Lifetime Data Anal
DOI 10.1007/s10985-015-9352-x

A case-base sampling method for estimating recurrent event intensities

Olli Saarela[1]

Received: 23 March 2015 / Accepted: 5 October 2015

Abstract Case-base sampling provides an alternative to risk set sampling based methods to estimate hazard regression models, in particular when absolute hazards are also of interest in addition to hazard ratios. The case-base sampling approach results in a likelihood expression of the logistic regression form, but instead of categorized time, such an expression is obtained through sampling of a discrete set of person-time coordinates from all follow-up data. In this paper, in the context of a time-dependent exposure such as vaccination, and a potentially recurrent adverse event outcome, we show that the resulting partial likelihood for the outcome event intensity has the asymptotic properties of a likelihood. We contrast this approach to self-matched case-base sampling, which involves only within-individual comparisons. The efficiency of the case-base methods is compared to that of standard methods through simulations, suggesting that the information loss due to sampling is minimal.

Keywords Case-base sampling · Conditional logistic regression · Hazard regression · Recurrent events · Self-matching

1 Introduction

Hanley and Miettinen (2009) proposed case-base sampling, combined with a logistic regression form likelihood expression, as a method for fitting flexible hazard regression models to survival data, with applications for example in prognostic

✉ Olli Saarela
olli.saarela@utoronto.ca

[1] Dalla Lana School of Public Health, University of Toronto, 155 College Street, Toronto, ON M5T 3M7, Canada

Published online: 14 October 2015

modeling. Their method is a generalization of one originating from Mantel (1973). Rather than through categorizing time, as in e.g. Arjas and Haara (1987), in this framework the logistic form likelihood expression is obtained by selecting a discrete set of person-time coordinates ('person-moments') in continuous time from all observed follow-up experience constituting the study base. In particular, all person-moments where an outcome event occurred are selected as the case series, complemented by a randomly sampled base series of person-moments serving as controls.

Saarela and Arjas (2015) generalized this approach to multiple competing causes, and noted that the sampling mechanism can be formulated more generally as a non-homogeneous Poisson process. However, neither Hanley and Miettinen (2009) nor Saarela and Arjas (2015) showed that the estimating function resulting from the sampling procedure actually has the asymptotic properties of a likelihood, or studied the finite-sample performance through simulations, issues which we aim to rectify herein. The main theoretical objectives of the present paper are to show that the case-base sampling estimating function can be characterized as a partial likelihood, and to show that it possesses the usual likelihood properties, namely a zero mean score process, and information equality, the latter enabling using the inverse of the observed information for variance estimation. Establishing these theoretical properties opens the way for more practical applications utilizing the case-base sampling methodology.

The context for developing the theory and methods are vaccination safety studies where the exposures are time-dependent (e.g. a pre-specified time period after the vaccination), and the adverse outcome events of interest may be recurrent. The self-controlled case series method of Farrington (1995) is often applied in the vaccination safety context, as it automatically controls for any time-invariant individual-level characteristics. As a variation of this, Saarela and Hanley (2015) suggested a self-matched case-base sampling approach, which results in a conditional logistic form likelihood expression. Under settings where both self-matched and unmatched case-base sampling methods are applicable, it is of interest to compare the efficiency of these, as the self-matched method only uses information from individuals with at least one outcome event.

In the present paper, we generalize the unmatched case-base sampling method to recurrent event data and time-varying exposures, as well as show that the resulting likelihood expression is still of a logistic regression form with an offset term. This enables fitting flexible intensity models using standard logistic regression software, with time effects fitted through suitable regression splines. We show that the resulting estimating function can be characterized as a partial likelihood, and show through a martingale representation for the corresponding score process that it has the asymptotic properties of a likelihood (Sect. 2). We then demonstrate how the self-matched case-base sampling method arises through the same sampling mechanism, but different conditioning (Sects. 3, 4). The efficiencies of the two case-base sampling methods are compared to standard methods through simulations (Sect. 5). We conclude with a discussion in Sect. 6.

2 A partial likelihood for estimating recurrent event intensities

2.1 Notation, assumptions, and the object of inference

We approach the modeling in the general framework of event history analysis and counting processes (e.g. Aalen et al. 2008). First, let $N_i(t) \in \{0, 1, 2, \ldots\}$ be a counting process for the adverse events of interest, while $Z_i(t) \in \{0, 1\}$ records the time-dependent exposure status (for instance, receiving value of one during a fixed time interval after a vaccination) for individual i. In addition, we have a collection of recorded baseline characteristics x_i, possibly relevant for controlling for confounding. Random, possibly covariate-dependent, censoring at time C_i is accommodated by introducing the at-risk process $Y_i(t) \equiv \mathbf{1}_{\{C_i \geq t\}}$ indicating that the individual is still under observation at time t, with the observation otherwise right censored at a fixed time τ. The observed history on individual i until just before time t is denoted as $\mathcal{F}_{it^-} \equiv \{N_i(u), Y_i(u): 0 \leq u < t; Z_i(u): 0 \leq u \leq t; x_i\}$. We also introduce the notations $\mathcal{N}_{it^-} \equiv \{N_i(u): 0 \leq u < t\}$ and $\mathcal{Z}_{it^-} \equiv \{Z_i(u): 0 \leq u < t\}$ for the observed outcome event and exposure process histories separately. In addition, to define the intensity function of interest, we introduce the latent outcome process $\tilde{N}_i(t)$, in the absence of censoring, and the corresponding latent history $\tilde{\mathcal{F}}_{it^-} \equiv \{\tilde{N}_i(u): 0 \leq u < t; Z_i(u): 0 \leq u \leq t; x_i\}$. We assume a continuous time model, which implies that the counting process jumps are less or equal to one. We are interested in modeling of the recurrent event intensity function $\lambda_i(t) \equiv \lim_{\Delta t \to 0} P(\Delta \tilde{N}_i(t) = 1 \mid \tilde{\mathcal{F}}_{it^-})/\Delta t$; this can be equivalently characterized through $\lambda_i(t)\,\mathrm{d}t = E[\mathrm{d}\tilde{N}_i(t) \mid \tilde{\mathcal{F}}_{it^-}]$. We assume non-informative censoring which requires that the expected jump for the observed outcome process is given by $E[\mathrm{d}N_i(t) \mid \mathcal{F}_{it^-}] = Y_i(t)\lambda_i(t)\,\mathrm{d}t$.

2.2 Case-base sampling

We refer to all observed follow-up experience in the interval $[0, \tau]$ for individuals $i = 1, \ldots, n$ as the 'study base'. In case-base sampling, as discussed by Hanley and Miettinen (2009), a sample from the study base is obtained by first selecting the 'case series' as the 'person-moments' [that is, person-time coordinates (i, t)] corresponding to all outcome events ($\mathrm{d}N_i(t) = 1$). To complement this, a sample of 'base series' person-moments is drawn randomly from the study base to serve as controls. The base series sampling mechanism can be represented through non-homogeneous Poisson processes $R_i(t) \in \{0, 1, 2, \ldots\}$, with the person-moments where $\mathrm{d}R_i(t) = 1$ constituting the base series. In addition, we introduce the process $Q_i(t) = N_i(t) + R_i(t)$ counting both the case and base series person-moments contributed by individual i. The corresponding observed histories are denoted as $\mathcal{Q}_{it^-} \equiv \{Q_i(u): 0 \leq u < t\}$ and $\mathcal{R}_{it^-} \equiv \{R_i(u): 0 \leq u < t\}$.

Since the 'observation' of the sampling process is also considered censored at time C_i, we introduce the latent version $\tilde{R}_i(t)$, which is characterized by the intensity function $\rho_i(t) \equiv \lim_{\Delta t \to 0} P(\Delta \tilde{R}_i(t) = 1 \mid \mathcal{Z}_{it}; x_i)/\Delta t$. This does not depend on the past history of the sampling process, as this is a Poisson process, and is restricted to not depend on the history of the outcome process. This agrees with the

terminology of Miettinen (2011, p. 115), where a base series is drawn as a fair (representative) sample from the study base. However, the sampling intensity may depend on the exposure history and the covariates. The non-informative censoring condition $E[\mathrm{d}R_i(t) \mid \mathcal{Z}_{it}; x_i] = Y_i(t)\rho_i(t)\,\mathrm{d}t$ is satisfied automatically, since $\rho_i(t)$ is user-chosen.

To complete the definitions, a latent process $\tilde{Q}_i(t) = \tilde{N}_i(t) + \tilde{R}_i(t)$ is counting all sampled person-moments in the absence of random censoring. This, assuming that under continuous time the outcome and sampling processes cannot jump simultaneously, is characterized by the intensity $\lambda_i(t) + \rho_i(t)$. The observed process $Q_i(t)$ is then characterized by $E[\mathrm{d}Q_i(t) \mid \mathcal{F}_{it^-}] = Y_i(t)\lambda_i(t)\,\mathrm{d}t + Y_i(t)\rho_i(t)\,\mathrm{d}t$. The likelihood analogue (to be considered in the following section) arising from such a sampling mechanism is motivated by the conditional probability $P(\mathrm{d}N_i(t) = 1 \mid \mathrm{d}Q_i(t) = 1, \mathcal{F}_{it^-}) = \lambda_i(t)/[\lambda_i(t) + \rho_i(t)]$.

Here we concentrate on settings where the vaccination and outcome information is available on a large population of individuals through routinely maintained health care databases. Therefore, the motivation for sampling from the study base is not to collect further exposure information, but to obtain a convenient estimating function. A graphical illustration of the case-base sampling mechanism is given in Fig. 1 in Sect. 5.

2.3 Inference methods

If the intensity function $\lambda_i(t;\theta)$ is parametrized in terms of θ, the estimation could proceed through maximization of the likelihood expression

$$L_0(\theta) = \prod_{i=1}^{n} \exp\left\{-\int_0^{\tau} Y_i(t)\lambda_i(t;\theta)\,\mathrm{d}t\right\} \prod_{i=1}^{n} \prod_{t\in[0,\tau)} \lambda_i(t;\theta)^{\mathrm{d}N_i(t)},$$

where $\prod_{t\in[0,u)}$ represents a product integral from 0 to u (Gill and Johansen 1990). Instead of using this expression for estimation, in what follows we will consider likelihood analogues that eliminate the integral over time in the survival contribution of $L_0(\theta)$. This integral is a computational inconvenience, especially in fitting non-parametric Bayesian hazard models using Markov chain Monte Carlo. A likelihood analogue arising from the case-base sampling mechanism is obtained by conditioning on a sampled person-moment, resulting in terms of the form

$$P\left(\mathrm{d}N_i(t) \mid \mathrm{d}Q_i(t) = 1, \mathcal{F}_{it^-}\right) \overset{\theta}{\propto} \frac{\lambda_i(t;\theta)^{\mathrm{d}N_i(t)}}{\lambda_i(t;\theta) + \rho_i(t)},$$

with the corresponding estimating function

$$L(\theta) = \prod_{i=1}^{n} \prod_{t\in[0,\tau)} \left(\frac{\lambda_i(t;\theta)^{\mathrm{d}N_i(t)}}{\lambda_i(t;\theta) + \rho_i(t)}\right)^{\mathrm{d}Q_i(t)}. \tag{1}$$

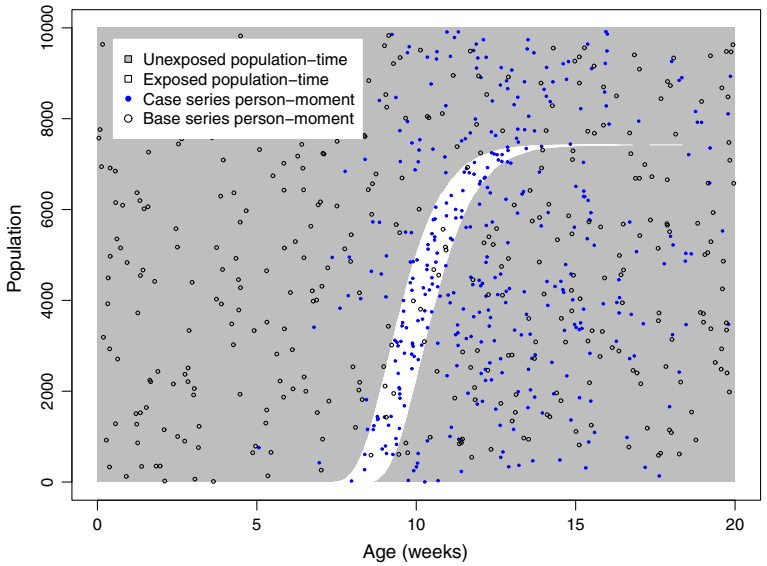

Fig. 1 A population-time plot illustration of case-base sampling based on one simulation round. The study base represented by the rectangle of 10,000 × 20 weeks of population-time is divided into unexposed (*gray area*) and exposed (*white area*; exposure defined as vaccination during the previous 7 days) population-time. The individuals on y-axis are ordered by their simulated vaccination times, with unvaccinated individuals distributed randomly over the top part of the y-axis. 377 case series person-moments are represented by *filled circles* and 377 base-series person-moments are represented by *unfilled circles*

When a logarithmic link function is used for modeling the intensity function, (1) is of a logistic regression form with an offset term $\log(1/\rho_i(t))$. In fact, the multiplicative terms in (1) are of the same form as conditional likelihood contributions arising from a case-control sampling mechanism where the controls are selected as independent Bernoulli draws from an underlying cohort, and the condition is that an individual was selected into the study either through a case or a control (cf. Langholz and Goldstein 2001, pp. 69–70; Cox 2006, pp. 155–156). Despite these similarities, the sampling units selected in the case-base sampling mechanism are person-moments, rather than individuals, and the parameters to be estimated are hazards or hazard ratios rather than odds or odds ratios. Generally, an individual can contribute more than one person-moment, and thus the terms in the product integral of (1) are not independent. While Saarela and Arjas (2015) referred to an expression similar to (1) as a conditional likelihood, this is not an appropriate characterization because of the aforementioned dependency. This motivates us to study further the likelihood properties of the estimating function (1). In the rest of this section, we show that (1) can be characterized as a partial likelihood, and that it has the asymptotic properties of a likelihood.

We first demonstrate that (1) is a partial likelihood; for notational simplicity we first consider the situation in the absence of random censoring. The joint likelihood can be split into multiplicative components as

$$\begin{aligned}
&\prod_{i=1}^{n} P\left(\tilde{\mathcal{N}}_{i\tau}, \tilde{\mathcal{R}}_{i\tau}, \mathcal{Z}_{i\tau} \mid x_i\right) = \prod_{i=1}^{n} P\left(\tilde{\mathcal{N}}_{i\tau}, \tilde{\mathcal{Q}}_{i\tau}, \mathcal{Z}_{i\tau} \mid x_i\right) \\
&= \prod_{i=1}^{n} \prod_{t\in[0,\tau)} P\left(\mathrm{d}\tilde{N}_i(t), \mathrm{d}\tilde{Q}_i(t), \mathrm{d}Z_i(t) \mid \tilde{\mathcal{N}}_{it^-}, \tilde{\mathcal{Q}}_{it^-}, \mathcal{Z}_{it^-}, x_i\right) \\
&= \prod_{i=1}^{n} \prod_{t\in[0,\tau)} \left[P\left(\mathrm{d}\tilde{N}_i(t) \mid \mathrm{d}\tilde{Q}_i(t), \tilde{\mathcal{F}}_{it^-}\right) P\left(\mathrm{d}\tilde{Q}_i(t), \mathrm{d}\tilde{Z}_i(t) \mid \tilde{\mathcal{N}}_{it^-}, \tilde{\mathcal{Q}}_{it^-}, \mathcal{Z}_{it^-}, x_i\right)\right] \\
&\overset{\theta}{\propto} L(\theta) \prod_{i=1}^{n} \prod_{t\in[0,\tau)} \left[P(\mathrm{d}\tilde{Q}_i(t), \mathrm{d}\tilde{Z}_i(t) \mid \tilde{\mathcal{N}}_{it^-}, \tilde{\mathcal{Q}}_{it^-}, \mathcal{Z}_{it^-}, x_i)\right], \qquad (2)
\end{aligned}$$

where the second equality followed because the past history $\tilde{\mathcal{Q}}_{it^-}$ of the sampling process is not informative of the outcome process given $\tilde{\mathcal{F}}_{it^-}$, and where the last proportionality followed because $P(\mathrm{d}\tilde{N}_i(t) = 0 \mid \mathrm{d}\tilde{Q}_i(t) = 0, \tilde{\mathcal{F}}_{it^-}) = 1$. If we consider the transformation of the time-specific observations on individual i into $W_i(t) \equiv \mathrm{d}\tilde{N}_i(t)$ and $V_i(t) \equiv (\mathrm{d}\tilde{Q}_i(t), \mathrm{d}Z_i(t))$, with the corresponding histories denoted as $\mathcal{W}_{it^-} \equiv \{W_i(u): 0 \leq u < t\}$ and $\mathcal{V}_{it^-} \equiv \{V_i(u): 0 \leq u < t\}$, we can alternatively write (2) as

$$\begin{aligned}
&\prod_{i=1}^{n} P\left(\tilde{\mathcal{N}}_{i\tau}, \tilde{\mathcal{R}}_{i\tau}, \mathcal{Z}_{i\tau} \mid x_i\right) \\
&= \prod_{i=1}^{n} \prod_{t\in[0,\tau)} P\left(W_i(t) \mid \mathcal{V}_{it}, \mathcal{W}_{it^-}, x_i\right) \prod_{i=1}^{n} \prod_{t\in[0,\tau)} P\left(V_i(t) \mid \mathcal{V}_{it^-}, \mathcal{W}_{it^-}, x_i\right) \\
&\overset{\theta}{\propto} L(\theta) \prod_{i=1}^{n} \prod_{t\in[0,\tau)} P\left(V_i(t) \mid \mathcal{V}_{it^-}, \mathcal{W}_{it^-}, x_i\right). \qquad (3)
\end{aligned}$$

We can now see that the form (3) is a continuous-time analogue of the definition of partial likelihood given by Cox (1975, p. 270, Equation (2)). On the other hand, a conditional likelihood would be obtained if all observed data on individual i could be transformed into W_i and V_i so that the joint likelihood could be split into $\prod_{i=1}^{n} P(W_i \mid V_i) \prod_{i=1}^{n} P(V_i)$, with the first term referred to as the conditional likelihood. The partitioning (3) of the joint likelihood is not of this form, but rather, preserves the temporal ordering of the events. Thus (1), is not a conditional likelihood, but rather, a partial likelihood. In the presence of random censoring, the reasoning goes as above, by defining the transformation of the time-specific observations as $W_i(t) \equiv \mathrm{d}N_i(t)$ and $V_i(t) \equiv (\mathrm{d}Q_i(t), \mathrm{d}Z_i(t), \mathrm{d}Y_i(t))$.

Ignoring the second multiplicative term in (2) reflects the information lost through the sampling of the follow-up data; this can be minimized by choosing a 'large enough' base series size. Hanley and Miettinen (2009) provide some rules of thumb for this, based on the Woolf variance formula (Woolf 1955) for log-odds/log-rate ratio. However, the Woolf formula is also based on case-control type of sampling where the sampling units are individuals, while sampling multiple person-moments per individ-

ual may introduce dependencies between the sampling units. Thus, in Sect. 5 we will study the information loss through simulations, but before that we show that the partial likelihood $L(\theta)$ indeed has the asymptotic properties of a likelihood.

For this purpose, we note that since $\mathrm{d}N_i(t) = 1 \Rightarrow \mathrm{d}Q_i(t) = 1$ and $\mathrm{d}N_i(t)\mathrm{d}Q_i(t) = \mathrm{d}N_i(t)$, the log-likelihood corresponding to (1) may be expressed as

$$l(\theta) = \sum_{i=1}^{n} \int_0^\tau \log \lambda_i(t;\theta)\, \mathrm{d}N_i(t) - \sum_{i=1}^{n} \int_0^\tau \log\left[\lambda_i(t;\theta) + \rho_i(t)\right]\, \mathrm{d}Q_i(t),$$

and the resulting score process as

$$\begin{aligned} U(t;\theta) = &\sum_{i=1}^{n} \int_0^t \frac{\partial}{\partial\theta} \log \lambda_i(u;\theta)\, \mathrm{d}N_i(u) \\ &- \sum_{i=1}^{n} \int_0^t \frac{\partial}{\partial\theta} \log\left[\lambda_i(u;\theta) + \rho_i(u)\right]\, \mathrm{d}Q_i(u). \end{aligned}$$

Here we note that the two counting processes have the decompositions

$$\mathrm{d}N_i(t) = Y_i(t)\lambda_i(t;\theta)\, \mathrm{d}t + \mathrm{d}M_i(t) \tag{4}$$

and

$$\begin{aligned} \mathrm{d}Q_i(t) &= \mathrm{d}N_i(t) + \mathrm{d}R_i(t) \\ &= Y_i(t)\lambda_i(t;\theta)\, \mathrm{d}t + Y_i(t)\rho_i(t)\, \mathrm{d}t + \mathrm{d}M_i(t) + \mathrm{d}M_i^*(t), \end{aligned} \tag{5}$$

where the processes $M_i(t)$ and $M_i^*(t)$ are orthogonal martingales. Denoting $\frac{\partial}{\partial\theta}\lambda_i(t;\theta) \equiv \lambda_i'(t;\theta)$, the score process now has the martingale representation

$$\begin{aligned} U(t;\theta) = &\sum_{i=1}^{n} \int_0^t \frac{\partial}{\partial\theta} \log \lambda_i(u;\theta)\, \mathrm{d}M_i(u) + \sum_{i=1}^{n} \int_0^t Y_i(u)\lambda_i'(u;\theta)\, \mathrm{d}u \\ &- \sum_{i=1}^{n} \int_0^t \frac{\partial}{\partial\theta} \log\left[\lambda_i(u;\theta) + \rho_i(u)\right]\left[\mathrm{d}M_i(u) + \mathrm{d}M_i^*(u)\right] \\ &- \sum_{i=1}^{n} \int_0^t Y_i(u)\lambda_i'(u;\theta)\, \mathrm{d}u \\ = &\sum_{i=1}^{n} \int_0^t \frac{\partial}{\partial\theta} \left\{\log \lambda_i(u;\theta) - \log\left[\lambda_i(u;\theta) + \rho_i(u)\right]\right\}\, \mathrm{d}M_i(u) \\ &- \sum_{i=1}^{n} \int_0^t \frac{\partial}{\partial\theta} \log\left[\lambda_i(u;\theta) + \rho_i(u)\right]\, \mathrm{d}M_i^*(u), \end{aligned}$$

which has a zero mean at the true parameter value θ_0.

In "Appendix" we further show that the corresponding predictable variation process and the observed information process are equivalent in expectation, and sketch an argument for asymptotic normality of the sampling distribution of $\hat{\theta} \equiv \arg\max_\theta L(\theta)$.

3 Self-matched case-base sampling

For comparison to the method discussed in Sect. 2.3, we also show how the self-matched case-base sampling approach suggested by Saarela and Hanley (2015) results from the same sampling mechanism as described above. In this method, the base series is also sampled through a Poisson process, but the comparisons of the person-moments are matched by individual. Here we suppress the recorded baseline characteristics x_i from the notation, since these do not feature in the self-matched estimation. Also, again for notational simplicity, we consider the setting in the absence of random censoring. The base series sampling mechanism can be characterized through the same counting process $\tilde{R}_i(t)$, itself characterized by the non-homogeneous Poisson process intensity $\rho_i(t)$. However, the self-matched method requires certain additional assumptions, and can only be used for estimating relative intensities, rather than absolute ones. The first added assumption (i) is that the exposure process $Z_i(t)$ is external in the sense that

$$P\left(\mathrm{d}\tilde{N}_i(t) = 1 \mid \mathcal{Z}_{is}, \tilde{\mathcal{N}}_{it^-}\right) = P\left(\mathrm{d}\tilde{N}_i(t) = 1 \mid \mathcal{Z}_{it}, \tilde{\mathcal{N}}_{it^-}\right),$$

where $t \leq s$ (cf. Kalbfleisch and Prentice 2002, p. 196). The second added assumption (ii) is that the outcome event process itself is a non-homogeneous Poisson process, so that the past history of the process is non-informative of its future behaviour. These assumptions are the same as in the well-known self-controlled case series method (e.g. Farrington 1995; Whitaker et al. 2009).

Now only individuals with $\tilde{N}_i(\tau) = k > 0$ have a likelihood contribution, which is of the form

$$P\left(\mathrm{d}\tilde{N}_i(t_{i1}) = 1, \ldots, \mathrm{d}\tilde{N}_i(t_{ik}) = 1 \mid \tilde{N}_i(\tau) = k, \tilde{\mathcal{Q}}_{i\tau}, \mathcal{Z}_{i\tau}\right), \tag{6}$$

where $t_{i1}, \ldots, t_{ik}$ are the ordered event times. If we randomly sampled m base series person-moments for individual i through the jumps of the sampler process $\tilde{R}_i(t)$ at times $t_{i(k+1)}, \ldots, t_{i(k+m)}$, the counting process $\tilde{Q}_i(t)$ jumps only at times $t \in A \equiv \{t_{i1}, \ldots, t_{i(k+m)}\}$, with $\tilde{Q}_i(\tau) = m + k$. Thus, the condition in (6) implies that k of the times in A are event times, and the likelihood contribution is the probability that the events occurred specifically at the times $t_{i1}, \ldots, t_{ik}$.

For example, when $k = 1$, (6) simplifies into

$$P\left(\mathrm{d}\tilde{N}_i(t_{i1}) = 1 \mid \tilde{N}_i(\tau) = 1, \tilde{\mathcal{Q}}_{i\tau}, \mathcal{Z}_{i\tau}\right)$$
$$= \frac{P\left(\mathrm{d}\tilde{N}_i(t_{i1}) = 1, \tilde{N}_i(\tau) = 1 \mid \tilde{\mathcal{Q}}_{i\tau}, \mathcal{Z}_{i\tau}\right)}{\sum_{t \in A} P\left(\mathrm{d}\tilde{N}_i(t) = 1, \tilde{N}_i(\tau) = 1 \mid \tilde{\mathcal{Q}}_{i\tau}, \mathcal{Z}_{i\tau}\right)}$$

$$= \frac{P\left(\tilde{\mathcal{Q}}_{i\tau} \mid \mathrm{d}\tilde{N}_i(t_{i1})=1, \tilde{N}_i(\tau)=1, \mathcal{Z}_{i\tau}\right) P\left(\mathrm{d}\tilde{N}_i(t_{i1})=1, \tilde{N}_i(\tau)=1 \mid \mathcal{Z}_{i\tau}\right)}{\sum_{t\in A} P\left(\tilde{\mathcal{Q}}_{i\tau} \mid \mathrm{d}\tilde{N}_i(t)=1, \tilde{N}_i(\tau)=1, \mathcal{Z}_{i\tau}\right) P\left(\mathrm{d}\tilde{N}_i(t)=1, \tilde{N}_i(\tau)=1 \mid \mathcal{Z}_{i\tau}\right)}.$$

Here, from assumptions (i) and (ii),

$$P(\mathrm{d}\tilde{N}_i(t)=1, \tilde{N}_i(\tau)=1 \mid \mathcal{Z}_{i\tau}) = \lambda_i(t;\theta)\exp\left\{-\int_0^\tau \lambda_i(u;\theta)\,\mathrm{d}u\right\},$$

where crucially the survival contribution does not depend on t. Further, because of the definition $\tilde{Q}_i(t) = \tilde{N}_i(t) + \tilde{R}_i(t)$, we have that

$$\begin{aligned} &P\left(\tilde{\mathcal{Q}}_{i\tau} \mid \mathrm{d}\tilde{N}_i(t_{i1})=1, \tilde{N}_i(\tau)=1, \mathcal{Z}_{i\tau}\right) = P\left(\tilde{\mathcal{R}}_{i\tau} \mid \mathcal{Z}_{i\tau}\right) \\ &= \exp\left\{-\int_0^\tau \rho_i(u)\,\mathrm{d}u\right\} \prod_{t\in A\setminus\{t_{i1}\}} \rho_i(t), \end{aligned}$$

where again, because the outcome process is a non-homogeneous Poisson process, the survival contribution does not depend on which one of the times in A is the event time. Thus, finally, both survival contributions cancel out from the ratio, leaving

$$P\left(\mathrm{d}\tilde{N}_i(t_{i1})=1 \mid \tilde{N}_i(\tau)=1, \tilde{\mathcal{Q}}_{i\tau}, \mathcal{Z}_{i\tau}\right) = \frac{\lambda_i(t_{i1};\theta)\prod_{t\in A\setminus\{t_{i1}\}}\rho_i(t)}{\sum_{t\in A}\lambda_i(t;\theta)\prod_{u\in A\setminus\{t\}}\rho_i(u)}. \quad (7)$$

If the base series sampling intensity is constant over time so that $\rho_i(t) \propto 1$, the sampling weights also cancel out, leaving an expression which is of the usual conditional logistic form, and can be fitted using standard conditional logistic regression software. The advantage of the likelihood expression (7) is the same as that of (6), namely, it does not involve an integral over time (unlike the expression considered by Ghebremichael-Weldeselassie et al. 2014). In the case of multiple events per individual (7), generalizes in the usual way, as in other matched case-control designs (e.g. Clayton and Hills 1993, pp. 295–296).

Similarly to the self-controlled case-series method, the advantage of self-matched case-base sampling is that it automatically controls for any time-invariant individual level characteristics u_i (measured or not). This is important even if these characteristics are not confounders. If for example the outcome events are generated by non-homogeneous Poisson process intensities of the form

$$\lambda_i(t;\theta) = U_i \exp\{\alpha + f(t,\beta) + \eta Z_i(t)\},$$

where $U_i \in [0,\infty)$ is latent, omitting these characteristics from the model will induce dependencies between the successive events, even if these are independent conditional on u_i. The approaches discussed in Sect. 2.3 would then require either estimation of these dependencies by introducing covariates corresponding to the past history of the event process, or frailty models.

Modeling of the effect of the past history could proceed through the marginal event intensities

$$E_{U_i|\tilde{\mathcal{F}}_{it^-}}[\lambda_i(t;\theta)] = E_{U_i|\tilde{\mathcal{F}}_{it^-}}[U_i]\exp\{\alpha + f(t,\beta) + \eta Z_i(t)\}.$$

If indeed U_i is not a confounder in the sense that $\mathcal{Z}_{it} \perp\!\!\!\perp U_i \mid \tilde{\mathcal{N}}_{it}$, we could specify a parametric model for the latent variable as, say,

$$E_{U_i|\tilde{\mathcal{F}}_{it^-}}[U_i] = \exp\left\{\kappa + \gamma_1 \mathbf{1}_{\left\{\tilde{N}_i(t^-)=1\right\}} + \gamma_2 \mathbf{1}_{\left\{\tilde{N}_i(t^-)=2\right\}} + \ldots\right\}.$$

This would result in the marginal intensity model

$$\begin{aligned} &E_{U_i|\tilde{\mathcal{F}}_{it^-}}[\lambda_i(t;\theta)] \\ &= \exp\left\{\alpha^* + f(t,\beta) + \eta Z_i(t) + \gamma_1 \mathbf{1}_{\{\tilde{N}_i(t^-)=1\}} + \gamma_2 \mathbf{1}_{\{\tilde{N}_i(t^-)=2\}} + \ldots\right\}, \end{aligned}$$

where $\alpha^* = \alpha + \kappa$. Here the conditional interpretation of the exposure log intensity ratio η is preserved, but at the cost of having to model the effect of the past history of the outcome process, which is subject to model misspecification; for instance, it may well be that not only the current value of the counting process $\tilde{N}_i(t)$, but also the past jump times and current survival experience are important. In contrast, in self-matching, the characteristics U_i cancel out from the ratio in (7), retaining the conditional interpretation even if U_i is a confounder.

Since

$$\begin{aligned} &\prod_{\{i:\tilde{N}_i(\tau)>0\}} P\left(\tilde{\mathcal{N}}_{i\tau}, \tilde{\mathcal{Q}}_{i\tau}, \mathcal{Z}_{i\tau}\right) \\ &= \prod_{\{i:\tilde{N}_i(\tau)>0\}} P\left(\mathrm{d}\tilde{N}_i(t_{i1}) = 1, \ldots, \mathrm{d}\tilde{N}_i(t_{ik}) = 1 \mid \tilde{N}_i(\tau) = k, \tilde{\mathcal{Q}}_{i\tau}, \mathcal{Z}_{i\tau}\right) \\ &\quad \times \prod_{\{i:\tilde{N}_i(\tau)>0\}} P\left(\tilde{N}_i(\tau) = k, \tilde{\mathcal{Q}}_{i\tau}, \mathcal{Z}_{i\tau}\right), \end{aligned}$$

the product of the terms (6) is a true conditional likelihood, unlike (1). However, in contrast to the expression (1), the self-matched approach does not allow for estimation of absolute intensities, and requires added assumptions on the outcome and exposure processes. In situations where both are applicable, it is of interest to compare their efficiency in estimating the exposure effect η through simulations (Sect. 5).

4 Sampling of the base series

Hanley and Miettinen (2009) suggested a two-step mechanism for the sampling of the base series. They first selected the person coordinates from a multinomial distribution with probabilities proportional to the follow-up periods of the individuals, and

then sampled the corresponding time coordinates uniformly from the follow-up time contributed by the sampled individuals.

This procedure can be generalized as follows. Because $R_i(t) \sim \text{Poisson}(\mu_i(t))$, where $\mu_i(t) \equiv \int_0^t Y_i(u)\rho_i(u)\,\mathrm{d}u$, the expected base series size is given by $\sum_{i=1}^n \mu_i(\tau_i)$, where we denote $\tau_i \equiv \min(C_i, \tau)$. Given an intensity $\rho_i^*(t)$ that captures the desired variation between individuals and over time, this can be rescaled to give an expected base series size b by taking the actual sampling intensity to be

$$\rho_i(t) = \frac{b\rho_i^*(t)}{\sum_{j=1}^n \mu_j^*(\tau_j)}.$$

Because the sampling probability may be approximated by

$$\rho_i(t)\,\mathrm{d}t \approx 1 - \left(1 - \frac{\mu_i^*(\tau_i)}{\sum_{j=1}^n \mu_j^*(\tau_j)} \frac{\rho_i^*(t)\,\mathrm{d}t}{\mu_i^*(\tau_i)}\right)^b,$$

a base series of a fixed (instead of expected) size b can be selected using the two-step procedure where first the number of person-moments contributed by individual i is sampled from a multinomial distribution with probabilities $\mu_i^*(t)/\sum_{j=1}^n \mu_j^*(t)$, with the corresponding times then sampled independently from the density $\rho_i^*(t)/\mu_i^*(t)$. If the chosen sampling intensity is constant over both individuals and time, that is, $\rho_i^*(t) = \rho^*$, the rescaled sampling intensity reduces to $\rho_i(t) = b/\sum_{j=1}^n \tau_j$, which is the setting considered by Hanley and Miettinen (2009).

5 Simulation study

Our simulation setting mimics the follow-up of a cohort of newly born infants until right censoring the age of $\tau = 140$ days, with scheduled vaccinations after 50 days of age. The adverse outcome events are sampled from a Poisson process with Gamma distribution shaped baseline incidence as given in Fig. 2 (solid line), scaled to give an expected 300 background events in a population of size 10,000 during the 140-day follow-up. This baseline incidence was modified using an exposure period of 7 days after the vaccination, with log intensity ratios of $\eta = 0, 0.5$, and 1.5. The probability of ultimately receiving the vaccination was taken to be 0.75, with the vaccination times v_i for the vaccinated individuals simulated from $(v_i - 50)/7 \sim \text{Gamma}(5, 2)$, giving a smooth distribution of vaccination times centered at 67.5 days.

We simulated 1000 datasets, and used the methods proposed in Sects. 2.3 (unmatched case-base) and 3 (self-matched case-base) to estimate η. For comparison, we estimated η also using the standard Mantel–Haenszel (M–H) and conditional logistic regression methods, sampling the full riskset at each event time. In the case-base sampling methods, we selected alternatively $b = 10c$, $50c$ and $100c$ base series person-moments, where c is the total number of events. The sampling intensity was taken to be uniform over the follow-up time and individuals, scaled to produce the aforementioned base series sample sizes, as discussed in Sect. 4. The case-base sam-

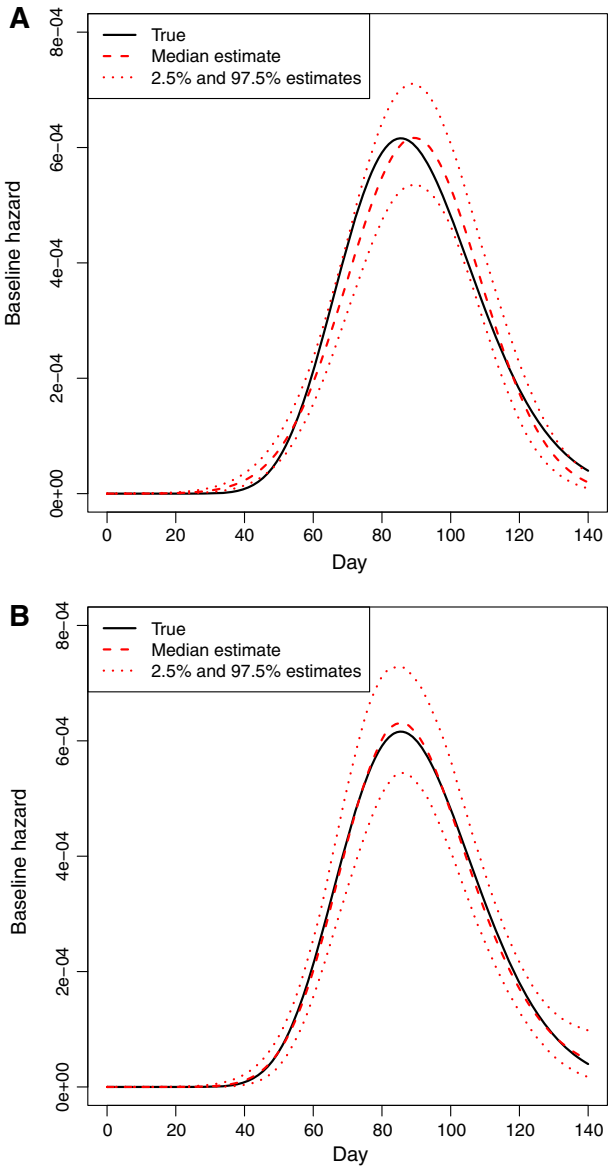

Fig. 2 **a** Summary statistics for pointwise baseline intensity estimates from the model assuming quadratic effect of age, **b** summary statistics for pointwise baseline intensity estimates from the model assuming cubic effect of age

pling from the study base is illustrated in Fig. 1, which features one simulated dataset with $\eta = 1.5$, $c = 377$, and $b = 377$, presented in a population-time plot.

In both case-base methods, the fitted intensity model was specified as

$$\lambda_i(t;\theta) = \exp\left\{\alpha + f(t,\beta) + \eta \mathbf{1}_{\{t-7<v_i\leq t\}}\right\},$$

where the intercept term α cancels out from likelihood expression (7). The effect of age $f(t, \beta)$ was modeled with low-dimensional parametric functions, using quadratic and cubic polynomials. However, any regression spline specification would be applicable for modeling of the age effect.

The numerical results are presented in Table 1. These demonstrate that with a base series size of 100 times the number of events, both case-base sampling methods give efficiency comparable to that of M–H and conditional logistic methods which involved sampling of the entire risk set at each event time. In the null $\eta = 0$ and $\eta = 0.5$ scenarios, there was little difference in efficiency between the methods. However, the case-base sampling method was more efficient than the self-matched method (and M–H and conditional logistic methods) in the scenario with the effect size $\eta = 1.5$, owing to the fact that it estimates the absolute intensities, and uses information also from the individuals without an event. With 50 base series person-moments per case, the efficiency of the case-base sampling methods is still similar to the M–H and conditional logistic methods, but with 10 base series person-moments per case, the efficiency penalty due to the sampling begins to show.

The results also demonstrate that under this simulation setting, two parameters are not enough to capture the age effect on the event intensity, resulting in some bias in the η estimates. However, the bias is effectively removed by allowing for one further parameter for the age effect. This is illustrated in Fig. 2 which shows summary statistics for the pointwise estimates of the baseline intensities from the models with quadratic and cubic age effects; the latter model is sufficient to capture the true age effect used in the simulation.

6 Discussion

As far as we know, the likelihood properties of the estimating function resulting from case-base sampling of person-moments have not been established before, especially in the recurrent event context. We studied these properties through a martingale representation for the corresponding score process. We demonstrated that the partial likelihood expression resulting from unmatched sampling of person-moments can be used as a likelihood. The self-matched version of this procedure could be motivated as a conditional likelihood. This raises the question of whether the two different kind of comparisons could be combined in a single estimating function, which is a topic for further study.

In this paper, we motivated the case-base sampling method as a means to obtain likelihood analogues that do not involve integrals of intensity functions over time. This provides computational benefits especially in Bayesian inferences using MCMC where the likelihood function needs to be evaluated numerous times. However, the computational expense of these likelihood expressions can also be contrasted to the standard Cox partial likelihood/conditional logistic likelihood involving the sampling of the entire riskset at each event time. With type I censoring at a fixed time point τ, the evaluation of the Cox partial likelihood involves evaluation of $c \times n$ linear predictors, which may be time-dependent. (Naturally, this can be reduced to $c + c \times m$ evaluations by sampling m controls from each riskset, but with an efficiency penalty.) Based on the

Table 1 Results for the point estimators of the log intensity ratio parameter η and the corresponding standard errors

η	b	Estimator			
		M–H	SE	Cond. logistic	SE
0.0	–	−0.021 (0.240)	0.239 (0.022)	−0.021 (0.240)	0.238 (0.022)
0.5	–	0.492 (0.193)	0.195 (0.013)	0.492 (0.192)	0.194 (0.013)
1.5	–	1.499 (0.140)	0.142 (0.005)	1.497 (0.138)	0.139 (0.005)
η	b	Estimator			
		Self-matched (quadratic)	SE	Self-matched (cubic)	SE
0.0	$10c$	0.105 (0.262)	0.263 (0.021)	−0.012 (0.264)	0.267 (0.021)
	$50c$	0.097 (0.242)	0.240 (0.022)	−0.011 (0.245)	0.243 (0.022)
	$100c$	0.096 (0.239)	0.237 (0.022)	−0.011 (0.242)	0.241 (0.022)
0.5	$10c$	0.625 (0.223)	0.222 (0.013)	0.508 (0.227)	0.227 (0.012)
	$50c$	0.613 (0.195)	0.196 (0.012)	0.505 (0.199)	0.201 (0.012)
	$100c$	0.610 (0.189)	0.193 (0.012)	0.503 (0.193)	0.198 (0.012)
1.5	$10c$	1.637 (0.175)	0.173 (0.007)	1.521 (0.181)	0.179 (0.007)
	$50c$	1.622 (0.140)	0.143 (0.005)	1.514 (0.146)	0.149 (0.005)
	$100c$	1.619 (0.138)	0.140 (0.005)	1.512 (0.145)	0.146 (0.005)
η	b	Estimator			
		Case-base (quadratic)	SE	Case-base (cubic)	SE
0.0	$10c$	0.099 (0.263)	0.253 (0.021)	−0.012 (0.266)	0.257 (0.021)
	$50c$	0.095 (0.241)	0.235 (0.022)	−0.010 (0.243)	0.238 (0.022)
	$100c$	0.094 (0.236)	0.232 (0.022)	−0.009 (0.238)	0.236 (0.022)
0.5	$10c$	0.614 (0.222)	0.211 (0.012)	0.504 (0.226)	0.216 (0.012)
	$50c$	0.607 (0.189)	0.190 (0.013)	0.503 (0.193)	0.195 (0.012)
	$100c$	0.606 (0.189)	0.188 (0.013)	0.503 (0.193)	0.192 (0.012)
1.5	$10c$	1.623 (0.154)	0.159 (0.005)	1.514 (0.162)	0.165 (0.005)
	$50c$	1.614 (0.131)	0.135 (0.004)	1.511 (0.137)	0.141 (0.005)
	$100c$	1.610 (0.128)	0.132 (0.004)	1.508 (0.133)	0.137 (0.005)

The numbers are mean point and standard error estimates over 1000 replications (the numbers in brackets are Monte-Carlo standard deviations of the point and standard error estimates). b is the number of base series person-moments, given by a multiple of the total number of cases c

simulation study, a comparable efficiency can be obtained in the case-base sampling method through $c + c \times m$ evaluations of linear predictors, where $m \geq 50$. In part this is due to the parametric estimation of the baseline intensity, but here flexible model specifications can be applied when necessary.

Acknowledgments The author acknowledges the support of the Natural Sciences and Engineering Research Council (NSERC) of Canada, and thanks Prof. Elja Arjas (University of Helsinki) for helpful comments.

Appendix

Because $M_i(t)$ and $M_i^*(t)$ are orthogonal, the predictable variation process of the score process can be expressed as

$$
\begin{aligned}
&\langle U\rangle(t;\theta)\\
&= \sum_{i=1}^{n}\int_0^t \left(\frac{\partial}{\partial\theta}\left\{\log\lambda_i(u;\theta) - \log\left[\lambda_i(u;\theta)+\rho_i(u)\right]\right\}\right)^{\otimes 2} Y_i(u)\lambda_i(u;\theta)\,\mathrm{d}u\\
&\quad + \sum_{i=1}^{n}\int_0^t \left(\frac{\partial}{\partial\theta}\log\left[\lambda_i(u;\theta)+\rho_i(u)\right]\right)^{\otimes 2} Y_i(u)\rho_i(u)\,\mathrm{d}u\\
&= \sum_{i=1}^{n}\int_0^t \left(\frac{\lambda_i'(u;\theta)}{\lambda_i(u;\theta)} - \frac{\lambda_i'(u;\theta)}{\lambda_i(u;\theta)+\rho_i(u)}\right)^{\otimes 2} Y_i(u)\lambda_i(u;\theta)\,\mathrm{d}u\\
&\quad + \sum_{i=1}^{n}\int_0^t \left(\frac{\lambda_i'(u;\theta)}{\lambda_i(u;\theta)+\rho_i(u)}\right)^{\otimes 2} Y_i(u)\rho_i(u)\,\mathrm{d}u\\
&= \sum_{i=1}^{n}\int_0^t \left(\frac{\lambda_i'(u;\theta)^{\otimes 2}}{\lambda_i(u;\theta)^2} - \frac{2\lambda_i'(u;\theta)^{\otimes 2}}{\lambda_i(u;\theta)\left[\lambda_i(u;\theta)+\rho_i(u)\right]} + \frac{\lambda_i'(u;\theta)^{\otimes 2}}{\left[\lambda_i(u;\theta)+\rho_i(u)\right]^2}\right)\\
&\quad \times Y_i(u)\lambda_i(u;\theta)\,\mathrm{d}u + \sum_{i=1}^{n}\int_0^t \frac{\lambda_i'(u;\theta)^{\otimes 2}}{\left[\lambda_i(u;\theta)+\rho_i(u)\right]^2} Y_i(u)\rho_i(u)\,\mathrm{d}u\\
&= \sum_{i=1}^{n}\int_0^t \left(\frac{\lambda_i'(u;\theta)^{\otimes 2}}{\lambda_i(u;\theta)} - \frac{\lambda_i'(u;\theta)^{\otimes 2}}{\lambda_i(u;\theta)+\rho_i(u)}\right) Y_i(u)\,\mathrm{d}u.
\end{aligned}
$$

The observed information process is given by

$$
\begin{aligned}
J(t;\theta) &= -\sum_{i=1}^{n}\int_0^t \frac{\partial^2}{\partial\theta\partial\theta^\top}\log\lambda_i(u;\theta)\mathrm{d}N_i(u)\\
&\quad + \sum_{i=1}^{n}\int_0^t \frac{\partial^2}{\partial\theta\partial\theta^\top}\log\left[\lambda_i(u;\theta)+\rho_i(u)\right]\mathrm{d}Q_i(u)\\
&= -\sum_{i=1}^{n}\int_0^t \left(\frac{\lambda_i''(u;\theta)\lambda_i(u;\theta) - \lambda_i'(u;\theta)^{\otimes 2}}{\lambda_i(u;\theta)^2}\right)\mathrm{d}N_i(u)\\
&\quad + \sum_{i=1}^{n}\int_0^t \left(\frac{\lambda_i''(u;\theta)\left[\lambda_i(u;\theta)+\rho_i(u)\right] - \lambda_i'(u;\theta)^{\otimes 2}}{\left[\lambda_i(u;\theta)+\rho_i(u)\right]^2}\right)\mathrm{d}Q_i(u),
\end{aligned}
$$

where we denoted $\lambda_i''(u;\theta) \equiv \frac{\partial^2}{\partial\theta\partial\theta^\top}\lambda_i(u;\theta)$.

Using the decompositions (4) and (5), the observed information process can be further written as

$$
\begin{aligned}
&J(t;\theta)\\
&\quad= -\sum_{i=1}^{n}\int_0^t\left(\frac{\lambda_i''(u;\theta)\lambda_i(u;\theta)-\lambda_i'(u;\theta)^{\otimes 2}}{\lambda_i(u;\theta)^2}\right)Y_i(u)\lambda_i(u;\theta)\,\mathrm{d}u\\
&\qquad+\sum_{i=1}^{n}\int_0^t\left(\frac{\lambda_i''(u;\theta)[\lambda_i(u;\theta)+\rho_i(u)]-\lambda_i'(u;\theta)^{\otimes 2}}{[\lambda_i(u;\theta)+\rho_i(u)]^2}\right)Y_i(u)[\lambda_i(u;\theta)+\rho_i(u)]\,\mathrm{d}u\\
&\qquad+\mathcal{E}(t)\\
&\quad=\sum_{i=1}^{n}\int_0^t\left(\frac{\lambda_i'(u;\theta)^{\otimes 2}}{\lambda_i(u;\theta)}-\frac{\lambda_i'(u;\theta)^{\otimes 2}}{\lambda_i(u;\theta)+\rho_i(u)}\right)Y_i(u)\,\mathrm{d}u+\mathcal{E}(t),
\end{aligned}
$$

where we denoted

$$
\begin{aligned}
\mathcal{E}(t)\equiv&-\sum_{i=1}^{n}\int_0^t\left(\frac{\lambda_i''(u;\theta)\lambda_i(u;\theta)-\lambda_i'(u;\theta)^{\otimes 2}}{\lambda_i(u;\theta)^2}\right)\mathrm{d}M_i(u)\\
&+\sum_{i=1}^{n}\int_0^t\left(\frac{\lambda_i''(u;\theta)\,[\lambda_i(u;\theta)+\rho_i(u)]-\lambda_i'(u;\theta)^{\otimes 2}}{[\lambda_i(u;\theta)+\rho_i(u)]^2}\right)\left[\mathrm{d}M_i(u)+\mathrm{d}M_i^*(u)\right].
\end{aligned}
$$

Therefore, $E[\langle U\rangle(t;\theta_0)]=E[J(t;\theta_0)]$. With these results, motivating the asymptotic normality of the maximum partial likelihood estimator $\hat{\theta}$ can proceed similarly as for parametric survival models (e.g. Kalbfleisch and Prentice 2002, p. 180). Briefly, assume a scalar θ for notational simplicity, and denote $U(\theta)\equiv U(\tau;\theta)$ and $J(\theta)\equiv J(\tau;\theta)$. From the martingale central limit theorem, it follows under the standard regularity conditions that

$$
\frac{\sqrt{n}}{n}U(\theta_0)\xrightarrow{d}N\left(0,\,\Sigma(\theta_0)\right),
$$

where the matrix $\Sigma(\theta_0)$ is such that $\frac{1}{n}\langle U\rangle(\tau;\theta_0)\xrightarrow{p}\Sigma(\theta_0)$. The Taylor expansion

$$
U(\hat{\theta})=U(\theta_0)-J(\theta_0)\left(\hat{\theta}-\theta_0\right)+\frac{1}{2}\frac{\partial^3 l(\theta^*)}{\partial\theta^3}\left(\hat{\theta}-\theta_0\right)^2
$$

can be used to motivate both the consistency and asymptotic normality of $\hat{\theta}$ by assuming that the third term on the right hand side is bounded in probability. In particular, we get

$$
\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{d}N\left(0,\,\Sigma(\theta_0)^{-1}\right),
$$

where $\Sigma(\theta_0)$ is in practice estimated by the average observed information $\frac{1}{n}J(\hat{\theta})$ at the maximum likelihood point.

References

Aalen O, Borgan Ø, Gjessing HK (2008) Survival and event history analysis: a process point of view. Springer, Berlin

Arjas E, Haara P (1987) A logistic regression model for hazard: asymptotic results. Scand J Stat 14:1–18

Clayton D, Hills M (1993) Statistical models in epidemiology. Oxford University Press, Oxford

Cox DR (1975) Partial likelihood. Biometrika 62:269–276

Cox DR (2006) Principles of statistical inference. Cambridge University Press, Cambridge

Farrington CP (1995) Relative incidence estimation from case series for vaccine safety evaluation. Biometrics 51:228–235

Ghebremichael-Weldeselassie Y, Whitaker HJ, Farrington CP (2014) Self-controlled case series method with smooth age effect. Stat Med 33:639–649

Gill RD, Johansen S (1990) A survey of product-integration with a view toward application in survival analysis. Ann Stat 18:1501–1555

Hanley JA, Miettinen OS (2009) Fitting smooth-in-time prognostic risk functions via logistic regression. Int J Biostat. doi:10.2202/1557-4679.1125

Kalbfleisch JD, Prentice RL (2002) The statistical analysis of failure time data, 2nd edn. Wiley, New York

Langholz B, Goldstein L (2001) Conditional logistic analysis of case–control studies with complex sampling. Biostatistics 2:63–84

Mantel N (1973) Synthetic retrospective studies and related topics. Biometrics 29:479–486

Miettinen OS (2011) Epidemiological research: terms and concepts. Springer, Dordrecht

Saarela O, Arjas E (2015) Non-parametric Bayesian hazard regression for chronic disease risk assessment. Scand J Stat 42:609–626

Saarela O, Hanley JA (2015) Case-base methods for studying vaccination safety. Biometrics 71:42–52

Whitaker HJ, Hocine MN, Farrington CP (2009) The methodology of self-controlled case series studies. Stat Methods Med Res 18:7–26

Woolf B (1955) On estimating the relationship between blood group and disease. Hum Genet 19:251–253