= 0\end{aligned}$$

and hence $S_{W,X}(T) = 0$ a.s. Thus $h_1(u) + h_2^T X(u) = c_{W,X}$, where $c_{W,X}$ is a constant that depends only on W and X . Hence by the form (b2) of assumption (b1), h_1 is a constant, so that $h_1 \equiv 0$ in BV/K , and $h_2 = 0$. Thus the information operator is continuously invertible.

The remaining condition is verified via Lemma 3.3.5 of Van der Waart & Wellner (1996). Let D denote the set of non-decreasing, absolutely continuous functions on $[a, b]$ with $\Psi(a) = 0, \Psi(b) = 1$ such that $\|\Psi - \Psi_0\|_\infty < \frac{1}{4}d$. Then for $\Psi \in D$, $\int_a^b W(s)d\Psi(s) \geq \int_a^b W(s)d\Psi_0(s) - 2\|\Psi - \Psi_0\|_\infty > \frac{1}{2}d$. Also, maps such as $(\beta, \Psi) \rightarrow I(\beta, \Psi) = \int_a^b W(s)e^{X(s)^\top \beta} d\Psi(s)$ are continuous and hence, by the continuous mapping theorem, the class $\left\{ \int_a^b W(s)e^{X(s)^\top \beta} d\Psi(s) : \beta \in B, \Psi \in D \right\}$ is Donsker. Since these integrals are positive and uniformly bounded away from zero, it follows that $\{(l_{1\beta\Psi}, l_{2\beta\Psi}) - (l_{1\beta_0\Psi_0}, l_{2\beta_0\Psi_0}) : \beta \in B, \Psi \in D\}$ is a Donsker class. Finally, for

$h \in BV/K \times R^p$ uniformly bounded by M , we have

$$|l_{1\beta\Psi}[h_1] - l_{1\beta_0\Psi_0}[h_1]| \leq C \int W(s) \left| e^{X(s)^\top \beta} \frac{d\Psi(u)}{d\Psi_0(u)} I(\beta_0, \Psi_0) - e^{X(s)^\top \beta_0} I(\beta, \Psi) \right| d\Psi_0(u)$$

which tends to 0 as $(\beta, \Psi) \rightarrow (\beta_0, \Psi_0)$. By the dominated convergence theorem, it follows that $\sup_h P_0 \{ (l_{1\beta\Psi}, l_{2\beta\Psi}) - (l_{1\beta_0\Psi_0}, l_{2\beta_0\Psi_0}) \}^2 \rightarrow 0$ as well.

Profile likelihood

To prove that the profile log likelihood ratio is asymptotically distributed χ_p^2 , we verify the conditions (8)-(10) and (18) of Murphy & Van der Waart (2000). Let h^* denote the least favourable direction at the true parameter (β_0, Ψ_0) , and define the path

$$d\Psi_t(\Psi, \beta) = \frac{(1 + (\beta - t)h^*d\Psi)}{\int_a^b (1 + (\beta - t)h^*(u))d\Psi(u)}.$$

The map $t \rightarrow l(t, \beta, \Psi) = \log L_{\mathfrak{S}}(t, \Psi_t(\Psi, \beta); Y)$ is twice continuously differentiable for all Y . Clearly, $\Psi_\beta(\Psi, \beta) = \Psi$, and

$$\left. \frac{\partial l(t, \beta_0, \Psi_0)}{\partial t} \right|_{\beta_0} = l_{2\beta_0\Psi_0}(Y) - l_{1\beta_0\Psi_0}(Y)[h^*],$$

the efficient score for β at (β_0, Ψ_0) . Thus, conditions (8) and (9) of Murphy & Van der Waart (2000) are satisfied.

Since the information operator is continuously invertible, and the score functions are uniformly bounded and of uniformly bounded variation, the conditions of Theorem 3.1 of Murphy & Van der Waart (1999) are satisfied and hence $\|\widehat{\Psi} - \Psi_0\|_\infty = O_p(\|\beta_n - \beta_0\|_2 + n^{-1/2})$. Thus conditions (10) and (18) of Murphy & Van der Waart (2000) are satisfied.

References

Aalen, O. O., Borgan, O., Keiding, N. & Thorman, J. (1980). Interaction between life history events: Nonparametric analysis for prospective and retrospective data in the presence of censoring., *Scandinavian Journal of Statistics* **7**: 161–171.

Andrews, N. J. (2002). Statistical assessment of the association between vaccination and rare events post-licensure, *Vaccine* **20**: S49–S53.

- Becker, N. G., Li, Z. & Kelman, C. W. (2004). The effect of transient exposures on the risk of an acute illness with low hazard rate, *Biostatistics* **5**: 239–248.
- Bickel, P., Klaassen, C., Ritov, Y. & Wellner, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, Baltimore.
- Black, C., Kaye, J. A. & Jick, H. (2003). MMR vaccine and idiopathic thrombocytopenic purpura, *British Journal of Clinical Pharmacology* **55**: 107–111.
- Borgan, O., Goldstein, L. & Langholz, B. (1995). Methods for the analysis of sampled cohort data in the cox proportional hazards model, *Annals of Statistics* **23**: 1749–1778.
- Cox, D. R. (1972a). Regression models and life-tables (with discussion), *Journal of the Royal Statistical Society Series B* **34**: 187–220.
- Cox, D. R. (1972b). The statistical analysis of dependencies in point processes, in P. A. W. Lewis (ed.), *Stochastic Point Processes: Statistical Analysis, Theory and Applications*, Wiley, New York, pp. 55–66.
- Farrington, C. P. (1995). Relative incidence estimation from case series for vaccine safety evaluation, *Biometrics* **51**: 228–235.
- Farrington, C. P., Miller, E. & Taylor, B. (2001). MMR and autism: Further evidence against a causal association, *Vaccine* **19**: 3632–3635.
- Farrington, C. P., Nash, J. & Miller, E. (1996). Case series analysis of adverse reactions to vaccines: A comparative evaluation, *American Journal of Epidemiology* **143**: 1165–1173. Erratum 1998;147:93.
- Feldmann, U. (1993). Epidemiologic assessment of risks of adverse reactions associated with intermittent exposure., *Biometrics* **49**: 419–428.
- Hocine, M., Guillemot, D., Tubert-Bitter, P. & Moreau, T. (2005). Testing independence between two poisson-generated multinomial variables in case series and cohort studies, *Statistics in Medicine* **24**: 4035–4044.
- Huang, J. & Rossini, A. J. (1997). Sieve estimation for the proportional odds failure-time regression model with interval censoring, *Journal of the American Statistical Association* **92**: 960–967.

- Hubbard, R., Farrington, P., Smith, C., Smeeth, L. & Tattersfield, A. (2003). Exposure to tricyclic and selective serotonin inhibitor antidepressants and the risk of hip fracture, *American Journal of Epidemiology* **158**: 77–84.
- Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd edn, Wiley, Hoboken, New Jersey.
- Lumley, T. & Levy, D. (2000). Bias in the case-crossover design: Implications for studies of air pollution, *Environmetrics* **11**: 689–704.
- Maclure, M. (1991). The case-crossover design: A method for studying transient effects on the risk of acute events, *American Journal of Epidemiology* **133**: 144–153.
- Madsen, K. M., Hviid, A., Vestergaard, M., Schendel, D., Wolfahrt, J., Thorsen, P., Olsen, J. & Mellbye, M. (2002). A population-based study of measles, mumps and rubella vaccination and autism, *New England Journal of Medicine* **347**: 1477–1482.
- McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models (2nd Ed.)*, Chapman and Hall, London.
- Murphy, S. A., Rossini, A. J. & Van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model, *Journal of the American Statistical Association* **92**: 968–976.
- Murphy, S. A. & Van der Vaart, A. W. (1999). Observed information in semiparametric models, *Bernoulli* **5**: 381–412.
- Murphy, S. A. & Van der Vaart, A. W. (2000). On profile likelihood, *Journal of the American Statistical Association* **95**: 449–465.
- Navidi, W. (1998). Bidirectional case-crossover designs for exposures with time trends, *Biometrics* **54**: 596–605.
- Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model, *The Annals of Statistics* **26**: 183–214.
- Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials., *Biometrika* **73**: 1–11.
- Prentice, R. L., Vollmer, W. M. & Kalbfleisch, J. D. (1984). On the use of case series to identify disease risk factors, *Biometrics* **40**: 445–458.

- Smeeth, L., Cook, C., Fombonne, E., Heavey, L., Rodrigues, L. C., Smith, P. G. & Hall, A. J. (2004). MMR vaccination and pervasive developmental disorders: A case-control study, *Lancet* **364**: 963–969.
- Smeeth, L., Thomas, S. L., Hall, A. J., Hubbard, R., Farrington, P. & Vallance, P. (2004). Risk of myocardial infarction and stroke after acute infection or vaccination, *New England Journal of Medicine* **351**: 2611–2618.
- Taylor, B., Miller, E., Farrington, C. P., Petropoulos, M.-C., Favot-Mayaud, I., Li, J. & Waight, P. A. (1999). Autism and measles, mumps and rubella vaccine: No epidemiological evidence for a causal association, *Lancet* **353**: 2026–2029.
- Thompson, R. (1977). The estimation of heritability with unbalanced data: I. observations available on parents and offspring, *Biometrics* **33**: 485–495.
- Van der Waart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*, Springer, New York.
- Vines, S. K. & Farrington, C. P. (2001). Within-subject exposure dependency in case-crossover studies, *Statistics in Medicine* **20**: 3039–3049.
- Whitaker, H., Farrington, C. P., Spiessens, B. & Musonda, P. (2006). The self-controlled case series method, *Statistics in Medicine (Published online DOI:10.1002/sim.2302)* **In Press**.
- Whitaker, H. J. & Farrington, C. P. (2005). On case-crossover methods for environmental time series data, *Technical Report 05/13*, Department of Statistics, The Open University.