Sequential case series analysis for pharmacovigilance

Mounia N. Hocine[1], Patrick Musonda[1],
Nick J. Andrews[2] and C. Paddy Farrington[1]
[1]Department of Statistics, The Open University,
Milton Keynes MK7 6AA, UK
[2]Statistics, Modelling and Bioinformatics Department, Centre for Infections,
London NW9 5EQ, UK

July 27, 2007

Abstract

The self-controlled case series method is used to evaluate drug safety, particularly the safety of paediatric vaccines with respect to rare adverse reactions. We propose a group sequential version of the method for prospective surveillance of drug safety. Although the method is more widely applicable, we focus on the surveillance of new vaccines. We develop a method based on the sequential probability ratio test applied at pre-determined surveillance intervals. We investigate the properties of the method analytically in a simple setting and by extensive simulations in more realistic scenarios. The method is applied to data on influenza vaccine and Bell's Palsy, and MMR vaccine and bleeding disorders.

Key words: self-controlled case series, sequential probability ratio test, surveillance, pharmacovigilance, vaccine safety

Corresponding Author: Paddy Farrington, Department of Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK.

Email: c.p.farrington@open.ac.uk.

This research was supported with funding from the Medical Research Council, the Engineering and Physical Sciences Research Council and the Fondation pour la Recherche Médicale.

1 Introduction

Pharmaceutical drugs undergo comprehensive testing for safety and efficacy before they are used on a large scale. Testing usually includes evaluation in randomized controlled clinical trials. However, even large trials lack power to detect association with rare events. Thus, safety monitoring is required after a new drug is licensed. In this paper we consider the statistical problem of monitoring prospectively (that is, as adverse events accrue) the safety of a new drug.

Although the methods described here apply more widely, we focus throughout on vaccines. Surveillance of the safety of vaccines is all the more important that they are often administered very widely to healthy individuals, and that public concern over vaccine safety can seriously undermine a vaccination programme, as occurred in recent years with measles mumps and rubella (MMR) vaccine and autism. It is therefore important to have in place effective surveillance systems that can detect problems early, or provide accumulating evidence of safety.

Surveillance systems for vaccine-associated adverse events include voluntary notification systems such as the Yellow Card system in the UK, run by the Medicines and Healthcare products Regulatory Agency (MRHA), and the Vaccine Adverse Event Reporting System (VAERS) in the United States (Chen et al 1994). In such systems, notification of an event is triggered by the suspicion that it is related to the vaccine. As a result, data from such systems are usually used primarily for hypothesis generation, though formal methods of analysis have been developed (Evans et al 2001, Rothman et al 2004, Waller et al 2004). Alternatively, surveillance using cohort and case-control methods may be undertaken using data from population based registers, such as the General Practice Research Database in the UK or the Vaccine Safety Datalink (VSD) in the US (Chen et al 1997). In 1995, new analysis techniques using the self-controlled case series method, which uses only cases, were proposed. These methods are particularly suitable for use with databases, including administrative databases detailing hospital episode statistics (Farrington et al 1995).

More recently, interest has focused on the rapid detection of emerging problems with new vaccines, by using sequential analysis methods applied to accumulating data from the VSD (Davis et al 2005). This approach is based on a risk-adjusted sequential probability ratio test (SPRT), derived using a cohort model in which event rates within 4 weeks of vaccination are compared with pre-vaccination rates. In the present paper, we develop sequential methodology for prospective monitoring of vaccine safety, and pharmacovigilance more widely, based on the self-controlled case series method.

This method is an alternative to traditional epidemiological designs, such as cohort and case-control methods, for investigating the association between an age-varying exposure and an acute event (Farrington 1995, Farrington & Whitaker 2006). Its two main advantages are that it requires only cases, and adjusts for multiplicative age-constant covariates and frailties. Consequently, the self-controlled case series method is easy and rapid to use, and avoids confounding, in particular confounding by indication. The method has been widely used in pharmacoepidemiology, most frequently to investigate the safety of paediatric vaccines, though there have been applications to drugs other than vaccines (Whitaker et al 2006; Hocine et al 2005).

These two features of the case series method recommend it for use in pharmacovigilance, as well as pharmacoepidemiology. In particular, the fact that the method requires only cases means that it can readily be applied to case databases, such as those recording hospital admissions and general practice consultations, without the need to sample suitable controls, and hence avoiding the selection biases inherent in such procedures.

However, using the method for surveillance, as distinct from epidemiological investigations, requires resolution of an awkward conundrum: while the surveillance is prospective, the method is retrospective. The case series method is derived from a cohort model by conditioning on the total number of events experienced by each individual in the cohort over a pre-determined observation period. Thus the total number of events and, crucially, the entire exposure history, are required over the observation period.

In consequence, a wholly sequential case series method, updated at each new event, appears not to be feasible. Instead, we consider a group sequential version of the method, applied at pre-specified, generally short surveillance intervals. We use the SPRT, applied to the log-likelihood of the self-controlled case series model. Our aim in this paper is not to develop new sequential methodology, but to apply the well-known SPRT method to the case series model, and investigate the feasibility of using the case series SPRT for surveillance of vaccine safety.

In Section 2, we describe the design of the case series SPRT. We derive expressions for the likelihood ratio statistic upon which the test is based, and discuss two contrasting methods of age adjustment. In Section 3 we investigate the properties of the case series SPRT analytically in a simple case, so as to gain a qualitative understanding of how the performance of the method is likely to vary as design parameters change. More complex and realistic scenarios are investigated by simulation in Section 4. In Section 5 we apply the proposed methods to data on influenza vaccine and Bell's Palsy, and to data on MMR vaccine and bleeding disorders. In Section 6 we discuss relaxing the assumptions and consider some variations on the basic model developed in the main body of the paper. We offer some recommendations in Section 7.

2 The case series SPRT

Throughout the paper, we shall focus on acute events, that is, events with a clearly defined onset, and transient age-varying exposures. We assume that the distribution of exposures is unaffected by previous events. The typical application we have in mind is to vaccine exposures, where the adverse event is not death or a contra-indication to vaccination. While the case series method may be used for long-term exposure effects, our focus here is on effects of short duration, typically a few days or weeks. The simplest case is when there is a single risk period of duration d stretching from the age of exposure v to age $v + d$. Outside this risk period, events are assumed to arise as a Poisson process with age-dependent rate $\lambda_i\gamma(t)$, where λ_i is the rate at some reference age $t = t_0$, which may vary between individuals i, and $\gamma(t)$ is an age-dependent effect common to all individuals. For each individual, within the risk period, the rate is increased by the multiplicative factor ρ to $\lambda_i(t; v) = \rho\lambda_i\gamma(t)$. The quantity ρ is the relative incidence.

In Section 6, we briefly consider some other settings. Specifically, we consider events that do not arise in a Poisson process, including deaths, multiple risk periods, and events that may alter subsequent exposures.

2.1 Outline of the surveillance system

In this subsection we briefly describe how the surveillance system is set up. We assume that surveillance begins at time 0. The analysis is repeated at calendar time intervals of fixed duration s. Thus self-controlled case series analyses are undertaken at calendar times s, $2s$, $3s$,... . No maximum number of analyses is specified. We base the system on the sequential probability ratio test developed by Wald (1945, 1947), implemented here as a group sequential test (Jennison and Turnbull 2000); for a brief review with applications to monitoring in a medical context see Grigg et al (2003).

Null and alternative hypotheses are defined as follows:

$$\begin{aligned} H_0 &: \rho = 1 \\ H_1 &: \rho = \rho_A \end{aligned}$$

where $\rho_A > 1$ is the value of the relative incidence that we wish to detect. At the k^{th} step, if the test has not terminated, the logarithm of the case series likelihood ratio test statistic Λ_k is calculated, and its running total is obtained:

$$Z_0 = 0, Z_k = Z_{k-1} + \Lambda_k.$$

Denote the true type I and type II error probabilities by α and β. Thus:

$$\begin{aligned} \alpha &= \text{Probability of rejecting } H_0 \text{ when } H_0 \text{ is true,} \\ 1-\beta &= \text{Probability of rejecting } H_0 \text{ when } H_1 \text{ is true.} \end{aligned}$$

Pre-determined test boundaries $\log A$ and $\log B$ are conventionally specified in terms of nominal error probabilities α^* and β^*, as follows:

$$\begin{aligned} A &= \frac{\beta^*}{1-\alpha^*} \\ B &= \frac{1-\beta^*}{\alpha^*}. \end{aligned}$$

The nominal error probabilities are upper bounds on the actual values: $\alpha^* \geq \alpha$ and $\beta^* \geq \beta$, the discrepancy being attributable to the discreteness of the observations (and, in our case, of the grouped nature of the data). Methods exist for calculating more accurate test boundaries, but these will not be considered here.

If $Z_k < \log A$ then the null hypothesis H_0 (corresponding to no vaccine effect) is accepted. If $Z_k > \log B$ then the alternative hypothesis H_1 is accepted. In either case, the sequential test terminates at step k. If $\log A \leq Z_k \leq \log B$, then no decision is made and the procedure is repeated at step $k+1$.

In the context of clinical trials, group sequential methods have been developed with other types of boundaries than the parallel boundaries we use here. In particular, boundaries may be chosen so as to ensure that the trial terminates within a given time period. However, in such situations, the investigator has some control over the recruitment rate to the trial. This is not the case for surveillance, and for this reason open boundaries are preferred.

2.2 Data required

It is important in what follows to distinguish between two time lines: calendar time, which determines the surveillance intervals, and age, which is the time line for analysis in the case series method. We denote age by t.

At the k^{th} analysis, events are ascertained within the calendar time period $((k-1)s, ks]$ (we shall refer to this as the k^{th} surveillance interval). We are often only interested in events arising within a pre-determined age range $(a, b]$. For example, if the exposure of interest is primary MMR vaccination, which in the UK is administered in the second year of life, then it makes sense to restrict attention to events arising within the second year of life. A case is an individual with one or more events. Let C_k denote the set of cases ascertained in the k^{th} surveillance interval. Thus:

$$C_k \quad : \quad \text{all individuals with one or more events within the } k^{th} \text{ surveillance interval and within the age range } (a, b].$$

Let n_k denote the number of cases ascertained in the k^{th} surveillance interval. Let $i = 1, ..., n_k$ label these cases. The observation period for case i is denoted $(a_{ik}, b_{ik}]$, and is the age range at risk for experiencing an event to be counted in C_k:

$$(a_{ik}, b_{ik}] = (a, b] \cap (\text{age of case } i \text{ at time } (k-1)s, \text{ age of case } i \text{ at time } ks].$$

Note that $(a_{ik}, b_{ik}]$ is nonempty since it contains at least one event experienced by case i.

A case i can have more than one event in $(a_{ik}, b_{ik}]$. However, one of the consequences of the Poisson assumption (to be relaxed later) is that multiple events are independent. Hence we can assume, without loss of generality, that each case in C_k has a single event, at age t_{ik} say. The reason is as follows. A case experiencing m events at times $t_1, ..., t_m$ may be replaced by m cases, the first with event at age t_1, the second with event at age t_2, and so on, and all with observation period $(a_{ik}, b_{ik}]$. Thus, without loss of generality, n_k is also the number of events ascertained in the k^{th} surveillance interval; this simplifies the presentation later on.

Finally, exposures are ascertained for each of the n_k cases. For a risk period of duration d starting from time of exposure, exposures must be sought in the age interval $(a_{ik} - d, b_{ik}]$ for case i to ensure that all time in $(a_{ik}, b_{ik}]$ can be classified as at risk or otherwise. Denote v_{ik} the vector of ages at exposure for case i in surveillance interval k.

This completes the data definition for surveillance interval k. The detailed and rather formal explanation given above is necessary to ensure a correct specification of the data required. However, it can be summed up informally thus: ascertain events that occur in individuals aged $(a, b]$ within the k^{th} surveillance interval, define the observation periods for these cases based on age and time

boundaries, and ascertain exposure status at all times within the observation periods. The selection of cases and calculation of observation periods may readily be programmed within standard databases and spreadsheets.

2.3 Likelihood ratio test

At the end of each surveillance interval k, the cases C_k together with their event times and exposure histories are used to obtain the log likelihood ratio test statistic Λ_k. The self-controlled case series log-likelihood contribution for a case i with observation period $(a_{ik}, b_{ik}]$, single event time t_{ik} and exposure history v_{ik} is

$$l_i = I(t_{ik}; v_{ik}) \log \rho + \log \gamma(t_{ik}) - \log \int_{a_{ik}}^{b_{ik}} \{\rho I(t; v_{ik}) + 1 - I(t; v_{ik})\} \gamma(t) dt \quad (1)$$

where $I(t; v)$ is an exposure indicator function, taking the value 1 if an individual with exposure vector v is exposed at age t and 0 otherwise. Note that the individual effects λ_i do not feature in this log likelihood, which involves only the relative incidence ρ and the age effect $\gamma(t)$.

Suppose first that $\gamma(t)$ is known. Let

$$h_{ik}(t) = \frac{\gamma(t)}{\int_{a_{ik}}^{b_{ik}} \gamma(u) du}$$

denote the within-individual density of age at event. It follows from (1) that the log likelihood ratio test contribution from a single case is

$$\Lambda_{ik} = I(t_{ik}; v_{ik}) \log \rho_A - \log (\rho_A \omega_{ik} + 1 - \omega_{ik})$$

where

$$\omega_{ik} = \int_{a_{ik}}^{b_{ik}} I(t; v_{ik}) h_{ik}(t) dt$$

is the expected proportion of the observation period at risk for the ith case in surveillance interval k, the expectation being taken with respect to h_{ik}. Thus, when the age effect $\gamma(t)$ is known,

$$\Lambda_k = r_k \log \rho_A - \sum_{i=1}^{n_k} \log (\rho_A \omega_{ik} + 1 - \omega_{ik})$$

where r_k is the number of events within a risk period, among the n_k events in surveillance interval k.

In practice, it is unlikely that the age effect $\gamma(t)$ is known, though assuming it is establishes a useful benchmark against which to evaluate more general methods. When $\gamma(t)$ is not known, it must be estimated. If a long series of data

are available prior to the introduction of the vaccine, then $\gamma(t)$ can be replaced by an estimate based on these earlier data, and the calculations then continue as if $\gamma(t)$ were known with complete accuracy.

If no such pre-vaccination data exist, a natural way to proceed is to replace the log-likelihood by the profile log-likelihood for ρ (Barndorff-Nielsen & Cox 1994), as follows. Let $\gamma^{1k}(t)$ denote the maximum likelihood estimate of $\gamma(t)$ when $\rho = \rho_A$, and $\gamma^{0k}(t)$ the maximum likelihood estimate of $\gamma(t)$ when $\rho = 1$, both estimated from the data available in interval k. Similarly, let h^0_{ik} and h^1_{ik} denote the densities obtained by replacing $\gamma(t)$ with $\gamma^{0k}(t)$ and $\gamma^{1k}(t)$, respectively. Then the profile likelihood ratio statistic, now adorned with a tilda to denote that profiles are used, is

$$\widetilde{\Lambda}_k = r_k \log \rho_A - \sum_{i=1}^{n_k} \log\left(\rho_A \omega^1_{ik} + 1 - \omega^1_{ik}\right) + \log \frac{h^1_{ik}(t_{ik})}{h^0_{ik}(t_{ik})}$$

where

$$\omega^1_{ik} = \int_{a_{ik}}^{b_{ik}} I(t; v_{ik}) h^1_{ik}(t) dt.$$

Asymptotically as n_k grows large, the profile likelihood ratio $\widetilde{\Lambda}_k$ has similar distributional properties to Λ_k (Barndorff-Nielsen & Cox 1994). However, asymptotic results are unlikely to be relevant in the present application, and hence later in the paper the effect of profiling out the nuisance age effect parameter $\gamma(t)$ is investigated by simulation.

Other estimation schemes are possible when $\gamma(t)$ is not known. For example, at surveillance interval k, the value Z_k may be obtained directly using the profile log-likelihood for ρ based on the data for all periods $1, 2, ..., k$. In this case, standard asymptotic theory is valid as k becomes large. However, this scenario is not particularly relevant either as we generally require k to be small, and furthermore the increments Λ_k are no longer independent. We did not consider this approach further.

3 Analytical investigation in a simple case

The investigator has some control over the design parameters, which include the surveillance interval length s, the risk period d, the alternative hypothesis relative incidence ρ_A, and the boundary values $\log A$ and $\log B$. Also, the order of magnitude of the frequency of the adverse events under consideration, and of the proportion likely to be exposed during a given period, may be known. The practical issue of interest is then to design the surveillance system in such a way that it is sensitive and has a high probability of terminating within a reasonable time. This requires some understanding of how the performance of the system is affected by varying the key design parameters.

In order to gain some qualitative understanding of how performance depends on the design parameters, we consider a simplified scenario analytically. We assume that it is known that there is no age effect, so that $\gamma(t) = 1$. We

also assume that all observation periods $(a_{ik}, b_{ik}]$ are of length s; this is approximately true when s is small compared to $b - a$. Finally, we assume that, when the true relative incidence is ρ and the surveillance interval length is s, the expected proportion of time under exposure is ω in a proportion $\pi(\rho, s, \omega)$ of cases (which we call exposed cases) and 0 in the remainder (the unexposed cases). Finally, we assume that the number of events in a surveillance interval of duration s, when the true relative incidence is ρ and the proportion of time exposed (in vaccinees) is ω, is distributed Poisson with mean $\lambda(\rho, s, \omega)$.

Since there is no age effect, unexposed cases do not contribute any information so can be ignored. Let m_k denote the number of events in exposed cases in surveillance interval k. Then m_k is distributed Poisson with mean $\pi(\rho, s, \omega)\lambda(\rho, s, \omega)$, and the log-likelihood ratio statistic in surveillance interval k is

$$\Lambda_k = r_k \log \rho_A - m_k \log (\rho_A \omega + 1 - \omega)$$

where, conditionally on m_k, r_k is binomial $B(m_k, \rho\omega(\rho\omega + 1 - \omega)^{-1})$.

3.1 Operating characteristic

We first investigate the operating characteristic of the SPRT in this simplified setting. Fix the values ρ_A, s and ω. The operating characteristic $OC(\rho)$ represents the probability of accepting H_0, when the true relative incidence is ρ. As only qualitative insights are sought, we derive $OC(\rho)$ using Wald's approximation (Wald 1947). This approximation improves as the effect of discreteness reduces.

For a single surveillance interval k, the case series likelihood ratio is

$$\frac{L_{1k}}{L_{0k}} = \exp(\Lambda_k) = \frac{(\rho_A)^{r_k}}{(\rho_A \omega + 1 - \omega)^{m_k}}$$

which can be written as

$$\frac{L_{1k}}{L_{0k}} = \left(\frac{\rho_A}{\rho_A \omega + 1 - \omega}\right)^{r_k} \times \left(\frac{1}{\rho_A \omega + 1 - \omega}\right)^{m_k - r_k}. \tag{2}$$

As m_k is Poisson and r_k is binomial conditionally on m_k, then, unconditionally, r_k and $m_k - r_k$ are independent Poisson:

$$r_k \sim P\left(\pi(\rho, s, \omega)\lambda(\rho, s, \omega)\frac{\rho\omega}{\rho\omega + 1 - \omega}\right) \equiv P(\phi_1)$$

and

$$m_k - r_k \sim P\left(\pi(\rho, s, \omega)\lambda(\rho, s, \omega)\frac{1 - \omega}{\rho\omega + 1 - \omega}\right) \equiv P(\phi_2).$$

Let $f(r, m - r; \rho)$ denote the joint distribution of these two independent Poisson variates, and let E_ρ denote expectation with respect to f. Let $h \equiv h(\rho)$ denote the non-zero value such that

$$E_\rho\left[\left(\frac{L_{1k}}{L_{0k}}\right)^h\right] = 1. \tag{3}$$

Note that the likelihood ratio is based on conditional likelihoods, whereas the expectation is with respect to the full density f. However, Wald's method still works, and in particular it is easily checked that under H_0, $h = 1$ and under H_1, $h = -1$. The operating characteristic depends on h (and hence on ρ) as follows:

$$OC(\rho) = \begin{cases} (1-A^h)(B^h - A^h)^{-1}, \text{if } h \neq 0; \\ (\log A)(\log A - \log B)^{-1}, \text{if } h = 0. \end{cases}$$

Under the simplified model, using equation (2) and the fact that r_k and $m_k - r_k$ are independent, equation (3) becomes:

$$E_\rho\left[\theta_1(h)^{r_k}\right] \times E_\rho\left[\theta_2(h)^{m_k - r_k}\right] = 1$$

where

$$\begin{aligned} \theta_1(h) &= \left(\frac{\rho_A}{\rho_A\omega + 1 - \omega}\right)^h, \\ \theta_2(h) &= \left(\frac{1}{\rho_A\omega + 1 - \omega}\right)^h. \end{aligned}$$

It follows that

$$\exp\left[\phi_1\left\{\theta_1(h) - 1\right\}\right] \times \exp\left[\phi_2\left\{\theta_1(h) - 1\right\}\right] = 1.$$

Solving for ρ, we find the following expression for ρ as a function of h:

$$\rho(h) = \frac{(\omega - 1)\left(\theta_2(h) - 1\right)}{\omega\left(\theta_1(h) - 1\right)}.$$

As $h \to 0$, this becomes

$$\frac{(1-\omega)\log\left(\rho_A\omega + 1 - \omega\right)}{\omega\log[\rho_A\left(\rho_A\omega + 1 - \omega\right)^{-1}]}.$$

For example, taking the particular values $\rho_A = 3$, $\omega = \frac{1}{12}, \alpha = \beta = 0.05$, the true relative incidence ρ and the operating characteristic $OC(\rho)$ may be calculated for different choices of h, with results as shown in Table 1.

Table 1. Values of ρ and OC for different values of h.

h	$-\infty$	-1	0	1	$+\infty$
ρ	$+\infty$	3	1.8	1	0
$OC(\rho)$	0	0.05	0.5	0.95	1

The operating characteristic curve for different values of ω is shown in Figure 1(a). Clearly, varying ω has very little impact on $OC(\rho)$. Figure 1(b) shows $OC(\rho)$ for different values of ρ_A. Not surprisingly, the choice of ρ_A has a very big influence. For example, if $\rho_A = 5$ but in fact $\rho = 2$ then there is a high probability of accepting the null hypothesis.

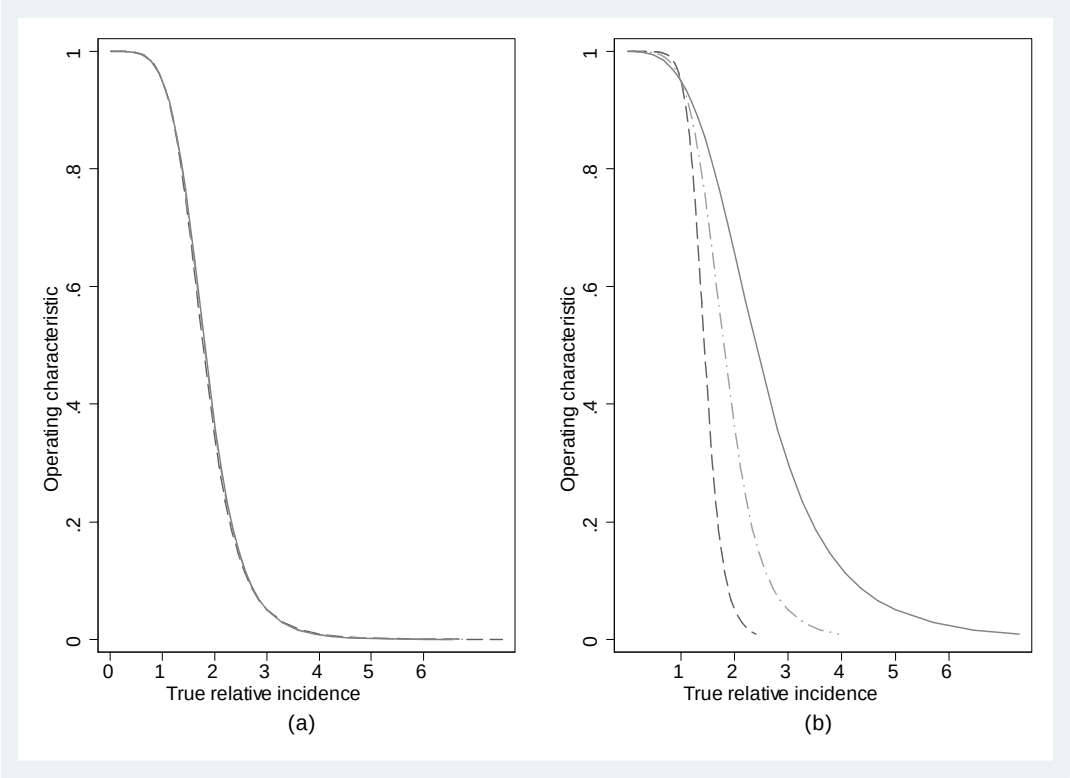

Figure 1: Operating characteristic $OC(\rho)$: (a) $\omega = 1/6$ (- - -), $1/12$ (- – -) and $1/24$ (—) with $\rho_A = 3$; (b) $\rho_A = 2$ (- - -), 3 (- – -) and 5 (—) with $\omega = 1/12$.

3.2 Mean time to decision

Let N denote the random number of surveillance intervals until a boundary is crossed. Thus the total time to reach a decision (in years) is Ns. Following Wald (1947), the expected value of N is given approximately by:

$$E(N;\rho,s,\omega) \simeq \frac{\log A\ [1-OC(\rho)] + \log B\ OC(\rho)}{E_\rho(\Lambda_k)}.$$

For the simplified model, the expectation $E_\rho(\Lambda_k)$ is:

$$\begin{aligned} E_\rho(\Lambda_k) &= E_\rho(r_k)\,(\log \rho_A) - E_\rho(m_k)\log(\rho_A\omega + 1 - \omega) \\ &= \pi(\rho,s,\omega)\lambda(\rho,s,\omega)\left[\frac{\rho\omega}{\rho\omega+1-\omega}\,(\log\rho_A) - \log(\rho_A\omega+1-\omega)\right]. \end{aligned}$$

It follows that

$$E(N;\rho,s,\omega) = \frac{\log A\ [1-OC(\rho)] + \log B\ OC(\rho)}{\pi(\rho,s,\omega)\lambda(\rho,s,\omega)\left[\rho\omega(\rho\omega+1-\omega)^{-1}\,(\log\rho_A) - \log(\rho_A\omega+1-\omega)\right]}.$$

The expectation $E(N;\rho,s,\omega)$ depends on ρ, ρ_A, ω and on the product $\pi(\rho,s,\omega)\lambda(\rho,s,\omega)$. The functions $\pi(\rho,s,\omega)$ and $\lambda(\rho,s,\omega)$ are difficult to specify. We thus fix ρ, s and ω and study how $E(N;\rho)$ varies in relative terms as a function of ρ_A. Specifically, we consider the ratio

$$R(\rho_A,\rho) = \frac{E(N;\rho_A;\rho,s,\omega)}{E(N;\rho_A=2;\rho,s,\omega)}.$$

Figure 2 illustrates the relative time to decision $R(\rho_A,\rho)$ as a function of the alternative hypothesis relative incidence ρ_A for three values of ρ.

The vertical location of the three curves is determined by our choice of $\rho_A = 2$ as reference, and hence is arbitrary. However, the shapes of the curves are of interest. These indicate that the expected time to decision rises rapidly as ρ_A declines for $\rho_A < 2$. This suggests that 'playing safe' by choosing a lower value ρ_A carries the risk of greatly increasing the time to decision when $\rho_A < 2$. In contrast, if say $\rho \geq 3$, then choosing $\rho_A = 3$ is not unreasonable as the curves flatten out.

4 Simulation study

In the previous section, we considered a very simplified scenario and derived approximations to the operating characteristic and the time to decision. These approximations improve as the effect of discreteness reduces, namely as the expected number of events per surveillance interval reduces and the time to decision increases. Such calculations, though approximate, are useful in that

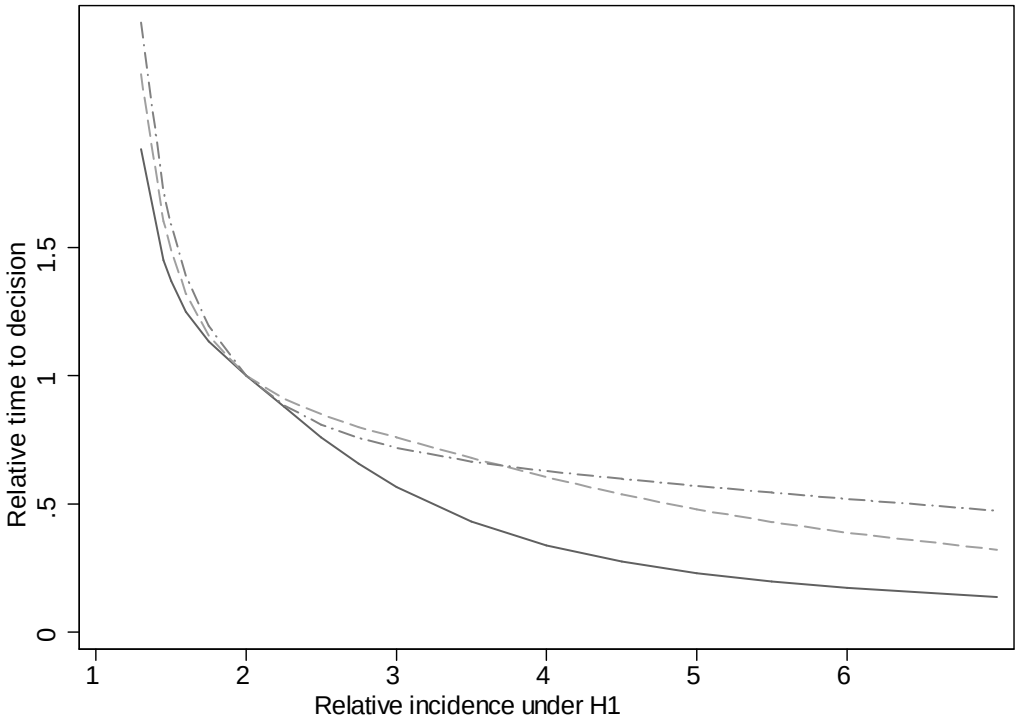

Figure 2: Relative time to decision for $\rho = 5$ (- – -), 3 (- - -) and 2 (—), with $\rho_A = 2$.

they help to understand in qualitative terms the properties of the system and how these vary with the parameters.

However, analytical investigations are impractical in more general and realistic settings. In particular, it is not feasible to investigate the effect of age adjustments, or how changing the duration of the surveillance interval alters the time to decision. We thus turn to simulations to investigate the performance of the surveillance system in practical settings.

4.1 Design of the simulation study

We considered adverse events arising within the age group $(a, b] = (0, 364]$ days. We assumed that the distribution of age at exposure $f(v)$ in the population is proportional to a gamma density with mean 120 days, shape parameter 1.2 and probability of remaining unexposed at 364 days equal to 0.1.

For a given annual incidence λ for events within the age group $(0, 364]$ days, and surveillance interval length s, we simulated sequences of event counts $n_k \sim$Poisson(λs).

Consider surveillance interval k. For each of the n_k events occurring within this interval, we simulated the calendar time of occurrence within the surveillance interval $((k-1)s, ks]$ by means of a uniform distribution, and the conditional distribution of exposure, given that an event has occurred within the age group $(0, 364]$ days. This conditional distribution is

$$f(v|T \in (0, 364]) = \frac{f(v)\int_0^{364} \{\rho I(t; v) + 1 - I(t; v)\}\gamma(t)dt}{\int_0^{\infty}\int_0^{364} \{\rho I(t; v) + 1 - I(t; v)\}\gamma(t)f(v)dtdv}.$$

We simulated the age at event, conditional on the age at exposure v just simulated, using the density

$$\frac{\{\rho I(t; v) + 1 - I(t; v)\}\gamma(t)}{\int_0^{364} \{\rho I(t; v) + 1 - I(t; v)\}\gamma(t)dt}.$$

Given the simulated calendar time and simulated age at event, we then obtained the case's date of birth and hence the observation period for the case, namely the intersection of $(0, 364]$ and the interval (age at $(k-1)s$, age at ks]. These procedures were repeated for each of the n_k events in the k^{th} surveillance interval.

Case series likelihood ratios were then calculated using the data on the n_k cases, which comprise their observation periods, age at event, and exposure histories. Two ratios were obtained: the value Λ_k assuming the true age effect $\gamma(t)$ known, and the profile log-likelihood ratio $\widetilde{\Lambda}_k$ in which $\gamma(t)$ is obtained by maximum likelihood under the null and alternative hypotheses. The traces $\Lambda_1 + ... + \Lambda_k$ and $\widetilde{\Lambda}_1 + ... + \widetilde{\Lambda}_k$ were computed.

This process was repeated for $k = 1, 2, ...$ until both of the traces had crossed one of the test boundaries. For each of the two traces, the value of k at the first crossing, and which boundary was crossed, were recorded. A total 1000 traces of each type were simulated for each combination of parameters. The proportions of traces crossing each boundary, and the mean and standard deviation of the time to decision (in years) were obtained.

The nominal error probabilities were taken to be $\alpha^* = \beta^* = 0.05$, giving test boundaries $\log A = 2.94$, $\log B = -2.94$. The parameters used in the simulations were as follows:

$$\begin{aligned} \rho_A, \rho &= 2, 3, 5, 10 \\ s &= 0.25, 0.5, 1 \text{ year} \\ d &= 1, 2, 4 \text{ weeks} \\ \lambda &= 25, 50, 100 \text{ cases per year} \end{aligned}$$

The age effect $\gamma(t)$ was taken as piecewise constant on thirteen 28-day intervals. Three age effects were tried: (a) constant age effect, (b) age effect increasing by the factor 1.2 in successive age groups, and (c) age effect decreasing by the factor 1.2 in successive age groups.

4.2 Results of the simulation study

We began by investigating a baseline scenario with $\rho = \rho_A$, $s = 0.5$ year, $d = 2$ weeks, and $\lambda = 50$ cases per year. The results, for different values of ρ_A, three age effects $\gamma(t)$, and two methods of age adjustment ($\gamma(t)$ assumed known, and profiled out as described above), are shown in Table 2. The power, that is the probability that the SPRT crosses the upper boundary, is always in excess of the nominal value $1 - \beta^* = 0.95$. The excess power is least when the mean time to decision is greatest, that is, when the effect of discreteness is least marked, as expected. The mean time to decision is only marginally longer when the age effect is profiled out, compared to when it is known (for $\rho = 10$, smaller values were obtained in some cases, attributable to Monte Carlo error). In contrast, there is substantial variation in the mean time to decision, and its standard deviation, according to which type of age effect is present. The mean and standard deviation of the decision times are lowest for decreasing age effects, since most events occur close to the age at vaccination (which is right-skew), thereby increasing the information available on the exposure effect.

Table 2. Simulation results for the baseline scenario: power, mean and standard deviation (SD) of the time to decision (years).

		Age effect profiled out			Age effect known		
	Type of	Power	Time to decision		Power	Time to decision	
$\rho = \rho_A$	age effect	$1-\beta$	Mean	SD	$1-\beta$	Mean	SD
2	constant	0.969	6.30	2.73	0.964	5.98	2.51
	↗	0.962	10.2	8.28	0.953	9.42	7.04
	↘	0.964	4.67	1.89	0.958	4.48	1.92
3	constant	0.978	2.33	1.56	0.975	2.24	1.53
	↗	0.975	3.65	2.29	0.968	3.39	2.24
	↘	0.979	1.68	1.15	0.978	1.65	1.04
5	constant	0.988	1.04	0.91	0.984	1.01	0.87
	↗	0.981	1.43	1.08	0.979	1.40	1.02
	↘	0.989	0.87	0.42	0.987	0.86	0.37
10	constant	0.995	0.59	0.22	0.993	0.60	0.23
	↗	0.995	0.74	0.38	0.986	0.73	0.35
	↘	0.998	0.56	0.18	0.998	0.57	0.19

Table 3. Simulation results: power, means and standard deviation (SD) of the time to decision (years) for $s = 0.5$ years, $d = 2$ weeks and $\lambda = 25$ cases per year.

		Age effect profiled out			Age effect known		
	Type of	Power	Time to decision		Power	Time to decision	
$\rho = \rho_A$	age effect	$1-\beta$	Mean	SD	$1-\beta$	Mean	SD
2	constant	0.961	12.7	9.08	0.956	11.4	8.79
	↗	0.963	19.7	14.4	0.951	17.2	12.6
	↘	0.975	8.87	6.35	0.969	8.31	6.10
3	constant	0.972	4.35	3.14	0.966	3.92	2.77
	↗	0.975	6.75	4.71	0.973	6.00	4.32
	↘	0.978	3.29	2.45	0.971	3.09	2.34
5	constant	0.980	1.86	1.22	0.976	1.76	1.22
	↗	0.981	2.79	1.94	0.967	2.480	1.81
	↘	0.982	1.47	0.97	0.978	1.41	0.95
10	constant	0.992	0.86	0.46	0.983	0.84	0.48
	↗	0.987	1.17	0.75	0.978	1.07	0.68
	↘	0.993	0.77	0.39	0.997	0.75	0.41

Table 3 gives the results for $\rho = \rho_A$, $s = 0.5$ year, $d = 2$ weeks, and $\lambda = 25$ cases per year. The times to decision have larger mean and standard deviation compared to the results in Table 2 for $\lambda = 50$, and the difference between the results when the age effect $\gamma(t)$ is known compared to when it is unknown is more marked, especially for $\rho = \rho_A = 2$. When $\lambda = 100$, all other parameters remaining the same, the times to decision have smaller mean and standard deviation than those in Table 2 (not shown; typically they are less than 60% of the values in Table 2).

Table 4 shows the effect on the time to decision of varying the risk period length d; in these simulations, $\rho = \rho_A$, $s = 0.5$ year, $\lambda = 50$ cases per year, there was no underlying age effect, that is $\gamma(t) = 1$, and the profile method was used to adjust for age.

Table 4. Mean and standard deviation (SD) of time to decision (years) for different risk periods d.

d (weeks)	1		2		4	
$\rho = \rho_A$	Mean	SD	Mean	SD	Mean	SD
2	10.9	8.19	6.30	2.73	3.89	2.77
3	4.09	3.08	2.33	1.28	1.55	0.99
5	1.60	1.11	1.04	0.91	0.78	0.41
10	0.77	0.43	0.59	0.22	0.54	0.14

The mean and standard deviation of the time to decision decreases as the length of the risk period increases (within this range). This is not unexpected, since the information about the exposure effect is greatest when the expected number of events in and outside the risk period is similar (Musonda et al 2007).

We also studied the effect of varying the length of the surveillance interval s. Table 5 shows the results, with $\rho = \rho_A$, $d = 2$ weeks, $\lambda = 50$ cases per year, no underlying age effect, and the profile method.

Table 5. Mean and standard deviation (SD) of time to decision (years) for different surveillance interval lengths s.

s (years)	0.25		0.5		1	
$\rho = \rho_A$	Mean	SD	Mean	SD	Mean	SD
2	7.00	5.08	6.30	2.73	6.49	4.43
3	2.38	1.56	2.33	1.28	2.52	1.59
5	1.08	0.74	1.04	0.91	1.30	0.58
10	0.46	0.25	0.59	0.22	1.03	0.16

As shown in Table 5, the relationship between the time to decision and the length of the surveillance interval is complex. On the one hand, shorter surveillance intervals reduce the impact of the discreteness of the inspections and hence will tend to reduce time to decision, especially when this is rapid. For example, in Table 5, the minimum mean time to decision when $\rho = 10$ is when $s = 0.25$ year. However, increasing the surveillance interval increases the amount of information about the exposure effect, since the observation periods

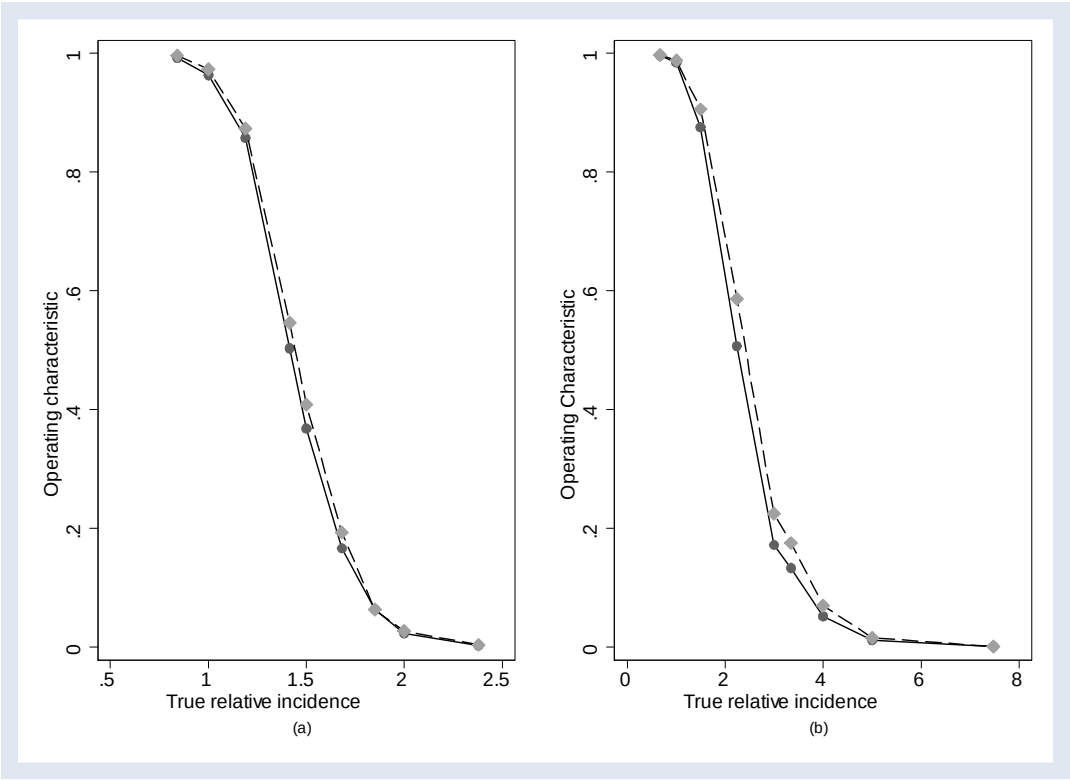

Figure 3: Simulated operating characteristic curve $OC(\rho)$, with age effect known (- - -) and profiled out (——): (a) $\rho_A = 2$; (b) $\rho_A = 5$.

in the case series analysis are longer and hence the chance of overlapping with a risk period is greater. For example, an individual who has an event in the first 0.25 years, but is exposed in the next 0.25 years, contributes more information on the exposure effect if $s = 0.5$ (covering both event and exposure) than if $s = 0.25$. Thus, in Table 5, increasing s from 0.25 to 0.5 reduces the mean time to decision when $\rho \leq 5$.

So far, all reported simulations have taken $\rho = \rho_A$. We now fix ρ_A and let ρ vary. We set $\lambda = 50$ cases per year, $d = 2$ weeks and $s = 0.5$ year, and assumed that there was no underlying age effect, that is $\gamma(t) = 1$. Figure 3 shows the operating characteristic curve $OC(\rho)$ for $\rho_A = 2$ and 5 when the age effect is known and profiled out.

Figure 3 shows the same qualitative features as displayed in Figure 1. If a large value, $\rho_A = 5$ say, is chosen, then the system is insensitive for detecting lower values of ρ. On the other hand, if a low value $\rho_A = 2$ is chosen, the system remains highly sensitive for detecting larger values of ρ. However, there is a price to pay in time to decision, since more data are required to decide

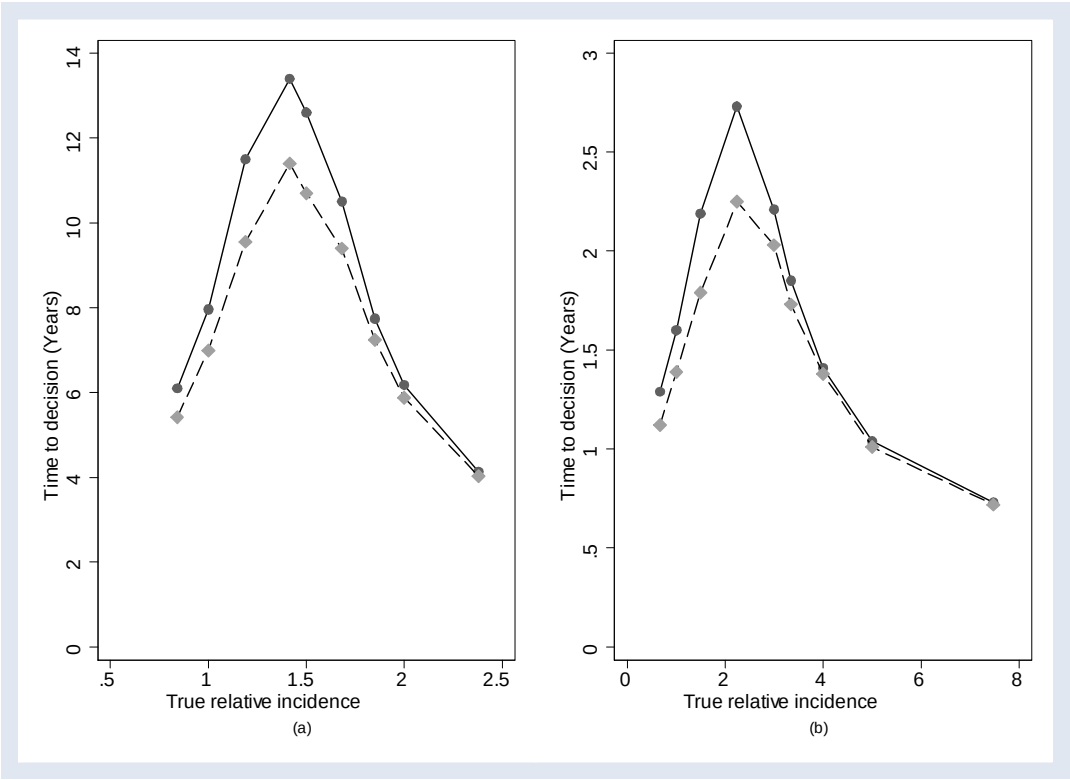

Figure 4: Simulated time to decision, with age effect known (- - -) and profiled out (——): (a) $\rho_A = 2$; (b) $\rho_A = 5$.

between the null and alternative hypotheses.

Figure 4 shows the mean time to decision (in years). If $\rho_A = 2$, then when $\rho = 2$ the expected time to decision is about 6 years - rather long if it was considered important to detect such an event. On the other hand, if $\rho_A = 5$ then all expected times to decision are under 3 years - though of course if $\rho < 5$, the chance of accepting the null hypothesis increases.

5 Examples

In this section we present two examples where we retrospectively monitor vaccine safety data using the SPRT.

5.1 Influenza vaccine and Bell's Palsy

This analysis was undertaken following evidence that an inactivated nasal formulation of the influenza vaccine caused Bell's palsy, an acute facial paralysis affecting the seventh facial nerve. An increased number of Bell's palsy cases were reported after the introduction of this form of influenza vaccine in Switzerland in October 2000. A self-controlled case series analysis was carried out, and the relative incidence within $31-60$ day post-vaccination was estimated to be 35.6, 95%CI $(14.1-89.8)$. Following the Swiss study, an analysis of cases of Bell's palsy and the standard influenza vaccine in the UK was undertaken using data from the GPRD from July $1^{st}, 1992$ to 30^{th} June 2005 (Stowe et al 2006). This dataset includes a total of 2263 episodes of Bell's palsy. Using a self-controlled case series analysis, the estimated relative incidence of Bell's palsy in the 3 months following parenteral inactivated influenza vaccine was 0.92, 95%CI $(0.78-1.08)$.

We reanalysed the GPRD data as if undertaking prospective monitoring using the case series SPRT. We used a six month surveillance interval. We considered two risk periods, $1-60$ days and $1-7$ days after any dose of influenza vaccine. In view of possible temporal confounding from the highly seasonal administration of influenza vaccine, the analyses were performed using a parametric case series model with 12 one month seasonal periods (this represents a slight departure from the model presented above, in which age rather than seasonal effects were allowed for).

Figure 5 shows the SPRT traces using the $1-60$ day risk period for the two values $\rho_A = 5$ and 1.5 and an alert threshold with $\alpha^* = \beta^* = 0.01$. With the alternative hypothesis $\rho_A = 5$, the null hypothesis is accepted at the end of the 3^{rd} surveillance interval, that is, after 18 months. However, if $\rho_A = 1.5$ then the issue remains for 12 years. This illustrates the sensitivity of the SPRT to the choice of the alternative hypothesis relative incidence ρ_A. To gain further insight, we calculated the time to decision for a range of values of ρ_A, and for $1-7$ day and $1-60$ day risk periods. The results, using the alert boundary and a further alarm boundary ($\alpha^* = \beta^* = 0.001$) are shown in Figure 6.

This figure shows that the time to decision drops very rapidly for $\rho_A < 3$, and is insensitive to the choice of ρ_A for larger values, suggesting that the value $\rho_A = 3$ is appropriate in this instance.

5.2 MMR vaccine and bleeding disorders

The causal association between MMR vaccine and Idiopathic Thrombocytopenic Purpura (ITP), a potentially recurrent bleeding disorder, is well known (Farrington et al 1995). A self-controlled case series study found a significantly increased relative incidence in a six-week post-vaccination risk period: 3.01, 95%CI $(1.38, 6.54)$ (Miller et al 2001).

The present data include 41 cases with 48 ITP admissions during nine years between April 1^{st}, 1997 and May 15^{th} 2006. We reanalyse these data prospectively as in a surveillance setting. Figure 7 shows the SPRT traces for different

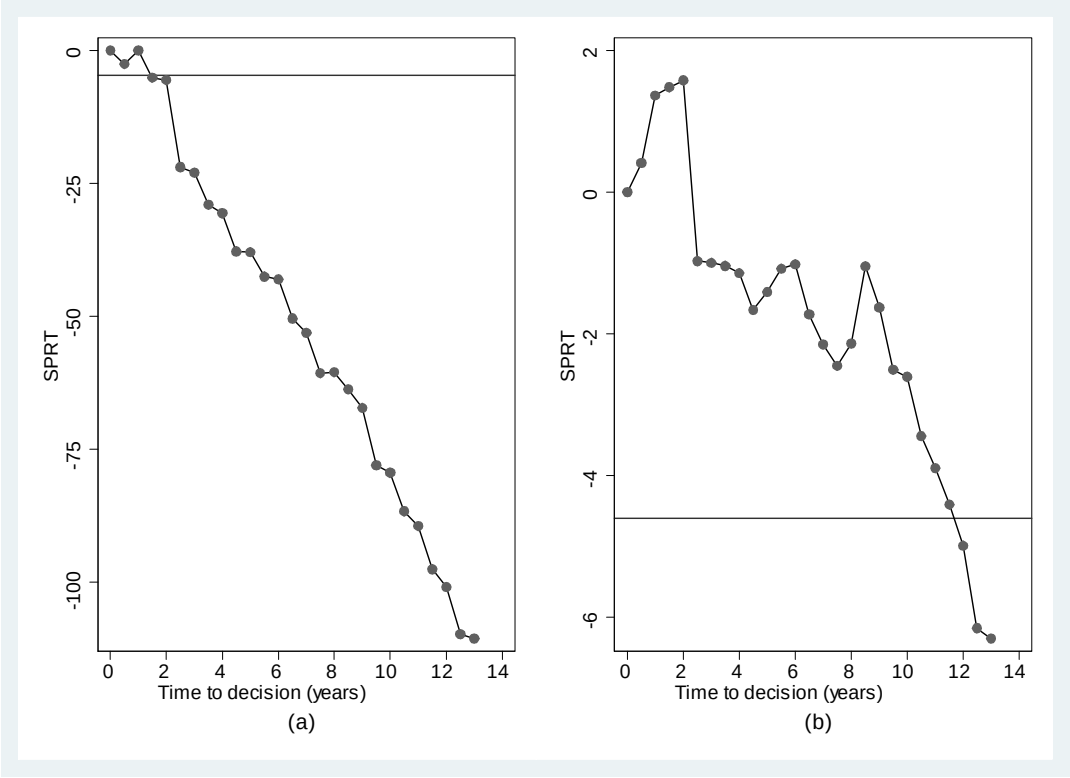

Figure 5: Influenza and Bell's Palsy, SPRT for $1-60$ day risk period. Lower boundary -4.6 ($\alpha^* = \beta^* = 0.01$); upper boundary not shown. (a) $\rho_A = 5$; (b) $\rho_A = 1.5$.

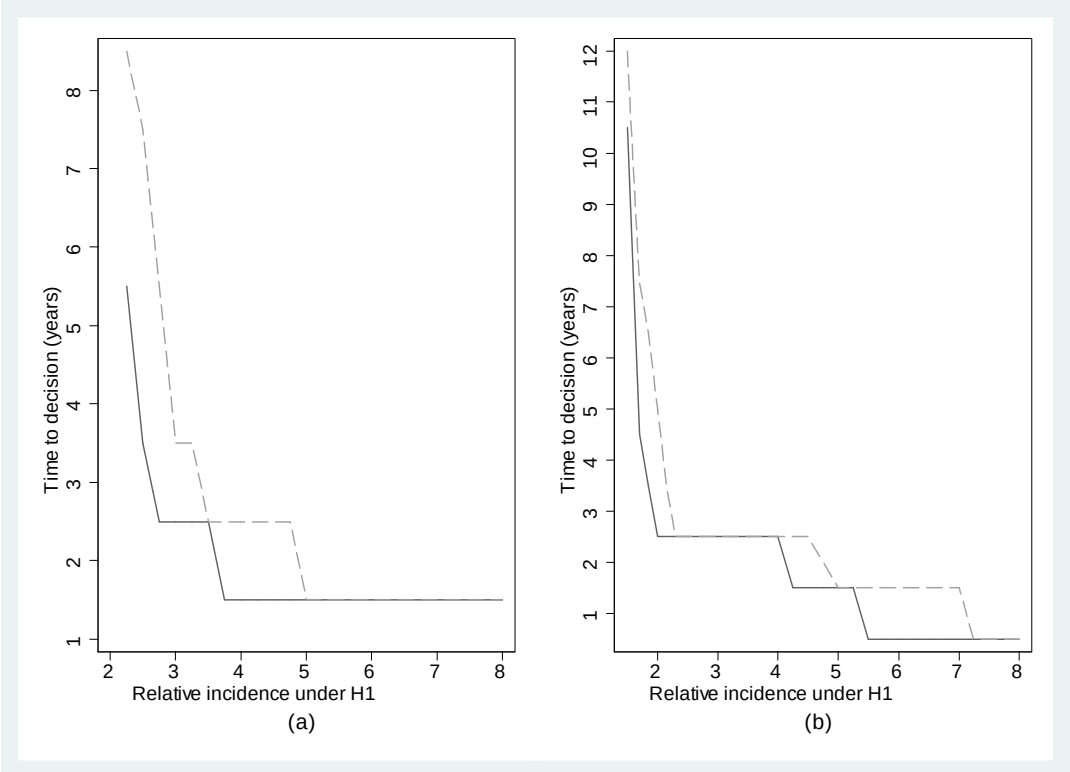

Figure 6: Time to decision for $\alpha^* = \beta^* = 0.01$ (alert threshold —) and $\alpha^* = \beta^* = 0.001$ (alarm threshold - - -): (a) $1-7$ day risk period; (b) $1-60$ day risk period.

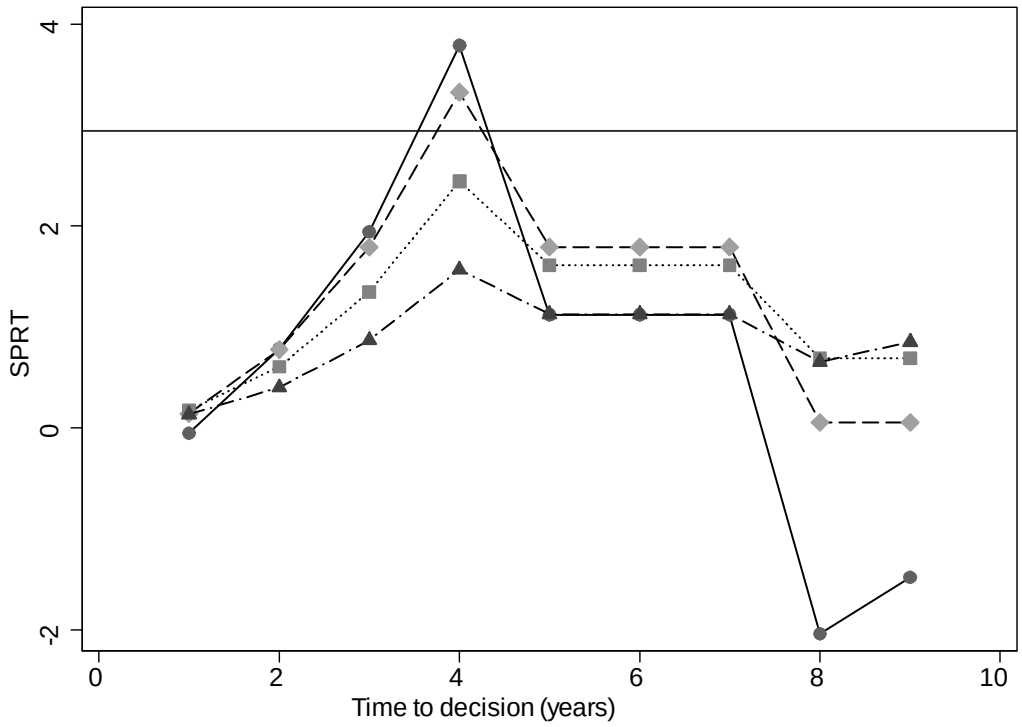

Figure 7: SPRT traces for ITP data with $0-41$ day risk period; $\rho_A = 5$(———), 3 (- - - - - -), 2 (.....) and 1.5 (- – - –). The upper boundary at 2.94 is shown.

values of ρ_A, $0-41$ day risk period, four three-month age groups, one year surveillance interval, and alert thresholds with $\alpha^* = \beta^* = 0.05$ corresponding to boundaries at ± 2.94. The upper boundary is crossed for alternative hypothesis relative incidences $\rho_A = 3$ and $\rho_A = 5$ after 4 years. The results with a $0-27$ day risk period are similar.

It is interesting to note that thereafter the SPRT values drop. This is unexpected; the reason for the change is the subject of further investigations.

6 Relaxing the assumptions

In this section we briefly indicate how some of the assumptions of the case series method, as applied to the SPRT, can be relaxed; see also Whitaker et al (2006) and Farrington and Whitaker (2006) for further details.

The first assumption is that adverse events occur as a non-homogeneous

Poisson process within an individual's lifetime. This assumption fails when the event is non-recurrent. However, the method is valid approximately to a high degree of accuracy when the event of interest is rare. So, in this case, the method can be applied as described above.

Second, we have assumed a single risk period. In fact the case series SPRT can be used with several risk periods, in which case the null hypothesis is that $\rho = 1$ in each risk period, and the alternative hypothesis is that $\rho = \rho_{Ak}$ in the k^{th} risk period. However, it is advisable to use a single risk period otherwise the results might be difficult to interpret. The risk period need not necessarily start at time of vaccination, but some time later.

Third, if the occurrence of an event changes the probability of subsequent vaccination, or is a contra-indication to vaccination, then the method can be used for single dose vaccines with the following modification: each individual's observation period should start from age at vaccination. The same approach can be used when the event of interest is death. For multi-dose vaccines, the method described in this paper cannot be applied straightforwardly when the event alters the subsequent likelihood of vacination. While a case series analysis method has been developed for such a situation, it is not likelihood-based, so does not fit within the SPRT framework, and in any case is complicated to implement (Farrington et al 2007). Instead, we recommend focusing on one of the doses as of primary interest and monitoring this particular vaccine dose alone, taking observation periods from that dose as described ealier in this paragraph.

7 Discussion and recommendations

We have proposed a framework for implementing prospective surveillance of vaccine safety, based on the sequential probability ratio test applied to the self-controlled case series method. This has the advantage of controlling for all fixed confounders, whether known or not, and whether measured or not. The method requires a surveillance database of adverse events and independent ascertainment of vaccination histories. Some analytic results were derived in a simple scenario in order to describe the performance of the system, and extensive simulations were undertaken, some of which were reported here.

We recommend that surveillance intervals of 6 months or 1 year are used. There is little benefit in using shorter intervals. We recommend 6 months, except when events are so rare that there is an appreciable probability of no cases arising within a 6-month period, in which case we suggest using 1-year surveillance intervals (as was the case with the MMR and ITP example). For most vaccination programmes, it is essential to control for age effects (as in the MMR and ITP example) or, occasionally, for seasonal effects (as in the influenza vaccine and Bell's palsy example). We recommend that this is done by pre-selecting age groups (or seasonal intervals) in which the effect can reasonably be assumed constant, and using profile likelihoods to eliminate the nuisance parameters. We found that this only marginally increased the time to decision. Failing to allow

for temporal effects (whether age or season) may produce erroneous results.

The relative incidence ρ that can be detected within a reasonable time period depend critically on the underlying incidence rate of the event λ, that is, the annual mean number of events. If this is in excess of 100, then reasonable times to decision can be achieved with ρ as low as 2. However, for $\lambda \leq 50$, reasonably short times to decision are really only achievable with ρ in excess of 3. If it is believed that ρ could be much greater, say $\rho > 5$, then we urge caution and propose that the alternative hypothesis value ρ_A be set conservatively at, say 3. This provides insurance against accepting the null hypothesis if ρ turns out not to be so large after all, while not increasing the time to decision unduly. However, we do not recommend that this logic is pursued to the point of setting a value $\rho_A < 3$, unless large numbers of events are expected to occur, since the time to decision could be adversely affected.

When it is not wholly clear what unique value of ρ_A to choose, we recommend that several values of ρ_A are monitored, say $\rho_A = 2$, 3 and 5, so as to be able to make a broader judgement about the evidence available for and against a range of hypotheses, as shown in the examples presented here. This is particularly useful with the SPRT, since the way evidence builds up in relation to contrasting hypotheses is perhaps less obvious than in the conventional setting of a clinical trial.

In this paper, we have considered the issue of prospective surveillance of a new vaccine. This provides a specific focus for surveillance. One might also be interested in longer term surveillance of an established vaccination programme, using continuous monitoring methods such as cumulative sum techniques (CUSUM). We shall report on the application of case series methodology in this context in a separate publication.

8 References

Barndorff-Nielsen O.E. and Cox D. R. (1994). *Inference and Asymptotics*. Chapman and Hall, London.

Chen R.T., Rastogi S.C. and Mullen J.R. (1994). The Vaccine Adverse Event Reporting System (VAERS). *Vaccine* **12**, 542 - 550.

Chen R. T. , Glasser J. W., Rhodes P. H. et al. (1997). Vaccine Safety Datalink Project: a new tool for improving vaccine safety monitoring in the United States. *Pediatrics* **99**, 765 - 773.

Davis R. L., Kolczak M., Lewis E. et al. (2005). Active surveillance of vaccine safety: a system to detect early signs of adverse events. *Epidemiology* **16**, 336 - 341.

Evans S. J. W., Waller P. C. and Davis S. (2001). Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and drug safety* **10**, 483 - 486.

Farrington C. P. (1995). Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics* **51**, 228 - 235.

Farrington C. P., Pugh S., Colville A. et al. (1995). A new method for active surveillance of adverse events from diphtheria tetanus pertussis and measles mumps rubella vaccines. *Lancet* **345**, 567 - 569.

Farrington C. P. and Whitaker H. J. (2006). Semiparametric analysis of case series data (with Discussion). *Journal of the Royal Statistical Society, Series C*, **55**, 553 - 594.

Farrington C.P., Whitaker H.J. and Hocine M.N. (2007). Case series analysis of censoring events. Research Report 2006 06/12, available from:

http://statistics.open.ac.uk/TechnicalReports/TechnicalReportsIntro.htm

Grigg O. A., Farewell V. T. and Spiegelhalter D. J. (2003). Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. *Statistical Methods in Medical Research* **12**, 147 - 170.

Hocine M., Guillemot D., Tubert-Bitter P. and Moreau T. (2005) Testing independence between two Poisson-generated multinomial variables in case-series and cohort studies. *Statistics in Medicine* **24**, 4035 - 4044.

Jennison C. and Turnbull B. W. (2000). *Group sequential methods with applications to clinical trials*. Chapman and Hall/CRC Press, London.

Miller E., Waight P., Farrington P., Andrews N., Stowe J. and Taylor B. (2001). Idiopathic thrombocytopenic purpura and MMR vaccine. *Archives of Disease in Childhood* **84**, 227 - 229.

Musonda P., Hocine M. N., Whitaker H. J. and Farrington C. P. (2007). Self-controlled case series analyses: small sample performance. *Computational Statistics and Data Analysis*, DOI: 10.1016/j.csda.2007.06.016.

Rothman K. J., Lanes S. and Sacks S. T. (2004). The reporting odds ratio and its advantages over the proportional reporting ratio. *Pharmacoepidemiology and drug safety* **13**, 519 523.

Stowe J. Andrews N., Wise L. and Miller E. (2006). Bell's palsy and parenrteral inactivated influenza vaccine. *Human vaccines* **2**, 110 - 112.

Wald A. (1945) Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics* **16**, 117 - 186.

Wald A. (1947). *Sequential Analysis*. Wiley, New York.

Waller P. C., van Puijenbroek E. P., Egberts A. C. G. and Evans S. J. W. (2004) The reporting odds ratio versus the proportional reporting ratio: 'deuce'. *Pharmacoepidemiology and drug safety* **13**, 525 - 526.

Whitaker H. J., Farrington C. P., Spiessen B. and Musonda P. (2006). Tutorial in Biostatistics: The self-controlled case series method. *Statistics in Medicine* **25**, 17687 - 1798.