The self-controlled case series method: performance and design in studies of vaccine safety

Patrick Musonda

Department of Statistics
Faculty of Mathematics and Computing
The Open University

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy
October 2006

Abstract

The self-controlled case series method (SCCSM) is a novel study design to investigate associations between acute responses with transient point exposures (for example vaccination). The method provides an attractive alternative to cohort and case-control designs. The method is unusual in that it requires data only on individuals who experience a response (the 'cases'). The method works as follows. Prior to the study a post-exposure risk period is defined, which corresponds to the period in which responses causally related to exposure are likely to occur. An observation period is also defined, and individuals with responses arising within this observation period are sampled. The data are then analysed using a Poisson model, conditional on the total number of events occurring for each individual. This conditioning ensures that including only cases does not bias the relative risk estimator.

The self-controlled case series method has been used to good effect in many settings, particularly in investigating putative associations between adverse events and paediatric vaccines. However, so far only limited research has been undertaken on the statistical properties of the method in finite samples, and virtually no work has been undertaken on design issues. The method also needs to be extended in various directions, for example application in surveillance methods.

This thesis provides detailed investigations of these topics. To this end, expressions for the asymptotic bias, variance and mean square error of the log-relative incidence are derived. Simulation studies taking account of age are carried out to study small and medium sample performance. Sample size formulae are obtained and validated

via simulations, thus improving the design of self-controlled case series studies. The method is extended to applications in surveillance and simulation studies are conducted to evaluate this use of the method. The methods are illustrated using data on intussusception and oral polio vaccine.

Acknowledgements

I greatly appreciate the contributions of my supervisor Prof. Paddy Farrington, not only for his guidance and invaluable advice on the thesis, but also for his thoughtfulness, care and concern through out the last three years. In addition, I am extremely grateful for his help on computing aspects of the project, also with his many suggestions for improvement on the drafts of all the chapters in the thesis. Our weekly meetings during the last three years were very enjoyable, there was never a dull moment in our meetings, and the discussions of theoretical issues arising from the research were very interesting. I was always looking forward to the meetings so as to reveal my findings following his advice.

I am indebted to my other co-supervisors Prof. Frank Critchley (Open University), Dr Thomas Verstraeten (GlaxoSmithKline (GSK) in Belgium) and my third-party monitor, Dr Karen Vines (Open University). I thank the EPSRC and GSK Biologicals for funding this research via a CASE studentship. My visits to GSK in Belgium always gave me another perspective of doing research in industry. I received various help while visiting GSK in the last two years and I am very grateful to you all. To mention but a few, thanks to Dr Thomas Breuer, Dr Bart Spiessens, Dr Claire Newbern, and Mrs Jenny Schelfhout.

I thank Dr Heather Whitaker my office mate for her help in various situations; she was always my first point of call for help whenever I got stuck in one way or the other. Many thanks to Mrs Val Spearman for sorting out administration issues that I

needed in the last three years. I also remain truly grateful for the wealth of support and talent I shared with so many researchers, friends and colleagues.

Finally, I wish to thank most sincerely my wife, Fleur and our beloved two children Martha and Frankie, for their understanding and patience during the last three years. This thesis would not have been possible without their help and support.

Contents

Tables and Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 6

Chapter 7

Chapter 8

Chapter 1

Introduction, background and literature review

1.1 Introduction

Research has traditionally been classified into two types: pure and applied. Philips et al [1] have considered a threefold classification of research: exploratory, testing-out and problem-solving, which applies to both quantitative and qualitative research.

Exploratory research is said to be the type of research that involves tackling a new problem/issue/topic about which little is known, so the research idea cannot at the beginning be formulated very well. The problem may come from any part of the discipline; it may be a theoretical research puzzle or have an empirical basis. In this type of research, a researcher will need to examine what theories and concepts are appropriate, developing new ones if necessary, and whether existing methodologies can be used. Exploratory research involves pushing out the frontiers of knowledge in the hope that something useful will be discovered.

Philips et al describe testing-out research as the type of research in which a researcher is trying to find the limits of previously proposed generalisations. In this type of research, one might ask questions such as: Does the theory of previously proposed generalisation apply in different situations? Can the theory apply in new technology? Under what circumstances does the theory fail? What bits of the theory might need extending? In this type of research, all sorts of questions can be tested, the amount of

testing out to be done is endless and continuous. By doing this, in the process the researcher will be able to improve previously proposed theories or generalisations by specifying, modifying, extending and clarifying.

As for problem-solving research, the research starts from a particular problem in the real world, and bring together all the intellectual resources that can be brought to bear on its solution. The problem has to be defined and the method of solution has to be discovered. The person working in this research may have to create and identify original problem solutions every step of the way. This will usually involve a variety of theories and methods, often ranging across more than one discipline since real-world problems are likely to be 'messy' and not soluble within the narrow confines of an academic discipline.

With respect to the Philips et al research classification, this thesis can be described as one of testing-out research, with some elements of problem-solving, as applied to a statistical method in epidemiology called the self-controlled case series method. We will begin by first describing what this method is, its advantages and limitations, its application, and the aims of the thesis.

1.2 Background

The self-controlled case series method (SCCSM), or case series method for short is a modified cohort method for estimating the relative incidence of specified events in a defined period after a point exposure. While the method was originally developed to investigate associations between vaccination and acute adverse events [2, 3], it has

subsequently been applied in other settings for example in pharmacoepidemiology (Hubbard et al [4] and Hocine et al[5]). Becker et al [6] have independently derived and applied the case series method in other areas of epidemiology. A step-by-step account of the theory, applications, modelling issues are given by Whitaker et al [7]. The same paper by Whitaker et al describes how the method can be implemented in various statistical software packages. The case series method in its semiparametric form [8] can be applied to continuous exposures but in the thesis we shall consider point exposures only.

1.3 Motivation

The self-controlled case series model was developed in order to analyse vaccine safety record linkage data relating to measles, mumps and rubella (MMR) vaccination and aseptic meningitis [9, 10]. For this study, episodes of aseptic meningitis arising in children aged 1-2 years over a defined calendar time period were obtained from laboratory and hospital records. The age and calendar time window determined by the period of event ascertainment defines an observation period for each child. From now onwards, the term ‘case’ refers to an individual who has experienced one or more events of interest over his or her observation period. Vaccination records were linked to cases resulting in a combined data set that consisted of cases and their exposures. The difficulty with such data sets is that usually they do not comprise accurate denominators. Furthermore, it may not be wholly clear from what population the cases arise and most likely the catchment areas of the hospitals from which the cases were obtained may not be clearly defined. Thus using methods such as cohort and case-control studies which are population based methods would require ingenuity,

especially since vaccine coverage in the population is unlikely to be uniform. If answers are required quickly about a possible association between an event of interest with the vaccine exposure, employing a cohort study may not be a good idea as it may take a long time, it would be very expensive to undertake and would require a large sample size. The self-controlled case series method was developed to deal with such difficulties. In the MMR and convulsions data set, a positive association between vaccination with the Urabe mumps strain and aseptic meningitis in the period 15-35 days post-vaccination was confirmed, and the composition of MMR vaccines used in the UK was changed [9, 10].

The self-controlled case series method will be described in technical terms in chapter 2. Briefly, a retrospective Poisson cohort model is specified, and the case series model is derived from this by conditioning on the total number of events experienced by each individual in the observation period.

1.4 Advantages and limitations

The following are the main advantages of the self-controlled case series method. The method uses cases only and provides consistent estimates (as the number of cases becomes large) of the relative incidence. It controls implicitly for all fixed multiplicative confounders, that is, confounders that act multiplicatively on the baseline rates and do not vary (or not vary much) with time over the observation period, such as variables relating to genetics, location, socio-economic status, gender, individual frailty, severity of underlying disease, etc. Age or temporal variation in the baseline incidence is controlled for in the model. Further under certain circumstances,

the method has high efficiency relative to the retrospective cohort method from which it is derived by conditioning [2]. Assembling the required data is much more likely to be easier in self-controlled case series method than cohort or case-control studies.

Like any other method, the self-controlled case series method has limitations which we now give. The most restrictive limitation is that the method requires that the probability of exposure is not affected by the prior occurrence of an outcome; sometimes this condition may not be fulfilled. For non-recurrent events, the method works only when the event risk is small over the observation period. The method does not produce estimates of absolute incidence, only estimates of relative incidence. A further assumption is that the observation period is independent of the timing of events. A less severe limitation of the method is that it requires variability in the time or age of the event: if all events were to happen at exactly the same age, which is very unlikely but not impossible, then the method would fail.

1.5 Why use the self-controlled case series method

Investigations of suspected or hypothesised associations of adverse outcomes with transient exposures, such as vaccination, usually require epidemiological studies such as cohort studies and case control studies. A disadvantage of a cohort study is that for rare events it has to be very large to achieve sufficient power. This may not be practical and can be very expensive. Sometimes researchers have got round this problem by reconstructing large retrospective cohorts (Ray et al [11]) using data sets assembled for other purposes. Case control studies require smaller sample sizes. The main disadvantages with case control studies is that they are more prone to selection

bias, recall bias, and ascertainment bias (Altman [12]). Confounding by variables related both to avoidance of vaccination and to the outcome of interest is a major problem for both cohort and case-control studies as noted by Farrington et al [3]. For example Fine et al [13] found that parental education, ethnic group, age of the mother, maternal smoking, birth weight, evolving neurological disorders, and conditions predisposing to seizures are related to both vaccination and to sudden infant death syndrome or encephalopathy and hence may be confounding factors. Both the cohort method and case-control method are data-intensive, involving large cohorts or careful selection and matching of controls [14]. The self-controlled case series method aspires to control for fixed confounders by using cases only. This helps to reduce the data collection effort, and concentrates it on the cases.

Observations of clustering or troughs of events shortly after exposure leads to speculation about associations with exposure. There are several methodological difficulties involved in carrying out epidemiological studies to monitor such associations. Such studies are prone to many biases, for example, Fine et al [13] found that there is often differential ascertainment of cases in recently vaccinated and unvaccinated individuals and differential vaccination rates in individuals at higher or lower risk. Both would lead to bias in cohort and case control studies, whereas case series studies may escape bias from the latter. The cohort method is based on comparisons of incidence rates for person-time aggregated both across and within individuals. But the self-controlled case series method removes the contribution of comparisons between individuals, focussing attention on event rates in different periods within each individual's observation time (Farrington et al [3]). For this reason, individuals who experience no events contribute no information about the

association between vaccination and outcome. Such individuals can be ignored without introducing any bias. On the other hand individuals who experience one or more events do contribute information on the risk period and age group in which the events occurred. The self-controlled case series method thus combines aspects of the case control and cohort methods, using retrospectively ascertained vaccination histories in cases to estimate the relative incidence in different intervals after vaccination relative to a control period.

1.6 Other case-only methods

Looking at cases only to detect risk factors for diseases is not new. Various studies have been conducted in which cases only are used, for example, a Markov chain method using cases only was used by Aalen et al [15] and a similar method modified as survival analysis was used by Prentice, et al [16]. However, it has been argued [2, 8] that the methods of Aalen et al and Prentice et al give a valid test for no association but do not yield readily interpretable effect estimates. Another use of cases only can be seen in the case-crossover model developed by Maclure [17]. Maclure's method resembles a case-control method with referents selected from the case's own history. It has been argued [18] that although the case-crossover method is self-matched, it only yields consistent estimates when the distribution of exposure in case and control time intervals is exchangeable, in particular implying stationarity of exposures. There are several variants of this method, reviewed by Greenland [19], and the case-crossover approach has been used in many settings [20]. Another method in which cases only are used is that of Feldmann [21]. In this method, a constant base-line incidence is assumed. Feldmann's method does give consistent estimates, though it is

only approximately self-matched for rare events. The earlier approaches of using cases only have characteristics which the self-controlled case series method incorporates, in particular, it coincides with Feldmann's method when the disease is rare and the base-line incidence is constant. The self-controlled case series method is similar to Prentice's in that it also controls for age, it is similar to Maclure's in that it also controls for fixed confounders. The main difference between the case series method and the method of Maclure is that it is derived from the same statistical model as a cohort study design, and hence can handle non-exchangeable exposures and in particular controls for age effects. Furthermore in this method one does not need to specify the prior probabilities for exposure as required in some other case-crossover designs (Marshall et al [22]). Smeeth et al [14] describe the advantages and disadvantages of case-control and case-only study designs.

1.7 Where the self-controlled case series method has been used.

This method has been used in various situations, but the main area it has been used is in modelling adverse events in vaccine studies. Table 1.1 below is adapted from Whitaker et al [7] which documents published applications of the case series method. A review of applications to vaccine safety is given by Andrews [23] and also by Farrington [24]. Independently, Navidi [25] proposed what is essentially a case series method, with time-varying exposures, for application in studies of air pollution. This method is described as a bi-directional or ambidirectional case-crossover method. The case series version of this method is that in which the entire observation period is used as controls. A similar approach has also been discussed by Lumley and Levy [26].

Farrington and Whitaker [8] describe a generalisation of this approach, in which residual seasonality is controlled.

Table 1.1 Studies using the case series method

Exposure	Outcome	Reference
DTP vaccine	Febrile convulsion	[9]
MMR vaccine	Febrile convulsion	[9]
MMR vaccine	Idiopathic thrombo-cytopenic purpura	[9, 27]
MMR vaccine	Aseptic meningitis	[9, 28]
MMR vaccine	Autism	[29, 30]
MMR vaccine	Invasive bacterial infection	[31]
MMR vaccine	Gait disturbance	[32]
Influenza vaccine	Asthma	[33, 34]
Influenza vaccine	Bell's palsy	[35]
Oral polio vaccine	Intussusception	[36, 37]
Oral rotavirus vaccine	Intussusception	[38]
DTP, MMR, HBV, HIB, OPV vaccine	Wheezing	[39]
Antidepressants	Hip fracture	[4]
Antidepressants	Myocardial infarction	[40]
Long-haul air travel	Venous thromboembolism	[6]
Influenza vaccine	Any medical visits	[41]
Common vaccines and infections	Myocardial infarction and stroke	[42]

DTP=diphtheria, tetanus, pertusis
MMR=measles, mumps, rubella
HBV=hepatitis B vaccine, HIB=haemophilias influenza type B
OPV=oral polio vaccine

A comparative evaluation of the self-controlled case series method has been undertaken by Farrington et al [3] and also by Glanz et al [43]. In Farrington et al's comparisons, estimates of the relative incidence of febrile convulsions associated with Measles Mumps and Rubella (MMR) vaccine were obtained using the case series method, the case-control method and the cohort method. Theoretical arguments about the efficiency of the self-control case series method were presented. Overall the findings were that the self-controlled case series method produced results similar to the cohort method, whereas the 1-1 matched case-control estimates had wider confidence intervals reflecting the lower power of the method for a given number of cases. In conclusion Farrington et al [3] noted that the cohort study remains the

"ideal" design for the study of adverse reaction to vaccines, and should be used whenever feasible. However, for studies of rare adverse events or for routine surveillance purposes, large-scale cohort studies may be costly, impractical, or prone to confounding. In such circumstances, the case series method provides a powerful and practical alternative to cohort and case-control studies.

1.8 Issues to explore and outline of the thesis

The self-controlled case series method is relatively new, and some statisticians and epidemiologists are naturally sceptical. This scepticism is a barrier to its use, in spite of its benefits such as good power, reduced confounding and practicability. Testing-out and extension of the method will hopefully contribute to a better understanding of the method amongst the epidemiological community as a whole, including the pharmaceutical industry. In this thesis the following issues will be explored:

- Further statistical properties of the method
- Evaluation of its small sample performance
- Improvement in the design of self-controlled case series studies by obtaining and validating sample size formulae
- Extending the method's application to prospective surveillance

We now give the outline of the thesis. In chapter two, we present the case series method, and derive some expressions of its theoretical properties. The case series method involves fitting a particular log-linear model using maximum likelihood. Thus, the asymptotic performance of the method is guaranteed by statistical theory. Expressions for the asymptotic bias, variance, and the asymptotic mean square error

of the estimate of relative incidence are derived. A graphical study of the bias, variance and asymptotic mean square error are given. In chapter three we present extensive simulations to study the validity of asymptotic results in finite samples under different situations. We describe how the simulations were carried out. Results from the simulations are given starting with what we call the standard scenario with varying number of cases and a range of true relative incidences. We then explore different risk periods, the effect of age, and different distributions of age at exposure. We also investigate indefinite risk periods and the presence of unexposed cases. We explore the effects of age using several contrasting scenarios.

Chapters 4 and 5 concern the estimation of sample sizes for case series studies. So far little work has been done on the design of self-controlled case series studies. Sample size formulae are developed and validated using simulations. The impact of age effects on power and sample size are studied. In Chapter 4, we study an earlier published sample-size formula [3]. We find that this formula is not accurate, and investigate several alternative approaches. In Chapter 5, we extend one successful approach to take account of the effect of age (Musonda et al [44]).

Chapters 6 and 7 relate to applying the self-controlled case series method in a prospective surveillance context. The issue of interest is how to apply the self-controlled case series method, which is a retrospective method, in a prospective way so that possible adverse outcomes with a new vaccine (or several vaccines in routine use) can be detected early so that remedial action can be taken. This constitutes a new application of the case series method. Following Wald [45] and Page [46] we use the sequential probability ratio test (SPRT) and cumulative sum (CUSUM) based on the

self-controlled case series method so as to apply the self-controlled case series method in a prospective situation. These approaches along with extensive simulations to demonstrate their performance under different situations are presented.

In chapter 8 we analyse a data set on oral polio vaccine and intussusception, provided to us by GlaxoSmithKline Biologicals (Belgium). This study was undertaken in preparation for field trials of a new oral rotavirus vaccine. These data require some ingenuity in how one applies the self-controlled case series method owing to censoring of exposure histories. We describe how to analyse such data. We go on to discuss how the findings of the thesis throw light on the results, and how they may inform the design of future studies and surveillance programmes based on the case series method.

The conclusions of the thesis and its contribution to knowledge about the self-controlled case series method are presented in chapter 9.

Chapter 2

The self-controlled case series model

2.1 Introduction

In this chapter we introduce some notation, present the self-controlled case series model and derive some of its large sample properties in a simple setting. We present the likelihood in section 2.2. In section 2.3 we study the asymptotic bias of the relative incidence estimator. In section 2.4 we present a graphical study of the asymptotic bias. The asymptotic variance of the estimator is derived in section 2.5. In section 2.6 we present a graphical study of the variance. We derive the asymptotic mean square error (AMSE) of the relative incidence estimator in section 2.7 and present a graphical study of AMSE in section 2.8. We conclude the chapter with a brief discussion in section 2.9.

2.2 The self-controlled case series model

The self-controlled case series method is a conditional cohort method for estimating the relative incidence of specified events in a defined period after a point exposure. In this method, first an observation period is defined. Time within the observation period is classified as at risk or as control time in relation to point exposures that are regarded as fixed. We then condition on the number of events experienced by each individual over the observation period. As mentioned in chapter one, the method

allows valid inference about the relative incidence of events in risk periods relative to the control period, using data on cases only.

We now derive the general form of the likelihood of the self-controlled case series model. The pictorial configuration (Figure 2.1) below will help to understand the general form of the likelihood described.

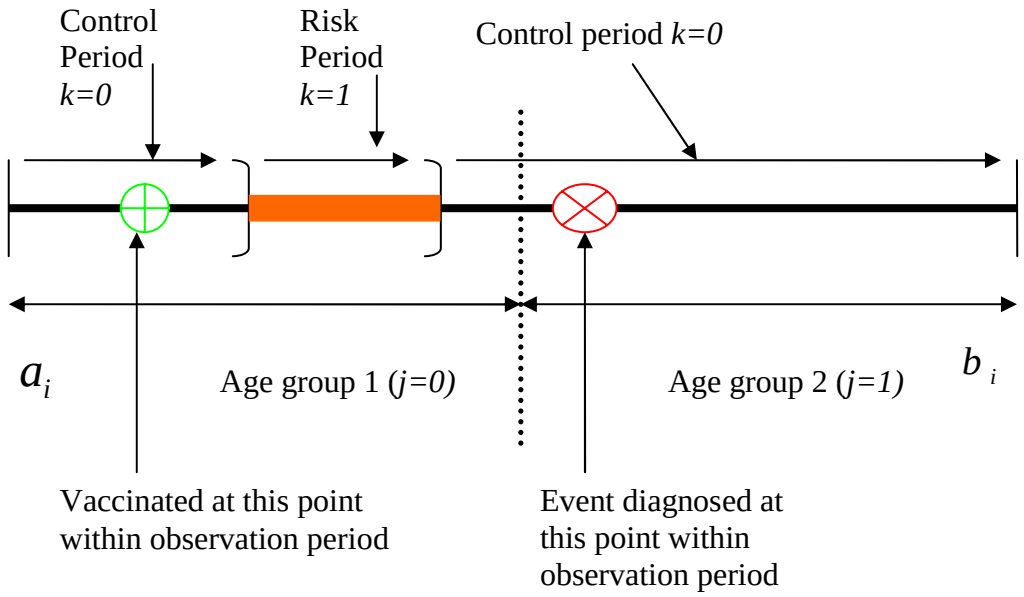


Figure 2.1 Possible case series configuration

In Figure 2.1 we see a possible configuration in which an observation period $(a_i, b_i]$ is defined within which an individual i was exposed (vaccinated) and a risk period (red line) is defined shortly after the exposure. It is possible to have several risk periods depending on prior knowledge of what time intervals are important. For example Griffin et al [47] assumed that the effect of DTP on febrile convulsions or encephalopathy had risk periods of 0-3, 4-7, 8-14, and 15-29 days after any dose of DTP. The observation period is further divided in age groups; in Figure 2.1 there are

two age groups. As with risk periods, it is possible to define several age groups. The period outside the risk period is known as the control period. In Figure 2.1 it comprises of the period before vaccination, a period shortly after vaccination, and a period after the risk period. The event in this case was diagnosed some time after the risk period in age group 2, but could have occurred anywhere within the observation period. The observation period and the location of the risk period within it will generally vary between individuals.

In general, we assume that events arise within individuals as a non-homogeneous, age-dependent Poisson process. In what follows, a proportional incidence model is used to describe the relation between vaccination and the outcome of interest (Farrington et al [3]).

Let individuals be indexed by $i = 1, 2, ..., N$, age groups be indexed by $j = 0, 1, ..., J-1$ (0 denoting the reference age group) and the risk periods be indexed by $k = 0, 1, ..., K-1$ (0 denoting the control period).

Further suppose we let the symbols $\lambda_{ijk}, e_{ijk}, n_{ijk}$ respectively denote incidence, length of time at risk, and number of events experienced by an individual i, in age group j and risk period k during the observation period $(a_i, b_i]$. The log-linear model [3]

$$\ln(\lambda_{ijk}) = \phi_i + \alpha_j + \beta_k$$

is used to parameterise the incidence of an event for an individual effect ϕ_i, age effect α_j, and exposure effect β_k (with $\alpha_0 = \beta_0 = 0$). Thus the incidence function during the baseline period is simply $\lambda_{i00} = \exp(\phi_i)$. The Poisson probability model is given

by: $\Pr[r] = \dfrac{e^{-\lambda} \times \lambda^r}{r!}$ where $r = 0, 1, 2, \ldots$ and for the underlying cohort model,

$n_{ijk} \sim \text{Poisson}(\lambda_{ijk} \times e_{ijk})$.

For the cohort model, $\phi_i = x_i^T \gamma$ for fixed covariates x_i, and the Poisson log-likelihood kernel (which is equal to the log-likelihood up to an additive constant) is

$$\ell_{\text{co}}(\alpha, \beta, \gamma) = \sum_i \sum_{jk} n_{ijk} (x_i^T \gamma + \alpha_j + \beta_k) - \sum_i \sum_{jk} \exp(x_i^T \gamma + \alpha_j + \beta_k) e_{ijk}.$$

The self-controlled case series model is derived from the cohort model with the ϕ_i unrestricted by conditioning on the $n_{i..}$, (the total number of events experienced), thus giving a product multinomial distribution as described by McCullagh and Nelder [48]. So the log-likelihood kernel for the self-controlled case series model is

$$\ell(\alpha, \beta) = \sum_i \sum_{jk} n_{ijk} \log \left(\frac{\exp(\alpha_j + \beta_k) e_{ijk}}{\sum_{rs} \exp(\alpha_r + \beta_s) e_{irs}} \right).$$

We can see from above that the individual effects $\phi_i = x_i^T \gamma$ cancel out. This is because incidence rates are contrasted within the same individual's person-time, so that, in this sense, the method is self-controlled. Thus, provided the model is correct, inferences from a case series analysis cannot be confounded by fixed multiplicative individual effects, which might include genetic factors, location, socio-economic status, sex, underlying health status, individual frailty, and so on [7]. Individual effects can nonetheless modify the exposure effect but this can be modelled by including suitable interaction terms. Note that self-control applies to fixed covariates only and not age or time-dependent covariates.

In much of what follows we shall only need the log likelihood in the following simplified situation. We suppose that there are no age effects, and that all individuals

are observed over the same observation period, comprising two adjacent periods of duration e_1 (the risk period) and e_0 (the control period). Suppose that all individuals are vaccinated at the start of period e_1 and subsequently at increased risk during this period. Suppose that in a sample of n events, n_0 occur in period e_0 and n_1 in period e_1 with $n = n_0 + n_1$. Let the ratio of the risk period to the observation period be r, that is, $r = \frac{e_1}{e_0 + e_1}$. Usually the risk period and the observation period will be specified in advance. However, only their ratio r is required. Let ρ be the relative incidence e^{β} (so that $\beta = \log(\rho)$).

In this simple situation the log-likelihood kernel is equal to:

$$\ell(\beta) = n_1\beta - n\log(e_1 e^{\beta} + e_0)$$

Note that this is the same log-likelihood kernel as for the binomial model

$$n_1 \sim B(n, p)$$

with

$$p = \frac{e^{\beta} e_1}{e^{\beta} e_1 + e_0}.$$

The maximum likelihood estimator $\hat{\beta}$ of β is obtained by setting $\frac{\partial \ell}{\partial \beta} = 0$, that is ,

$$\frac{\partial \ell}{\partial \beta} = n_1 - n\frac{e_1 e^{\beta}}{e_1 e^{\beta} + e_0} = 0$$

giving

$$\hat{\beta} = \log\left(\frac{n_1 / e_1}{n_0 / e_0}\right) \text{ and } \hat{\rho} = e^{\hat{\beta}} = \frac{n_1 / e_1}{n_0 / e_0}.$$

The likelihood ratio statistic for the test of $H_0 : \beta = 0$ for this simplified situation is

then $D = 2\left[\ell\left(\hat{\beta}\right) - \ell(0)\right] = 2\left[n_1\hat{\beta} - n\left\{\log\left(e_1 e^{\hat{\beta}} + e_0\right) - \log(e_1 + e_0)\right\}\right]$,

where $\hat{\beta}$ is the maximum likelihood estimator.

2.3 Derivation of the bias of the estimator $\hat{\beta}$

In this section we derive an expression of the asymptotic bias up to the second order of the estimator of the relative incidence in the simple situation described in the previous section when there are no age effects and all individuals have the same observation period. The maximum likelihood estimator may be written as

$$\hat{\beta} = \log\left(\frac{n_1}{n - n_1}\right) + \log\left(\frac{e_0}{e_1}\right).$$

Let the function $f(x) = \log\left(\frac{x}{n-x}\right) + \log\left(\frac{e_0}{e_1}\right)$

This is equivalent to:

$f(x) = \log(x) - \log(n-x) + \log\left(\frac{1-r}{r}\right)$, where $r = \frac{e_1}{e_0 + e_1}$.

The random variable X is the number of events occurring in the risk period. This follows a binomial distribution

$X \sim Binomial(n, p)$

where $p = \frac{e_1\rho}{e_0 + \rho e_1} = \frac{\rho r}{(1-r) + \rho r}$.

It follows that the expectation of X is $E(X) = \mu = np$ and $f(\mu) = \log(\rho) = \beta$.

By Taylor expansion of $\hat{\beta} \equiv f(X)$ about μ, we get:

$$\hat{\beta} = f(\mu) + (X-\mu)\, f'(\mu) + \frac{1}{2}(X-\mu)^2\, f''(\mu) + \frac{1}{6}(X-\mu)^3\, f'''(\mu)$$
$$+ \frac{1}{24}(X-\mu)^4\, f^{iv}(\mu) + \text{residue} \qquad (2.1)$$

Taking expectations of the above expression, we have:

$$E\left(\hat{\beta}\right) = \beta + \frac{1}{2}\text{var}(X)\, f''(\mu) + \frac{1}{6}E\left(X-\mu\right)^3 f'''(\mu) + \frac{1}{24}E\left(X-\mu\right)^4 f^{iv}(\mu) + \text{residue.}$$

Now we have:

$$f'(x) = \frac{1}{x} + \frac{1}{n-x} = \frac{n}{x(n-x)},\ \text{so } f'(\mu) = \frac{1}{npq} \text{ with } q = 1-p;$$

$$f''(x) = \frac{-1}{x^2} + \frac{1}{(n-x)^2} = \frac{x^2-(n-x)^2}{x^2(n-x)^2},\ \text{so } f''(\mu) = \frac{1}{n^2}\left[\frac{p-q}{p^2q^2}\right];$$

$$f'''(x) = \frac{2}{x^3} + \frac{2}{(n-x)^3} = 2\frac{x^3+(n-x)^3}{x^3(n-x)^3},\ \text{so } f'''(\mu) = \frac{2}{n^3}\left[\frac{p^3+q^3}{p^3q^3}\right];$$

$$f^{iv}(x) = 2\left[\frac{-3}{x^4} + \frac{3}{(n-x)^4}\right] = 6\left[\frac{x^4-(n-x)^4}{x^4(n-x)^4}\right],\ \text{so } f^{iv}(\mu) = \frac{6}{n^4}\left[\frac{p^4-q^4}{p^4q^4}\right] = \frac{6}{n^4}\frac{(p-q)\left(p^2+q^2\right)}{p^4q^4}.$$

We know that for a binomial, (Johnson et al [49]), the variance, the third and fourth moment about the mean are: $npq,\ npq(q-p),$ and $npq\left[1+3pq(n-2)\right]$ respectively. Moments of higher order contribute terms that are $O\left(n^{-3}\right)$ at most. Replacing these values in (2.1) above and only considering terms of order up to $O(n^{-2})$ we have:

$$E\left(\hat{\beta}\right) = \beta + \frac{npq}{2}\frac{1}{n^2}\left[\frac{p-q}{p^2q^2}\right] + \frac{npq(q-p)}{6}\frac{2}{n^3}\left[\frac{p^3+q^3}{p^3q^3}\right] + \frac{3n^2p^2q^2}{24}\frac{6}{n^4}\frac{(p-q)\left(p^2+q^2\right)}{p^4q^4} + O(n^{-3})$$

Simplifying, we get the asymptotic bias of β up to second order in terms of p and q as given below:

$$E\left(\hat{\beta}\right)-\beta=\frac{1}{2n}\frac{(p-q)}{pq}+\frac{1}{n^2}\frac{(p-q)}{p^2q^2}\left[\frac{3}{4}\left(p^2+q^2\right)-\frac{\left(p^3+q^3\right)}{3}\right]+O(n^{-3}) \qquad (2.2)$$

For interpretation purposes, we substitute

$$p=\frac{\rho r}{(1-r)+\rho r}=\frac{e^{\beta}r}{(1-r)+e^{\beta}r},\ q=\frac{1-r}{(1-r)+e^{\beta}r}\ ,$$

and note that $p^3+q^3=(p+q)(p^2-pq+q^2)=(p^2-pq+q^2)$

thus

$$E\left(\hat{\beta}\right)-\beta=\frac{1}{2n}\left(re^{\beta}-(1-r)\right)\left(\frac{1}{1-r}+\frac{1}{re^{\beta}}\right)+$$
$$\frac{1}{12n^2}\left(re^{\beta}-(1-r)\right)\frac{\left(re^{\beta}+1-r\right)}{\left(re^{\beta}\right)^2(1-r)^2}\left[5\left(re^{\beta}\right)^2+4re^{\beta}(1-r)+5(1-r)^2\right]+O(n^{-3}) \quad (2.3)$$

Note that in (2.3) above, the expression $\frac{1}{2n}\left(re^{\beta}-(1-r)\right)\left(\frac{1}{1-r}+\frac{1}{re^{\beta}}\right)$ is the asymptotic bias up to the first order. We can further factorise (2.3) above to get the following expression:

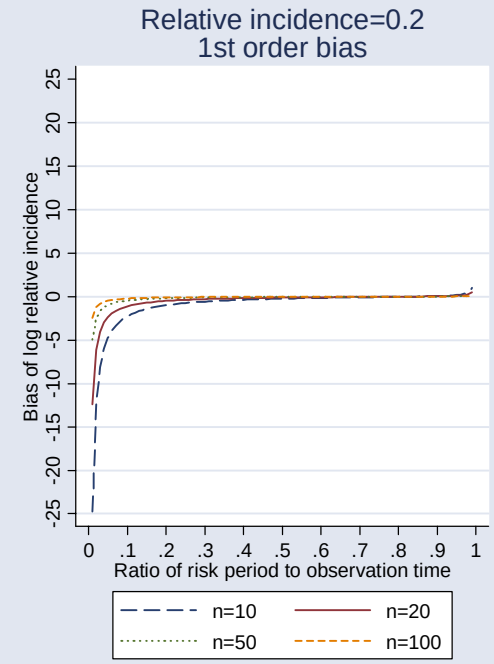
$$E\left(\hat{\beta}\right)-\beta=\frac{1}{2n}\left(re^{\beta}-(1-r)\right)\left(\frac{1}{1-r}+\frac{1}{re^{\beta}}\right)\left\{1+\frac{1}{6n(1-r)re^{\beta}}\left[5\left(re^{\beta}\right)^2+4re^{\beta}(1-r)+5(1-r)^2\right]\right\}$$
$$+O(n^{-3}) \qquad (2.4)$$

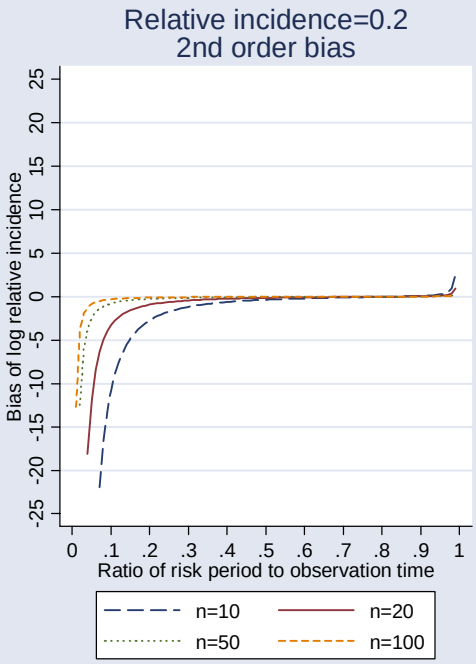
We can see in (2.4) above that the expression $\left\{1+\frac{1}{6n(1-r)re^{\beta}}\left[5\left(re^{\beta}\right)^2+4re^{\beta}(1-r)+5(1-r)^2\right]\right\}$ is always greater than 1. So the $\operatorname{sgn}\left(\left[E\left(\hat{\beta}\right)-\beta\right]\right)=\operatorname{sgn}\left(\left[re^{\beta}-(1-r)\right]\right)$ and $|\text{2nd order bias}|>|\text{1st order bias}|$. Both the first and second order asymptotic bias are zero when $re^{\beta}=1-r$. This occurs when

the risk period is chosen such that the same (expected) number of cases occurs in the risk period as outside it. Further we note that the asymptotic bias is negative when $e^{\beta} < \frac{1-r}{r}$ and positive when $e^{\beta} > \frac{1-r}{r}$.

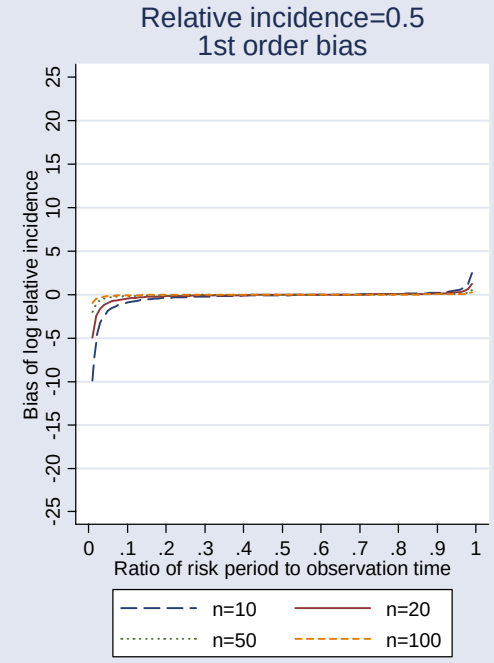
2.4 Graphical study of the asymptotic bias

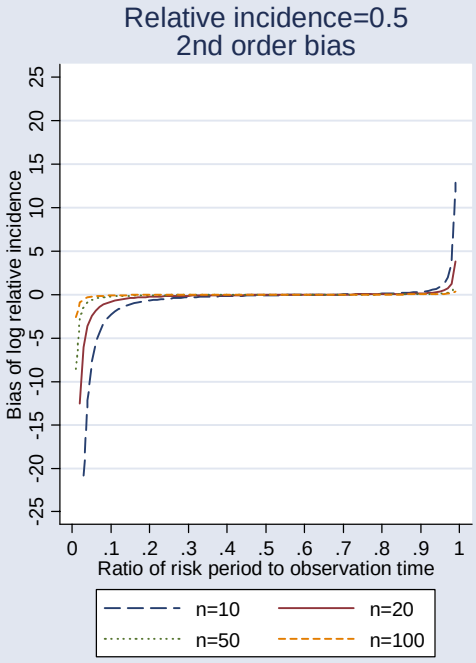
In this subsection, we explore the behaviour of the asymptotic bias (2.4) graphically. Note that an estimate with small bias and small variance is generally preferable to one with zero bias and large variance [50]. In addition to exploring the behaviour of the bias of $\hat{\beta}$, we shall later explore under what circumstances the estimator from the self-controlled case series method has small bias and small variance. Figure 2.2 (a), (b), … (f) below shows the asymptotic bias of the first and second order varying with the ratio of the risk period to the observation period at fixed relative incidence of 0.5, 1, 2, 5 and 10. We present the asymptotic bias for n=10, 20, 50, 100 cases; asymptotically as $n \to \infty$ the bias is zero (explored but results not shown) when there are a lot (n > 100) of cases.


Relative incidence=0.2
1st order bias
Bias of log relative incidence
-25 -20 -15 -10 -5 0 5 10 15 20 25
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
Ratio of risk period to observation time
n=10
n=20
n=50
n=100

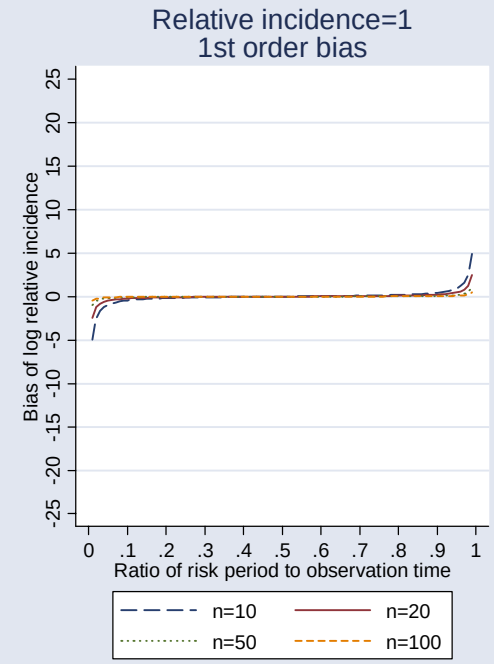

Relative incidence=0.2
2nd order bias
Bias of log relative incidence
-25 -20 -15 -10 -5 0 5 10 15 20 25
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
Ratio of risk period to observation time
n=10
n=20
n=50
n=100

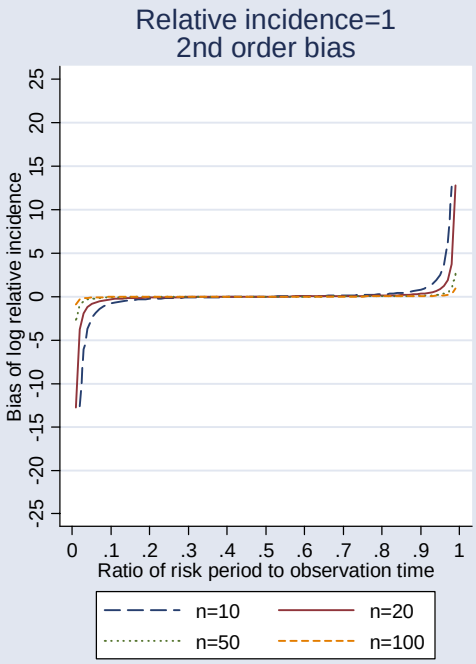
(a)


Relative incidence=0.5
1st order bias
Bias of log relative incidence
-25 -20 -15 -10 -5 0 5 10 15 20 25
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
Ratio of risk period to observation time
n=10
n=20
n=50
n=100

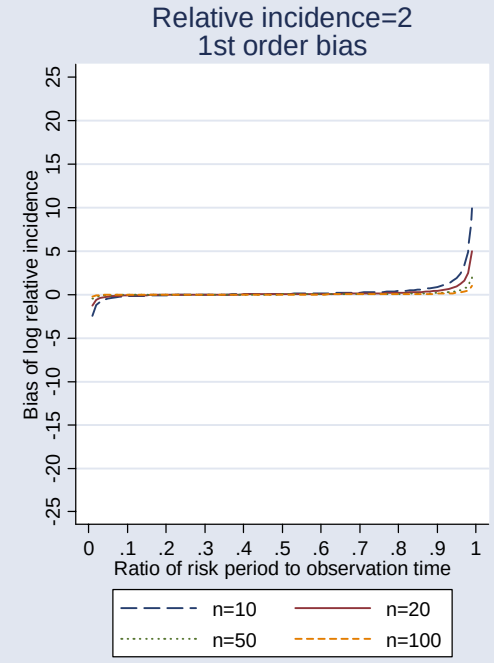

Relative incidence=0.5
2nd order bias
Bias of log relative incidence
-25 -20 -15 -10 -5 0 5 10 15 20 25
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
Ratio of risk period to observation time
n=10
n=20
n=50
n=100

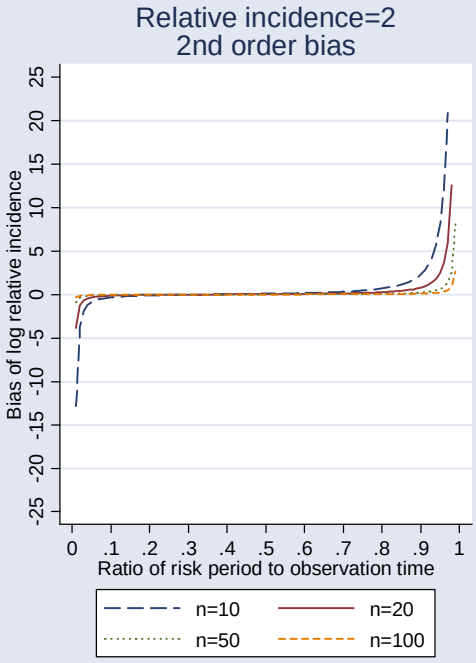
(b)


Relative incidence=1
1st order bias
Bias of log relative incidence
-25 -20 -15 -10 -5 0 5 10 15 20 25
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
Ratio of risk period to observation time
n=10
n=20
n=50
n=100


Relative incidence=1
2nd order bias
Bias of log relative incidence
-25 -20 -15 -10 -5 0 5 10 15 20 25
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
Ratio of risk period to observation time
n=10
n=20
n=50
n=100

(c)


Relative incidence=2
1st order bias
Bias of log relative incidence
-25 -20 -15 -10 -5 0 5 10 15 20 25
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
Ratio of risk period to observation time
n=10
n=20
n=50
n=100


Relative incidence=2
2nd order bias
Bias of log relative incidence
-25 -20 -15 -10 -5 0 5 10 15 20 25
0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1
Ratio of risk period to observation time
n=10
n=20
n=50
n=100

(d)

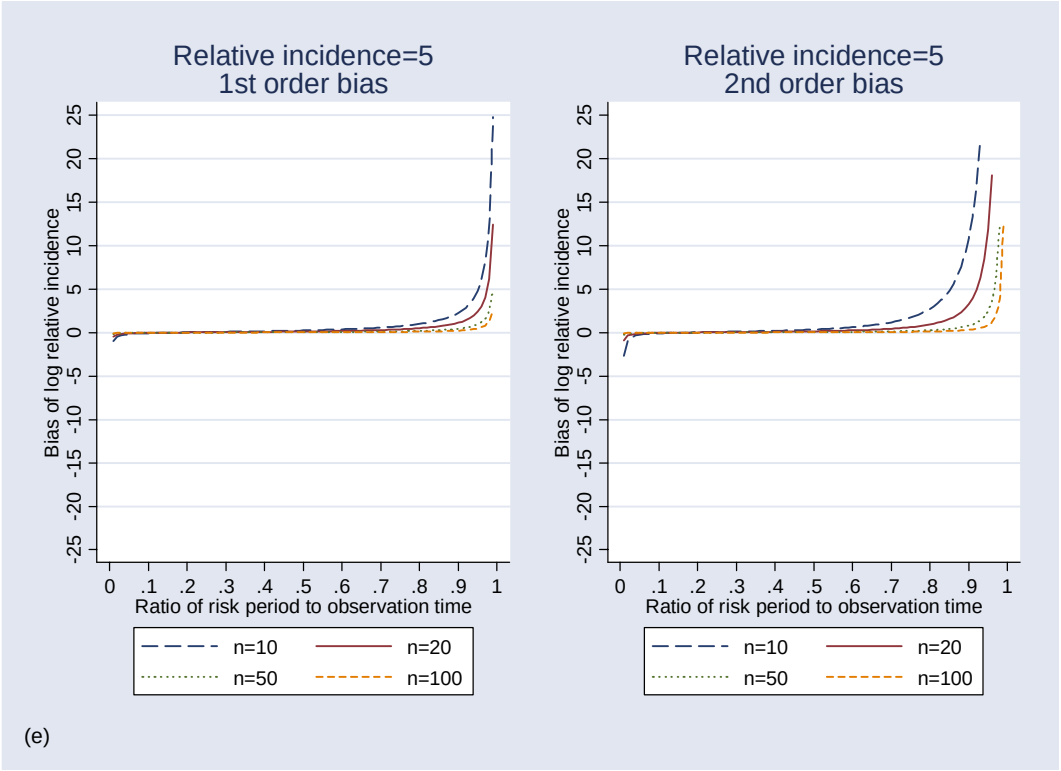
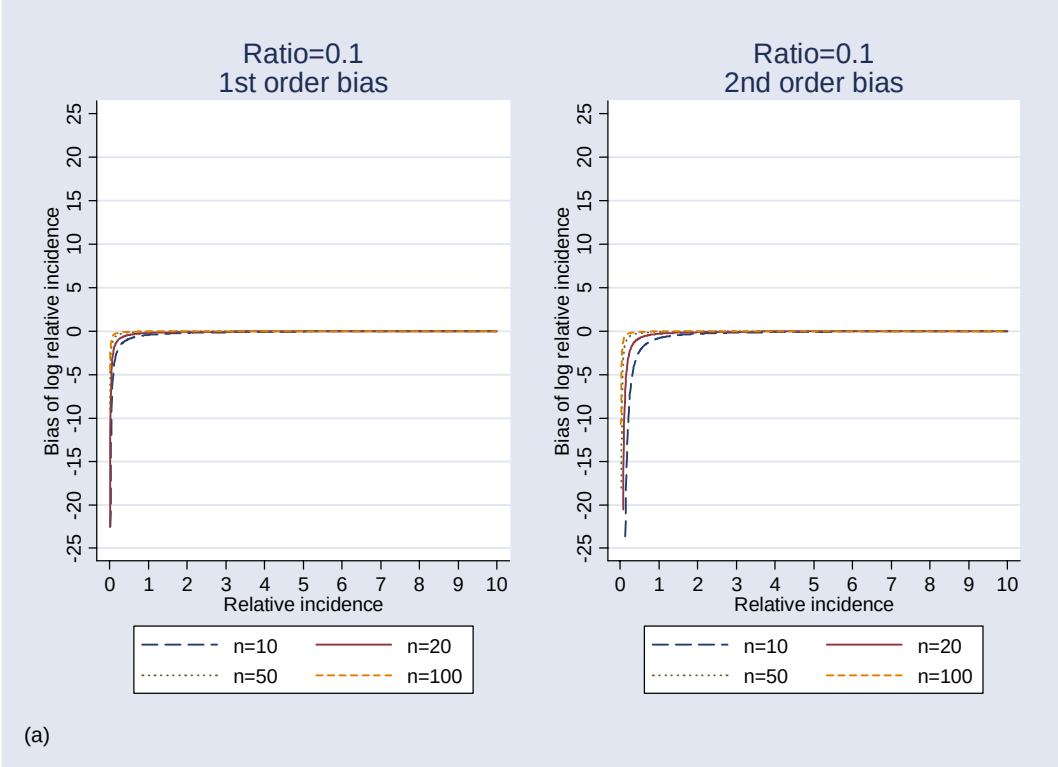
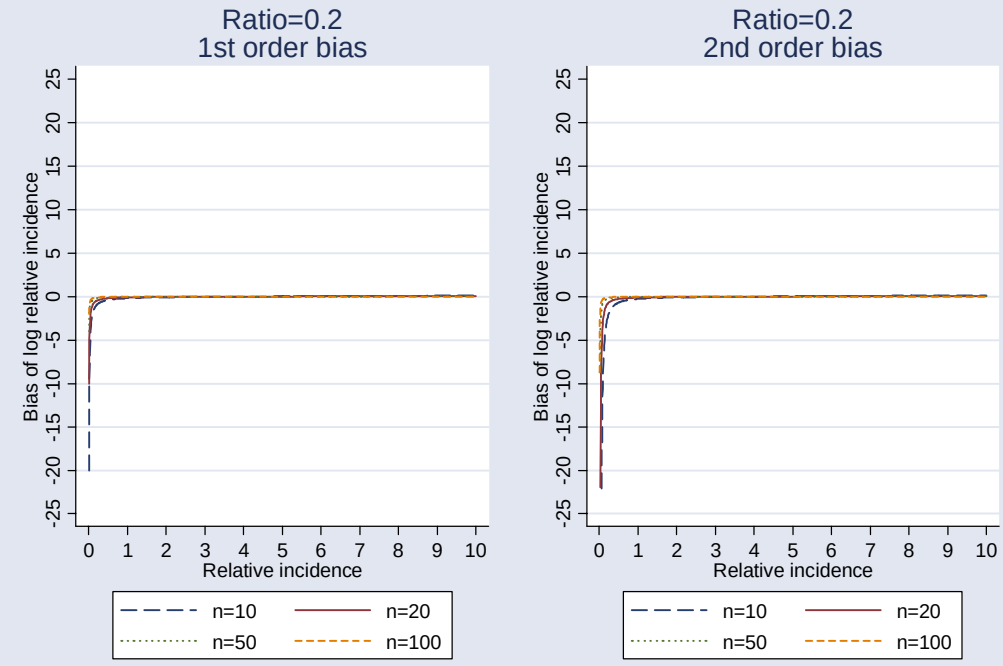


Figure 2.2 First and second order asymptotic bias varying with the ratio of the risk period to the observation period.

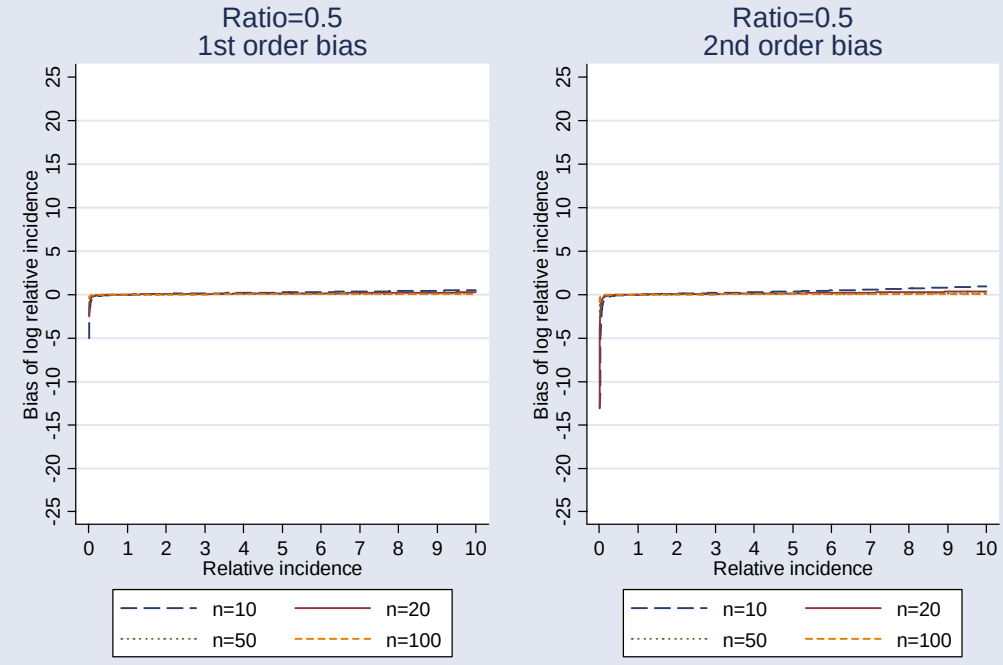
Similarly Figure 2.3 (a), (b), (c) and (d) below shows the bias varying with relative incidence at fixed ratios of the risk period to the observation $r = 0.1, 0.2, 0.5$ and 0.9. We can see that the bias decreases as one would expect with increasing sample size. There is little difference between the first order bias and the second order bias except for small sample sizes (e.g, n=10). The asymptotic bias is greatest for small ratio of risk period to the observation period (ratio less than 0.1 Figure 2.2 (a), and (b)) and long ratio of risk period to observation period (ratio greater than 0.9, Figure 2.2 (d), (e), (f)). Varying the relative incidence with fixed ratio of risk period to observation period (ratios 0.1, 0.2, 0.5, and 0.9) and fixed number of cases (n=10, 20, 50 and 100) there is an appreciable bias when the relative incidence is less than 1 and $r \leq 0.5$ (Figure 2.3. (a), (b), (c)) and for large relative incidences when $r = 0.9$ (Figure 2.3 d).



(a)


Ratio=0.2
1st order bias
Ratio=0.2
2nd order bias
Bias of log relative incidence
-25 -20 -15 -10 -5 0 5 10 15 20 25
0 1 2 3 4 5 6 7 8 9 10
Relative incidence
n=10
n=20
n=50
n=100

(b)


Ratio=0.5
1st order bias
Ratio=0.5
2nd order bias
Bias of log relative incidence
-25 -20 -15 -10 -5 0 5 10 15 20 25
0 1 2 3 4 5 6 7 8 9 10
Relative incidence
n=10
n=20
n=50
n=100

(c)

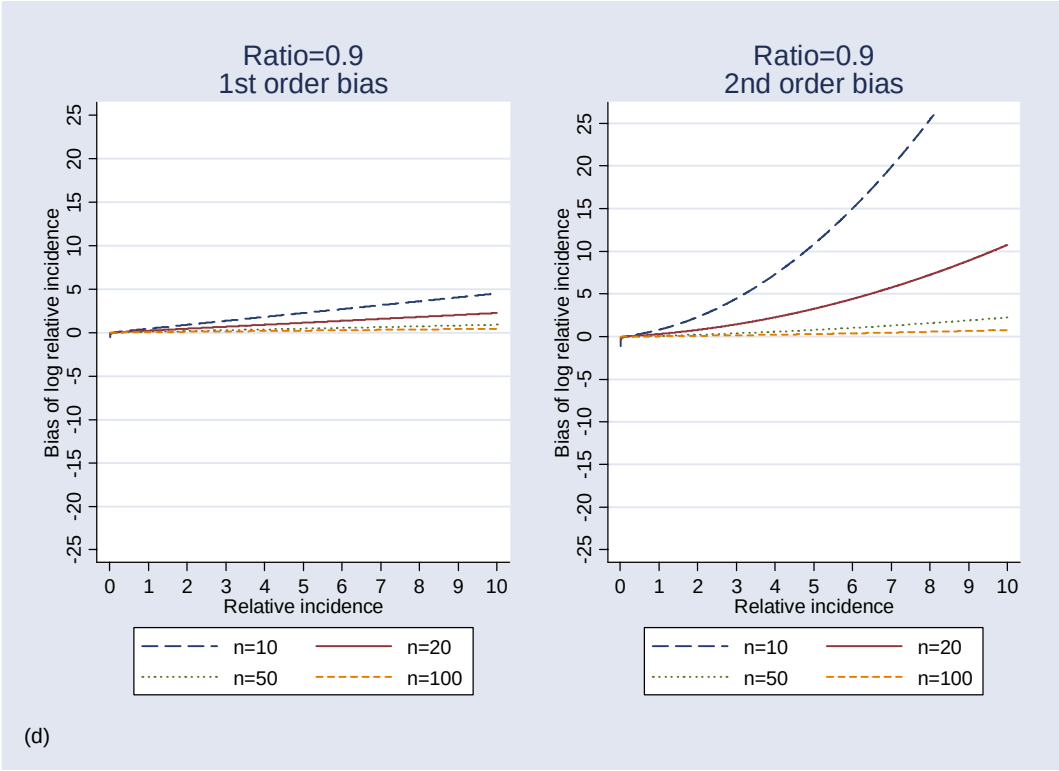


Figure 2.3 First and second order asymptotic bias varying with relative incidence

2.5 The asymptotic variance of $\hat{\beta}$

Farrington et al [3] found the variance of an estimator $\hat{\beta}$ up to the first order in the simplified situation described in section 2.3. In this section, we extend the calculation up to the second order. The variance of $\hat{\beta}$ up to the second order is derived as follows.

Squaring both sides of (2.1) and simplifying, keeping powers up to order 4, we get:

$$\hat{\beta}^2 = f(\mu)^2 + 2f(\mu)(X-\mu)f'(\mu) + (X-\mu)^2\left[f(\mu)f''(\mu) + (f'(\mu))^2\right] +$$

$$(X-\mu)^3\left[\frac{1}{3}f(\mu)f'''(\mu) + f'(\mu)f''(\mu)\right] + (X-\mu)^4\left[\frac{f(\mu)f^{iv}(\mu)}{12} + \frac{f'(\mu)f'''(\mu)}{3} + \frac{(f''(\mu))^2}{4}\right]$$

$+$ residue

From above, we know $f(\mu) = \beta$. Substituting this value in the expression above we get:

$$\hat{\beta}^2 = \beta^2 + 2\beta(X-\mu) f'(\mu) + (X-\mu)^2 \left[\beta f''(\mu) + (f'(\mu))^2 \right] +$$
$$(X-\mu)^3 \left[\frac{1}{3}\beta f'''(\mu) + f'(\mu) f''(\mu) \right] + (X-\mu)^4 \left[\frac{\beta f^{iv}(\mu)}{12} + \frac{f'(\mu) f'''(\mu)}{3} + \frac{(f''(\mu))^2}{4} \right]$$
$$+ \text{ residue}$$

Taking expectations of both sides:

$$E\left(\hat{\beta}^2\right) = \beta^2 + \text{var}(X) \left[\beta f''(\mu) + (f'(\mu))^2 \right] +$$
$$\text{E}(X-\mu)^3 \left[\frac{1}{3}\beta f'''(\mu) + f'(\mu) f''(\mu) \right]$$
$$+ E(X-\mu)^4 \left[\frac{\beta f^{iv}(\mu)}{12} + \frac{f'(\mu) f'''(\mu)}{3} + \frac{(f''(\mu))^2}{4} \right] + \text{residue}$$

Substituting for the other values we have:

$$E\left(\hat{\beta}^2\right) = \beta^2 + npq \left[\frac{\beta}{n^2} \frac{(p-q)}{p^2 q^2} + \frac{1}{(npq)^2} \right] + npq(q-p) \left[\frac{\beta}{3n^3} \frac{\left(p^3+q^3\right)}{p^3 q^3} + \frac{1}{npq} \frac{1}{n^2 p^2 q^2} (p-q) \right]$$
$$+ 3n^2 p^2 q^2 \left[\frac{\beta}{12} \times \frac{6}{n^4} \times \frac{(p-q)\left(p^2+q^2\right)}{p^4 q^4} + \frac{2}{3} \frac{1}{npq} \frac{p^3+q^3}{n^3 p^3 q^3} + \frac{1}{4n^4} \frac{(p-q)^2}{p^4 q^4} \right] + O(n^{-3})$$

Simplifying the above expression we get:

$$E\left(\hat{\beta}^2\right) = \beta^2 + \frac{1+\beta(p-q)}{npq} - \frac{(p-q)^2}{4n^2 p^2 q^2} + \frac{2\beta(p-q)}{n^2 p^2 q^2} \left[-\frac{1}{3}\left(p^3+q^3\right) + \frac{3}{4}\left(p^2+q^2\right) \right]$$
$$+ \frac{2\left(p^3+q^3\right)}{n^2 p^2 q^2} + O(n^{-3})$$

The asymptotic variance of $\hat{\beta}$ is given by:

$\text{var}\left(\hat{\beta}\right) = E\left[\hat{\beta}^2\right] - \left[E\left(\hat{\beta}\right)\right]^2$. We know from above (2.2) that

$E\left(\hat{\beta}\right) = \beta + \frac{1}{2n}\frac{(p-q)}{pq} + \frac{1}{n^2}\frac{(p-q)}{p^2 q^2}\left[\frac{3}{4}\left(p^2+q^2\right) - \frac{\left(p^3+q^3\right)}{3}\right] + O(n^{-3})$. Squaring both sides:

$$\left[E\left(\hat{\beta}\right)\right]^2 = \beta^2 + \frac{\beta}{n}\frac{(p-q)}{pq} + \frac{(p-q)^2}{4n^2p^2q^2} + \frac{2\beta}{n^2}\frac{(p-q)}{p^2q^2}\left[\frac{3}{4}\left(p^2+q^2\right) - \frac{\left(p^3+q^3\right)}{3}\right] + O(n^{-3})$$

The asymptotic variance up to the second order is given by:

$$\begin{aligned}\mathrm{var}\left(\hat{\beta}\right) = E\left[\hat{\beta}^2\right] - \left[E\left(\hat{\beta}\right)\right]^2 &= \frac{1}{npq} - \frac{(p-q)^2}{2n^2p^2q^2} + \frac{2\left(p^3+q^3\right)}{n^2p^2q^2} + O(n^{-3}) \\ &= \frac{1}{npq} - \frac{(p-q)^2}{2n^2p^2q^2} + \frac{2\left(p^2-pq+q^2\right)}{n^2p^2q^2} + O(n^{-3})\end{aligned}$$

Substituting the full expressions for p and q in the above we get:

$$\begin{aligned}\mathrm{var}\left(\hat{\beta}\right) = \frac{1}{n}\left(e^{\beta}r+1-r\right)\left(\frac{1}{1-r}+\frac{1}{e^{\beta}r}\right) - \frac{1}{2n^2}\left(\frac{\left(e^{\beta}r-(1-r)\right)\left(e^{\beta}r+1-r\right)}{e^{\beta}r(1-r)}\right)^2 \\ + \frac{2}{n^2}\left[\left(\frac{e^{\beta}r+1-r}{1-r}\right)^2 + \frac{\left(e^{\beta}r+1-r\right)^2}{e^{\beta}r(1-r)} + \left(\frac{e^{\beta}r+1-r}{e^{\beta}r}\right)^2\right] + O(n^{-3}) \qquad (2.5)\end{aligned}$$

Note from (2.5) above, the expression $\frac{1}{n}\left(e^{\beta}r+1-r\right)\left(\frac{1}{1-r}+\frac{1}{e^{\beta}r}\right)$ is the first order variance as was found by Farrington et al [3]. Note further that (2.5) can be factorised to give (2.6) below:

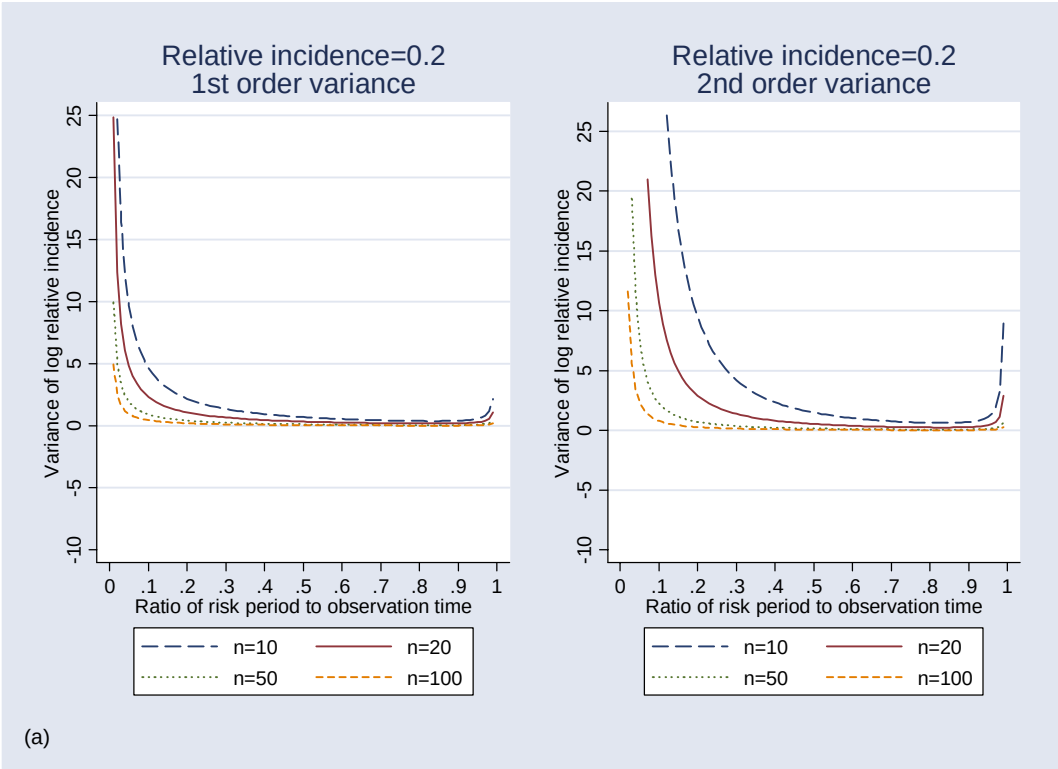
$$\mathrm{var}\left(\hat{\beta}\right) = \frac{1}{n}\frac{\left(e^{\beta}r+1-r\right)^2}{e^{\beta}r(1-r)}\left[1+\frac{3\left(e^{\beta}r\right)^2 - 2e^{\beta}r(1-r) + 3(1-r)^2}{2ne^{\beta}r(1-r)}\right] + O(n^{-3}) \qquad (2.6)$$

We see that the expression

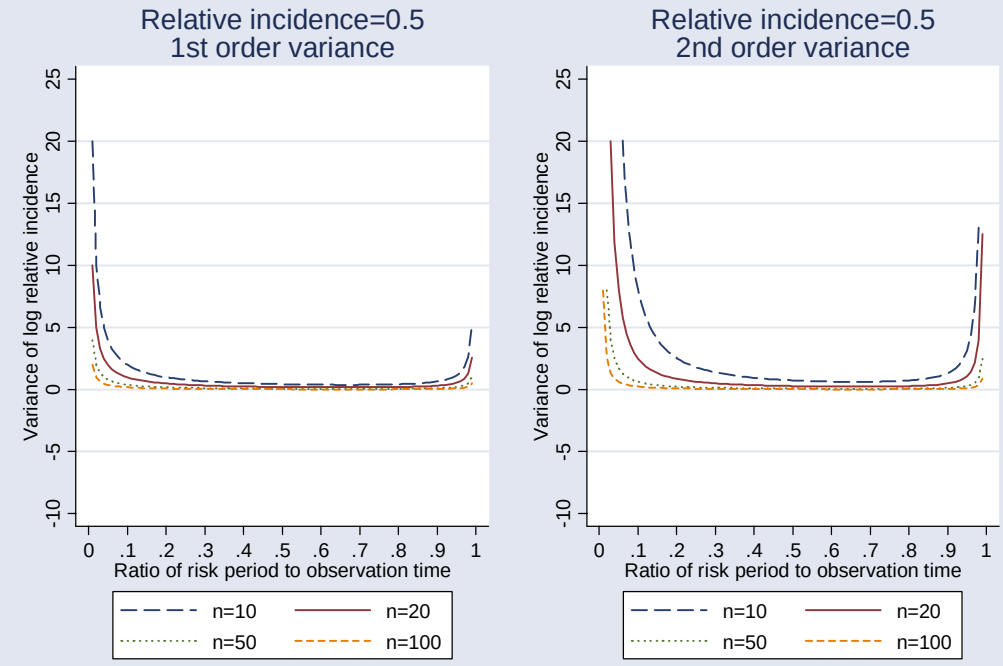
$\left[1+\frac{3\left(e^{\beta}r\right)^2 - 2e^{\beta}r(1-r) + 3(1-r)^2}{2ne^{\beta}r(1-r)}\right]$ is greater than 1, hence the second order variance is always greater than the first order variance. The asymptotic variance is minimised when $e^{\beta}r = 1-r$. This will be illustrated in the next section graphically.

2.6 Graphical study of the asymptotic variance of $\hat{\beta}$

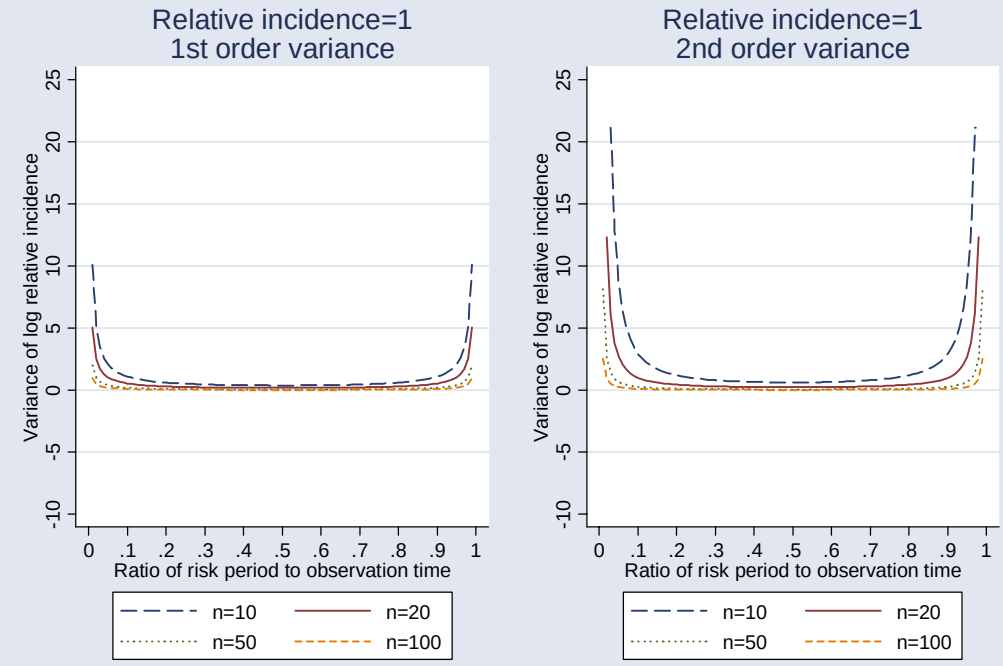
Figure 2.4 and Figure 2.5 below show how the asymptotic variance up to the first order and second order varies with the ratio of the risk period to the observation period and with the relative incidence at fixed sample sizes. Just as in the graphical study of the bias, we explored the behaviour of the asymptotic variance for $n = 10, 20, 50,$ and 100. We fixed the relative incidences at 0.2, 0.5, 1, 2, 5 and 10 and then varied the ratio of the risk period to the observation period (Figure 2.4 (a), (b), (c), (d), (e) and (f)). We also fixed the ratio of the risk period to the observation period at 0.1, 0.2, 0.5, and 0.9 and varied the relative incidence (Figures 2.5 (a), (b), (c) and (d)).



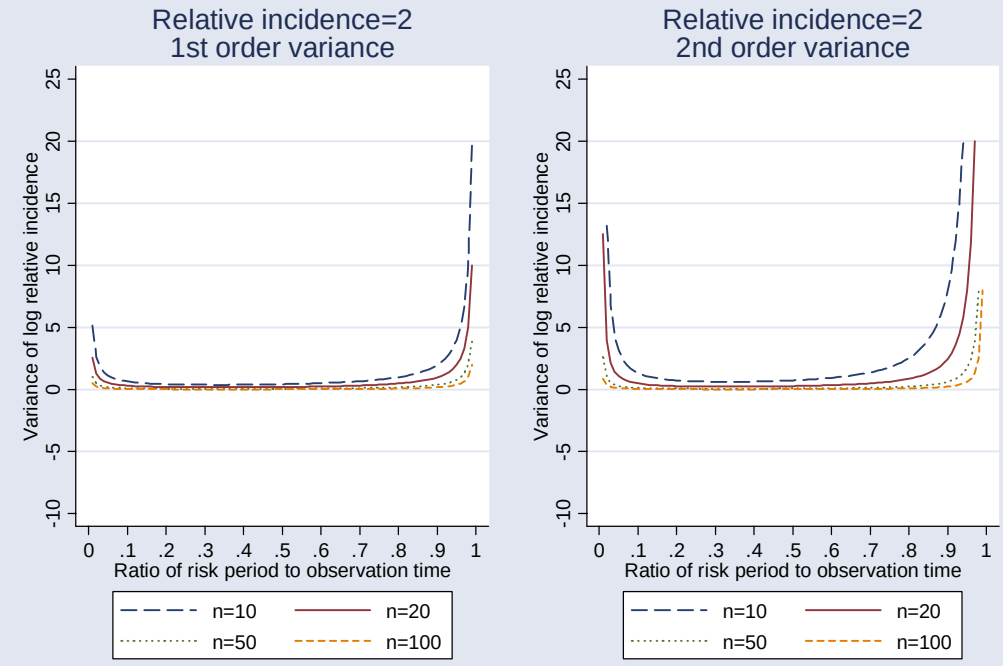
(a)


Relative incidence=0.5
1st order variance
Relative incidence=0.5
2nd order variance
Variance of log relative incidence
-10
-5
0
5
10
15
20
25
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Ratio of risk period to observation time
n=10
n=20
n=50
n=100

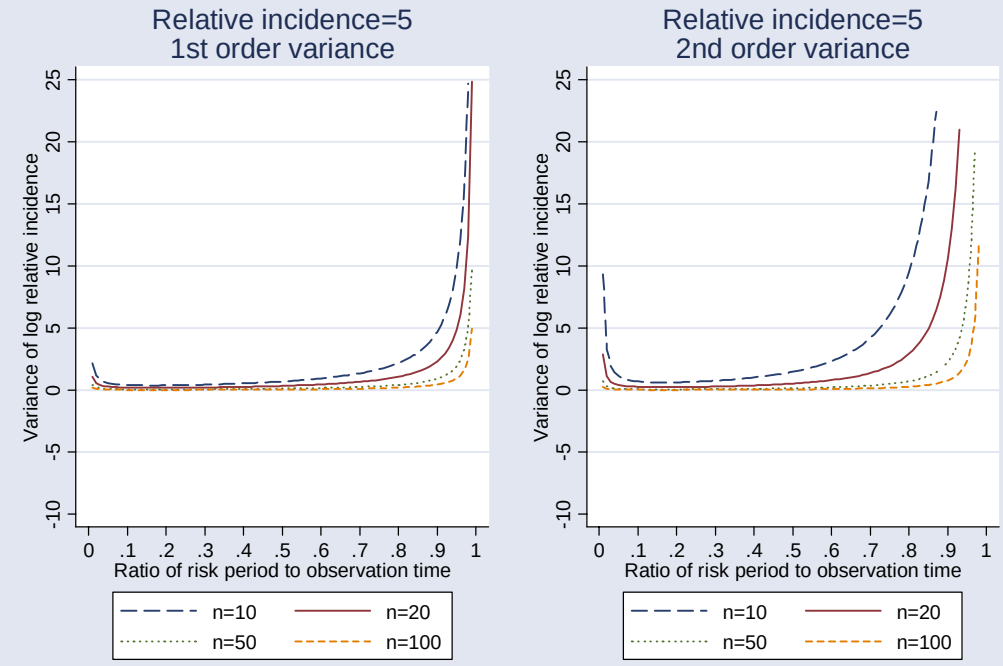
(b)


Relative incidence=1
1st order variance
Relative incidence=1
2nd order variance
Variance of log relative incidence
-10
-5
0
5
10
15
20
25
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Ratio of risk period to observation time
n=10
n=20
n=50
n=100

(c)


Relative incidence=2
1st order variance
Relative incidence=2
2nd order variance
Variance of log relative incidence
-10
-5
0
5
10
15
20
25
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Ratio of risk period to observation time
n=10
n=20
n=50
n=100

(d)


Relative incidence=5
1st order variance
Relative incidence=5
2nd order variance
Variance of log relative incidence
-10
-5
0
5
10
15
20
25
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Ratio of risk period to observation time
n=10
n=20
n=50
n=100

(e)

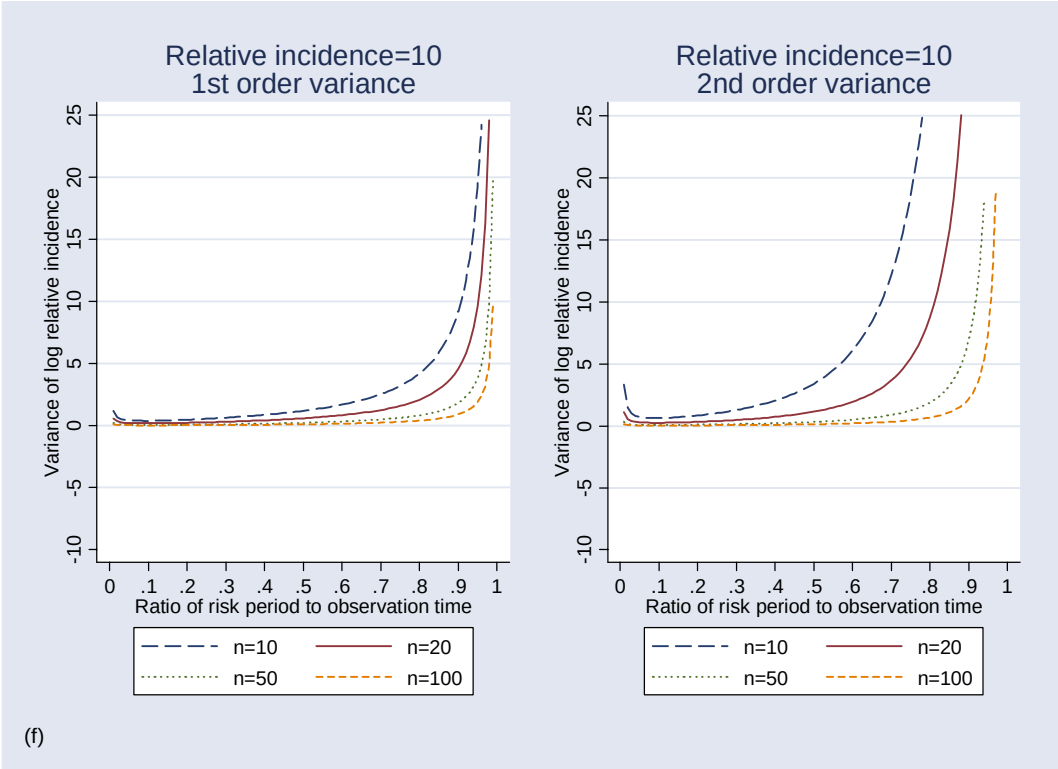
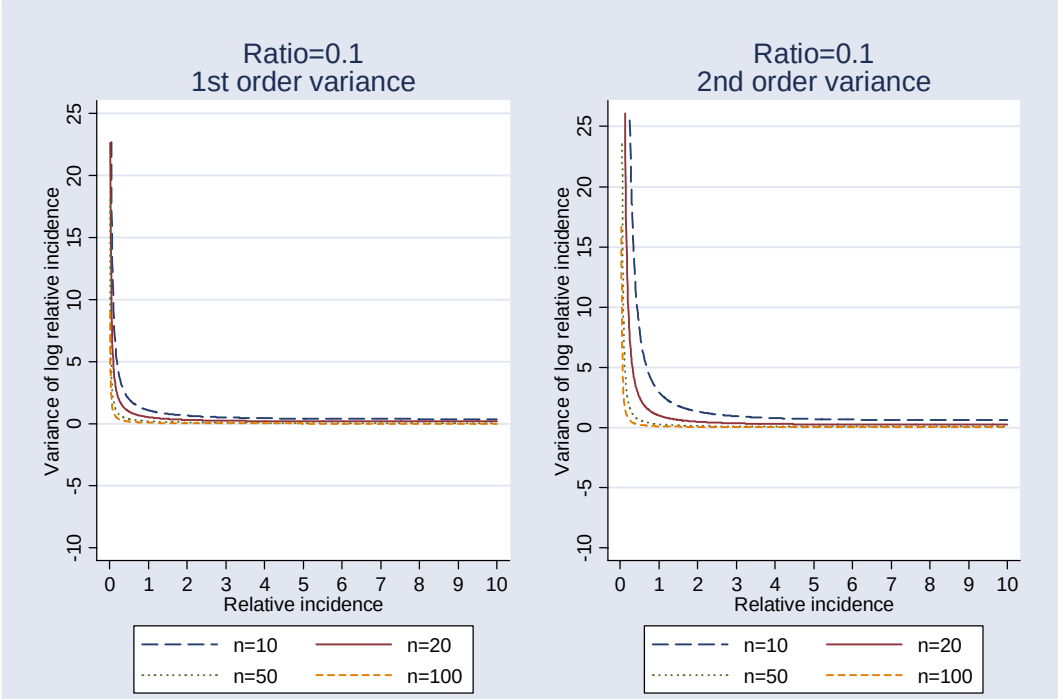


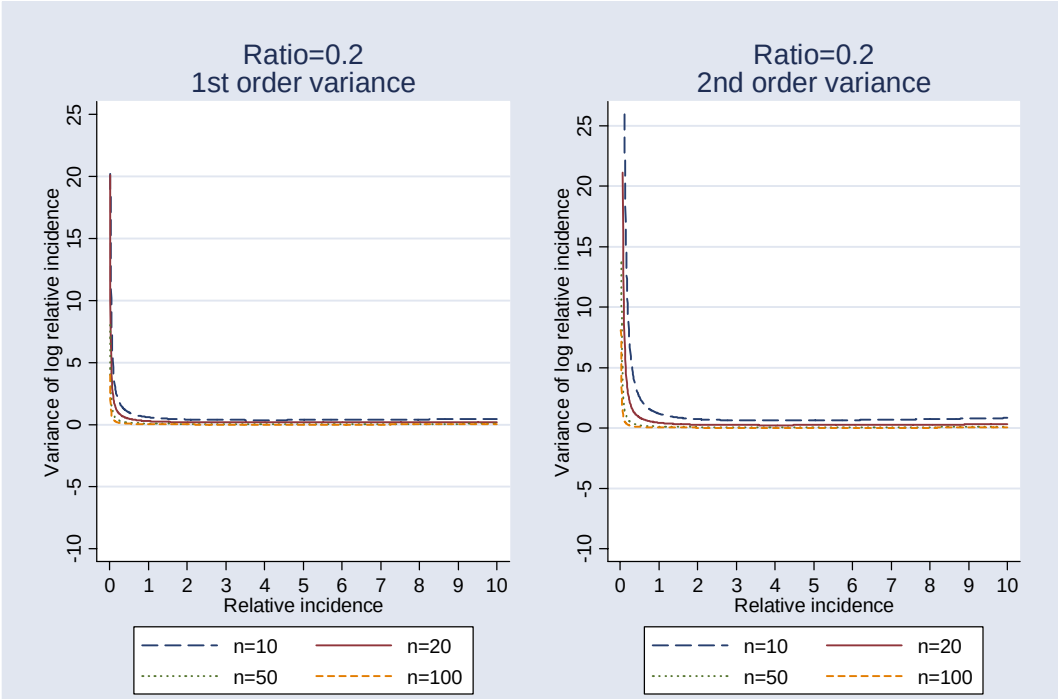
Figure 2.4 Asymptotic variance to first and second order varying with the ratio of the risk period to the observation period.

We note that in all graphs shown the first order variance was less than the second order variance as expected. However, there was little difference between the first order variance and the second order variance. As for the bias, the variance decreases with increasing sample sizes. The variance is largest for ratios of risk period to observation period that are less than 0.1 and greater than 0.9 (Figures 2.4 (a), (b), (c), (d), (e), and (f)). Varying the relative incidence and fixing the ratio of risk period to observation period, we obtain large variances for relative incidences less than 1 (see Figure 2.5 (a), (b), (c)). The asymptotic variance is also large when the ratio of the risk period to the observation period is high and the number of cases is small (n=10 and 20 see Figure 2.5 (d)) but this effect disappears as the number of cases increase (n=50 and 100). The parameter values that give large asymptotic bias tend to be the same that give large asymptotic variance, and conversely the parameter values that

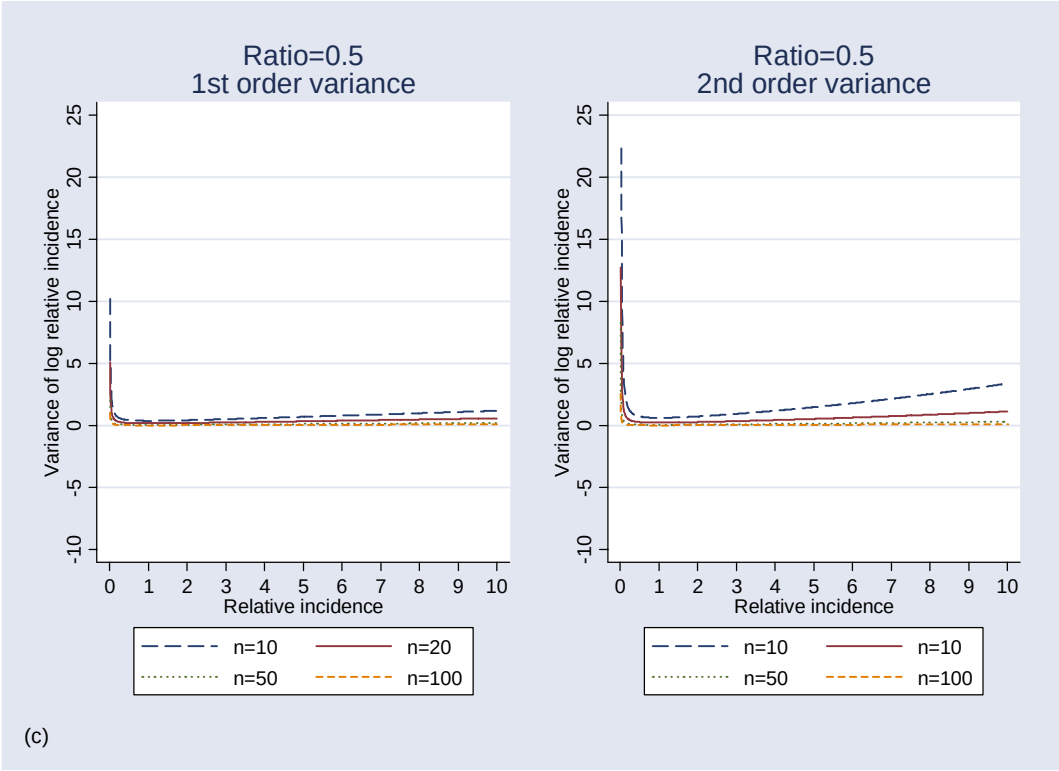
give small bias tend to give small variance (see Figure 2.2 compared to Figure 2.4, and Figure 2.3 compared to Figure 2.5).



(a)



(b)



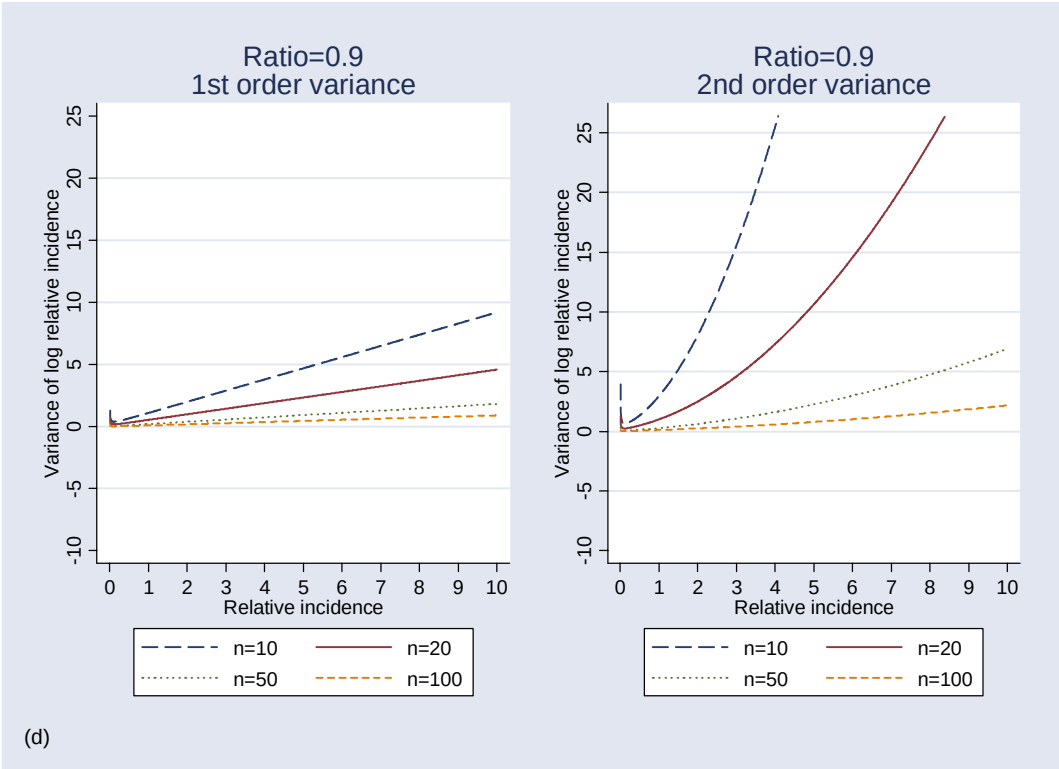


Figure 2.5 First and second order variance varying with relative incidence

2.7 Asymptotic mean square error (AMSE)

The asymptotic mean square error up to the second order is derived as shown below. Suppose $\hat{\beta}$ is an estimator of β with some bias, that is, $E\left(\hat{\beta}\right)=\beta+b$ where b is the bias. The variance of $\hat{\beta}$ is $\text{var}\left(\hat{\beta}\right)=E\left[\hat{\beta}-E\left(\hat{\beta}\right)\right]^2$. The quantity $E\left[\hat{\beta}-\beta\right]^2$ is called the mean square error and can be written as

$$E\left[\hat{\beta}-\beta\right]^2=E\left[\hat{\beta}-(\beta+b)+b\right]^2=E\left[\hat{\beta}-E\left(\hat{\beta}\right)+b\right]^2=E\left[\hat{\beta}-E\left(\hat{\beta}\right)\right]^2+b^2 \qquad (2.7)$$

$E\left[\hat{\beta}-E\left(\hat{\beta}\right)\right]^2$ is simply the variance. Hence if we use the asymptotic variance up to the second order and the asymptotic bias up to the first order in (2.7) above for the simple situation described earlier we obtain the asymptotic mean square error (AMSE) as given in the formula below.

$$\begin{aligned}\text{AMSE} &= \left(\text{asymptotic bias}\right)^2+\text{ asymptotic variance }+O\left(n^{-3}\right) \qquad (2.7)\\ &=\left[\frac{1}{2n}\times\left(re^{\beta}-(1-r)\right)\times\left(\frac{1}{1-r}+\frac{1}{re^{\beta}}\right)\right]^2\\ &\quad+\frac{1}{n}\frac{\left(e^{\beta}r+1-r\right)^2}{e^{\beta}r\left(1-r\right)}\left[1+\frac{3\left(e^{\beta}r\right)^2-2e^{\beta}r\left(1-r\right)+3\left(1-r\right)^2}{2ne^{\beta}r\left(1-r\right)}\right]+O(n^{-3})\end{aligned}$$

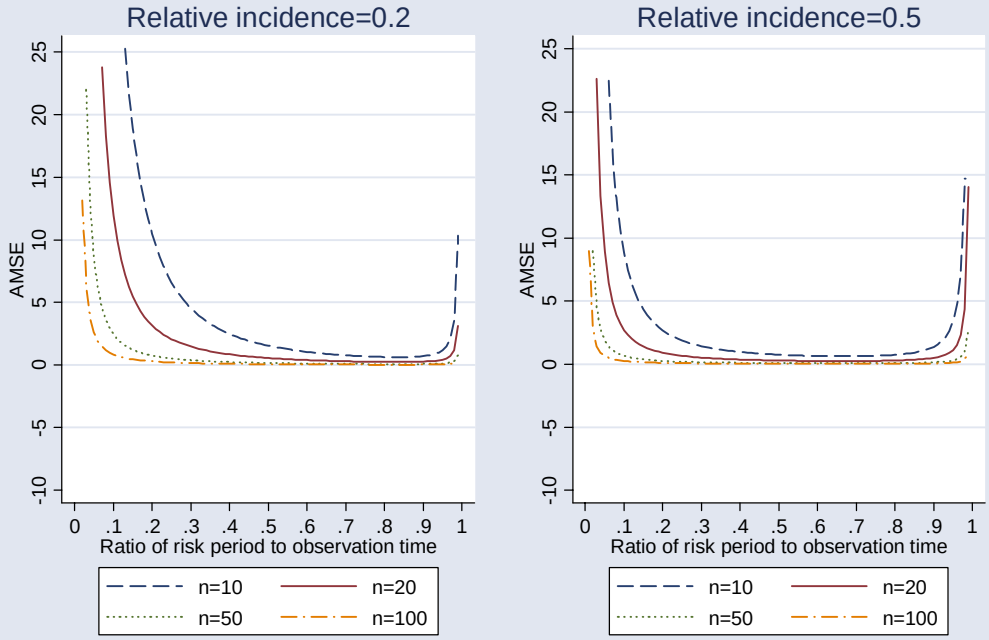
Simplifying the above, the AMSE up to the second order is:

$$\text{AMSE }=\frac{1}{n}\frac{\left(e^{\beta}r+1-r\right)^2}{e^{\beta}r(1-r)}\left[1+\frac{7\left(e^{\beta}r\right)^2-6e^{\beta}r\left(1-r\right)+7\left(1-r\right)^2}{4ne^{\beta}r(1-r)}\right]+O\left(n^{-3}\right).$$

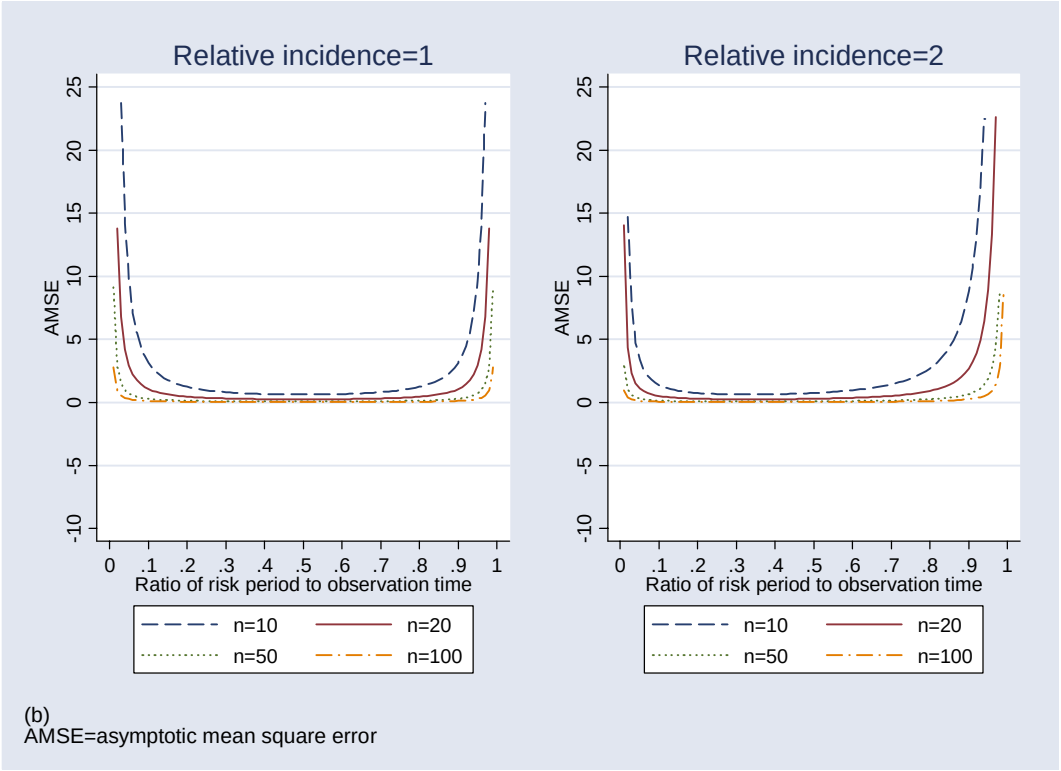
We can see that the expression $\frac{1}{n}\frac{\left(e^{\beta}r+1-r\right)^2}{e^{\beta}r(1-r)}$ is the variance up to the first order. We expect the graphical illustration of AMSE to be similar to that of the asymptotic variance, especially for those values where the bias is close to zero.

2.8 Graphical study of AMSE

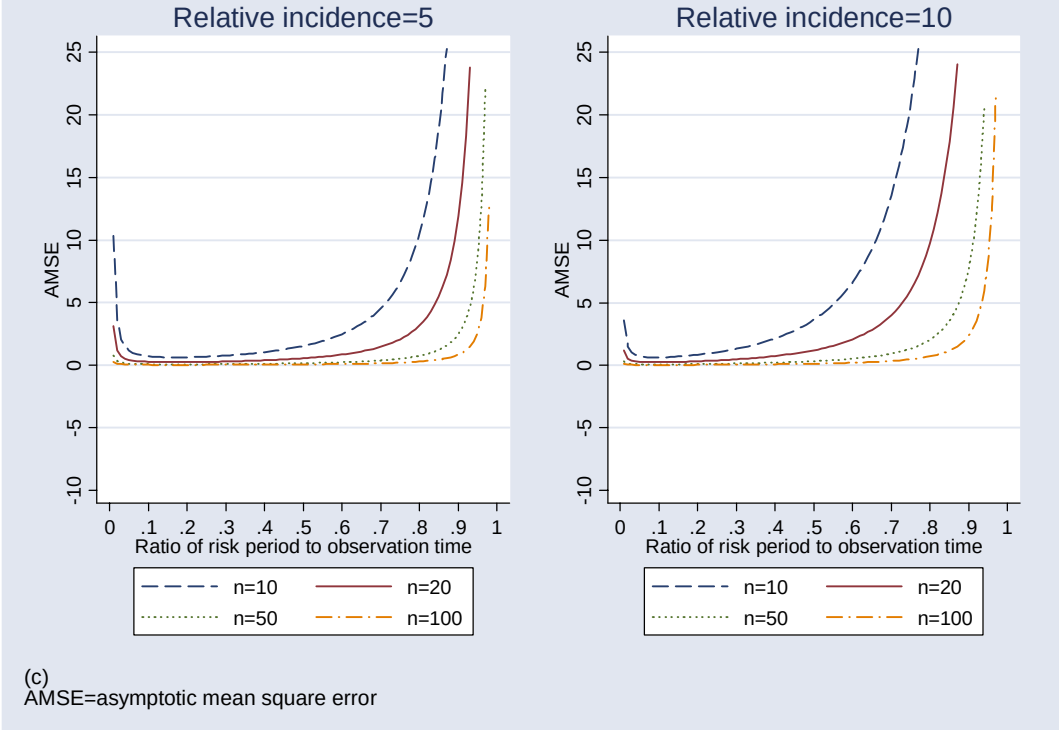
The AMSE varying with the ratio of the risk period to the observation at fixed relative incidences of 0.2, 0.5, 1, 2, 5 and 10 is represented in Figure 2.6 (a), (b) and (c). Similarly, Figure 2.7 (a) and (b) shows the AMSE varying with the relative incidence at fixed ratio of the risk period to the observation period with $r = 0.1, 0.2, 0.5, 0.9$.



(a)
AMSE=asymptotic mean square error

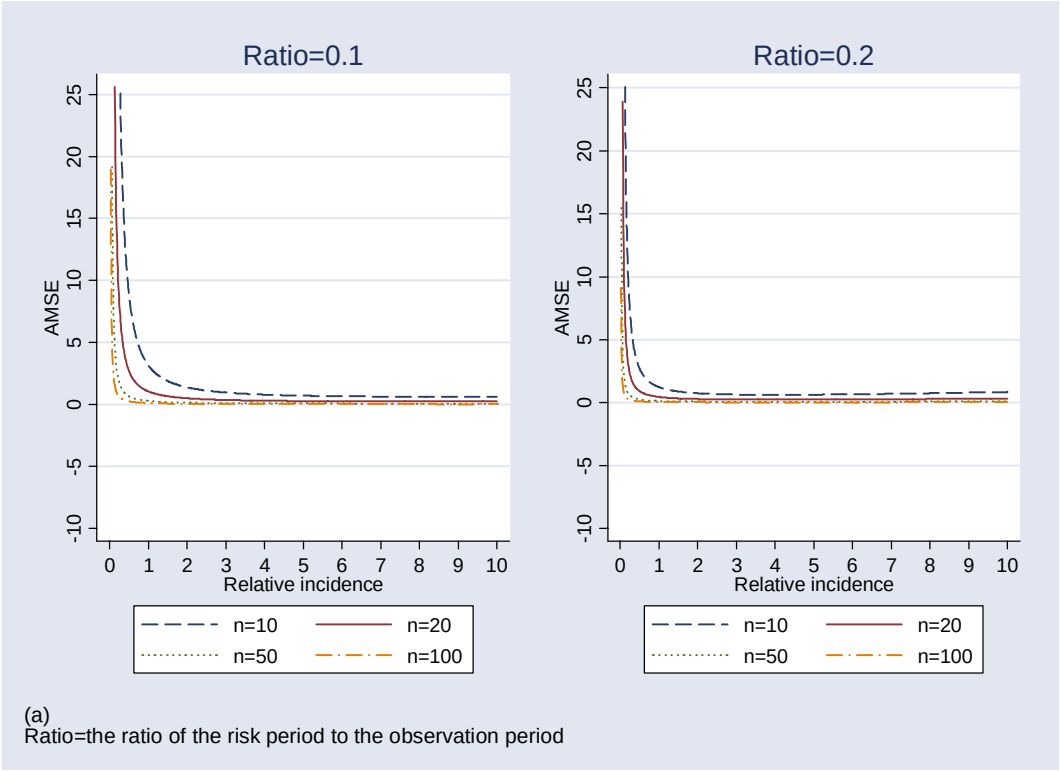


(b)
AMSE=asymptotic mean square error

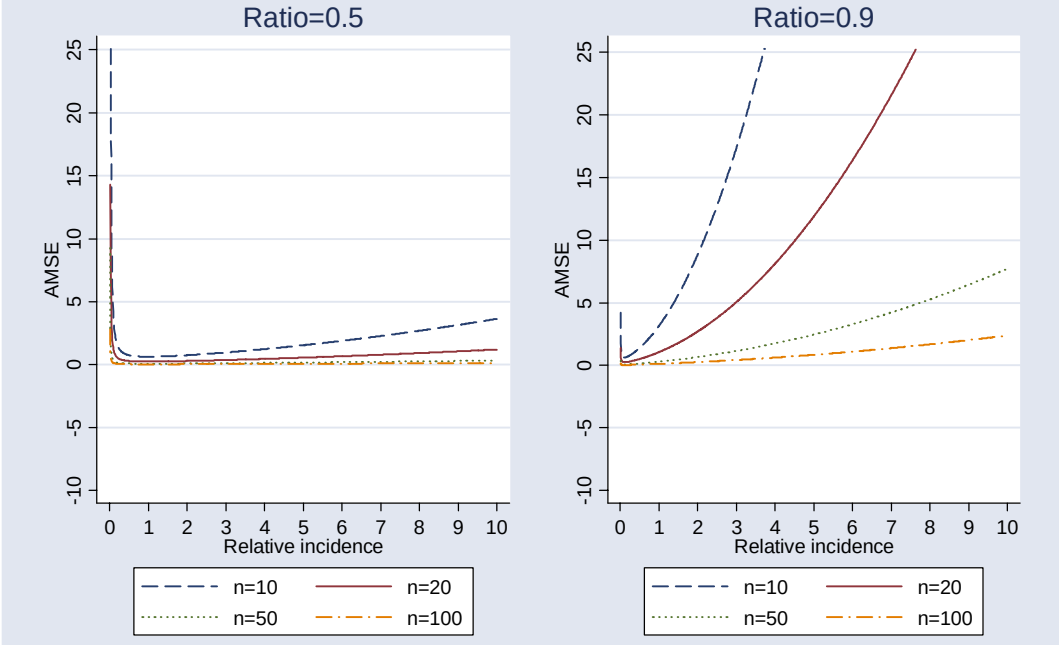


(c)
AMSE=asymptotic mean square error

Figure 2.6 Asymptotic mean square error as a function of the ratio of risk period to the observation period.



(a)
Ratio=the ratio of the risk period to the observation period



(b)
Ratio=the ratio of the risk period to the observation period

Figure 2.7 Asymptotic mean square error as a function of the relative incidence

As for the bias and variance the values of the AMSE were obtained for the sample sizes n=10, 20, 50, and 100. Just as for the asymptotic bias and variance, the AMSE decreases with increasing sample size. Comparing Figures 2.6 and 2.7 with those obtained for the asymptotic variance, we can see that the contribution of the bias is negligible in most situations.

2.9 Conclusion

In this chapter, we have described the self-controlled case series model and introduced the notation we will be working with from here onwards. The likelihood of the model was derived. The maximum likelihood estimator was obtained explicitly in a simple situation, for which the likelihood ratio statistic was also obtained. The asymptotic bias, variance and the AMSE of the estimator in the simplified scenario were calculated. We studied graphically how the asymptotic bias, variance, and the AMSE vary with the ratio of the risk period to the observation period and the relative incidence at fixed sample sizes.

The main finding in this chapter is to characterise how the asymptotic bias, variance and AMSE vary. It is important to know under what situations the self-controlled case series method is going to yield biased and/or imprecise estimates. As illustrated, the bias is zero when there is same number of cases expected in the risk period as out of it. The asymptotic bias and the variance are large when the ratio or the risk period to observation period is less than 0.1 and when it is greater than 0.9. They are also large for relative incidences less than 1 and when the relative incidence is greater than 8. However outside these extremes, the asymptotic bias is close to zero and the

asymptotic variance is close to its minimum. In other words, the bias and variance do not depend sensitively on the parameters r and β within their central range.

Nevertheless, in many applications to vaccines, short risk periods are required and substantial bias can arise in such circumstances unless the relative incidence is high. Further investigation, based on simulation rather than asymptotic theory, is therefore required. A further limitation of our asymptotic results is that we have explored the model without taking age into account. Age is a well known confounder in adverse outcomes with respect to vaccines in which the self-controlled case series model is widely used. We did not take age into account because the calculation for the asymptotic bias, variance and the AMSE become unwieldy. We shall explore the effects of age at event, age at exposure, risk periods (fixed or indefinite) and small samples by simulation in the next chapter.

Chapter 3

Performance of the self-controlled case series method: Simulation study

3.1 Introduction.

In previous chapters the theory behind the self-controlled case series method has been described and its properties outlined. In chapter 2, the asymptotic variance and bias and asymptotic mean square error up to the second order were derived in a simplified scenario.

The estimates obtained by the method are valid asymptotically by virtue of likelihood theory. What we now need is to explore the performance of the method in small samples and under different conditions. This chapter explores various simulations in which we generated data where the true population value of the relative incidence is known. We use the self-controlled case series model to analyse the simulated data and compare the estimate with the true value. The simulations were set up to mimic those scenarios that typically occur in studies of paediatric vaccines.

In section 3.2 we describe how the simulations were carried out. The results from the simulations are given in section 3.3. We first present the results from what we call the standard scenario with varying number of cases and a range of true relative incidences. In subsequent subsections we vary the risk periods, the effect of age at event, and the distribution of age at exposure. We also investigate indefinite risk periods and the presence of unexposed cases. Results exploring the effects of age in

form of age groups labelled as strong symmetric, weak monotone increasing age effect and strong monotone increasing age effect are given. In all the simulations we obtain central estimates and the coverage probabilities for 90%, 95%, and 99% confidence intervals. Also given are the percentage of simulations for which the true value of β was below the lower 90%, 95%, 99% confidence limits, and the percentage for which the true value of β was above the upper 90%, 95%, and 99% confidence limits. The conclusions of the chapter are given in section 3.4. To reduce clutter, most of the detailed results are given in an Appendix.

3.2 The structure of the simulation study

Figure 3.1 shows the structure and stages of the simulation that were carried out. For a given set of parameters (described below), sample size *n* and random seed, a set of *n* exposure times were generated, together with *n* marginal total numbers of events. These marginal totals are generated using a truncated Poisson distribution (excluding zero), conditionally on the exposure history.

The exposures and marginal totals vary randomly between runs. However, in each run of 10,000 simulations, the exposures and marginal totals remain fixed. This is to mimic the fact that the case series method is conditional on exposures and marginal totals.

Within a run of 10,000, the events for each individual are randomly reallocated 10,000 times to the age/exposure categories within each individual's person time. This is done according to the case series model, using a multinomial distribution.

The output from each run includes the median of the relative incidence estimates. We also quote the median of their logarithms. The median is chosen rather than the mean, since in finite samples there is a non-zero probability that the estimated log relative incidence is $\pm\infty$ and hence its expectation does not exist. The median, compared to the true value, provides an appropriate measure of central tendency of the finite sample bias. Note that all runs are based on 10,000 independent samples. With this run size, the Monte Carlo standard error for the coverage probability of a 95% confidence interval is about 0.002 (or 0.2 %, see chapter 4 page 101 for how it is calculated).

3.2.1 The parameters

Each simulation requires the following parameters to be specified.

- Observation period, always taken to be 500 days
- Risk period following exposure (described in section 3.2.4)
- Relative incidence: the true relative incidence took values 0.5, 1, 1.5, 2, 5, 10
- Exposure distribution (section 3.2.3)
- Age groups and age-specific relative incidences (section 3.2.2, Figure 3.2 below)
- Baseline rate always taken to be $\phi = 2\times10^{-7}$ per day, or one per hundred thousand over 500-day observation period. Thus the event is assumed to be rare, and with high probability, a case has only a single event.
- Sample size: we did simulations with 10, 20, 50, 100, 200, 500, 1000 cases

3.2.2 Age at the event

In most case series analyses, one needs to control for age. We varied the effect of age on the event incidence according to practically realistic scenarios. Thus we explored the self-controlled case series model's performance in the presence of what we call weak symmetric, strong symmetric, weak monotone increasing, and strong monotone increasing age effects. These are defined as follows.

- Weak symmetric age effect. Age groups (in days, with age associated relative incidence in brackets) are: 1-100 (1), 101-200 (1.2), 201-300 (1.5), 301-400 (1.2), and 401-500 (1).
- Strong symmetric age effect. Age groups (in days, with age associated relative incidence in brackets) are: 1-50 (1), 51-100 (2), 101-150 (3), 151-200 (4), 201-250 (5), 251-300 (5), 301-350 (4), 351-400 (3), and 401-500 (1).
- Weak monotone increasing age effect. Age groups (in days, with age associated relative incidence in brackets) are: 1-100 (1), 101-200 (1.1), 201-300 (1.2), 301-400 (1.3), 401-500 (1.4)
- Strong monotone increasing age effect. Age groups (in days, with age associated relative incidence in brackets) are: 1-50 (1), 51-100 (1.5), 101-150 (2), 151-200 (2.5), 201-250 (3), 251-300 (3.5), 301-350 (4), 351-400 (5), 451-500 (5.5).

Figure 3.2 below shows bar charts representing each of these four choices.

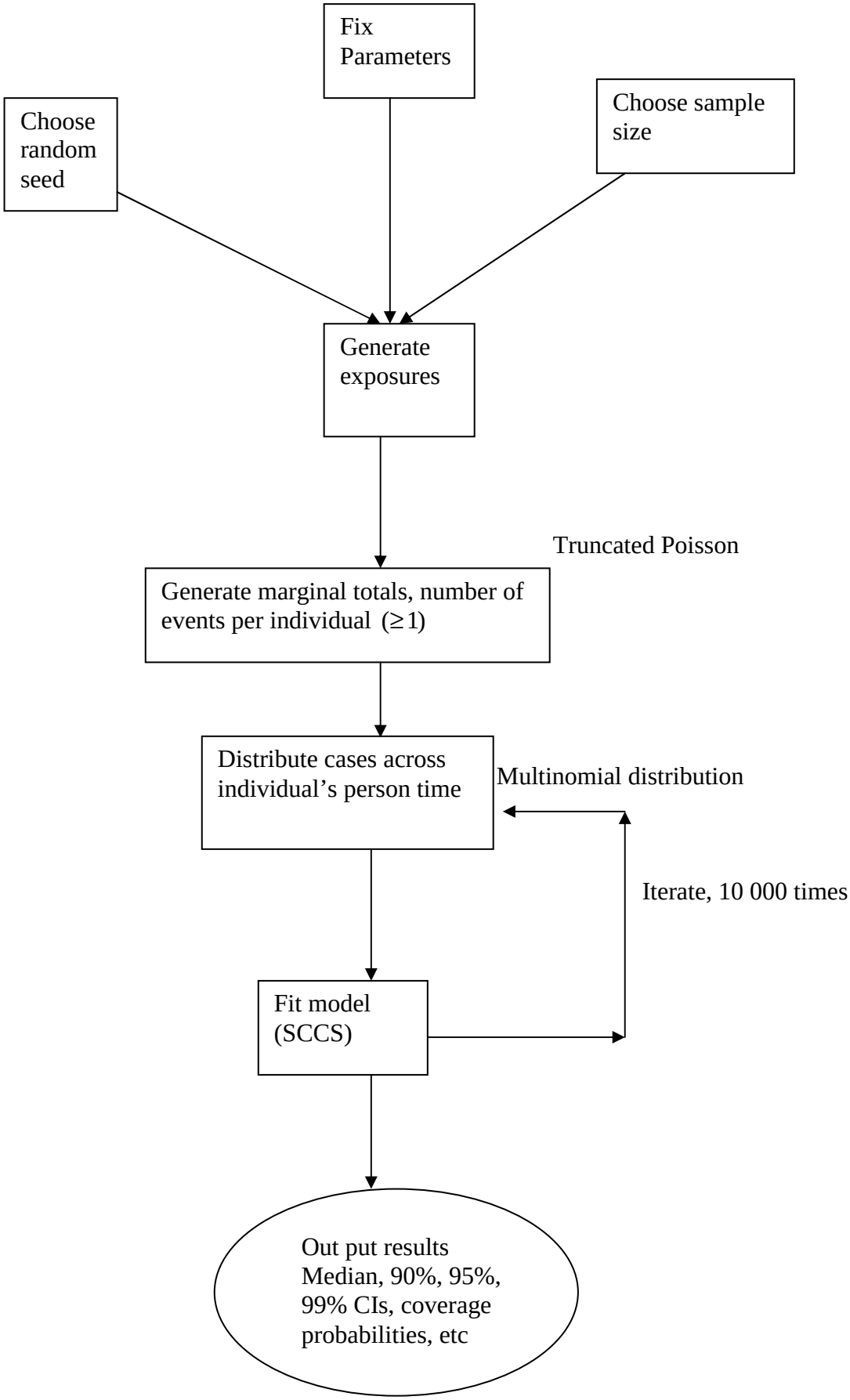


Figure 3.1 Overview structure of the simulation study.

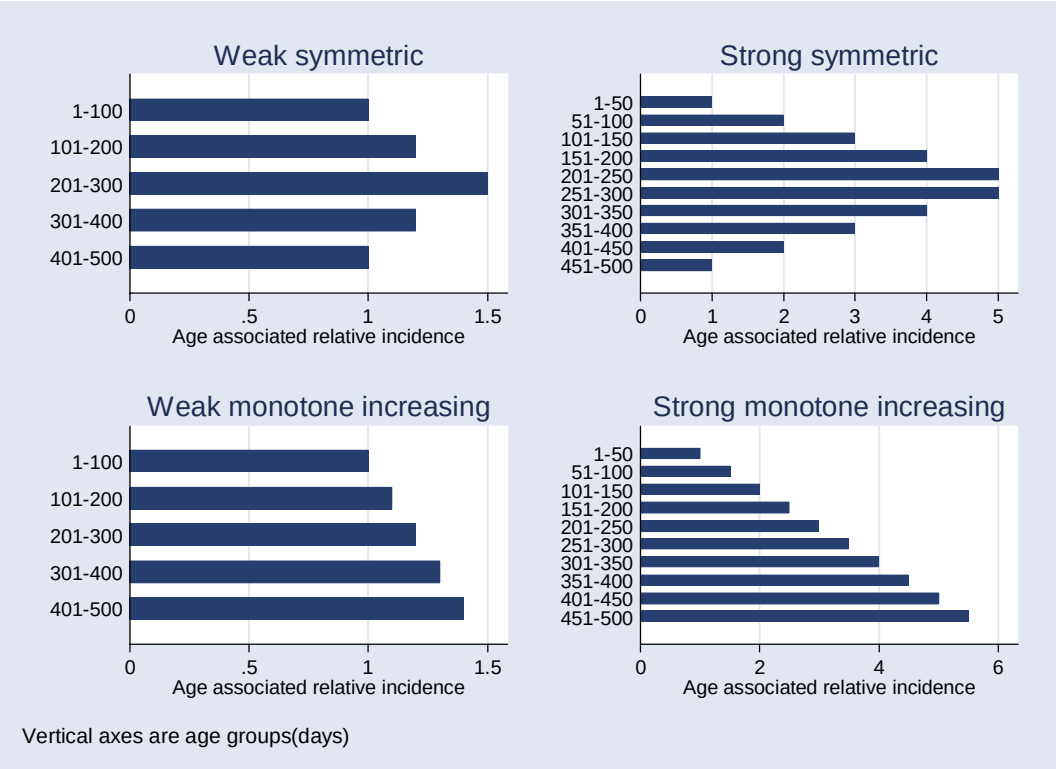


Figure 3.2 The four types of age effect.

3.2.3 Exposure distribution

The precision of the relative incidence estimator depends on the extent of between individual variation in exposure. We used a beta distribution on $[0,500]$ to generate age at exposure. The following distributions of age at exposure were investigated.

- Mean age of 250 days and standard deviation of 100 days
- Mean age of 250 days and standard deviation of 50 days
- Mean age of 125 days and standard deviation of 100 days
- Mean age of 125 days and standard deviation of 50 days

These distributions are shown in Figure 3.3.

For some simulations, much more highly peaked distributions of age at exposure were considered, with mean age of 125 days and standard deviation of 10, 20, 30, and 40 days. Figure 3.4 below shows graphs of these more extreme distributions of age at exposure.

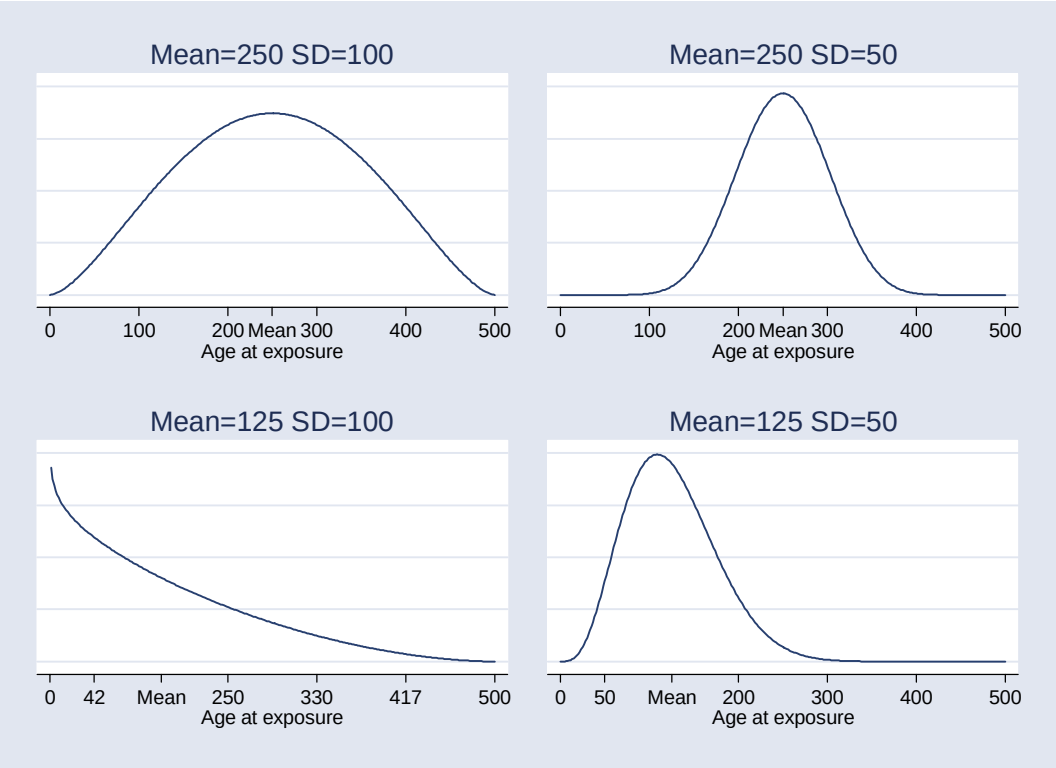


Figure 3.3 Distribution of age at exposure

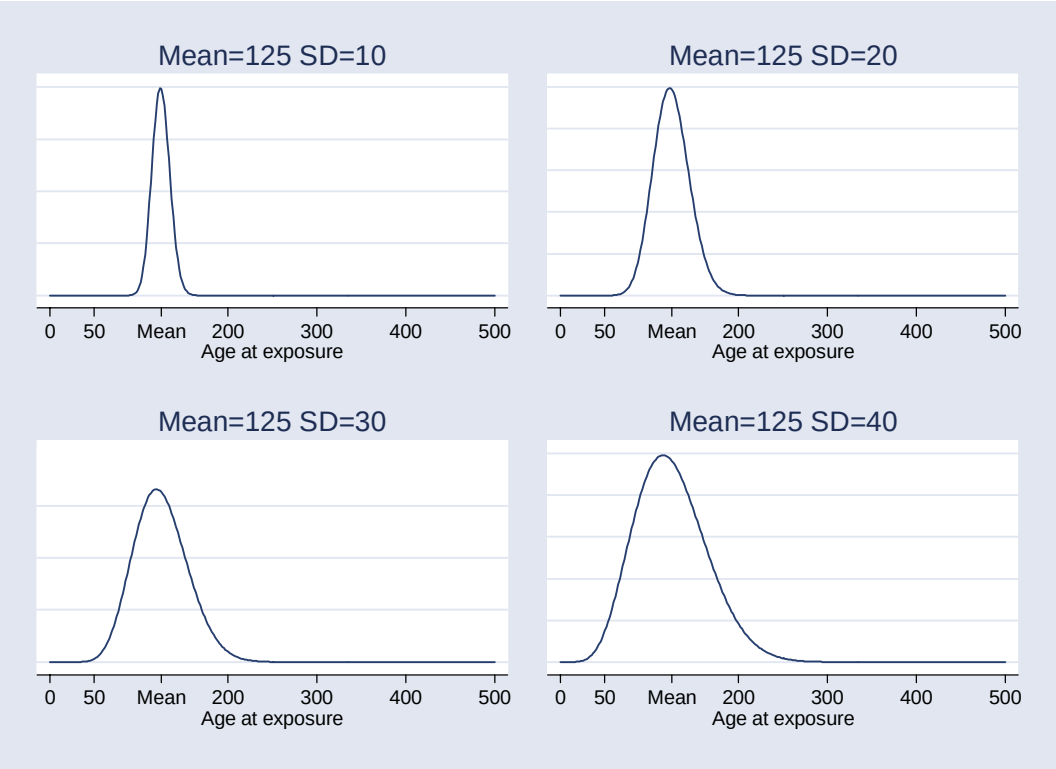


Figure 3.4 Unusually peaked distributions of age at exposure

3.2.4 Risk periods

In the self-controlled case series method, a major issue one has to consider before doing any analysis is to define the risk period. Generally speaking the risk period is elicited from experts. Different studies need different risk periods. These range from very short (one or a few days) to very long (and occasionally indefinite). Typically, in vaccine studies, risk periods of a few days/weeks are used, for example Farrington et al [9] defined risk periods in three groups (0-3, 4-7, and 8-14 days) when they investigated whether there was any association of diphtheria tetanus pertussis (DTP) vaccine with febrile convulsion, whereas to study a putative association of measles mumps rubella (MMR) vaccine with febrile convulsion they defined two risk periods of 6-11 and 15-35 days after vaccination. In each case these choices were based on prior knowledge of the biology of the relevant bacteria and viruses.

We used different risk periods in order to investigate the effect of risk periods on the performance of the model. We looked at risk periods of 1, 5, 10, 25, 50, 100 and 200 days. We also investigated indefinite risk periods. Owing to potentially strong confounding between age and exposure effects with indefinite risk periods, we considered this scenario separately, and also varied the proportion of cases exposed (in other simulations all cases are exposed).

3.3 Results from the simulation study

3.3.1 The standard scenario

We shall now define what we are considering as the reference point, or the standard (default) values that are typical in studies of childhood vaccination. The standard scenario is one in which the risk period is 25 days, all cases have experienced the exposure (vaccination), and the age effect is weak symmetric (see Figure 3.2). The standard distribution of age at exposure has mean 250 days and standard deviation 100 days (see Figure 3.3 above).

For each run, ten thousand samples of 10 cases, 20 cases, 50 cases, 100 cases, 200 cases, 500 cases, 1000 cases with relative incidences 0.5, 1, 1.5, 2, 5, and 10 were simulated (a total of $7 \times 6 = 42$ runs). Tables 3.1 and 3.2 below show the results of the data simulated under the standard scenario.

True value RI Log(RI)	10 000 samples of 10 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 20 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 50 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*
0.500 -0.693	**0.000 −∞** *(4%, 96%, 0%)* *(3%, 97%, 0%)* *(2%, 98%, 0%)*	**0.000 −∞** *(8%, 92%, 0%)* *(3.5%, 96%, 0.5%)* *(1.7%, 98%, 0.3%)*	**0.378 -0.973** *(4.9%, 95%, 0.1%)* *(4%, 96%, 0%)* *(1.%, 99%, 0%)*
1.000 0.000	**0.000 −∞** *(7%, 93%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**0.887 -0.120** *(7%, 93%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.994 -0.006** *(5.8%, 94%, 0.2%)* *(3.8%, 96%, 0.2%)* *(1%, 99%, 0%)*
1.500 0.405	**1.411 0.541** *(4.8%, 95%, 0.2%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**1.414 0.347** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**1.478 0.391** *(5.7%, 94%, 0.3%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*
2.000 0.693	**2.004 0.695** *(6%, 94%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**1.908 0.646** *(6%, 94%, 0%)* *(3.6%, 96%, 0.4%)* *(1%, 99%, 0%)*	**1.967 0.676** *(6%, 92%, 2%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*
5.000 1.609	**4.875 1.584** *(5%, 95%, 0%)* *(2%, 98%, 2%)* *(0%, 100%, 0%)*	**5.037 1.617** *(6%, 92%, 2%)* *(3%, 97%, 0%)* *(0%, 100%, 0%)*	**5.008 1.611** *(6%, 91%, 3%)* *(3%, 96%, 1%)* *(1%, 99%, 0%)*
10.000 2.303	**11.189 2.415** *(5%, 94%, 1%)* *(1%, 99%, 0%)* *(0%, 100%, 0%)*	**10.667 2.367** *(6%, 90%, 4%)* *(3%, 96%, 1%)* *(0%, 100%, 0%)*	**10.224 2.325** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(1%, 99%, 0%)*

Table 3.1 Simulation results for standard scenario. **RI**= True value of Relative incidence. **Log (RI)**= Logarithm of the relative incidence
%low=percentage where true value was below lower limit of 90%, 95%, 99% confidence interval.
%covered=actual percentage coverage of 90%, 95%, 99% confidence interval.
%hi=percentage where true value was above the upper limit of 90%, 95%, 99% confidence interval.

Table 3.2, Simulation results for the standard scenario (continued from Table 3.1)

True Value **RI** **Log(RI)**	**10 000 samples of 100 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	**10 000 samples of 200 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	**10 000 samples of 500 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	**10 000 samples of 1000 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*
0.500 -0.693	**0.509 -0.676** *(5%, 95%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.474 -0.823** *(5.6%, 92%, 2.4%)* *(2.9%, 97%, 0.1%)* *(1%, 99%, 0%)*	**0.495 -0.703** *(5.7%, 91%, 3.3%)* *(3.1%, 96%, 0.9%)* *(0.7%, 99%, 0.3%)*	**0.496 -0.701** *(5.5%, 90%, 4.5%)* *(3.1%, 95%, 1.9%)* *(0.7%, 99%, 0.3%)*
1.000 0.000	**0.995 -0.005** *(5.7%, 92%, 2.3%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.989 -0.011** *(5.8%, 91%, 3.2%)* *(3.3%, 95%, 1.7%)* *(1%, 99%, 0%)*	**0.996 -0.004** *(5.1%, 90%, 4.9%)* *(3.1%, 95%, 1.9%)* *(1%, 99%, 0%)*	**1.000 -0.004** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0%, 99%, 1%)*
1.500 0.405	**1.409 0.380** *(5.9%, 91%, 3.1%)* *(3.2%, 96%, 0.8%)* *(1%, 99%, 0%)*	**1.492 0.400** *(5.9%, 90%, 4.1%)* *(2.9%, 95%, 2.1%)* *(0.7%, 99%, 0.3%)*	**1.493 0.401** *(5.5%, 90%, 4.5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**1.497 0.404** *(5%, 90%, 5%)* *(2%, 96%, 2%)* *(1%, 99%, 0%)*
2.000 0.693	**1.976 0.681** *(5.5%, 91%, 3.5%)* *(2.9%, 96%, 1.1%)* *(1%, 99%, 0%)*	**1.992 0.689** *(5.7%, 90%, 4.3%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**2.000 0.693** *(5.8%, 89%, 5.2%)* *(3.1%, 95%, 1.9%)* *(0.7%, 99%, 0.3%)*	**1.995 0.691** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(1%, 99%, 0%)*
5.000 1.609	**5.011 1.612** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(1%, 99%, 0%)*	**5.012 1.612** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(1%, 99%, 0%)*	**5.005 1.610** *(3%, 90%, 5%)* *(2%, 95%, 3%)* *(1%, 99%, 0%)*	**5.004 1.610** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(1%, 99%, 0%)*
10.000 2.303	**10.123 2.315** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(1%, 99%, 0%)*	**10.035 2.306** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(1%, 99%, 0%)*	**10.016 2.304** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(1%, 99%, 0%)*	**10.020 2.305** *(5%, 90%, 5%)* *(2.9%, 95%, 2.1%)* *(0.6%, 99%, 0.4%)*

The first column of Tables 3.1 and 3.2 shows the value of the true relative incidence followed by the logarithm of the true relative incidence. The next columns shows median estimate (values in bold) for 10 000 samples of the simulated data for different number of cases, followed by the logarithm of the median estimate. Below the median estimates (values in italics), we have three rows of values corresponding to the percentages of 90%, 95%, and 99% confidence intervals were the true value was below *(%low)*, within *(% covered*) and above *(%hi)* the 90%, 95%, and 99% confidence intervals respectively. Figure 3.5 below shows the relative bias, defined as the ratio

$$\frac{(\text{median relative incidence}) - (\text{true relative incidence})}{(\text{true relative incidence})},$$

as a function of sample size for relative incidences of 0.5, 1, 1.5, 2, 5, 10. The numbers of cases considered were: 10, 20, 50, 100, 200, 500 and 1000.

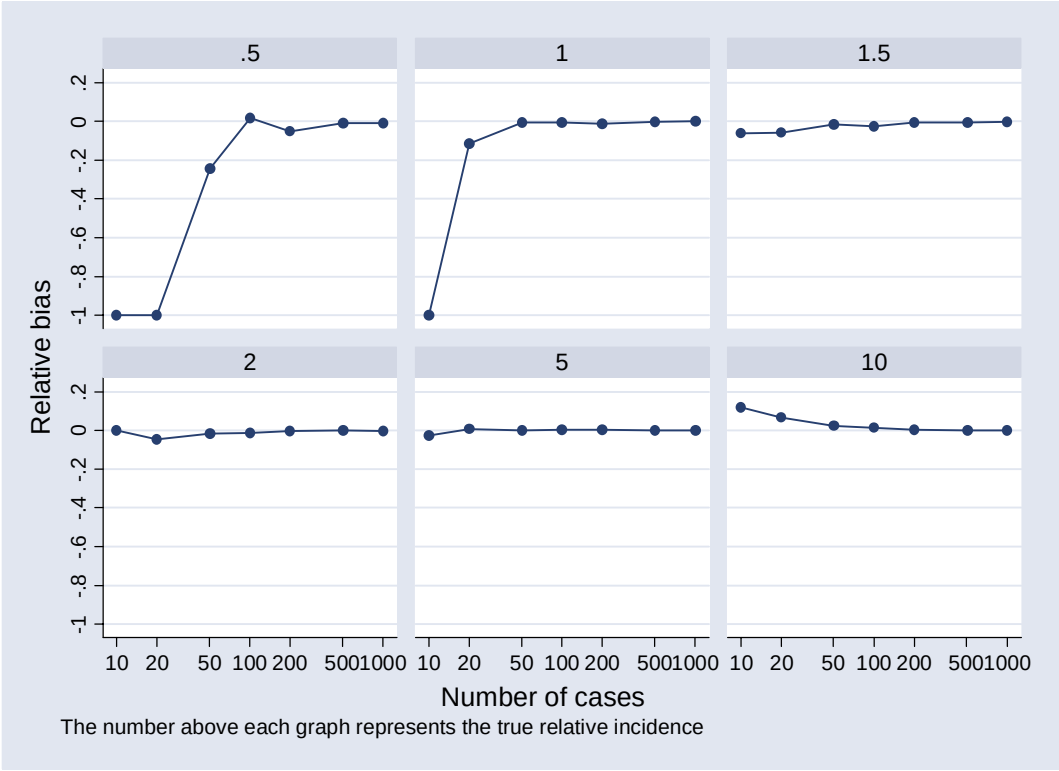


Figure 3.5 Relative (median) bias against median estimates for samples of 10, 20, 50, 100, 200, 500, and 1000 cases for true relative incidences of 0.5, 1, 1.5, 2, 5, 10.

The relative bias is used here for comparison purposes between different true relative incidences. Note that we would expect the relative bias to be equal to zero if the estimates were not biased.

We can see in Figure 3.5 above that the median estimates obtained using the self-controlled case series model with 10 and 20 cases were biased when estimating true relative incidences of 0.5 and 1. This bias largely disappears as the number of cases increases say for numbers of cases greater or equal to 20. The bias was small when estimating the true relative incidences of 1.5, 2, 5 and 10. In chapter two, we showed that the bias (when there are no age effects) is negative when $e^{\beta} < \frac{1-r}{r}$ and positive when $e^{\beta} > \frac{1-r}{r}$. A similar phenomenon is demonstrated in Figure 3.5 above: the bias tends to be negative for small values of the relative incidence, and positive for large values of the relative incidence (here $r = \frac{25}{500} = 0.05$).

The other pattern we can see in Tables 3.1 and 3.2 is that the coverage probabilities for the confidence intervals were reasonably close to the nominal values even with small numbers of cases. Figure 3.6 below illustrates the percentages of 90%, 95%, and 99% confidence intervals that contained the true relative incidence. The figures indicate that overall, for the standard scenario described above, the coverage probabilities tend to be slightly conservative (that is, higher than nominal values) for small sample sizes. As one would expect, the larger the number of cases, the more accurate the coverage probabilities are. Note finally that the confidence intervals tend to be non-central, and are systematically shifted upwards so that the percentage of simulations in which the lower confidence limit falls above the true parameter value is

greater than the percentage of simulations in which the upper confidence limit falls below the true value.

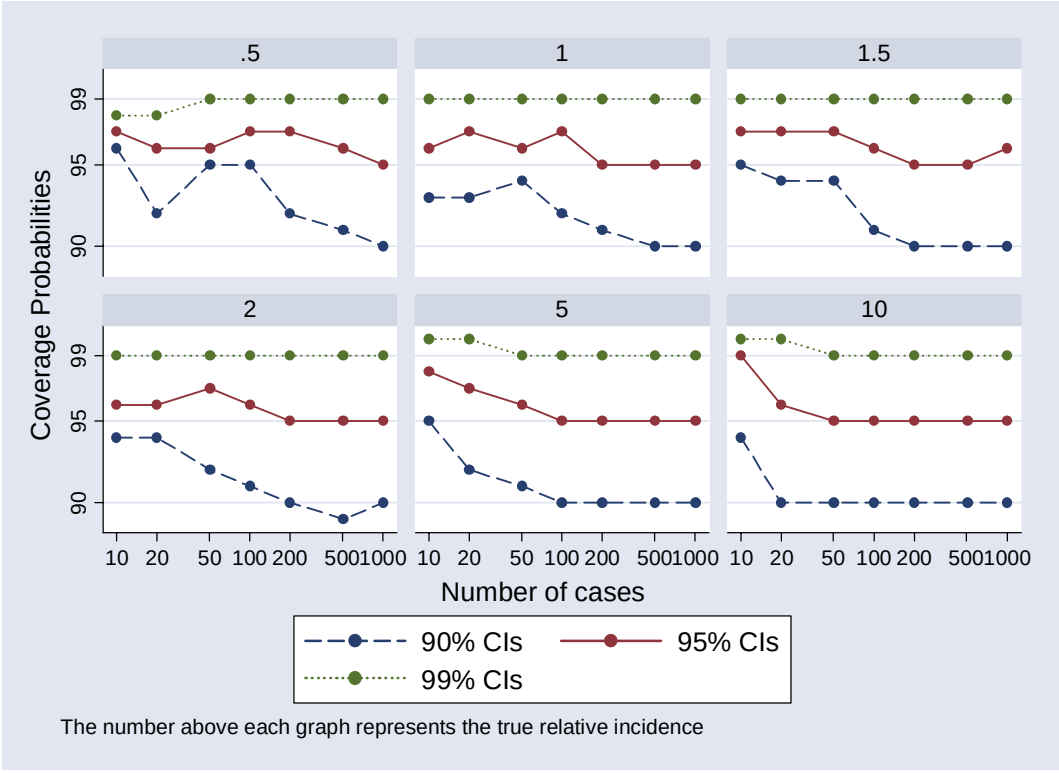


Figure 3.6 Percentages of 90% (in blue/dash), 95% (in red/solid) and 99% (in green/dots) confidence intervals that contained the true relative incidence of 0.5,1, 1.5, 2, 5, 10 for samples of 10, 20, 50, 100, 200, 500, and 1000 cases.

3.3.2 Varying the risk period

In this section we present results from simulations for the standard scenario defined above, except that instead of keeping the risk period fixed at 25 days, we varied it. The risk periods we looked at were: 1 day, 5 days, 10 days, 50 days, 100 days, and 200 days. We shall classify the risk periods as; ‘short risk period’ for 1 day and 5 days risk periods, ‘typical risk period’ for 10 days and 50 days, and ‘long risk period’ for 100 and 200 days risk periods. The interest is in observing the effect of the risk period on the estimates obtained by the model. Further, to reduce the output, we have

restricted our simulations to 20, 100, and 500 cases. Tables 3.3 to 3.8 shows the results. In order to reduce clutter, we only present Table 3.3 and 3.4 here; Tables 3.5 to 3.8 are presented in APPENDIX 1. Table 3.3 shows results for a 1 day risk period, Table 3.4 for a 5 days risk period and so on in increasing order up to Table 3.8 which shows results for a 200 days risk period. We summarise the results given in Tables 3.3 to 3.8 via the relative bias graph (see Figures 3.7 (a), (b), (c)) for true relative incidence of 0.5, 1, 1.5, 2, 5, 10. We can see from Tables 3.3 to 3.8 that the coverage probabilities for the 90%, 95%, and 99% confidence intervals are generally conservative, and become closer to their nominal values as the risk period increases in the range considered.

RI	Log(RI)	10 000 samples of 20 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 100 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 500 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*
0.500	**-0.693**	**0.000** −∞ *(2%, 98%, 0%)* *(2%, 98%, 0%)* *(2%, 98%, 0%)*	**0.000** −∞ *(1%, 98%, 1%)* *(2%, 98%, 0%)* *(2%, 98%, 0%)*	**0.000** −∞ *(10%, 90%, 0%)* *(2%, 98%, 0%)* *(2%, 99%, 0%)*
1.000	**0.000**	**0.000** −∞ *(4%, 96%, 0%)* *(4%, 96%, 0%)* *(2%, 98%, 0%)*	**0.000** −∞ *(2%, 98%, 0%)* *(2%, 98%, 0%)* *(2%, 98%, 0%)*	**0.928** **-0.074** *(10%, 90%, 0%)* *(1%, 99%, 0%)* *(2%, 98%, 0%)*
1.500	**0.405**	**0.000** −∞ *(6%, 94%, 0%)* *(6%, 94%, 0%)* *(1%, 99%, 0%)*	0.000 −∞ *(4%, 96%, 0%)* *(4%, 96%, 0%)* *(2%, 98%, 0%)*	**0.961** **-0.040** *(8%, 92%, 04%)* *(2%, 98%, 0%)* *(1%, 99%, 0%)*
2.000	**0.693**	**0.000** −∞ *(8%, 92%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	0.000 −∞ *(7%, 93%, 0%)* *(3%, 96%, 1%)* *(1%, 99%, 0%)*	**1.864** **0.623** *(7%, 93%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*
5.000	**1.609**	**0.000** −∞ *(5%, 95%, 0%)* *(2%, 98%, 0%)* *(2%, 98%, 0%)*	**4.636** **1.534** *(9%, 91%, 0%)* *(3%, 97%, 0%)* *(2%, 98%, 0%)*	**4.732** **1.554** *(5%, 93%, 2%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*
10.000	**2.303**	**0.000** −∞ *(7%, 93%, 0%)* *(2%, 96%, 2%)* *(1%, 99%, 0%)*	**9.397** **2.240** *(6%, 94%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**9.669** **2.269** *(6%, 90%, 4%)* *(3%, 96%, 1%)* *(1%, 99%, 0%)*

Table 3.3 Simulation results for 1 day risk period.

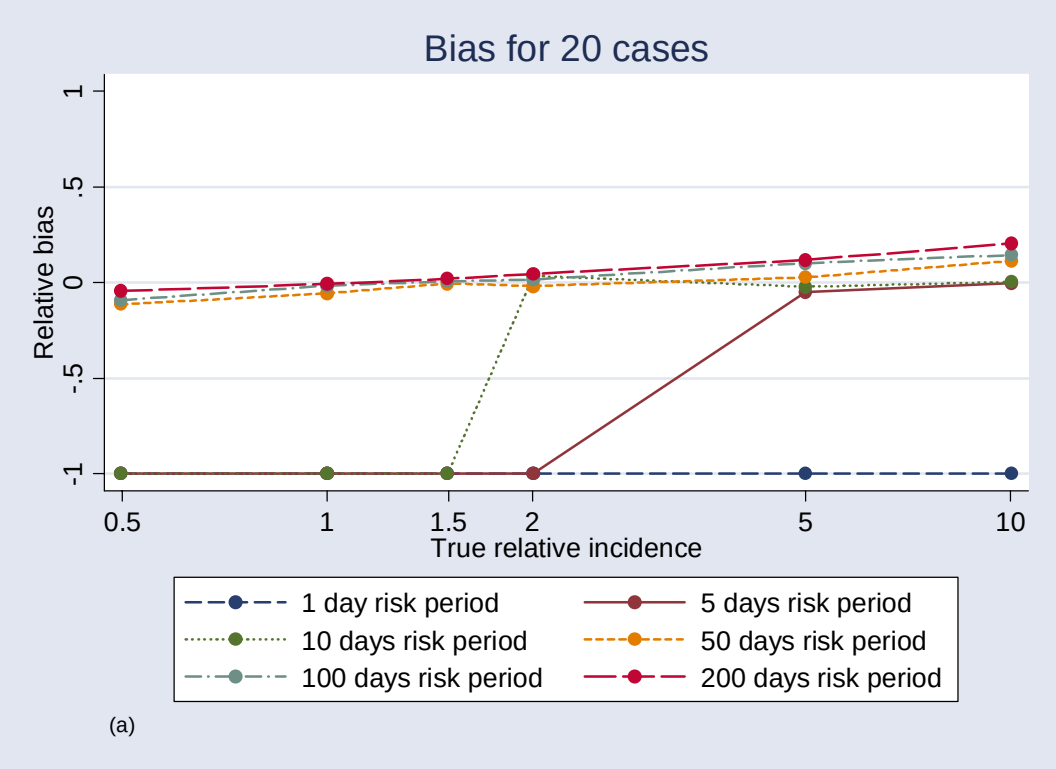
(see Table 3.1 for details)

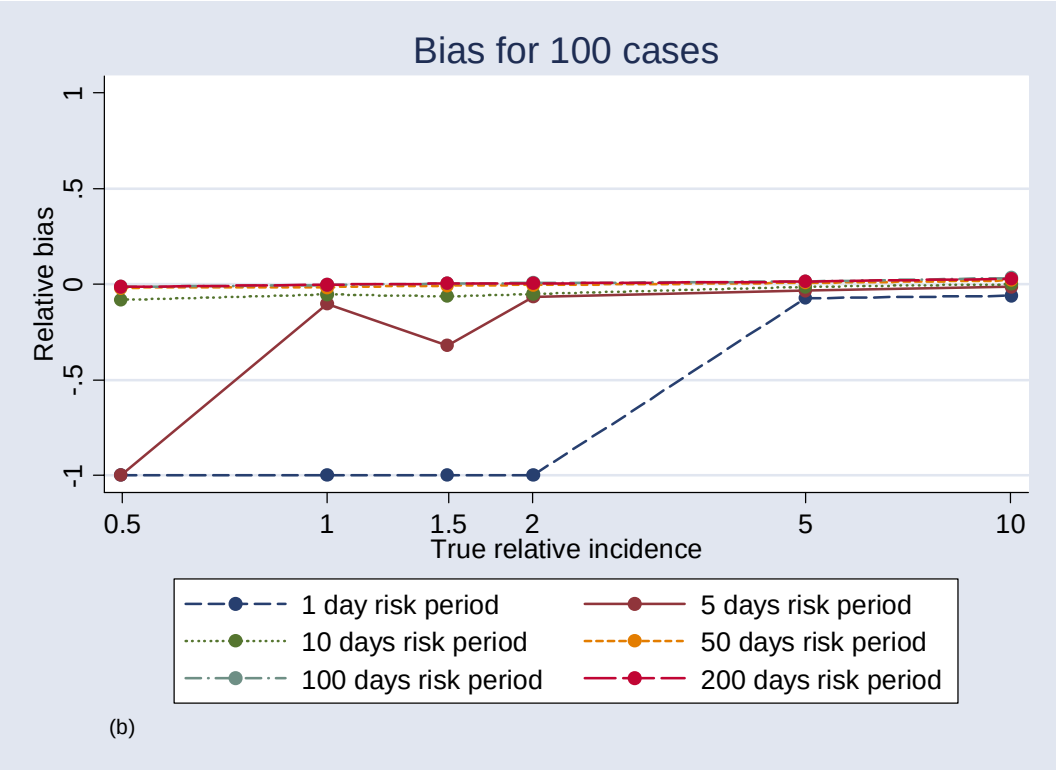
RI	Log(RI)	10 000 samples of 20 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 100 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 500 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*
0.500	**-0.693**	**0.000 −∞** *(10%, 90%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.000 −∞** *(10%, 90%, 0%)* *(2%, 98%, 0%)* *(2%, 98%, 0%)*	**0.531 -0.634** *(6%, 94%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*
1.000	**0.000**	**0.000 −∞** *(5%, 95%, 0%)* *(2%, 98%, 0%)* *(2%, 98%, 0%)*	**0.898 -0.108** *(8%, 92%, 0%)* *(3%, 97%, 0%)* *(2%, 98%, 0%)*	**0.943 -0.058** *(5%, 93%, 2%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*
1.500	**0.405**	**0.000 −∞** *(4%, 96%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**1.018 0.018** *(7%, 93%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**1.477 0.390** *(6%, 90%, 4%)* *(3%, 96%, 1%)* *(1%, 99%, 0%)*
2.000	**0.693**	**0.000 −∞** *(6%, 94%, 0%)* *(6%, 94%, 0%)* *(1%, 99%, 0%)*	**1.866 0.625** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**2.000 0.646** *(5%, 91%, 4%)* *(2%, 96%, 2%)* *(1%, 99%, 0%)*
5.000	**1.609**	**4.745 1.557** *(7%, 93%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**4.840 1.577** *(6%, 92%, 2%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**4.986 1.607** *(6%, 91%, 3%)* *(3%, 95%, 2%)* *(0.6%, 99%, 0.4%)*
10.000	**2.303**	**9.969 2.299** *(5%, 95%, 0%)* *(3%, 97%, 2%)* *(1%, 99%, 0%)*	**9.882 2.291** *(6%, 90%, 4%)* *(3%, 96%, 1%)* *(1%, 99%, 0%)*	**9.963 2.299** *(5%, 91%, 4%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*

Table 3.4 Simulation results for 5 days risk period.

(see Table 3.1 for details)

The graphs in Figures 3.7 (a), (b) and (c) show that, for each sample size, the relative bias was greatest for 1 day risk period, particularly for low relative incidences. For a given relative incidence, the bias is negative for short risk periods and positive for long risk periods. The sign of the bias essentially depends on the number of cases in the risk and control periods. As the number of cases increases, the relative bias decreases and is close to zero even with 1 day risk period.





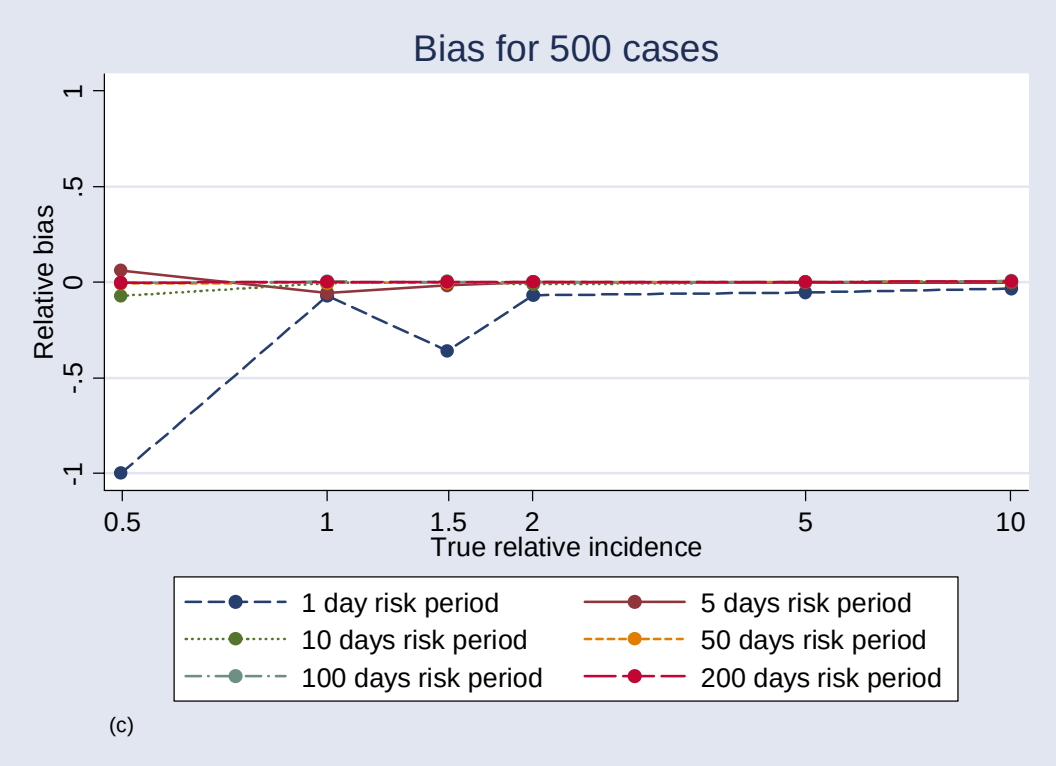


Figure 3.7 Relative median bias for 10 000 samples of 20, 100, and 500 cases at true relative incidences of 0.5, 1, 1.5, 2, 5, 10

3.3.3 Varying the age effect

In this section, we explore the effects of age for the strong symmetric, weak monotone, and strong monotone increasing age effects (see Figure 3.2 for the different age groups and their corresponding age specific relative incidences). We use sample sizes 20, 100 and 500 cases as in section 3.3.2, and use risk periods of 10 days, 25 days, and 50 days with relative incidences of 1, 2 and 5. The distribution of age at exposure is the standard one with mean 250 days and standard deviation 100 days.

Tables 3.9 to 3.11 in APPENDIX 1 shows the results. Tables 3.9 corresponds to the strong symmetric age effect, Table 3.10 to the weak monotone increasing age effect, and Table 3.11 to the strong monotone increasing age effect. The results in these tables are summarised in Figures 3.8 to 3.10. We can see that the relative bias is largest for 20 cases, being the smallest number of cases considered. As for the findings in section 3.3.2, we note that it is for the shorter risk period (10 days risk period in blue/dash) that the relative bias is most prominent. In contrast, the age structure, (strong monotone, strong symmetric, or weak monotone) has little influence on the relative bias. It appears that the performance of the model is mainly influenced by the number of cases and the length of the risk period. Likewise, Tables 3.9 to 3.11 show that the coverage probabilities of the 90%, 95%, and 99% confidence intervals are not substantially affected by the age structure.

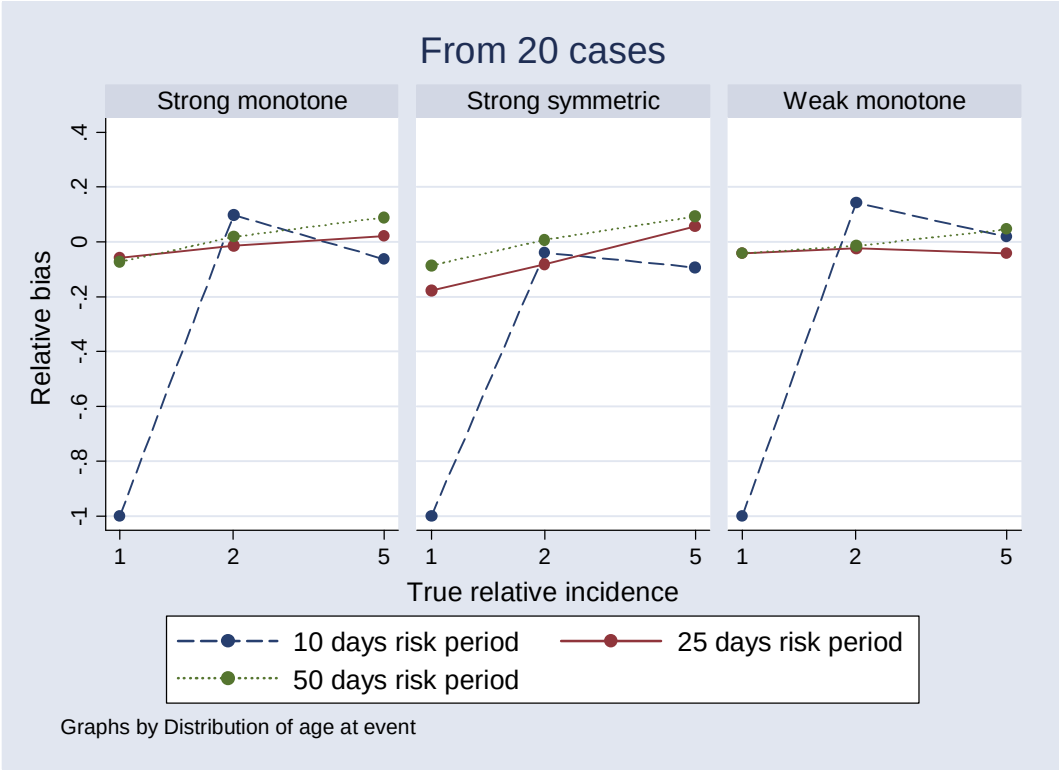


Figure 3.8 Relative median bias for 10 000 samples of 20 cases at true relative incidences of 1, 2, 5 with age effects for strong monotone, strong symmetric and weak monotone. The risk periods are 10 days, 25 days and 50 days.

Figure 3.9 Relative median bias for 10 000 samples of 100 cases at true relative incidences of 1, 2, 5 with age effects for strong monotone, strong symmetric and weak monotone. The risk periods are 10 days, 25 days and 50 days.

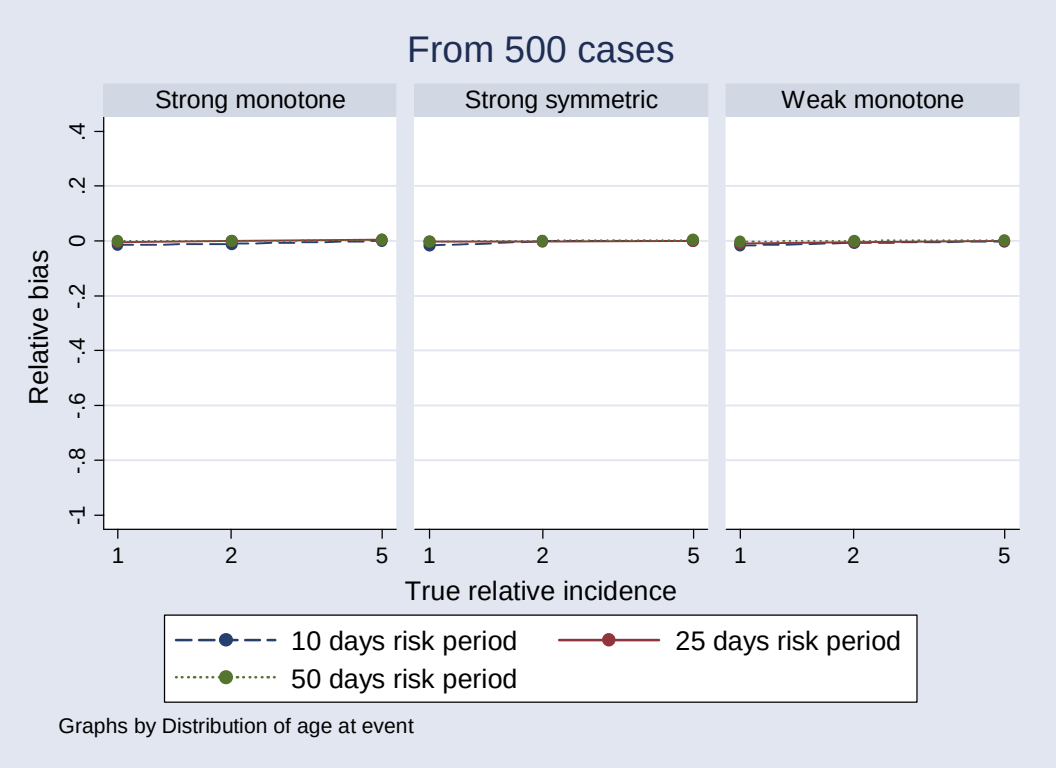


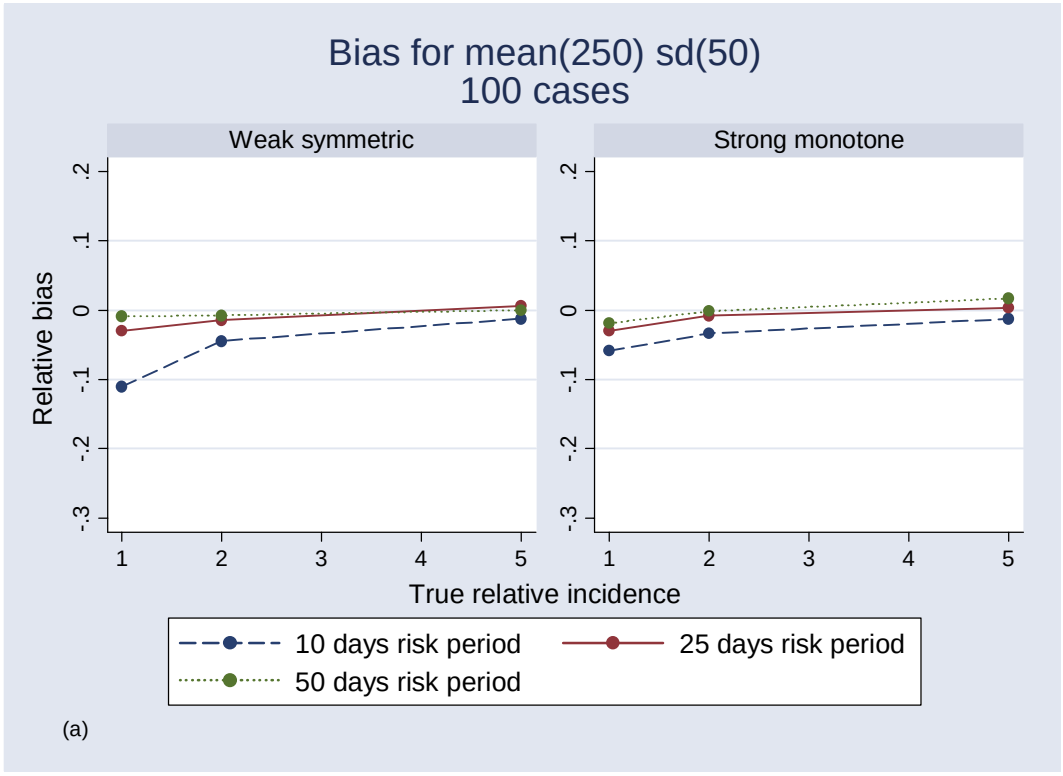
Figure 3.10 Relative median bias for 10 000 samples of 500 cases at true relative incidences of 1, 2, 5 with age effects for strong monotone, strong symmetric and weak monotone. The risk periods are 10 days, 25 days and 50 days.

3.3.4 Varying the age at exposure.

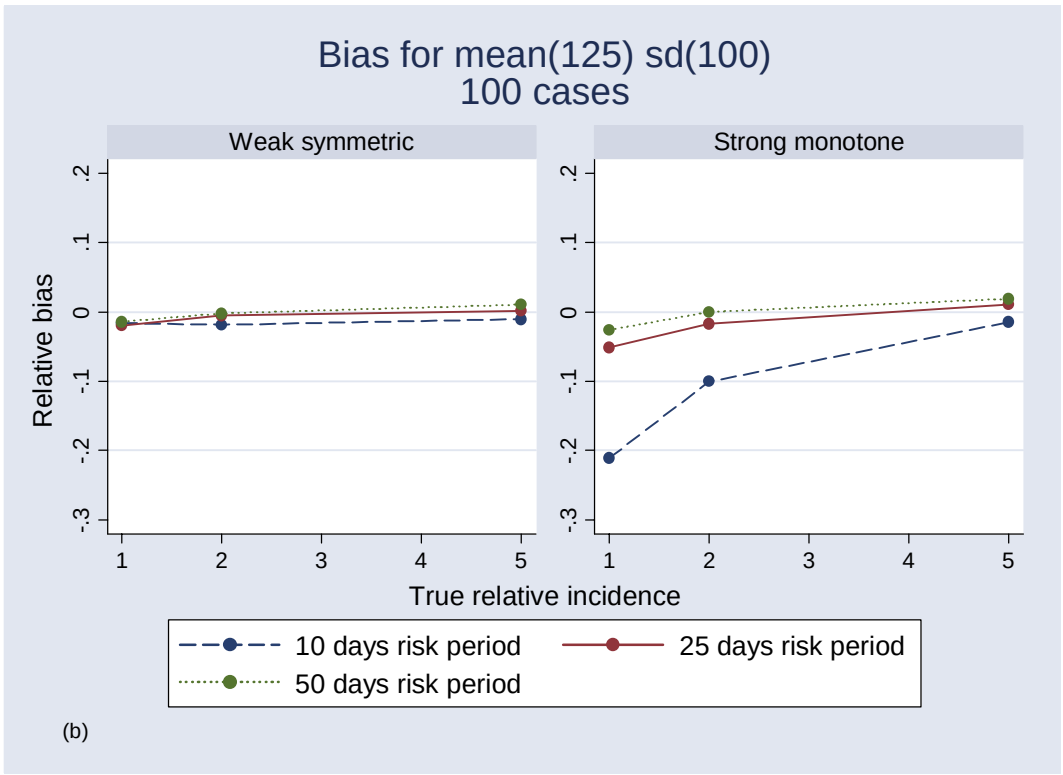
In the last few sections, we varied the relative incidence, risk periods and the age dependence but in all scenarios, the distribution of age at exposure was kept fixed to a symmetrical beta distribution with mean 250 days and standard deviation 100 days (see Figure 3.3). Here we explore the performance of the model when we vary the distribution of age at exposure. We present results for the simulations of 100 cases only, and restrict attention to relative incidences of 1, 2, and 5 and risk periods of 10 days, 25 days and 50 days as in section 3.3.3. We present results for weak symmetric

age effects (the standard scenario) and strong monotone increasing age effects. We investigate distributions of age at exposure with mean 250 days and standard deviation of 50 days, mean age of 125 days and standard deviation of 100 days, and mean age 125 and standard deviation 50 days, as shown in Figure 3.3. The results are given in Tables 3.12 and 3.13 (APPENDIX 1). The graphs in Figures 3.11, (a), (b) and (c) below show the relative bias for different age distribution at exposure and different age effects.

Figure 3.11 (a) shows that a symmetric distribution of age at exposure that is more peaked than that of standard scenario (standard deviation 50 days and mean 250 days) has little effect on the results. However Figures 3.11 (b) and (c) reveal some differences when the distribution of age at exposure is skewed. In this situation, the bias is greater when the age effect is strong monotone than when it is weak symmetric. When the age effect is strong monotone, events are most likely to occur at older ages, whereas most risk periods span younger ages. Thus the imbalance between expected numbers of events in risk period and control periods is greater and this leads to greater bias. This effect is most pronounced for short risk periods. A further reason for the bias may be confounding between age and exposure effects.


Bias for mean(250) sd(50)
100 cases
Weak symmetric
Strong monotone
Relative bias
-.3
-.2
-.1
0
.1
.2
1
2
3
4
5
True relative incidence
10 days risk period
25 days risk period
50 days risk period

(a)


Bias for mean(125) sd(100)
100 cases
Weak symmetric
Strong monotone
Relative bias
-.3
-.2
-.1
0
.1
.2
1
2
3
4
5
True relative incidence
10 days risk period
25 days risk period
50 days risk period

(b)

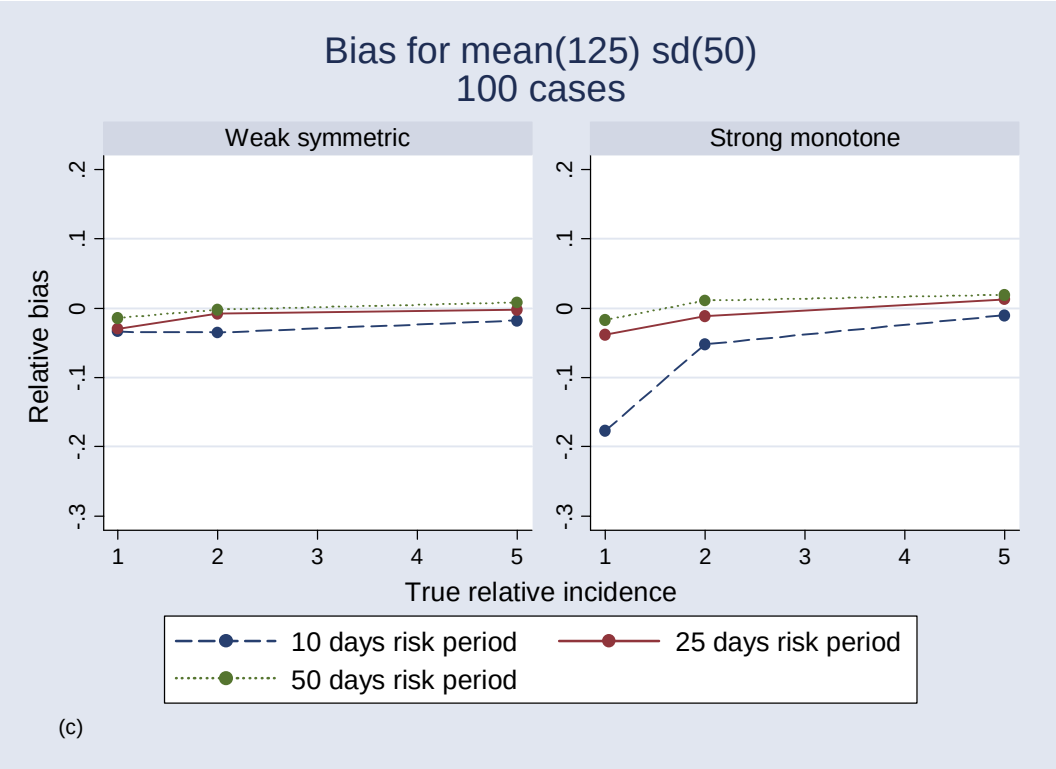


Figure 3.11 Relative bias against relative incidence, for risk periods 10, 25, 50 days and two age effects, when the mean age at exposure is 250 days, standard deviation 50 days (a), mean age at exposure is 125 days, standard deviation 100 days (b), and mean age at exposure is 125 days and standard deviation 50 days (c).

3.3.5 Indefinite risk periods

In all the scenarios explored so far, the risk periods were of pre-determined length. The self-controlled case series method can be used even when the risk period following an exposure is indefinite. However, the effects of exposure and age may then be substantially confounded. The confounding can be controlled by including unexposed cases, which contribute exclusively to the estimates of the age effect. This was explored in the simulations described below.

In this section, we present results using indefinite post-exposure risk periods when 100% of cases are exposed, five sixth of the cases are exposed, two thirds of the cases are exposed and half of the cases are exposed. We first used two exposure

distributions (for those exposed): mean 250 days with 100 days standard deviation and with mean 125 days with 50 days standard deviation. We used weak symmetric and strong monotone increasing age effect with true relative incidences of 1, 2, and 5. All runs include 100 cases exposed. The results from the simulations are presented in Tables 3.14 to 3.17 (APPENDIX 1). The coverage probabilities are very close to the nominal values. Figure 3.12 summarises the relative bias for all the simulations presented in Tables 3.14 to 3.17.

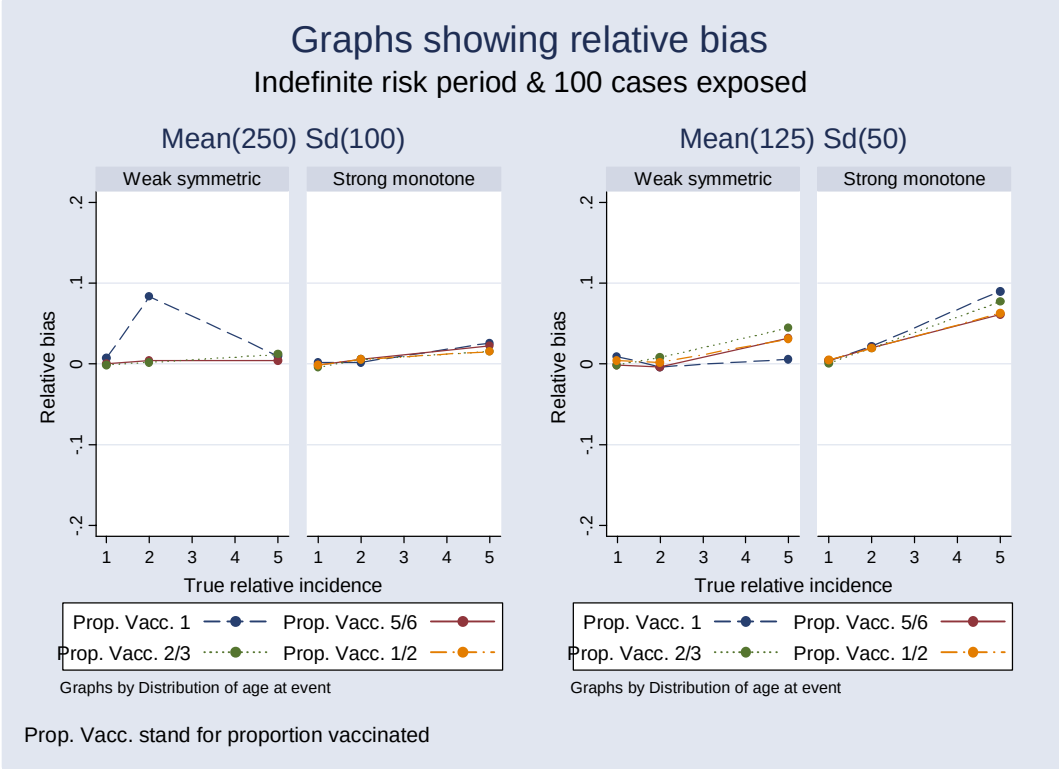


Figure 3.12 Relative median bias for 10 000 samples of different exposed proportions such that 100 cases were exposed for the true relative incidences of 1, 2, 5 with age effects for strong monotone and weak symmetric age groups.

Figure 3.12 shows that the relative bias is small for relative incidences of 1 and 2, whatever the age at exposure and age at event distributions. Note that there was one exception to this for the relative incidence of 2 with the weak symmetric age effect. The outlier could be due to random variation.

However, for relative incidence of 5 the relative bias is greater for strong monotone increasing age effects than for weak symmetric age effects and greater when the age at exposure distribution is skewed than when it is symmetric. This may be explained by confounding of age and exposure effects when the risk period is indefinite.

The presence of unexposed cases reduces the bias to some degree, though perhaps less than anticipated. To explore these effects further, more peaked and asymmetric distributions of age at exposure were investigated. In these additional simulations, the relative incidences were 1, 2, and 5 as before, and we used the strong monotone increasing age effect. The distributions of age at exposure had mean 125 days and standard deviations of 10, 20, 30, and 40 days respectively (see Figure 3.4). All simulations were done with indefinite risk periods, and the same proportions exposed as those used earlier.

The results are in Tables 3.18 to 3.20 of APPENDIX 1 and are summarised in Figure 3.13 below. As expected, the bias tends to increase as the standard deviation of the age at exposure decreases. The presence of a small proportion of unvaccinated cases greatly reduces the bias: the bias with $\frac{5}{6}$ cases vaccinated is much the same as with $\frac{1}{2}$ cases vaccinated.

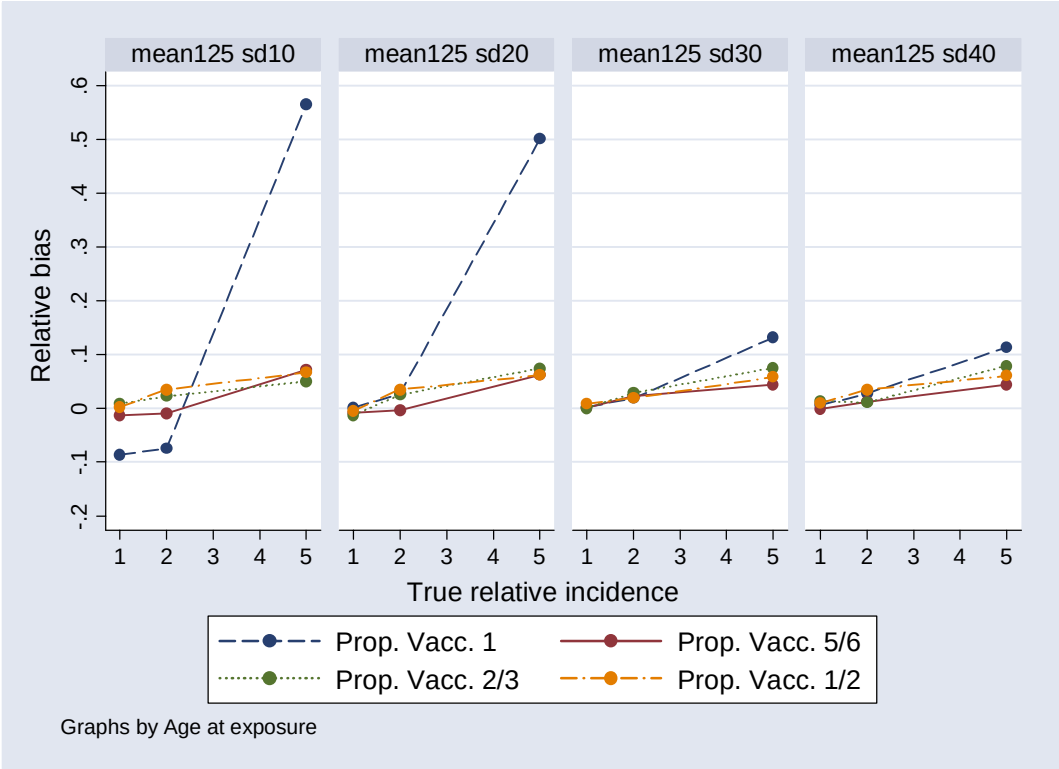


Figure 3.13 Relative bias for strong monotone increasing age effect.

3.4 Conclusion

In this last section of the chapter, we bring together the main findings. We have investigated a very broad range of scenarios, based on variations of a 'standard scenario' which is a representative of many studies of paediatric vaccines. In the standard scenario we found that the estimates were substantially biased for sample sizes of 20 or less, when the true relative incidence was ≤ 1. However for relative incidence ≥ 1.5 the biases were moderate even with sample sizes of 10 cases, and very small when the number of cases was ≥ 50.

In section 3.3.2, risk periods as short as 1 day and up to a maximum of 200 days were investigated. In these situations, the estimates were biased for short risk periods. For

example when the risk period was 1 day, the bias was large when the relative incidence was 0.5 even with 500 cases. Similarly the bias was large for relative incidence of 2 and 100 cases. Generally speaking, the longer the risk period in the range considered (up to 200 days), the less biased the estimates were.

Different age effects classified as weak symmetric, strong symmetric, weak monotone increasing, strong monotone increasing were explored in section 3.3.3. There was little evidence that these age effects affected the performance of the self-controlled case series model with fixed risk periods. The effect of different distributions of age at exposure was explored in section 3.3.4. As with the age at event, the distribution of age at exposure did not have much bearing on the results for fixed risk periods.

In section 3.3.5 we looked at the effect of indefinite risk periods. Some researchers [51, 52] have argued that the self-controlled case series model may not be effective if one is looking at a situation were adverse events may manifest themselves a long time after exposure. We explored this issue by extending the risk periods to indefinite length. Results showed that overall there was little bias except for large relative incidences and distributions of age at event and age at exposure that induce confounding between exposure and age effects. This confounding and the bias it generates can be controlled by including unvaccinated cases. Some bias remains, but it is not large. Including 20% unexposed cases appears sufficient to reduce the bias to acceptable levels.

In all situations explored, the coverage probabilities from ten thousand samples of different number of cases were in excess of their nominal values, even in the presence of substantial bias.

In chapter 2 we found that when there are no age effects, the magnitude of the asymptotic bias depends largely on the balance of expected numbers of events in the risk and control periods: when the expected number of events in the risk period is less than that in the control period, the bias is negative, and vice versa. When the two expectations are equal, the bias is zero.

In this chapter we have explored more complex situations and finite samples by simulation. Qualitatively similar results emerge: for a given sample size, the bias is greatest in magnitude when the expected number of events in the risk period is much smaller than the expected number in the control period. In practice, bias is only a real problem when the risk period is very short or relative incidence is low. In other circumstances, sample sizes in excess of 20 appear to give reliable results.

Another point to mention is the coverage probabilities. These are generally reasonably accurate even in the presence of extreme bias: this is not surprising, since when the expected number of event in the risk period is very small, the variance of $\hat{\beta} = \log(\hat{\rho})$ (where $\rho = e^{\beta}$ is the relative incidence) is very large, as may be seen from the asymptotic calculations of chapter 2. Hence the confidence intervals will themselves be very wide. Confidence intervals based on profile likelihood methods may perhaps be better in such circumstances.

The overall conclusion is that estimates and confidence intervals based on asymptotic theory are reliable except in extreme scenarios (namely very small sample size, very short risk period, low relative incidence).

Chapter 4

Sample size formulae for the self-controlled case series method; first attempts

4.1 Introduction

When designing a study, one of the most important questions to address is the required sample size. In this chapter we propose various sample size formulae for designing a study that will use the self-controlled case series method. In [3] a sample size formula for use in case series studies was derived based on a normal approximation to the distribution of the estimated relative incidence. However, the performance of this sample size formula has not been evaluated. Furthermore it is only valid when there are no age effects. In this chapter the accuracy of this sample size formula is assessed by simulation. Other sample size formulae are explored and results from simulation studies to evaluate their performance are given. All these formulae assume a simplified situation without age effect. In chapter 5, a formula incorporating age effects will be presented.

In section 4.2, we briefly present some background and the notation to be used in this chapter. In section 4.3, we derive four sample size formulae based on different asymptotic arguments. These formulae are derived under the assumption that there are no age effects, and are evaluated in section 4.4. We conclude with a brief discussion in section 4.5.

4.2 Background and notation

In this chapter we will be concerned only with situations where the underlying (or baseline) incidence of an event is constant, that is, does not vary with age (or time, if time is the relevant time line). At each time point, an individual is categorized as exposed or unexposed. Typically, the times at which an individual is considered to be exposed occur within a defined time interval following a point exposure, for example receipt of a vaccine. The period of exposure is called the risk period.

We further assume that all individuals are followed up for an observation period of the same length, and that a proportion ν of individuals in the population experience the exposure during this observation period. We will assume, also for simplicity, that all exposed individuals spend the same time exposed. As mentioned before in chapter 2, in practice, the observation and exposure periods vary between individuals, but this variation can reasonably be ignored for the purposes of sample size calculations. If e_1 is the length of the risk period and e_0 is the duration of the control period, then $r = e_1/(e_0 + e_1)$ is the proportion of observation time for which an individual is exposed. Usually, e_1 and $e_0 + e_1$ will be specified in the design. However, only their ratio r is required.

During the risk period, the baseline incidence of an adverse event is increased by a multiplicative factor $\rho = e^{\beta}$, where ρ is the relative incidence. The value of parameter ρ (or β) which may be considered clinically important is the focus of

inference. Under the null hypothesis, $\rho = 1$ whereas under the alternative hypothesis we specify some value for $\rho \neq 1$, the value we wish the study to detect.

A case is an individual who experiences at least one event during the observation period. Suppose that a sample of cases is available, and that a total of n events arise in these cases. Note that n refers to events, not individuals: the case series method allows multiple events per individual, provided these events are independent. Our sample size formulae will generally relate to numbers of events, though in the next chapter we briefly touch upon estimating the number of cases required. If the event of interest is non-recurrent, then the case series method still applies provided the event is rare. Of these n events, suppose that n_e arise in exposed individuals, that is, individuals who were exposed at some time during the observation period. Suppose also that n_u events arise in unexposed individuals, that is, individuals who were not exposed during the observation period. Of the n_e events in exposed individuals, suppose that x arise in a risk period. Then the case series log likelihood for the parameter ρ can be shown (see chapter 2) to be

$$\ell(\rho) = x \log\left(\frac{\rho r}{\rho r + 1 - r}\right) + (n_e - x)\log\left(\frac{1-r}{\rho r + 1 - r}\right). \qquad (4.1)$$

Note that (4.1) is equivalent to a binomial likelihood with binomial proportion $\pi = \rho r / (\rho r + 1 - r)$ and index n_e, and that it does not involve n_u: only exposed individuals contribute to the log likelihood when there are no age effects. The likelihood ratio statistic for the test of $H_0 : \rho = 1$ (or equivalently $\beta = 0$) is thus

$$D = 2\{\ell(\hat{\rho}) - \ell(1)\} = 2\{x\log(\hat{\rho}) - n_e \log(\hat{\rho}r + 1 - r)\}. \quad (4.2)$$

where $\hat{\rho}$ is the maximum likelihood estimator of ρ. Finally, note that, in large samples, we have

$$n \approx n_e \frac{1+\nu r(\rho-1)}{\nu(\rho r+1-r)} . \quad (4.3)$$

In particular, if $\rho = 1$ then $n \approx n_e / \nu$ in large samples, and if $\nu = 1$, then $n = n_e$.

We present four sample size formulae based on different asymptotic approximations, assuming there is no age effect. In what follows, the significance level is denoted α, and to avoid confusion with the parameter β (the log relative incidence) we shall denote power to be $1-\gamma$. Thus $Z_{1-\alpha/2}$ is the $(1-(\alpha/2))-$quantile of the standard normal distribution, and Z_γ is its $\gamma-$quantile. For simplicity, the formulae quoted in this chapter are for n_e, the total number of events required in exposed individuals.

4.3 Sample size formulae without age effects

4.3.1 Sample size formula based on the asymptotic sampling distribution of ρ

In this subsection, we describe the sample size formula which was first published by Farrington et al [3]. The idea behind the derivation of the formula is to use $\hat{\rho}$ as the test statistic, and base the sample size formula on its asymptotic normal distribution. The asymptotic variance of $\hat{\rho}$ may be obtained by twice differentiating (-1) times expression (4.1) with respect to ρ, taking expectations, and inverting the result. This gives the expression for the variance as

$$\text{var}(\hat{\rho}) \approx \frac{1}{n_e}\frac{\rho(\rho r+1-r)^2}{r(1-r)}.$$

A general sample size formula for a normally distributed test statistic where one assumes that the variance under the null hypothesis is different from that under the alternative hypothesis is given by Armitage et al [53] as

$$n=\left(\frac{\sigma_0 Z_{1-\alpha/2}+\sigma_1 Z_{\gamma}}{\mu_1-\mu_0}\right)^2. \qquad (4.4)$$

where the test statistic under H_0 is distributed

$$N\left(\mu_0, \frac{\sigma_0^2}{n}\right)$$

and under H_1 is distributed

$$N\left(\mu_1, \frac{\sigma_1^2}{n}\right)$$

Note that $\mu_1-\mu_0$ is the clinically important difference to be detected in the population. Thus under the null hypothesis,

$$\hat{\rho} \approx N(1, 1/n_e r(1-r))$$

and under the alternative,

$$\hat{\rho} \approx N\left(\rho, \rho(\rho r+1-r)^2/n_e r(1-r)\right).$$

Replacing the parameters from the two approximate distributions of $\hat{\rho}$ under the null and alternative hypothesis in the general expression (4.4) leads to the following sample size formula:

$$n_e = \frac{1}{r(1-r)(\rho-1)^2}\times\left[Z_{1-\alpha/2}+Z_{\gamma}(\rho r+1-r)\sqrt{\rho}\right]^2. \qquad (4.5)$$

The above is a special case of the formula given by Farrington et al [3] with everyone exposed. Note that in all sample size formulae given here we round n_e up to the next integer that is greater or equal to the number of events needed.

4.3.2 Sample size formula based on the asymptotic sampling distribution of β

A concern about (4.5) is that the sampling distribution of $\hat{\rho}$ may not be symmetric in small samples. Thus we derived a sample size formula based on the sampling distribution of $\hat{\beta} = \log(\hat{\rho})$ in the hope that this might be less skewed. In chapter two we showed that the asymptotic variance of $\hat{\beta}$ up to the first order in n_e (see formula (2.6)) is

$$\text{var}\left(\hat{\beta}\right) \approx \frac{1}{n_e}\frac{\left(\rho r + 1 - r\right)^2}{\rho r(1-r)}.$$

Under the null hypothesis, $\hat{\beta} \approx N\left(0, 1/n_e r\left(1-r\right)\right)$ whereas under the alternative, $\hat{\beta} \approx N\left(\beta, \left(\rho r + 1 - r\right)^2 / n_e \rho r\left(1-r\right)\right)$. Using (4.4) leads to the following expression for the sample size formula:

$$n_e = \frac{1}{r\left(1-r\right)\log\left(\rho\right)^2} \times \left[Z_{1-\alpha/2} + Z_{\gamma}\left(\rho r + 1 - r\right)/\sqrt{\rho} \right]^2. \qquad (4.6)$$

4.3.3 Sample size formula using second order variance of $\hat{\beta}$

As has been mentioned above, sample size formula (4.6) was obtained in an effort to try and improve on the possible non-symmetric distribution of $\hat{\rho}$ in small samples. Evaluation of formulae (4.5) and (4.6) under simulation showed that the two formulae were not accurate (results shown in section 4.4): the observed power was not as expected in some cases. The results from the simulation were particularly poor when the risk period was short and the clinically important relative incidence to be detected was in the extremes, for example relative incidences of 0.1, or 10. Having identified this problem, we extended the asymptotic variance of the estimate $\hat{\beta}$ up to the second order in n_e, in the hope that the sample size formula derived would better take account of the variation in the estimate of the relative incidence. Below is how the sample size formula using second order asymptotic variance was derived.

Let $p = \dfrac{\rho r}{\rho r + 1 - r}$ be the risk of an event in the risk period

and $q = 1 - p = \dfrac{1-r}{\rho r + 1 - r}$ the risk of the event occurring in the control period. As before r is the ratio of the risk period to the observation period. The second order asymptotic variance was found in chapter 2 as:

$$\text{var}(\hat{\beta}) = \frac{1}{n_e pq} - \frac{(p-q)^2}{2n_e^2 p^2 q^2} + \frac{2\left(p^3 + q^3\right)}{n_e^2 p^2 q^2}$$

Under H_0 : $\mu_0 = \beta = 0$, σ_0^2 □ n_e var($\hat{\beta}$), and $\rho = 1$. Substituting ρ, p, and q in $\text{var}\left(\hat{\beta}\right)$ above and simplifying, we have:

$$\sigma_0^2 = \frac{1}{r(1-r)} - \frac{(2r-1)^2}{2n_e r^2 (1-r)^2} + \frac{2\left(r^3 + (1-r)^3\right)}{n_e r^2 (1-r)^2}.$$

Under $H_1 : \mu_1 = \beta, \quad \sigma_1^2 = n_e \text{ var}(\hat{\beta}), \quad \therefore \ \sigma_1^2 = \frac{1}{pq} - \frac{(p-q)^2}{2n_e p^2 q^2} + \frac{2\left(p^3+q^3\right)}{n_e p^2 q^2}$

Substituting σ_0 and σ_1 in (4.4) and simplifying in terms of r, and ρ, we get the following sample size formula:

$$n_e = \frac{\left[Z_{1-\alpha/2}\sqrt{A} + Z_\gamma \rho^{-1}\sqrt{B}\right]^2}{(r(1-r))^2\left(\log(\rho)\right)^2} \tag{4.7}$$

where

$$A = r(1-r) - \frac{(2r-1)^2}{2n_e} + \frac{2(r^3+(1-r)^3)}{n_e}$$

$$B = \rho r(1-r)(\rho r+1-r)^2 - \frac{(\rho r-1+r)^2(\rho r+1-r)^2}{2n_e} + \frac{2\left((\rho r)^3+(1-r)^3\right)(\rho r+1-r)}{n_e}$$

Note that the above formula is implicit since n_e occurs both on the left and right hand sides of (4.7). We obtained n_e by an iterative search from a particular starting point. We used the lowest value of n_e obtained from sample size formulae (4.5) and (4.6) as the starting values.

4.3.4 Sample size formula based on the binomial proportion

In the simplified setting as described earlier, events in exposed individuals occur either in the risk period or in the control period. Hence the events can be considered to follow a binomial distribution with proportion $\pi = \rho r / \left(\rho r + 1 - r\right)$; therefore, we can use binomial proportions to derive a sample size formula for the self-controlled case series method. The binomial probabilities under H_0 and H_1 are:

$\pi_0 =$ P(event in risk period $\big| H_0) = r$

$$\pi_1 = \text{ P(event in risk period} \mid H_1) = \frac{r\rho}{r\rho + 1 - r} = p$$

Let x denote the number of events in the risk period,

thus $x \sim \text{Bin}(n_e, \pi_0)$ under H_0 and $x \sim \text{Bin}(n_e, \pi_1)$ under H_1.

By the normal approximation to the binomial (Fleiss et al [54]), under

H_0, $\frac{x - n_e\pi_0}{\sqrt{n_e\pi_0(1-\pi_0)}} \approx N(0,1)$ and under H_1, $\frac{x - n_e\pi_1}{\sqrt{n_e\pi_1(1-\pi_1)}} \approx N(0,1)$. The sample size formula based on the normal approximation to the binomial distribution (Fleiss et al [54]) is given by:

$$n_e = \left[\frac{Z_{1-\alpha/2}\sqrt{\pi_0(1-\pi_0)} + Z_\gamma\sqrt{\pi_1(1-\pi_1)}}{\pi_1 - \pi_0} \right]^2.$$

Substituting the values of π_0 and π_1 we get:

$$n_e = \left[\frac{Z_{1-\alpha/2}\sqrt{r(1-r)} + Z_\gamma\sqrt{\frac{r\rho(1-r)}{(r\rho+1-r)^2}}}{\frac{r(\rho-1)(1-r)}{r\rho+1-r}} \right]^2$$

$$= \left[\frac{(r\rho+1-r)Z_{1-\alpha/2}\sqrt{r(1-r)} + Z_\gamma\sqrt{r\rho(1-r)}}{r(\rho-1)(1-r)} \right]^2. \qquad (4.8)$$

In the next section we present a comparative evaluation by simulation of sample size formulae (4.5), (4.6), (4.7) and (4.8).

4.4 Comparative evaluation of sample size formulae (4.5), (4.6), (4.7) and (4.8)

So far, we have derived four sample size formulae for the self-controlled case series method. We now present results from a simulation study to evaluate them.

4.4.1 Simulation study

The simulation study was carried out as follows. First we specified the observation period (500 days), the relative incidence (ρ), the value r , that is the ratio of the risk period to the observation period, the power (set at either 80% or 90%), the significance level (set at 5%) and we assumed everybody was exposed. The values of r , the ratio of the risk period to the observation period were 0.01, 0.05, 0.1, and 0.5 (corresponding to 5, 25, 50, and 250 days, respectively). The relative incidences (ρ) to be detected at the two sided 5% significance level with 80% or 90% power were 0.1, 0.5, 1.2, 1.5, 2, 3, 5, 8 and 10.

After calculating the number of cases required using the specified parameters, we rounded the sample size n_e up to the next integer. We then generated 2000 random samples of n_e cases with a single event per case. Each of the n_e cases was obtained using a 500-day observation period, including a risk period of duration $500\times r$ days. Thus all cases were assumed to be exposed. The single event for each case was randomly allocated to the risk and control period based on the true value of ρ . Then ρ was estimated using the self-controlled case series method for each sample of n_e events. The observed power was found by calculating the overall proportion of

the 2000 samples for which the likelihood ratio test rejected the null hypothesis at the 5% significance level. The Monte Carlo standard error (MCSE) for the empirical power is found by

$$\text{MCSE} = \sqrt{\frac{\text{power} \times (1 - \text{power})}{\text{number of simulations}}}.$$

It is about 0.89 per cent at 80% power and 0.67 per cent at 90% power.

4.4.2 Results

Tables 4.1 and 4.2 below show the results obtained using the four sample size formulae. The combination of four values of r, nine values of ρ, two powers, and four sample size formulae thus required 288 different simulations of 2000 samples. The different sample sizes calculated using the different formulae are given in columns headed $N5, N6, N7, N8$ corresponding to sample size formulae numbered $(4.5), (4.6), (4.7), (4.8)$ respectively. The corresponding observed power (in %) under simulation for each sample size is given in columns headed $P5$, $P6$, $P7$, $P8$.

The sample sizes produced by the four formulae are generally of a similar order of magnitude, with some very noticeable exceptions, particularly with $r = 0.01$ and $\rho \geq 5$. For $r = 0.01, \rho = 10$ and power 0.8, the sample sizes ranged from 29 ($N8$) to 170 ($N7$), a greater than 5-fold difference. This variation indicates that some of these formulae, at least, must be inaccurate over this range.

It is clear that all the four formulae $N5, N6, N7$ and $N8$ are inaccurate for many parameter combinations, especially for extreme values of ρ. This could be most likely

due to skewness of the sampling distribution of $\hat{\rho}$ and $\hat{\beta}$ especially for sample size formulae $N5$ and $N6$. Formulae $N5$ and $N8$ tends to overestimate the sample size required (as seen by the empirical power) for relative incidence less than one and underestimate the sample size required for relative incidence greater than eight but the underestimation is not present for relative incidence greater than eight when the ratio of the risk period to the observation period was 0.5 for 80% power and when $r = 0.1$ and 0.5 for 90% power (see $P5$ and $P8$ in Tables 4.1 and 4.2). On the other hand, the power observed from formulae $N6$ and $N7$ does not seem to show a clear pattern; for some values of the relative incidence, the observed power is greater than the nominal value (see $P6$ and P7) and for others it is lower. The power observed from all four sample size formulae derived so far seem to be accurate when trying to detect relative incidences of $RI = 1.2$ and 1.5.

Table 4.1 Empirical power for 80 per cent nominal value

r	RI	$N5$	$N6$	$N7$	$N8$	$P5$ %	$P6$ %	$P7$ %	$P8$ %
0.01	0.10	617	403	819	609	96	67	98	88
	0.50	2632	2079	2265	2618	81*	80	81*	83
	1.20	21000	22644	22874	21031	80	80	80	79
	1.50	3627	4327	4448	3638	78	84	84	80
	2.00	1010	1379	1499	1015	76	87	88	78
	3.00	301	505	610	302	78	88	95	79
	5.00	97	216	306	98	80	95	97	80
	8.00	42	122	199	42	62	98	99	62
	10.0	30	97	170	29	79	98	99	79
0.05	0.10	128	81	161	119	96	67	99	89
	0.50	544	427	462	529	84	80	78	86
	1.20	4400	4741	4767	4431	80	80	80	80
	1.50	767	910	891	779	77	84	82	80
	2.00	217	293	274	223	76*	85	83	72*
	3.00	67	109	86	69	70	88	84	74
	5.00	24	48	56	24	80	95	96	80
	8.00	12	28	43	11	75	95	99	69
	10.0	9	22	36	8	63	98	99	56
0.1	0.10	67	41	79	58	96	62	98	85
	0.50	284	221	210	269	86	73*	80*	83
	1.20	2337	2517	2508	2371	78	81	80	80
	1.50	412	486	479	425	75	82	81	76
	2.00	119	159	153	125	75	85	81	75
	3.00	39	60	58	41	64	84	81	75
	5.00	15	27	15	15	69	89	69	69
	8.00	9	16	11	8	69*	92	80	81*
	10.0	7	13	9	6	79	96	79	69
0.5	0.10	22	9	14	9	99	80	97	80
	0.50	93	68	69	69	92	77	81	81
	1.20	886	947	948	948	77	80	80	80
	1.50	169	194	194	194	73	80	80	80
	2.00	57	68	69	69	77*	76*	81	81
	3.00	24	29	29	29	77	81	81	81
	5.00	15	16	14	16	76*	88	82*	88
	8.00	14	11	9	11	93	87	75	87
	10.0	14	9	11	9	97	81	93	81

*r = ratio of the risk period to the observation period, $v = 1$, the proportion vaccinated, RI = relative incidence to be detected, $N5$ = sample size (using formula (4.5)), $N6$ = sample size (using formula (4.6)), $N7$ = sample size(using formula (4.7)), $N8$ = sample size (using formula (4.8)), $P5$ = observed power for $N5$, $P6$ = observed power for $N6$, $P7$ = observed power for $N7$, $P8$ = observed power for $N8$. *Saw-tooth phenomenon (see text).*

Table 4.2 Empirical power for 90 per cent nominal value

r	RI	$N5$	$N6$	$N7$	$N8$	$P5$ %	$P6$ %	$P7$ %	$P8$ %
0.01	0.10	696	681	762	688	97	97*	99	95*
	0.50	3309	2978	3082	3297	91	87	90	90
	1.20	28622	2981	29950	28641	90	90	90	90
	1.50	5056	5573	5700	5062	90	91	90	87
	2.00	1452	1739	1855	1453	88	92	93	89
	3.00	451	617	720	450	86	94	97	84
	5.00	155	255	343	152	87	96	99	86
	8.00	71	140	217	68	91	97	99	89
	10.0	51	110	183	48	85	99*	98*	82
0.05	0.10	144	135	241	135	95*	97*	99	97
	0.50	681	609	648	669	92	90	93*	90*
	1.20	6006	6251	6277	6026	90	90	90	90
	1.50	1073	1178	1203	1079	88	90	92	89
	2.00	315	372	394	316	89	92	94	91
	3.00	102	135	154	101	84*	94	96	90*
	5.00	39	57	74	36	86*	93	99	90*
	8.00	21	33	47	17	90	98	99	81
	10.0	16	26	40	13	86*	99	99	87*
0.1	0.10	74	67	117	66	94	96*	99	97*
	0.50	354	314	332	341	92	90	92*	90*
	1.20	3196	3324	3336	3217	90	90	90	90
	1.50	579	633	644	585	89	91	93	88
	2.00	174	203	213	175	89	90	93	89
	3.00	60	75	84	59	85*	91	95	90*
	5.00	25	33	41	22	93	95	98	84
	8.00	15	20	27	11	91	96	99	84
	10.0	13	16	23	9	97	97	100	79
0.5	0.10	24	14	19	11	100	97	99	93
	0.50	112	92	94	91	95	91*	90*	91
	1.20	1228	1269	1271	1268	89	90	90	90
	1.50	247	260	262	259	90	89	90	89
	2.00	88	92	94	91	87	90	90	90
	3.00	41	40	42	38	95	90*	92	92*
	5.00	28	21	25	20	96	95*	95*	89
	8.00	28	15	20	13	100	91*	98	95*
	10.0	30	14	19	11	100	97	99	92

*r = ratio of the risk period to the observation period, $v = 1$, the proportion vaccinated, RI = relative incidence to be detected, $N5$ = sample size (using formula (4.5)), $N6$ = sample size (using formula (4.6)), $N7$ = sample size(using formula (4.7)), $N8$ = sample size (using formula (4.8)), $P5$ = observed power for $N5$, $P6$ = observed power for $N6$, $P7$ = observed power for $N7$, $P8$ = observed power for $N8$. *Saw-tooth phenomenon (see text).*

4.4.3 Saw-tooth phenomenon

The discreteness of the data induces a phenomenon known as saw-toothing. For example we can see from Table 4.1 that to detect a relative incidence of 2, at 80% power, when the ratio of the risk period to the observation period is 0.05, with all cases exposed (vaccinated), sample size formula $N5$ gives approximately 217 as the number of cases needed. The observed power by simulation was 76%, which is slightly less than the nominal 80% power. For the same parameters, using sample size formula $N8$ gives 223 as the approximate number of cases needed and the observed power was 72%. Thus the power is not a monotone increasing function of sample size. This was observed in several situations (marked by * in Tables 4.1 and 4.2) and with all the other formulae derived. This phenomenon has been observed by other researchers for example Chernic et al [55], Cesana et al [56] , Brown et al [57] and Hoehler [58].

The saw-tooth phenomenon means that the power function has the characteristic that it decreases slowly and then jumps up and then cyclically repeats the decreasing trend followed by an upward jump. The jump always occurs at a higher level of power than in the previous cycle. One consequence of this is that there may be no unique sample size.

Chernic et al [55] demonstrated that for continuous random variables, for a given significance level and alternative hypothesis, the power function increases monotonically as sample size increases, but that this is not the case for discrete random variables. Figure 4.1 below is adapted from Chernic et al [55] showing a saw-

toothed power function. Hoehler [58] argues that because the power function is non-monotonic with discrete data, calculation of a single required sample size is usually impossible. We can, however, specify a range of sample sizes over which a study will have a given power to reject the null hypothesis. To get round this problem similar methods have recently been used to determine exact sample sizes for comparative studies using Fisher's exact test (Thomas and Conlon [59]).

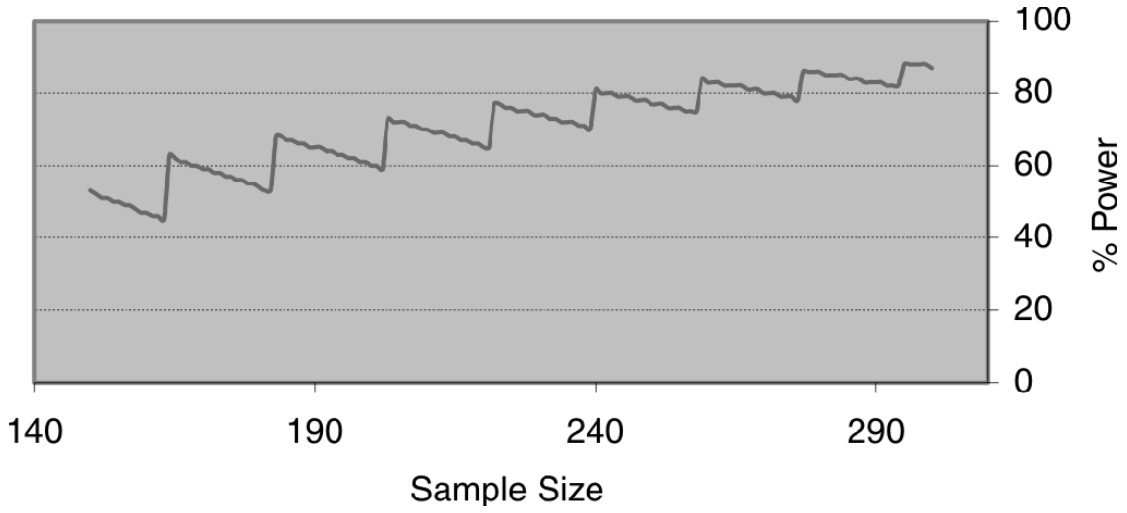


Figure 4.1 Saw-toothed behaviour

4.5 Discussion

The sample size formulae derived in this chapter are not accurate. In particular, the published sample size formula of Farrington et al [3] is inaccurate for $\rho < 1$ or $\rho \approx 1$. The least inaccurate of the four sample size formulae is perhaps that based on the binomial proportion. In chapter 5 we investigate further variants of the binomial-based sample size formula. We also explore other sample size formulae which can allow us to include the effect of age. Age is an important confounder in studies of

vaccine safety, so it is essential that we can have a sample size formula for the self-controlled case series that can take account of the effect of age.

Chapter 5

Improved sample size formulae for the self-controlled case series method

5.1 Introduction

The sample size formulae derived in the previous chapter are not accurate. Hence we sought other sample size formulae that we hoped would be more accurate. In this chapter, we present these other formulae.

In section 5.2, we derive a sample size formula based on binomial proportions but with continuity correction. In section 5.3 we derive a sample size formula also based on binomial proportions but based on the arcsine variance stabilizing transformation. In section 5.4 we derive a sample size formula based on the signed root likelihood ratio statistic. A comparative evaluation of the three formulae is given in section 5.5. The likelihood ratio statistic based sample size formula is then generalised to include allowance for age effects in section 5.6. This formula with age effects is evaluated in section 5.7. We discuss our findings in section 5.8.

5.2 Sample size formula based on the binomial proportion with continuity correction.

Fleiss et al [54] argue that the Type *I* error of sample size formula (4.8) is conservative, and this can be improved by a continuity correction. We thought this could be one of the reasons why formula (4.8) was not very accurate. With continuity

correction, according to Fleiss et al [54], sample size formula (4.8) can be written as shown below:

$$n_e^* = \frac{n_e}{4}\left[1+\sqrt{1+\frac{2}{n_e|\pi_1-\pi_0|}}\right]^2 = \frac{n_e}{4}\left[1+\sqrt{1+\frac{2}{n_e\left|\frac{r(\rho-1)(1-r)}{r\rho+1-r}\right|}}\right]^2 \qquad (5.1)$$

where n_e is the value obtained from sample size formula (4.8) and n_e^* is the sample size with continuity correction. As before r is the ratio of the risk period to the observation period. Note that (5.1) is derived under the same assumptions as those used to derive sample size formula (4.8).

5.3 Sample size formula based on the binomial proportion with arcsine transformation.

As described in chapter 4, in the simplified setting we are considering, the log likelihood is equivalent to that of a binomial with proportion $\pi = \rho r/(\rho r+1-r)$ and index n_e. A popular approach to improve the normal approximation to the binomial is to use the arcsine variance-stabilizing transformation [60]. In this situation the test statistic is

$$T = \arcsin\left(\sqrt{\hat{\pi}}\right);$$

under the null hypothesis

$$T \approx N\left(\arcsin\left(\sqrt{r}\right), 1/4n_e\right),$$

while under the alternative,

$T \approx N\left(\arcsin\left(\sqrt{\rho r/(\rho r+1-r)}\right), 1/4n_e\right)$. Thus we obtain the following expression for the sample size formula:

$$n_e = \frac{\left(Z_{1-\alpha/2} + Z_\gamma\right)^2}{4\left[\arcsin\left(\sqrt{\rho r/(\rho r+1-r)}\right) - \arcsin\left(\sqrt{r}\right)\right]^2} \tag{5.2}$$

5.4 Sample size formula based on the signed root likelihood ratio

A limitation of all sample sizes based on the binomial log likelihood is that they are not readily extended to handle age effects, since the likelihood is then multinomial. Furthermore, the most convenient test to use to decide whether the exposure is associated with the outcome is the likelihood ratio test. Thus it makes sense to base the sample size on the likelihood ratio statistic (4.2) given in chapter 4. Under the null hypothesis, the likelihood ratio statistic has the $\chi^2(1)$ distribution, asymptotically. To obtain an asymptotically normal test statistic, we use the signed root likelihood ratio derived as follows.

The log likelihood kernel for the parameter β given in (4.1) of chapter 4 can be written as,

$\ell(\beta) = x\beta - n_e \log\left(e^\beta r + 1 - r\right)$. The score function is

$\frac{\partial \ell}{\partial \beta} = U(\beta) = x - \frac{n_e e^\beta r}{e^\beta r + 1 - r}$, and solving $U\left(\hat{\beta}\right) = 0$ we have $\hat{\beta} = \log\left(\frac{x}{n_e - x}\frac{1-r}{r}\right)$

Hence the likelihood ratio statistic for the null hypothesis $H_0 : \beta = 0$ is

$D = 2\left\{\ell\left(\hat{\beta}\right) - \ell(0)\right\} = 2\left\{x\hat{\beta} - n_e \log\left(e^{\hat{\beta}} r + 1 - r\right)\right\}$

$$D = 2\left\{ x\log(x) + (n_e - x)\log(n_e - x) + x\log\left(\frac{1-r}{r}\right) - n_e \log(n_e(1-r)) \right\}.$$

Under $H_0, D \sim \chi_1^2$, so $\text{sgn}(\hat{\beta})\sqrt{D} \sim N(0,1)$ where

$$\text{sgn}(\hat{\beta}) = \begin{cases} +1 & \text{if } \hat{\beta} > 0 \\ 0 & \text{if } \hat{\beta} = 0 \\ -1 & \text{if } \hat{\beta} < 0 \end{cases}$$

Assume that under $H_1, \beta > 0,$ and that asymptotically $\text{sgn}(\hat{\beta})\sqrt{D} \sim N(\xi, \tau^2)$ for some ξ, τ to be determined later.

Write

$$f(x) = \text{sgn}(\hat{\beta})\sqrt{D} = \text{sgn}(\hat{\beta})\sqrt{2\text{P}(x)}$$

$$\text{P}(x) = x\log(x) + (n_e - x)\log(n_e - x) + x\log\left(\frac{1-r}{r}\right) - n_e \log(n_e(1-r))$$

Recall that $E(X)$ may be written

$$\mu = E(X) = n_e p$$

where $p = \dfrac{\rho r}{r\rho + 1 - r}$ and $\rho = e^{\beta}$.

Then, using the delta rule (Matthews[60]), to first order

$$f(X) \approx f(\mu) + (X - \mu) f'(\mu) \qquad (5.3)$$

hence

$$E(f(X)) \approx f(\mu).$$

Now

$$\text{P}(\mu) = n_e p \log(\rho) - n_e \log(\rho r + 1 - r)$$

and asymptotically, under H_1, $\text{sgn}(\hat{\beta}) = \text{sgn}(\beta)$. So

$$E(f(X)) \approx \text{sgn}(\beta)\sqrt{2P(\mu)}.$$

Also, from (5.3),

$$\left[f(X) - f(\mu) \right]^2 \approx (X-\mu)^2 \, f'(\mu)^2$$

hence

$$\operatorname{var}\left[f(X) \right] \approx \operatorname{var}(X) \, f'(\mu)^2 = n_e pq f'(\mu)^2 .$$

Now $f'(\mu)^2 = \frac{1}{2}\frac{P'(\mu)^2}{P(\mu)} = \frac{\left[\log(\rho)\right]^2}{2P(\mu)}$,

hence $\operatorname{var}\left[f(X) \right] \approx \frac{n_e pq\left(\log(\rho)\right)^2}{2P(\mu)} = \frac{pq\left(\log(\rho)\right)^2}{2\left[p\log(\rho) - \log(\rho r + 1 - r) \right]}$.

Thus the test statistic $\operatorname{sgn}\left(\beta\sqrt{D}\right)$ is distributed approximately $N\left(\xi, \tau^2\right)$ under H_1,

with

$$\xi = \operatorname{sgn}(\beta)\sqrt{2n_e\left[p\log(\rho) - \log(\rho r + 1 - r) \right]}, \quad \tau^2 = \frac{pq\left(\log(\rho)\right)^2}{2\left[p\log(\rho) - \log(\rho r + 1 - r) \right]}.$$

Let $A = 2\left[p\log(\rho) - \log(\rho r + 1 - r) \right], B = \frac{pq\left(\log(\rho)\right)^2}{A}$ and as before $p = \frac{\rho r}{\rho r + 1 - r}$

Also let C be the critical point of $\operatorname{sgn}(\hat{\beta})\sqrt{D}$, and for simplicity assume $\beta > 0$.

It follows that $1 - \frac{\alpha}{2} = \mathrm{P}\left(\operatorname{sgn}(\hat{\beta})\sqrt{D} > C \middle| H_0\right) = \mathrm{P}(Z > C)$ as the statistic is

approximately $N(0,1)$ under H_0. So $C = Z_{1-\frac{\alpha}{2}}$. Also, for power $\geq 1-\gamma$, the critical

point C must satisfy $1 - \gamma \leq P\left(\operatorname{sgn}(\hat{\beta})\sqrt{D} > C \middle| H_1\right) = P\left(Z > \frac{C - \xi}{\tau} \right)$.

So $\frac{C - \xi}{\tau} \leq -Z_\gamma$. It follows that $C = -Z_\gamma\sqrt{B} + \sqrt{n_e A}$.

Equating the two values of C and solving for n_e gives the following sample size

formula based on the signed root likelihood ratio:

$$n_e = \frac{\left(Z_{1-\alpha/2} + Z_\gamma \sqrt{B}\right)^2}{A}. \qquad (5.4)$$

If we include the proportion of the population vaccinated v, the above formula can be generalised to the formula shown below:

$$n = \frac{vr(\rho - 1) + 1}{v(r\rho + 1 - r)} \frac{\left(Z_{1-\alpha/2} + Z_\gamma \sqrt{B}\right)^2}{A}. \qquad (5.5)$$

Note that n and n_e from expressions (5.4) and (5.5) are in the ratio determined by (4.3) from chapter 4. We can see from (4.3) that if $v = 1$, that is everyone is vaccinated, then $n = n_e$.

5.5 Comparative evaluation of sample size formulae (5.1), (5.2), and (5.4)

In this section, we evaluate the performance of the sample size formulae (5.1), (5.2) and (5.4). We tested the performance of the sample size formulae using simulations in exactly the same way as described in section 4.4.1 of chapter 4. In all simulations we assumed all individuals were exposed, hence $v = 1$.

5.5.1 Results from the simulation study.

The results for 80% power are shown in Table 5.1, and those for 90% power are shown in Table 5.2. As in chapter 4, the numbers of cases calculated from sample size formulae (5.1), (5.2), (5.4) for a particular relative incidence to be detected are given in columns labelled $N1, N2,$ and $N4$. The corresponding observed powers (in %) from simulations are given in columns labelled $P1, P2,$ and $P4$. We can see a marked improvement in terms of the observed power for each sample size formula derived

here. Almost all parameter values gave accurate power (one exception is the combination of $r = 0.01$ and $RI = 0.1$). This shows that the sample size formula based on the binomial with continuity correction, or that using the arcsine variance stabilizing transformation, or that based on the signed root likelihood ratio statistic would give an accurate sample size. Note that the saw-toothed phenomenon mentioned in chapter 4 is still present for sample size formulae (5.1), (5.2) and (5.4). In the next section, we extend the sample size formula based on the signed root likelihood ratio statistic so as to take account of age.

Table 5.1 Empirical power for 80 per cent nominal value

r	RI	$N1$	$N2$	$N4$	$P1$ %	$P2$ %	$P4$ %
0.01	0.10	609	420	462	88*	67	92*
	0.50	2618	2299	2394	81	80	81
	1.20	21031	21801	21537	80	80	80
	1.50	3638	3944	3835	80	80	79
	2.00	1015	1167	1111	78	81	80
	3.00	302	377	348	78	79	80
	5.00	98	135	119	80*	77*	82
	8.00	42	63	54	82*	80*	81
	10.0	29	46	38	79	80	79
0.05	0.10	119	84	92	82	78	79
	0.50	529	469	487	84	80	82
	1.20	4431	4580	4529	79	81	80
	1.50	779	838	817	79	81	79
	2.00	223	253	242	81*	80*	79
	3.00	69	85	79	80	84	80
	5.00	24	32	29	80	82	80
	8.00	12	17	14	75	80*	81*
	10.0	9	13	11	78	79	78
0.1	0.10	58	42	46	85	82	83
	0.50	269	241	250	81*	82*	84
	1.20	2371	2441	2417	79	80	80
	1.50	425	453	443	76	80	79
	2.00	125	140	134	75	81	80
	3.00	41	49	46	80	81	80
	5.00	16	20	18	78	79*	84*
	8.00	8	11	10	81*	81	78*
	10.0	7	9	8	79	79	80
0.5	0.10	10	9	9	81	80	80
	0.50	69	68	69	81	79	81
	1.20	948	947	948	80	80	80
	1.50	194	194	194	80	80	80
	2.00	69	68	69	81	76	81
	3.00	30	29	29	77*	81*	81
	5.00	16	15	15	76	79	79
	8.00	11	10	10	80	80	80
	10.0	10	9	9	78*	81	81*

*r = ratio of the risk period to the observation period, $v = 1$, the proportion vaccinated, RI = relative incidence to be detected, $N1$ = sample size using formula (5.1), $N2$ = sample size using formula (5.2), $N4$ = sample size using formula (5.4), $P1$ = observed power for $N1$, $P2$ = observed power for $N2$, $P4$ = observed power for $N4$. *Saw-tooth phenomenon.*

Table 5.2 Empirical power for 90 per cent nominal value

r	RI	$N1$	$N2$	$N4$	$P1$ %	$P2$ %	$P4$ %
0.01	0.10	688	562	571	95	90*	89*
	0.50	1001	861	882	91	91*	90*
	1.20	28641	29185	28993	90	90	90
	1.50	5062	5279	5197	90	91	90
	2.00	1453	1562	1517	88	89	90
	3.00	450	505	480	86	89*	90*
	5.00	152	180	166	87	91	91
	8.00	68	85	76	91*	90*	90
	10.0	48	61	54	88	90	89
0.05	0.10	135	113	114	92	87	87
	0.50	669	628	639	92	90*	89*
	1.20	6026	6131	6094	90	90	90
	1.50	1079	1122	1106	88	90*	91*
	2.00	316	338	329	89*	90	89*
	3.00	101	113	108	84	91	91
	5.00	36	43	40	86	91	90
	8.00	18	22	20	90	93	90
	10.0	13	17	15	86	89*	93*
0.1	0.10	66	56	57	94	87	89
	0.50	341	323	328	92	90	92
	1.20	3217	3268	3250	90	90	90
	1.50	585	606	599	89	90	90
	2.00	176	187	182	89	89*	90*
	3.00	59	65	62	85	91	89
	5.00	23	26	25	93*	89*	91
	8.00	12	14	13	91*	94	90*
	10.0	9	11	10	93*	91*	89
0.5	0.10	12	12	11	92	92*	93*
	0.50	91	91	91	91	91	91
	1.20	1268	1268	1268	89	90	90
	1.50	259	260	259	90	90	90
	2.00	91	91	91	90	90	90
	3.00	38	39	38	92	91*	92*
	5.00	20	20	20	89	89	89
	8.00	14	14	13	91	91	90
	10.0	12	12	11	92	92	92

*r = ratio of the risk period to the observation period, $v = 1$, the proportion vaccinated, RI = relative incidence to be detected, $N1$ = sample size using formula (5.1), $N2$ = sample size using formula (5.2), $N4$ = sample size using formula (5.4), $P1$ = observed power for $N1$, $P2$ = observed power for $N2$, $P4$ = observed power for $N4$. *Saw-toothed phenomenon.*

5.6 Sample size formula with age effect

All the sample size formulae derived so far apply to a simplified situation in which there are no age effects. In practice, strong age effects may be present. Such age effects can have a big effect on study power, and must be taken into account in sample size calculations. We have seen so far that sample size formulae (5.1), (5.2), and (5.4) give accurate sample sizes in the simplified scenario. Expression (5.1) and (5.2) are based on binomial proportions, and thus cannot readily be extended to allow for age effects, since the likelihood then becomes product multinomial. However, sample size formula (5.4) based on the likelihood ratio test can be extended to allow for age effects.

In line with the parametric case series models described in chapter 2, in which age effects are modelled using a step function, we shall assume that the age-specific incidence is piecewise constant. In practical applications, we have found this approach for specifying the age effects both convenient and flexible.

5.6.1 Assumptions and notation

We again consider a simplified scenario, but involving age effects. We assume that all individuals are followed over the same observation period, which covers *J* age groups of duration e_j, $j = 0,1,2,...,J-1.$ Suppose that the probability that an individual is exposed in age group *j* is $p_j.$ The probability that an individual, randomly selected from the population, is exposed during the observation period is $v = \sum_{j=0}^{J-1} p_j.$ We

suppose furthermore that if an individual is exposed in age group j, the post-exposure risk period, of length e^*, is entirely contained within age group j. This assumption greatly simplifies the calculations, by avoiding any overlaps. It implies that $e^* \leq e_j$ for all age groups $j = 0,1,...,J-1$. This should not be too restrictive in practice, at least when the risk period is short.

Finally, let α_j denote the logarithm of the age-specific relative incidence, relative to age group 0, so that $\alpha_0 = 0$. We assume that these age effects are known. As before, let $\rho = e^{\beta}$ denotes the relative incidence associated with the exposure, and β its logarithm.

5.6.2 Sample size formula allowing for age effects

The full derivation is given in Appendix 2. The sample size formula involves the following intermediate quantities. First, let r_j denote the weighted ratio of time at risk to the overall risk period:

$$r_j = \frac{e^{\alpha_j} e^*}{\sum_{s=0}^{J-1} e^{\alpha_s} e_s}, \quad j = 0,1,...,J-1.$$

Note that if there are no age effects ($\alpha_j = 0$ for all j) then $r_j = r$, the ratio of the risk period to the observation period defined in chapter 4 and section 5.2 of this chapter. Second, let π_j denote the probability of an individual exposed in age group j that an event arising in age group j occurs during the exposure period:

$$\pi_j = \frac{r_j \rho}{r_j \rho + 1 - r_j},\ j = 0,1,...,J-1.$$

If there are no age effects, then $\pi_j = \pi$, the binomial probability defined in chapter 4.

Finally let v_j denote the probability that a case is exposed in age group j:

$$v_j = \frac{p_j\left(r_j \rho + 1 - r_j\right)}{p_0 + \sum_{s=0}^{J-1} p_s\left(r_s \rho + 1 - r_s\right)},\ \ j = 0,1,...,J-1. \tag{5.6}$$

Note that if there is no association between exposure and outcome, so that $\rho = 1$, then $v_j = p_j$, the population proportion exposed. If there is an association, however, the age distribution of exposure in the cases will usually differ from that of the general population. If there is no age effect, then $v_j = n_e / n$ from expression (4.3) of chapter 4.

Now define the following constants A and B:

$$\begin{aligned} A &= 2\sum_{s=0}^{J-1} v_s\left[\pi_s \beta - \log\left(r_s e^{\beta} + 1 - r_s\right)\right], \\ B &= \frac{\beta^2}{A}\sum_{s=0}^{J-1} v_s \pi_s\left(1 - \pi_s\right). \end{aligned} \tag{5.7}$$

Note that when there are no age effects and all individuals are exposed (so $v = 1$), then A reduces to the expression A and B reduces to the expression B given in section 5.4. The total number of events required for $100\gamma\%$ power at the $100\alpha\%$ significance level is

$$n_e = \frac{\left(Z_{1-\alpha/2} + Z_\gamma \sqrt{B}\right)^2}{A}. \qquad (5.8)$$

If there are no age effects, (5.8) reduces to expression (5.4).

5.6.3 Sample size formulae for the number of cases

So far we have presented formulae for n, the number of events. To obtain a sample size formula for the number of cases, an estimate of the cumulative incidence over the observation period is required. Let Λ denote this cumulative incidence. Then under the Poisson model, the number of cases required (that is, the number of individuals with one or more events), n_c is

$$n_c = n\left(\frac{1-e^{-\Lambda}}{\Lambda}\right).$$

Thus $n_c \leq n$. Generally, Λ is not known with any accuracy. In practice, most applications of the case series method are to situations where Λ is very small, in which case $n_c \approx n$. Furthermore, the independence of repeat events may be open to doubt. For these reasons, we would generally advise taking $n_c = n$.

5.7 Evaluation of sample size formula with age effects

5.7.1 Simulation study

We evaluated the sample size expression (5.8) as follows. As before, we assumed an observation period of 500 time units, but now partitioned into $J = 5$ age intervals of

100 units. We fixed the age-specific proportions p_j of the population exposed, and assumed that all individuals in the population are exposed, but varied the age effect: increasing, symmetric and decreasing. The parameter values we used are shown in Table 5.3 below.

The risk period duration e^* must be less than the shortest age group, and were set at 5, 10, and 50 units. For comparability with Tables 5.1 and 5.2, these are reported as proportions of the overall observation period and are denoted r. Thus $r = 0.01, 0.05$ and 0.1. The values for ρ were the same as in the previous simulations, but here we only present results from the values of $\rho = 0.5, 1.5, 2, 3, 5, 10$. We evaluated the sample size for powers of 80 and 90 per cent, at 5 per cent significance level. In this situation, the combination of three values of r, six values of ρ, two powers, and three age effects required 108 different simulations; each involved 5000 runs.

Table 5.3 Exposure and age effects used in the simulations**.**

Parameter	Age group j				
	0	1	2	3	4
Proportion exposed, p_j	0.35	0.30	0.20	0.10	0.05
Age effect, e^{α_j}					
Increasing	1	2	3	4	5
Symmetric	1	2	3	2	1
Decreasing	1	1/2	1/3	1/4	1/5

The sample sizes were calculated using expression (5.8) and were rounded up to the next integer. For each simulation, we randomly and independently allocated the exposure to an age group and the event to an age and exposure group combination. Since the simulations are conditional on an event occurring, we used the age-specific exposure probabilities defined by expression (5.6) to perform this allocation. We then

fitted the case series model with five age groups (and thus four age parameters), and carried out the likelihood ratio test of the null hypothesis $\rho = 1$. The Monte Carlo standard error for the empirical power is about 0.57 per cent at 80 per cent power and 0.42 per cent at 90 per cent power.

5.7.2 Results

The sample sizes and empirical powers are shown in Tables 5.4 and 5.5 for 80% and 90% power respectively. Note that, since $v = 1$, $n = n_e$. The empirical powers generally correspond closely to the nominal values, across the range of parameter values and age settings. There is one exception, namely the rather low (72-73 per cent) power obtained for $r = 0.01$ when $\rho = 10$. This occurred only for nominal power of 80 per cent, but not for 90 per cent power, with age effects, but not when there are no age effects. We have no definitive explanation for this observation, but we suspect it might be due to confounding with age when the expected number of events in the risk period is very small, or to the distribution of the data. In practice, it is most unlikely that a design value of ρ as high as 10 would be used.

Table 5.4 Sample sizes and empirical powers for 80% nominal power

Age effect

		Increasing		Symmetric		Decreasing	
r	ρ	n_e	Power	n_e	Power	n_e	Power
0.01	0.5	3267	81.2	2398	81.5	1825	81.2
	1.5	5219	80.8	3842	79.1	2936	77.6
	2	1509	78.2	1113	80.8	852	78.0
	3	471	79.4	348	79.0	268	79.9
	5	161	79.9	119	79.5	92	78.7
	10	51	72.2	38	73.1	30	72.5
0.05	0.5	667	80.2	491	81.5	379	80.6
	1.5	1103	80.0	825	80.0	649	79.2
	2	324	78.9	244	78.9	193	78.8
	3	104	81.1	80	78.6	64	78.5
	5	38	77.3	29	77.2	24	78.2
	10	13	79.5	11	79.5	10	81.1
0.1	0.5	343	78.9	254	79.4	200	78.6
	1.5	592	78.9	452	78.2	370	79.6
	2	177	80.3	137	79.8	114	80.2
	3	59	80.8	47	79.1	40	80.4
	5	23	78.8	19	79.4	16	78.8
	10	9	77.7	8	79.0	7	77.1

5.5 Sample sizes and empirical powers for 90 per cent nominal power.

		Age effect					
		Increasing		Symmetric		Decreasing	
r	ρ	n_e	Power	n_e	Power	n_e	Power
0.01	0.5	4276	90.6	3139	89.0	2390	91.3
	1.5	7073	89.7	5207	89.3	3978	89.5
	2	2062	91.1	1520	88.8	1163	90.2
	3	651	89.6	481	89.5	369	89.7
	5	224	89.9	167	91.1	128	90.7
	10	72	89.9	54	89.6	42	88.7
0.05	0.5	874	89.7	644	90.4	497	90.4
	1.5	1493	90.3	1116	89.5	877	89.2
	2	442	90.2	332	88.9	263	89.5
	3	143	91.2	109	89.9	87	87.4
	5	52	88.2	40	91.7	33	88.0
	10	19	88.9	15	90.7	13	89.5
0.1	0.5	450	89.7	334	89.9	263	90.2
	1.5	800	89.3	611	89.9	498	89.2
	2	241	89.4	186	90.6	154	89.5
	3	81	90.8	64	90.5	54	90.2
	5	31	90.5	25	90.1	22	89.7
	10	12	87.3	11	90.7	10	89.1

5.8 Conclusion

Sample size formulae for the self-controlled case series method have been discussed in chapter 4 as well as in chapter 5. In chapter 4, we saw that the sample size formula published by Farrington et al [3] was not accurate, neither were the sample size formulae based on the distribution of $\hat{\beta}$ with both first and second order approximations, or the sample size formula based on the binomial proportion without continuity correction or variance stabilizing transformation. In this chapter, we have shown that sample size formula based on the binomial with continuity correction and that with arcsine variance stabilizing transformation are accurate. Equally accurate is the sample size formula based on the signed root likelihood ratio statistic which can

be generalized to take account of age effects. We have seen that the type of age effects has a big impact on the sample size required, as shown in Table 5.5. For example (See also Musonda et al [44]), suppose the observation period includes the ages 366-730 days, divided into 4 age groups $J = 3$ with periods of lengths $e_0 = e_1 = e_2 = 91$ days, and $e_3 = 92$ days. Suppose we took the proportions vaccinated in each of the age intervals to be $p_0 = 0.6, p_1 = 0.2, p_2 = 0.05, p_3 = 0.05.$ Further take the age effects to be $e^{\alpha_0} = 1, e^{\alpha_1} = 0.6, e^{\alpha_2} = e^{\alpha_3} = 0.4,$ and the risk period $e^* = 42$ days, and set $\rho = 3, Z_{1-\alpha/2} = 1.96$ and $Z_\gamma = 0.8416$ for 80 per cent power to detect a relative incidence of 3 at the 5 per cent significance level. With these values, we find $n_c = 37$, but if we ignored the age effect, we would obtain $n_c = 45$. Thus it is important to allow for such age effects in calculating the sample size. In conclusion, we recommend the sample size formula based on the signed root likelihood ratio, as shown in expressions (5.4) and (5.8).

Our empirical power calculations were based on the likelihood ratio test. In practice, statistical significance is sometimes assessed by calculating the 95 per cent confidence interval for the relative incidence, and observing whether this confidence interval includes 1.We also evaluated our recommended sample size formula using this second criterion. The empirical powers were generally close to the nominal values, except for large relative risks and or very short risk periods when such confidence intervals can be markedly non-central as shown in chapter 3.

In calculating the sample size allowing for age effects, we assumed that the age effects were known, so as to obtain a one-parameter likelihood. In practice, the age

effects must be estimated. We had expected this to have some bearing on the results, in that some information in the sample is used to estimate the age effects. In the event, this effect is small.

A limitation for sample size formula (5.8) is the requirement that the risk period is shorter than the age groups involved. Another is that we have assumed that there is a single risk period. In practice, it is common to use several, usually rather short, risk periods. However, it is often possible to select a single, short risk period of special importance, on which to base the sample size calculations. If long risk periods are required in situations where age effects must be allowed for, our proposed sample size formula may not apply without further modification.

We recommend the sample size formula based on the signed root likelihood ratio statistic for use both when there is no age effect and where there is an age effect. In particular, the sample size formulae in this chapter help to emphasize the importance of taking age into account at the design stage.

Chapter 6

Application of the self-controlled case series method in surveillance

6.1 Introduction

In this chapter we aim to describe how we can use the self-controlled case series method in prospective surveillance. In section 6.2, we briefly describe the use of surveillance systems to monitor vaccine safety and relate this to the self-controlled case series method. We also identify the possible problems of using the self-controlled case series method for prospective surveillance. The background and review of various statistical methods used for surveillance are given in section 6.3. Section 6.4 describes the sequential probability ratio test (SPRT) and the theory behind the SPRT is given in section 6.5. Description of the application of the SPRT is given in section 6.6. Section 6.7 explores using the self-controlled case series adjusting for age. In section 6.8, we show results from a simulation study demonstrating possible parameter values for a surveillance system, and conclusions are drawn in section 6.9.

6.2 Surveillance systems for adverse events

Two examples of surveillance systems for vaccine–associated adverse events are: the Yellow Card system in the UK and the Vaccine Adverse Event Reporting System (VAERS) in the US [61]. The Yellow Card Scheme was introduced in 1964 to provide a straightforward route for a doctor or dentist and later any member of the

public to report a suspicion that a medicine could have harmed a patient. The Yellow Card Scheme (http://www.yellowcard.gov.uk) is run by the Medicines and Healthcare products Regulatory Agency (MHRA).

The Vaccine Adverse Event Reporting System (http://vaers.hhs.gov) is a cooperative program for vaccine safety of the Centres for Disease Control and Prevention (CDC) and the Food and Drug Administration (FDA) in the US. Like the Yellow Card system in the UK, VAERS collects information about adverse events that occur after the administration of US licensed vaccines. A similar scheme on the international level is conducted by the World Health Organisation (WHO) at the Uppsala Monitoring Centre [62].

The data collected in these systems are not independent of the exposure since only events occurring after exposure to the drug are collected. Data are usually collected from a mixture of populations, and there is no denominator data. There may be no confirmation of the reported adverse events. Such data sets may suffer from underreporting and differential reporting [63, 64]. There is no control group for comparison of adverse event rates [65, 66]. Over reporting may also occur because some reported conditions might not meet standard diagnostic criteria [64]. There is also lack of information on background incidence of adverse events in the general population and information concerning the total number of doses of vaccines or vaccine combinations actually administered. For these reasons surveillance systems such as VAERS and Yellow Card Scheme cannot be readily used to determine whether associations between vaccines and reported adverse events are causal.

Surveillance system such as VAERS and Yellow Card Scheme are nevertheless useful for generating hypotheses to be tested in other settings such as the Vaccine Safety Datalink (VSD), or through specially designed epidemiological studies [33, 67].

In this chapter and the next we discuss how the self-controlled case series method can be applied to the surveillance of adverse events following vaccination. The aim is to obtain better evidence of causality than available from systems such as the Yellow Card or VAERS, while remaining in a prospective setting.

There are possibly two ways in which the self-controlled case series method can be applied for surveillance. These are:

1. Prospective surveillance of a new vaccine.
2. Long term surveillance to identify changes in the performance of one or several existing vaccines.

In the first situation, there is a specific hypothesis to be tested, for example relating to intussusception following the introduction of a new rotavirus vaccine [68, 69]. In the second situation, one might be interested in monitoring a range of possible exposures, not a single one, to check that none of the vaccines in current use are associated with some adverse outcome. In this situation, monitoring may concern several outcomes with no specific hypotheses. For the second scenario, one needs to protect against false positive alarms more than for the first scenario.

Both the above scenarios can use data that are centrally collected in one or several databases, for example the Hospital Episode Statistics (HES) data, UK General Practice Research Database (GPRD), immunisation data which is independent from clinical records such as the Vaccine Safety Datalink (VSD) database, data from Patient Administration System (PAS), and Coded Clinical Records (CCR). Data collected routinely as in these databases are the best to use because ascertainment is likely to be independent from the exposure. The main difficulty is how to use the self-controlled case series method, which is a retrospective design, in a prospective surveillance context.

6.3 Background and review of some surveillance methods

Statistical methods can play an important role in detecting changes in many processes, including mortality and adverse event rates. Some surveillance methods have an established history of use with health care, while there is growing interest in others such as statistical process control (SPC) methods. The retrospective use of SPC by Spiegelhalter et al [70] provides an excellent example of the potential role that risk-adjusted control charts could have played in earlier detection of higher mortality rates in the Bristol Royal Infirmary and in the general practice of Harold Shipman.

Statistical control charts were first developed in the 1920s by Walter Shewhart at Bell Laboratories [71] and have been widely used by Deming [72]. Shewhart and Deming independently recognised the value of these methods for detecting statistical changes in many applications, though they were initially intended for use in industrial and chemical processes. As early as 1942, Deming [73] recognised their potential value

for disease surveillance and rare events. Important health care concerns in which control charts have been shown to be effective include surgical site infections, adverse drug events, needle stick injuries, and ventilator-associated pneumonia [74].

Some studies have attributed the growing use of SPC in the medical context to the staggering incidence and cost of medical mistakes [75], care induced injury, and hospital acquired infections. For example it is reported [75] that in the US alone, between 770 000 to 2 million patients are injured per year, 44 000 to 180 000 deaths, and the cost of all these accidents/incidents is estimated to be about $8.8 billion annually [76-78].

Surveillance systems such as the Yellow Card and VAERS all make use of suitable statistical methods to identify possible signals. For example measures of disproportionality including Proportional Reporting Ratio (PRR), Reporting Odds Ratio (ROR), and Yule's Q, along with more complex Bayesian methods are currently applied in various national spontaneous reporting centres [79-82].

Another approach for continuous systematic review of all combinations of drugs and suspected adverse reactions (ADRs) reported to a spontaneous reporting system to optimize signal detection makes use of Bayesian methods. This works by relating the prior and posterior probabilities before and after linking databases. It is currently being used by the Uppsala Monitoring Centre (Bayesian Confidence Propagation Neural Network analysis BCPNN) [83-85].

Cumulative monitoring approaches based on control charts of different kinds are widely used. For example it is well known [75] that the simplest types of statistical control charts, called Shewhart charts, perform fairly well for detecting moderate-to-large rate changes in the parameter of interest. In some industrial applications, more advanced tools such as sequential probability ratio test (SPRT), cumulative sum (CUSUM) charts are used to detect smaller changes, to monitor low rates, or in situations where sufficiently large sample sizes are not available. Examples of health care CUSUM applications include surveillance of seasonal influenza [86, 87] , community Salmonella [88], and fever curves in neutropenic patients. Various new SPC methods have been developed for non-standard applications dealing with rare events, infectious diseases and other event that naturally occur in clusters, overdispersion, naturally cyclic behaviour, and risk adjustment [70]. Related SPC methods have also been developed to handle non-homogenous event in manufacturing, such as for different production lines.

Another motivation for cumulative monitoring approaches is to accommodate rare events that otherwise would require large samples to yield adequate statistical sensitivity. SPRTs and CUSUMs are excellent for this purpose, and several other SPC methods also have been developed for rare events. Many of these are based on some variation of the idea of monitoring the number of cases or time between adverse events rather than the more traditional approach of monitoring the number of events or deaths within a fixed time period or accumulating sample size [89, 90].

The SPRTs and CUSUMs are the most adaptable cumulative monitoring methods to use with the self-controlled case series method. This is because they are based on the

likelihood ratio. We adapt them by using the likelihood of the self-controlled case series method. In this chapter, we shall concentrate on the SPRTs and in chapter seven, we will describe how to use the CUSUMs.

Charts derived from the sequential probability ratio test have been widely used in industry to monitor process performance. The SPRT is used both when the monitoring is continuous and items can be inspected one by one, and when items are inspected in a group after a fixed time interval. Studies have shown that charts based on the SPRT will signal an out-of-control process earlier than either the Shewhart p-chart or the CUSUM chart [91]. Recently there has been increased attention paid to the use of the CUSUM and SPRT charts in a medical context [74, 88, 92]. The SPRT is the most powerful method for discriminating between two hypotheses [70, 93], and was recommended well over 40 years ago in a medical context for clinical trial and clinical experiments [94, 95]. In the next section, we describe charts derived from the SPRT by first looking at the theory behind SPRT.

6.4 The sequential probability ratio test (SPRT)

Formal statistical methods for sequential analysis were developed in 1943 independently by Barnard in the UK and Wald in the US [93, 96]. Suppose we are in a situation where we have two hypotheses, the null hypothesis H_0 and the alternative hypothesis H_1. Interest is on deciding whether to accept the null hypothesis or reject the null hypothesis (hence accepting the alternative). The idea behind sequential testing is that we collect observations one at a time; when observation $X_i = x_i$ has been made, we choose between the following options:

- Accept the null hypothesis H_0 and stop observation.
- Accept the alternative hypothesis H_1 and stop observation.
- Defer decision until we have collected another piece of information X_{i+1}.

The challenge of course is to find out when to choose the above options. To do that, one has to control for two types of error:

$\alpha = P\{\text{Accepting } H_1 \text{when } H_0 \text{ is true}\}$ (Type I error), and

$\beta = P\{\text{Accepting } H_0 \text{ when } H_1 \text{ is true}\}$ (Type II error).

Note that it is common in this context to treat H_1 and H_0 symmetrically. More formally, suppose we consider a simple hypothesis $H_0 : \theta = \theta_0$ against a simple alternative $H_1 : \theta = \theta_1$. The standard likelihood ratio test has critical region of the form

$$Z_n = \log \frac{L(\theta_1; X_1, ..., X_n)}{L(\theta_0; X_1, ..., X_n)} > K$$

for some constant K and $X_1, ..., X_n$ are n independent observations on the random variable X. The expression $L(\theta_1; X_1, ..., X_n)$ represents the likelihood when H_1 is true and the expression $L(\theta_0; X_1, ..., X_n)$ represents the likelihood when H_0 is true. Note that assuming independence the log likelihood ratio Z_n is the cumulative sum

$$Z_n = \log\left(\frac{L(\theta_1; X_1)}{L(\theta_0; X_1)}\right) + ... + \log\left(\frac{L(\theta_1; X_n)}{L(\theta_0; X_n)}\right).$$

Now consider $X_1, X_2, \ldots$ being successive observations obtained sequentially. Wald's [45] sequential probability ratio test has the following form:

- If $Z_n \geq \log(A)$, decide that H_1 is true and stop;
- If $Z_n \leq \log(B)$, decide that H_0 is true and stop;
- If $\log(B) < Z_n < \log(A)$, collect another observation to obtain Z_{n+1},

where A and B are two constants such that $\log(B) < \log(A)$. The constants A and B are to be determined so that the test will have the prescribed strength (α, β).

It can be shown that the SPRT is optimal [45, 75, 91, 92] in the sense that it minimizes the average sample size before a decision is made among all sequential test which do not have larger error probabilities than the SPRT. An essential feature of the sequential test is that the number of observations required by the sequential test depends on the outcome of the observations and is, therefore, not predetermined, but a random variable [45]. This is because at any stage, the decision to terminate the process depends on the observations made so far.

6.5 Theoretical properties of the SPRT

6.5.1 The relations between the quantities α, β, A, B in an SPRT

Following Wald's [45] derivation, suppose we let $f(X,\theta)$ denote the density of the random variable X under consideration for some parameter θ. As before let H_0 be the hypothesis that $\theta = \theta_0$, and H_1 the hypothesis that $\theta = \theta_1$. We can thus denote $f(X,\theta_0)$ as the distribution of X given that H_0 is true and by $f(X,\theta_1)$ is distribution when H_1 is true. Successive observations on X shall be denoted by $X_1, X_2, ...,$.

Further suppose for any integer value m, the probability that a sample $X_1, X_2, ..., X_m$ is obtained is given by

$$p_{1m} = f(X_1,\theta_1)...f(X_m,\theta_1) \text{ when } H_1 \text{ is true,}$$

and by

$$p_{0m} = f(X_1,\theta_0)...f(X_m,\theta_0) \text{ when } H_0 \text{ is true.}$$

Suppose we say that the sample $(X_1, X_2, ..., X_n)$ is of type 0 if

$$B < \frac{p_{1m}}{P_{0m}} = \frac{f(X_1,\theta_1)...f(X_m,\theta_1)}{f(X_1,\theta_0)...f(X_m,\theta_0)} < A \text{ for } m = 1,...,n-1 \text{ and } \frac{p_{1n}}{p_{0n}} \le B .$$

Similarly, we shall say a sample $(X_1, ..., X_n)$ is of type 1 if

$$B < \frac{p_{1m}}{P_{0m}} = \frac{f(X_1,\theta_1)...f(X_m,\theta_1)}{f(X_1,\theta_0)...f(X_m,\theta_0)} < A \text{ for } m = 1,...,n-1 \text{ and } \frac{p_{1n}}{p_{0n}} \geq A .$$

Hence it follows from that a sample of type 0 leads to the acceptance of H_0 and a sample of the type 1leads to the acceptance of H_1. For any given sample $(X_1,...,X_n)$ of type1, the probability of obtaining such a sample is therefore at least A times as large under hypothesis H_1 as under H_0. As shown by Wald [45], the probability measure of the totality of all samples of type $1$is the same as the probability that the sequential process will terminate with acceptance of H_1. But the latter probability is equal to α when H_0 is true and to $1-\beta$ when H_1 is true. This is by definition of α and β and because the probability that the sequential process will eventually terminate is one. Hence,

$$1-\beta \geq A\alpha \quad .$$

The inequality above can be written as

$$A \leq \frac{1-\beta}{\alpha} \qquad (6.1)$$

and so $\frac{1-\beta}{\alpha}$ is an upper limit for A.

Similarly, a lower limit for B can be derived as follows.

For any given sample $(X_1,...,X_n)$ of type 0, the probability of obtaining such a sample under H_1 is at most B times as large as the probability of obtaining such a sample when H_0 is true. Thus, also the probability of accepting H_0 is at most B times

as large when H_1 is true as when H_0 is true. Since the probability of accepting H_0 is $1-\alpha$ when H_0 is true and β when H_1 is true, we obtain the inequality

$$\beta \leq (1-\alpha) B \quad .$$

It follows that

$$B \geq \frac{\beta}{1-\alpha} \tag{6.2}$$

and thus $\frac{\beta}{1-\alpha}$ is a lower limit for B.

The inequalities (6.1) and (6.2) have been derived under the assumption that the successive observations $X_1, X_2, ..., \text{etc}$, are independent. It can be shown [45] that the validity of the inequalities (6.1) and (6.2) is not restricted to the case of independent observations. They are generally valid also for dependent observations.

6.5.2 Calculating the constants *A* and *B*

Suppose that we wish to design a test procedure of strength (α, β). Then our problem is to determine the constants A and B such that the resulting test will have the desired strength (α, β). Let us denote by $A(\alpha, \beta)$ and $B(\alpha, \beta)$ the values of A and B, respectively, for which the test has the required strength (α, β). The exact determination of the values $A(\alpha, \beta)$ and $B(\alpha, \beta)$ is usually very laborious [45]. However, the inequalities (6.1) and (6.2) permit an approximate determination of A and B which will suffice for most practical purposes. From (6.1) and (6.2), it follows that

$A(\alpha,\beta) \leq \frac{1-\beta}{\alpha}$ and $B(\alpha,\beta) \geq \frac{\beta}{1-\alpha}$. It can be shown [45] that for most practical purposes, the constants A and B are approximately equal to:

$A \approx \frac{1-\beta}{\alpha}$ and $B \approx \frac{\beta}{1-\alpha}$.

Using approximate values of A and B instead of exact values results in some error in the Type I and Type II probabilities. Let us denote by α' and β' the resulting probabilities of errors of Type I and Type II respectively for using approximate values of A and B. From (6.1) and (6.2) it follows that

$$\frac{\alpha'}{1-\beta'} \leq \frac{\alpha}{1-\beta}, \text{ and} \qquad (6.3)$$

$$\frac{\beta'}{1-\alpha'} \leq \frac{\beta}{1-\alpha} . \qquad (6.4)$$

It follows from the above that

$$\alpha' \leq \frac{\alpha}{1-\beta} \text{ and} \qquad (6.5)$$

$$\beta' \leq \frac{\beta}{1-\alpha} . \qquad (6.6)$$

Multiplying (6.3) by $(1-\beta)(1-\beta')$ and (6.4) by $(1-\alpha)(1-\alpha')$ and adding the two inequalities, we obtain

$$\alpha' + \beta' \leq \alpha + \beta. \qquad (6.7)$$

The inequalities (6.5), (6.6), and (6.7) give useful upper limits for α' and β'. Wald [45] argued that since in practical applications, the values α and β will usually be

small, probably in the range 0.01 to 0.05, thus $\frac{\alpha}{1-\beta}$ and $\frac{\beta}{1-\alpha}$ will be very nearly equal α and β, respectively. Inequalities (6.5) and (6.6) indicates that the amount by which α' may exceed α, or β' may exceed β is small and can be neglected for practical purposes. In fact, inequality (6.7) implies that at least one of the inequalities $\alpha' \leq \alpha$ and $\beta' \leq \beta$ must hold. In other words, by using the approximate values of $A(\alpha, \beta)$ and $B(\alpha, \beta)$ instead of exact values of $A(\alpha, \beta)$ and $B(\alpha, \beta)$, respectively, at most one of the probabilities α and β may be increased.

Wald [45] concluded that the use of approximate values of $A(\alpha, \beta)$ and $B(\alpha, \beta)$, instead of exact values of $A(\alpha, \beta)$ and $B(\alpha, \beta)$ respectively, cannot result in large increase in the value of either α or β. This means, for all practical purposes the test corresponding to $A = \frac{1-\beta}{\alpha}$ and $B = \frac{\beta}{1-\alpha}$ provides at least the same protection against wrong decisions as the test corresponding to the use of the exact values of A and B.

The other possible consequence of using approximate values of $A(\alpha, \beta)$ and $B(\alpha, \beta)$ instead of exact values is that this may result in an appreciable decrease of either or both error probabilities. If this were so, it would mean only that the test based on the approximate values $A = \frac{1-\beta}{\alpha}$ and $B = \frac{\beta}{1-\alpha}$ would provide a better protection against wrong decisions than the test based on the exact values. The only possible disadvantage is an appreciable increase in the number of observations

required by the test. But this has been theoretically investigated [45] and it has been shown that such an increase in the number of observation is only slight and of no practical consequence. Thus the test based on the approximate values $A = \frac{1-\beta}{\alpha}$ and $B = \frac{\beta}{1-\alpha}$ serves the purpose just as well, and the determination of exact values is of little practical importance.

The ideas developed above relate to a situation in which a decision has to be made with observations taken singly, that is, by item-by-item inspection. An interesting question is therefore whether the sequential test works when items are inspected in groups, for example when items are inspected at a particular fixed time interval. Some researchers [91, 97] argue that the threshold for grouped data should be adjusted to take account of the grouping. But Wald [45] showed that taking observations in groups and applying the SPRT should lead to the same conclusions as item-by-item inspection. Wald's theoretical argument on using the SPRT with grouped data concluded that, for all practical purposes, grouping does not decrease the protection against wrong decisions provided by the test. Hence we shall use grouped data with A and B calculated as if the test was based on item-by-item inspection and be re-assured by Wald's findings, as used by Spiegelhalter et al [70], that we ought to make the same decisions as we would if we had singly collected data. Our strategy in any case will be to test performance by simulation.

6.6 Application of the SPRT to the self-controlled case series method.

6.6.1 Surveillance scenario

In this section, we introduce a hypothetical surveillance system which is set up to monitor the performance of a new vaccine, for example the introduction of a new rotavirus vaccine (see chapter 8). We shall describe how the self-controlled case series method can be used with the SPRT. For definiteness, we shall take the adverse event of interest to be intussuscetption in children aged under 2 years [68, 69, 98].

Cases of intussusception in children aged less than two years are reported to a central database. At regular time intervals (say every 6 months, or every 12 months) the vaccination records of the cases notified in the previous 6 or 12 monthly period are ascertained using a mechanism that is independent of the occurrence of the event. Such data might be obtained from databases such as hospital records, the GPRD or VSD [99].

The self-controlled case series method is then applied at the end of each successive 6 or 12-month calendar time interval. We call this interval the *monitoring interval*. The observation period for each case with an event during that time period includes all time spent in the defined age groups (in our example, 0-2 years) within the monitoring interval. If the adverse event is a contra-indication for subsequent vaccination, the observation period is further constrained to begin with vaccination. The risk period (in our example, this might be 2 weeks post-vaccination) will have been defined prior to

the start of surveillance. In what follows we assume for simplicity that there is a single risk period and that each child receives a single vaccine dose.

In the self-controlled case series method, fixed covariates are controlled for as was shown in chapter two. For simplicity we shall assume that there is no age effect.

We have two hypotheses: a null hypothesis H_0 corresponding to no association, hence the relative incidence $RI_0 = 1$, and H_1 corresponding to a relative incidence (say, $RI_1 = 2$) that is deemed important to detect. Note that, RI_0, α, β and RI_1 have to be defined in advance. The log-likelihood ratio is then calculated at the end of each fixed time interval.

From chapter two, the log-likelihood with no age effect is

$$\ell(\beta) = \sum_i \sum_k n_{ik} \log\left(\frac{\exp(\beta_k) e_{ik}}{\sum_s \exp(\beta_s) e_{is}} \right)$$

where i denotes individuals, and the risk period indexed by k with

$$k = \begin{cases} 0 \text{ if unexposed and } \beta_0 = 0 \\ 1 \text{ if exposed} \end{cases}.$$

The symbols e_{ik}, n_{ik} respectively denote length of time at risk and number of events experienced by an individual i in risk period k.

Under H_0, $\beta_0 = 0$. Let β_1 denote the value of β under H_1. Thus $\beta_0 = 0$ and $\beta_1 = \log(RI_1)$. Now let e_{ikt}, n_{ikt} denote, respectively, the time at risk and

number of events experienced by case i in risk period k during the tth monitoring interval. Also let L_{0t}, L_{1t} denote the likelihoods under H_0 and H_1, respectively, for time interval t.

It follows that:

$$\log\left(\mathrm{L}_{0t}\right) = \sum_{i,k} n_{ikt} \log\left(\frac{e_{ikt}}{\sum_{r=0}^{1} e_{irt}} \right), \quad \log\left(\mathrm{L}_{1t}\right) = \sum_{i,k} n_{ikt} \log\left(\frac{\exp(\beta_k) e_{ikt}}{\sum_{s=0}^{1} \exp(\beta_s) e_{ist}} \right)$$

and so

$$\Lambda_t = \log\left(\frac{\mathrm{L}_{1t}}{\mathrm{L}_{0t}}\right) = \sum_{i,k} n_{ikt} \log\left(\exp(\beta_k) \frac{\sum_{r=0}^{1} e_{irt}}{\sum_{s=0}^{1} \exp(\beta_s) e_{ist}} \right)$$

$$= \sum_{i,k} n_{ikt} \left(\beta_k - \log\left(\sum_{s=0}^{1} \exp(\beta_s) \frac{e_{ist}}{\sum_{r=0}^{1} e_{irt}} \right) \right).$$

Hence

$$\Lambda_t = n_{.1t} \log(RI_1) - \sum_i n_{i.t} \log\left(\omega_{i0t} + RI_1 \omega_{i1t}\right) \qquad (6.8)$$

where $\omega_{ikt} = \dfrac{e_{ikt}}{e_{iot} + e_{i1t}}$ is the proportion of time spent by case i in exposure category k during the tth monitoring interval.

In the simulations that follow we shall assume $\omega_{i0t} = 1 - r$ and $\omega_{i1t} = r$. Then (6.8) above reduces to:

$$\Lambda_t = n_{\cdot 1t}\beta_1 - n_{\cdot\cdot t}\log\left(1 - r + e^{\beta_1} r\right) \qquad (6.9)$$

where $n_{\cdot 1t}$ is the number of events during the monitoring interval (t) that occurred in the risk period, $n_{\cdot\cdot t}$ is the baseline incidence of the number of cases arising in the monitoring interval (t) and r is the ratio of the risk period to the observation period.

The SPRT chart involves plotting the pair (t, Z_t)

$$Z_t = Z_{t-1} + \Lambda_t, \quad t = 1, 2, 3, \ldots \qquad (6.10)$$

at the tth monitoring interval, where $Z_0 = 0$ and

$\Lambda_t = \log\left(\frac{L_{1t}}{L_{0t}}\right) = n_{\cdot 1t}\beta_1 - n_{\cdot\cdot t}\log\left(1 - r + e^{\beta_1} r\right)$ is the sample weight assigned to monitoring interval t.

In the SPRT chart, sampling should continue if the quantity Z_t lies between two thresholds $\log(A)$ and $\log(B)$. When Z_t exceeds $\log(A)$, stop and reject H_0 in favour of H_1 and vice versa when Z_t is less than $\log(B)$. Thus the boundaries take the form of horizontal lines.

6.6.2 Specifications in the SPRT chart

One of the most important specifications before carrying out such a surveillance exercise concerns values of α and β. The sizes of α and β should reflect the costs of making the two types of error. For example, if we wish to avoid falsely identifying an

adequate vaccine as being positively associated with an adverse outcome then α should be made very small, whereas if we consider it a serious mistake to miss a poor vaccine which is positively associated with an adverse outcome, then β should be made small. Both errors are serious, so we adopt a convention of using equal α and β. Spiegelhalter [70] advocates that instead of choosing a single value for α and β, a set of horizontal lines can be drawn on the chart to indicate different degrees of urgency: for example, a monitoring study might use $\alpha = \beta = 0.1$ as an 'alert' threshold and a more stringent $\alpha = \beta = 0.01$ for 'alarm'. Table 6.1 below gives some possible thresholds for various values of α and β.

Table 6.1 Thresholds for the SPRT for different values of α and β

α	β	Lower threshold $\log(B)$	Upper threshold $\log(A)$
0.05	0.05	-2.94	2.94
0.01	0.01	-4.60	4.60
0.01	0.02	-3.90	4.58
0.02	0.01	-4.58	3.90
0.005	0.005	-5.29	5.29
0.001	0.001	-5.91	5.91

It is also possible to set up several SPRT surveillances in different countries or institutions (hospitals or GP practices). If that were to be the case, more stringent boundaries may be appropriate because of the many comparisons being made. For example if we had 10 centres on surveillance, of 10 centres performing normally, we would expect one to cross the 'alert' boundary by chance alone. Some authors [70] propose a Bonferroni-like adjustment, for example when monitoring n institutions,

using values of $\alpha = \beta = 0.1/n$ and $\alpha = \beta = 0.01/n$ for 'alert' and 'alarm' respectively.

The original idea of the SPRT as conceived by Wald [45] was designed to carry out a test of hypothesis H_0 versus the alternative H_1 and then decide either to terminate because the threshold has been crossed or continue observing because the threshold has not been crossed. For long term surveillance we could modify the idea so as to restart the procedure when, say, we cross the lower boundary and so are confident there is no increase in the relative incidence of an adverse outcome. Modifying an SPRT in this way has the advantage that it is not possible to build up excessive 'credit' and so gains sensitivity changes in performance [70], but also has the disadvantage that the strict interpretation of α and β is lost. Such a loss is not too serious if the surveillance system as described here is only an aid to monitoring which should eventually trigger remedial action such as investigating by conducting a proper retrospective study to confirm or reject the 'signal' detected. Another variant is to introduce a third, vertical, boundary which effectively places a time limit on the surveillance. The rationale for using a third vertical boundary relate to the context in which we envisage to use the SPRT, namely focussed surveillance of a new vaccine. In such a situation it is not appropriate to wait indefinitely for evidence of safety or lack of it. Thus it is appropriate to build in a maximum surveillance time, and design the system so as to have a high probability of not hitting this vertical boundary. In the simulation study whose results and procedure is reported in section 6.8, we used this approach.

6.7 Adjusting for age in the SPRT

Both adverse events and vaccination are often highly age-dependent. Hence it is important to adjust for age. One way to control for age in the SCCS method could be to use profile likelihood where we profile out the age parameters as nuisance parameters [100].

Thus Z_t becomes $Z_t = \log\left(\frac{\left(L_{1t} \times L_{1t-1} \ldots L_{11}\right)^{\square}}{\left(L_{0t} \times L_{0t-1} \ldots L_{01}\right)^{\square}}\right)$

where $L^{\square}$ = profile likelihood for β (the logarithm of the relative incidence) having profiled out the age parameter.

There are three possible ways in which the age parameter could be profiled out within the surveillance system described earlier, and these are as follows:

1) First obtain the age parameter $\hat{\alpha}(\beta)$ values for the first monitoring interval, or for a baseline period, and keep these fixed thereafter or,
2) Re-estimate the age parameter $\hat{\alpha}(\beta)$ separately within each monitoring interval or,
3) Re-estimate the age parameter $\hat{\alpha}(\beta)$ at each monitoring time interval using all previous data.

We have left the investigation of these ideas about controlling for age for future research. In the next section we present results from simulations showing how the SPRT would work in a simplified scenario without controlling for age.

6.8 Simulation study: evaluating the performance of the case series SPRT

6.8.1 Description of the surveillance scenario

Let us assume that we have set up a surveillance system as described in section 6.6.1 to monitor a new vaccine every six months. In the surveillance system, the numbers of cases of a particular adverse outcome are collected at a central reporting centre. Note here that the monitoring interval can be of any length depending on prior knowledge of a particular vaccine being monitored.

We decided on a surveillance period of 10 years. This 10-year period determines the third vertical boundary discussed above. It is used primarily for design purposes, as we require that there should be good power to detect a problem within this period. In practice the surveillance could continue beyond this boundary. The choice of 10 years is arbitrary and could be varied according to requirements. In what follows, ‘power’ refers more precisely to operational power, namely the probability of detecting a genuine problem before the vertical boundary is reached. We carried out simulations for various lengths of surveillance periods, but here we only present results from a ten year surveillance period.

Recall that the SPRT chart involves plotting the pair (t, Z_t)

$$Z_t = Z_{t-1} + \Lambda_t, \quad t = 1, 2, 3, \ldots \qquad (6.10)$$

where $Z_0 = 0$, and t counts the monitoring interval. For the results presented here, we used a six months monitoring interval (for simplicity, all 'months' contain 4 weeks and all 'years' 48 weeks).

$\Lambda_t = \log\left(\frac{L_{1t}}{L_{0t}}\right) = n_{.1t}\beta_1 - n_{..t}\log\left(1 - r + e^{\beta_1} r\right)$ is the sample weight assigned to monitoring interval t, where $n_{.1t}$ is the number of events during the monitoring interval that occurred in the risk period, $n_{..t}$ is the number of events arising in the monitoring interval and *r* is the ratio of the risk period to the observation period. The risk period was varied: we used 1, 2 and 4 weeks. A range of relative incidences to be detected were investigated but here we only present results from relative incidence of 1.5, 2, 3, 3.5, 4, and 5.

It is important to distinguish between two uses of the relative incidence in the simulation. We shall denote $\mathrm{RI} = e^{\beta_1}$ the design value, that is, the value used in the SPRT. In addition, we shall denote $\mathrm{RI}_2 = e^{\beta_2}$ the actual value used to generate the data. The values of RI_2 included 1, 1.5, 2, 3, 3.5, 4, and 5.

We used a random number generator using SAS program version 8.2 [101] to generate the total number of cases in each monitoring interval arising from a Poisson distribution:

$n_{..t} \sim \mathrm{Poisson}(\lambda)$

where the underlying rate λ was fixed at one of the following values:

$\lambda = 5,10,20,50$.

The numbers of cases arising in the risk period were generated using the binomial distribution with the expression:

$$n_{.1t} \sim Binomial(n_{..t}, \pi)$$

where $\pi = \frac{e^{\beta_2} r}{e^{\beta_2} r + 1 - r}$ is the probability of a case being in the risk period.

We simulated a ten year surveillance period with six month monitoring time interval. So, if the process did not give any signal, we expect a total of 20 inspections in which the value of the SPRT is calculated every six months. For each combination of parameters we repeated the procedure 2000 times. We call a set of 2000 simulations a run. In each run, we observed the ability of the surveillance system to detect a particular relative incidence by finding the proportions of occasions on which the upper, lower and vertical thresholds were crossed. To check the speed of response of the surveillance system, we calculated the average time at which a particular boundary was crossed for those simulations in which the boundary was crossed.

Figure 6.1 below gives an example of the output. The cumulative value of SPRT is plotted at each monitoring interval. We see three realisations of the process, one in which the observed number of cases arising in each six month monitoring interval leads to the acceptance of alternative hypothesis, that the relative incidence is 5; this

happens in the second year. In such a situation, monitoring would have to be stopped and further investigations carried out.

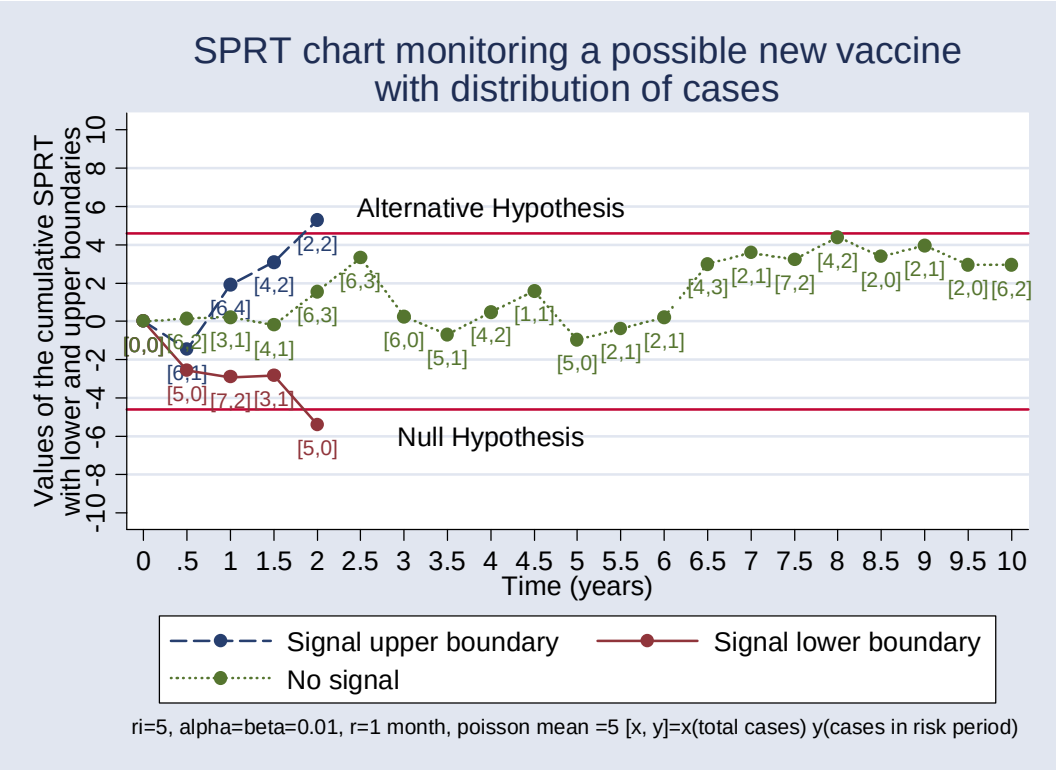


Figure 6.1. Example of three realizations with relative incidence 5, ratio of the risk period to the observation period $r = \frac{1}{6}, \lambda = 5, \alpha = \beta = 0.01$

The lower path is an example of realisation in which the process leads to the acceptance of the null hypothesis which also happens in the second year. The middle path is a realisation which does not lead to any signal all the way up to the end of the surveillance period (ten years). The numbers in the square brackets in Figure 6.1 represent the total number of events arising in the six month time interval and the number of events in the risk period. The nominal Type *I* and Type *II* errors were both set at 0.01.

6.8.2 Simulations based on the design values

Tables 6.2, 6.3, 6.4 below show results from the simulation with one week risk period, two weeks risk period, and one month risk period, for a range of relative incidences and different values of the baseline incidence λ, the mean number of cases per monitoring interval. We note some patterns in the results. The patterns noted can be described with respect to the risk periods whether short (Table 6.2 r=1/24), middling (Table 6.3, r=1/12) or long (Table 6.4 r=1/6). They can also be described in terms of the relative incidence to be detected whether it is low (RI=1.5, 2), middling (RI=3, 3.5), or large (RI=4, 5). The baseline incidence λ for the number of cases arising in each monitoring interval is also likely to have an effect as one would expect, for example a pattern emerges with respect to few cases (arising from Poisson mean of 5, and 10) and another emerges with respect to more cases (arising from Poisson mean of 20 and 50). There is a pattern with respect to the proportions out of 2000 ten year surveillance periods that crossed either boundaries or those that did not cross any boundary. Below, we describe the patterns observed.

6.8.3 Power and Type *II* error probabilities for design values.

The simulations were done with $\beta_2 = \beta_1$, so that in each case the true relative incidence was the relative incidence we wanted to detect (the design value). Throughout we set nominal Type *I* and Type *II* error probabilities at 0.01. The results are summarised in Figures 6.2, 6.3, and 6.4.

We measured the power by calculating the proportions of the 2000 ten year traces that gave a signal by crossing the upper boundary (in favour of the alternative hypothesis). We also calculated the proportions crossing the lower boundary (in favour of the null hypothesis). The proportions of traces crossing the upper and lower boundary are analogous to sensitivity and Type *II* error of the surveillance system.

Figure 6.2 shows that the power increased with the relative incidence, the baseline incidence of the number of cases arising in each six month monitoring interval and the risk period. For events arising with Poisson mean of 10 or more, the power is greater than 80% for relative incidences of 3 or more. For events arising with Poisson mean of 50 or more, the power is in excess of 95% for relative incidence of 2 or more. Figure 6.3 shows that the Type *II* error, that is, crossing the lower boundary in this case in favour of the null hypothesis given that the data arises from the distribution whose true relative incidence is the one we are trying to detect, is very low in all situations. In all cases the actual Type *II* error probability is much less than the design value of 0.01.

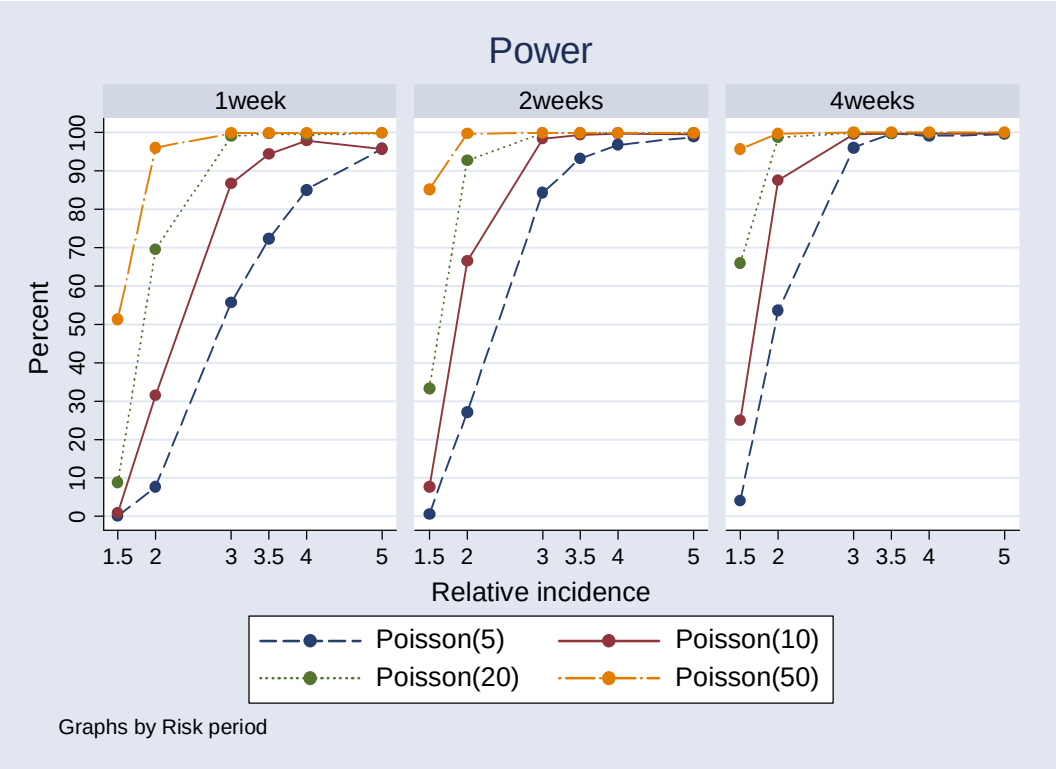


Figure 6.2. Power (percent) by relative incidence, risk period (1 week, 2 weeks, 4 weeks) and baseline incidence (Poisson mean of 5, 10, 20, 50).

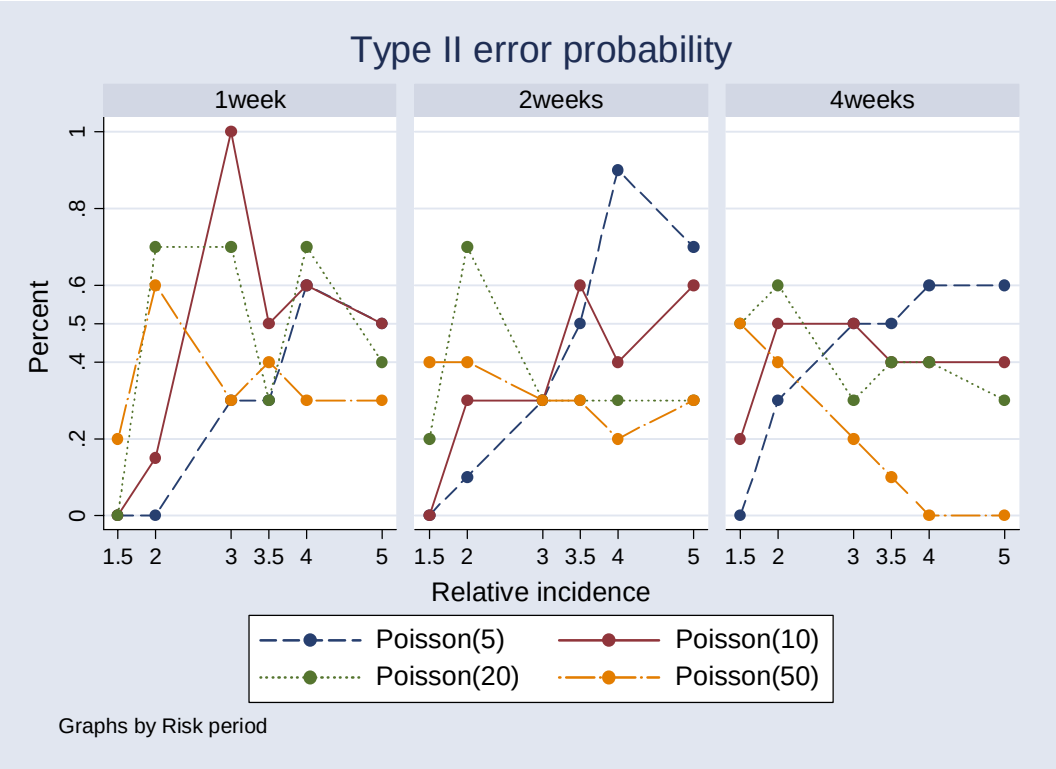


Figure 6.3. Type *II* error (percent) by relative incidence, risk period (1 week, 2 weeks, 4 weeks) and baseline incidence (Poisson mean of 5, 10, 20, 50).

Figure 6.4 shows a decreasing relationship between the proportions that did not cross either the upper or the lower boundary during the ten year surveillance period with relative incidence, risk period and baseline incidence. The proportions of the 2000 simulated values that did not cross the upper or lower boundary during the ten year surveillance period was very high (~100%) when trying to detect a small relative incidence, with small baseline incidence. For events with baseline incidence of 10 or more, relative risk of 3 or more and risk period of 2 weeks or more, the proportion not crossing the lower or upper boundary within 10 years is virtually zero.

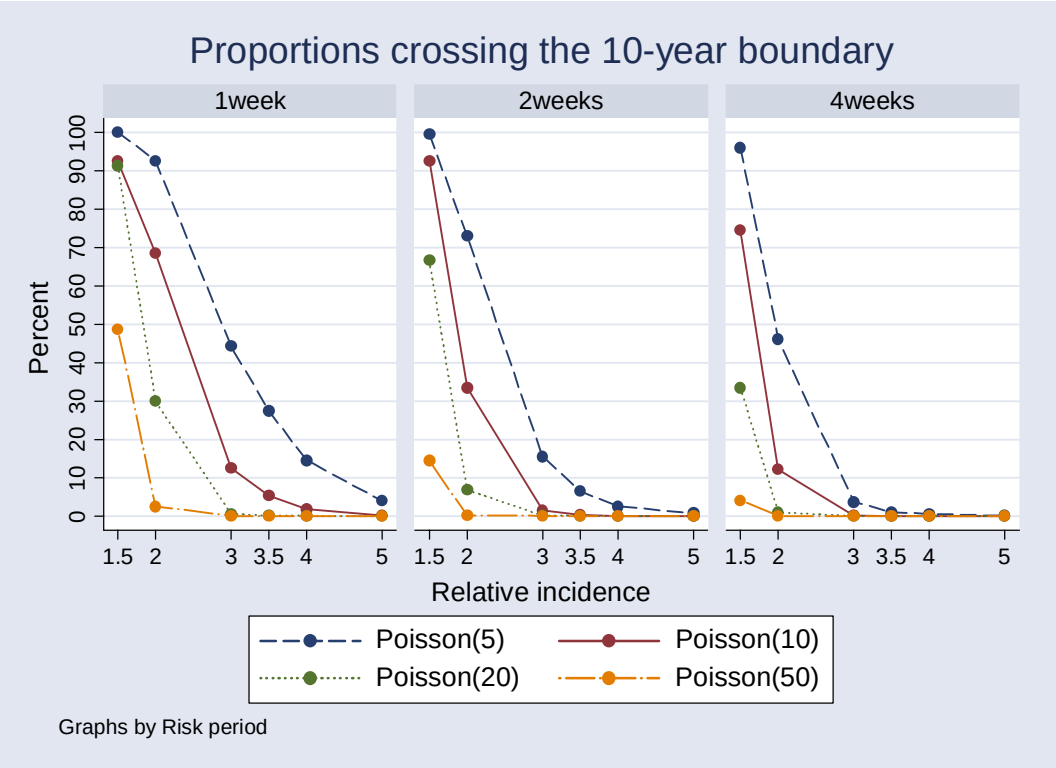


Figure 6.4. Proportions (percent) by relative incidence, risk period (1 week, 2 weeks, 4 weeks) and baseline incidence (Poisson mean of 5, 10, 20, 50).

So far we can see that the surveillance system is quite sensitive for detecting a relative incidence equal to the design value. Later, we investigate the performance of the surveillance system where data arises from a population whose true relative incidence is different from the design value.

Table 6.2. Results from 2000 simulations of 10 year surveillance period with six months monitoring interval and one week risk Period.

RI	Poisson Mean no. cases	Proportions of 2000 samples			Average year crossed			RI	Poisson Mean no. cases	Proportions of 2000 samples			Average year crossed	
		Pa	P b	Pc	Ya	Yb				Pa	P b	Pc	Ya	Yb
1.5	5	0.0000	0.0000	1.0000	-	-		3.5	5	0.7230	0.0030	0.274	5.40	5.00
	10	0.0065	0.0000	0.9245	8.18	-			10	0.9430	0.0050	0.0520	3.76	3.40
	20	0.0870	0.0000	0.9130	7.87	-			20	0.9960	0.0030	0.0010	2.30	3.11
	50	0.5130	0.0015	0.4855	5.31	4.03			50	0.9965	0.0035	0.0000	1.09	0.91
2.0	5	0.0755	0.0000	0.9245	7.42	-		4.0	5	0.8500	0.0055	0.1445	3.79 `	7.55
	10	0.3145	0.0015	0.6840	5.67	9.18			10	0.9765	0.0060	0.0175	3.20	3.70
	20	0.6950	0.0065	0.2985	5.63	7.70			20	0.9930	0.0070	0.0000	1.86	2.60
	50	0.9595	0.0060	0.0245	3.59	4.17			50	0.9970	0.0030	0.0000	0.90	0.66
3.0	5	0.5555	0.0030	0.4415	5.79	8.28		5.0	5	0.9565	0.0050	0.0385	3.79	5.64
	10	0.8655	0.0095	0.1245	4.65	5.78			10	0.9945	0.0045	0.0010	2.34	2.81
	20	0.9885	0.0070	0.0045	2.92	3.20			20	0.9965	0.0035	0.0000	1.30	1.36
	50	0.9970	0.0030	0.0000	1.96	2.65			50	0.9975	0.0025	0.0000	0.70	0.89

Pa, Pb, Pc=proportions of 2000 that crossed the upper boundary, lower boundary, vertical boundary, Ya, Yb=the average year when the upper, lower boundary was crossed conditional on having crossed the boundary. RI=relative incidence. $\alpha = \beta = 0.01$ Type I and Type II error probabilities.

Table 6.3. Results from 2000 simulations of 10 year surveillance period with six months monitoring interval and two weeks risk period

RI	Poisson Mean no. cases	Proportions of 2000 samples			Average year crossed			RI	Poisson Mean no. cases	Proportions of 2000 samples			Average year crossed	
		Pa	P b	Pc	Ya	Yb				Pa	P b	Pc	Ya	Yb
1.5	5	0.0050	0.0000	0.9950	7.60	-		3.5	5	0.9305	0.0050	0.0654	4.25	4.37
	10	0.0755	0.0000	0.9245	7.81	-			10	0.9915	0.0060	0.0025	2.59	3.14
	20	0.3330	0.0015	0.6655	7.00	8.76			20	0.9975	0.0025	0.0000	1.49	1.96
	50	0.8510	0.0040	0.1445	5.12	5.02			50	0.9970	0.0030	0.0000	0.76	1.02
2.0	5	0.2705	0.0010	0.7285	5.91	9.29		4.0	5	0.9665	0.0085	0.0250	3.54 `	4.75
	10	0.6635	0.0030	0.3335	5.78	5.95			10	0.9960	0.0040	0.0000	2.08	2.96
	20	0.9260	0.0065	0.0675	4.32	4.98			20	0.9975	0.0025	0.0000	1.23	1.51
	50	0.9955	0.0035	0.0010	2.20	2.26			50	0.9980	0.0020	0.0000	0.65	0.50
3.0	5	0.8420	0.0030	0.1550	4.98	4.85		5.0	5	0.9865	0.0070	0.0065	2.79	3.86
	10	0.9825	0.0025	0.0150	3.36	2.23			10	0.9945	0.0055	0.0000	1.58	1.31
	20	0.9970	0.0025	0.0005	1.90	2.53			20	0.9975	0.0025	0.0000	1.23	1.51
	50	0.9975	0.0025	0.0000	0.96	0.77			50	0.9975	0.0025	0.0000	0.58	0.60

Pa, Pb, Pc=proportions of 2000 that crossed the upper boundary, lower boundary, vertical boundary, Ya, Yb=the average year when the upper, lower boundary was crossed conditional on having crossed the boundary. RI=relative incidence. $\alpha = \beta = 0.01$ Type I and Type II error probabilities.

Table 6.4. Results from 2000 simulations of 10 year surveillance period with six months monitoring interval and one month risk Period.

RI	Poisson Mean no. cases	Proportion of 2000 samples Pa	P b	Pc	Average year crossed Ya	Yb		RI	Poisson Mean no. cases	Proportions of 2000 samples Pa	P b	Pc	Average year crossed Ya	Yb
1.5	5	0.0400	0.0000	0.9600	7.80	-		3.5	5	0.9955	0.0050	0.0095	3.28	2.62
	10	0.2500	0.0020	0.7440	5.87	7.06			10	0.9960	0.0040	0.0000	1.84	2.19
	20	0.6600	0.0050	0.3350	5.97	5.88			20	0.9960	0.0040	0.0000	1.08	1.02
	50	0.9555	0.0050	0.0395	3.81	5.26			50	0.9990	0.0010	0.0000	0.62	0.53
2.0	5	0.5364	0.0030	0.4610	5.24	8.90		4.0	5	0.9895	0.0055	0.0050	2.76 `	2.52
	10	0.8745	0.0045	0.1210	4.73	5.33			10	0.9960	0.0040	0.0000	1.55	2.18
	20	0.9860	0.0055	0.0085	3.08	2.99			20	0.9965	0.0035	0.0000	0.94	0.64
	50	0.9960	0.0040	0.0000	1.49	1.01			50	0.9995	0.0005	0.0000	0.57	0.50
3.0	5	0.9595	0.0045	0.0360	3.98	3.46		5.0	5	0.9940	0.0055	0.0005	2.10	1.91
	10	0.9940	0.0050	0.0010	2.36	1.78			10	0.9960	0.0040	0.0000	1.20	0.99
	20	0.9975	0.0025	0.0000	1.37	1.91			20	0.9975	0.0025	0.0000	0.74	0.73
	50	0.9985	0.0015	0.0000	0.71	0.50			50	0.9995	0.0005	0.0000	0.52	0.50

Pa, Pb, Pc=proportions of 2000 that crossed the upper boundary, lower boundary, vertical boundary, Ya, Yb=the average year when the upper, lower boundary was crossed conditional on having crossed the boundary. RI=relative incidence. $\alpha = \beta = 0.01$ Type I and Type II error probabilities

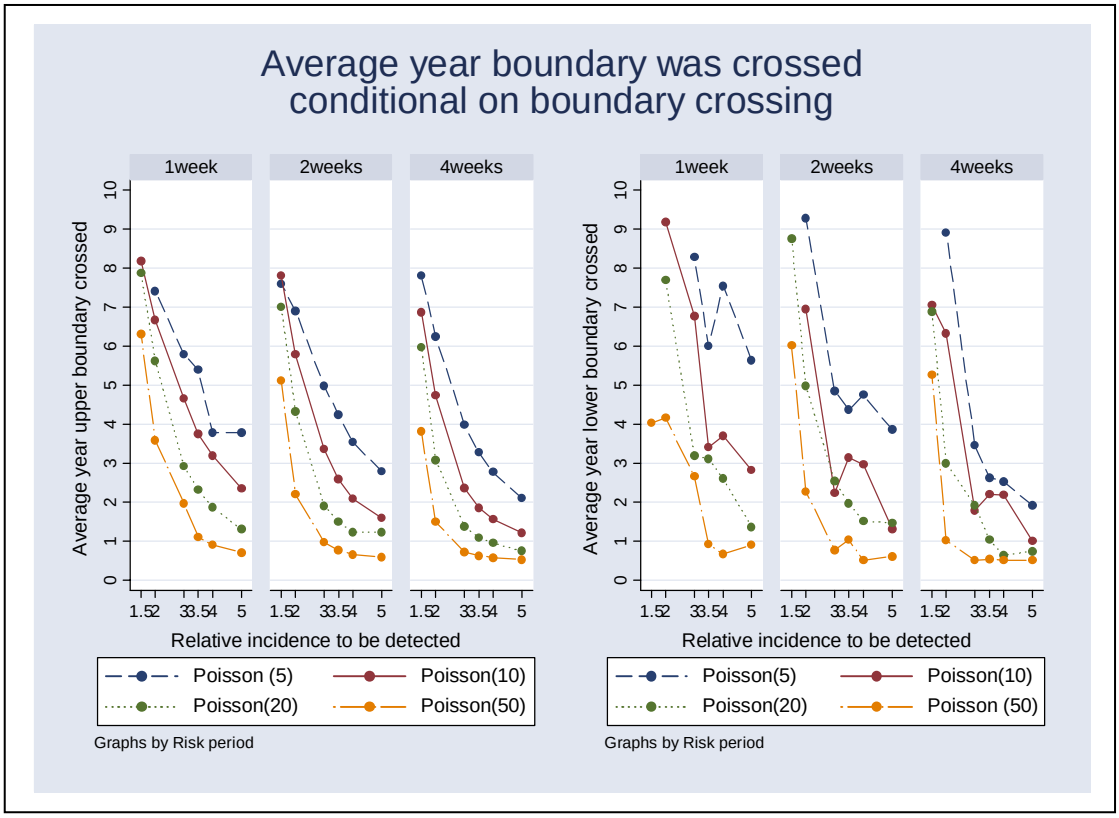


Figure 6.5. Effects of risk period, relative incidence and baseline incidence of the number of cases on the surveillance system from 2000 simulation of 10 year surveillance period with 6 months monitoring time interval when either boundaries were crossed.

6.8.4 Time to crossing a boundary

Figure 6.5 (left panel) shows the average time (years) to crossing the upper boundary, conditional on crossing it. Figure 6.5 (right panel) shows the corresponding results for lower boundary. Note that these graphs should be interpreted in conjunction with Figure 6.4.

In brief, figure 6.5 shows that, conditional on crossing, crossing either boundary occurs earlier for the following situations:

(a) As the risk period increases.

(b) As the relative incidence to be detected increases.

(c) As the baseline incidence increases (Poisson mean of 5, 20, 20, 50).

Most interest relates to Figure 6.5 (left part). For events with baseline incidence of 10 or more, and a relative incidence of 3 or more, detection occurs within 5 years on average in those detected. For events with an incidence of 50 or more, detection occurs within 2 years. This means that if there is a problem, then it is detected quickly.

6.8.5 Simulations for relative incidences other than the design value.

In the last three subsections, we have seen how the surveillance system performs when we simulated data using a relative incidence equal to the design value. We now present simulation results when the true relative incidence RI_2 associated with the event of interest differs from the design value of the SPRT (that is, the relative incidence RI the system is designed to detect).

We investigated similar situations as in those given in section 6.8.3 but here we present results for the 2 weeks risk period only. The design relative incidence RI is 1.5, 2, or 3 and the true relative incidence RI_2 is 1, 1.2, 1.5, 2, or 3.

Table 6.5. Results from 2000 simulations of 10 year surveillance period with six months monitoring interval and two weeks risk period.

RI RI2	Poisson Mean no. cases	Proportions of 2000 samples Pa	P b	Pc	Average year crossed Ya	Yb	RI RI2	Poisson Mean no. cases	Proportions of 2000 samples. Pa	P b	Pc	Average year crossed Ya	Yb
1.5 1.0							1.5 1.2						
	5	0.0005	0.0000	0.9995	9.50	-		5	0.0005	0.0000	0.9995	8.00	-
	10	0.0000	0.0210	0.9790	-	9.09		10	0.0050	0.0055	0.9895	7.44	9.39
	20	0.0035	0.2635	0.7330	5.72	7.80		20	0.0350	0.0445	0.9205	6.91	8.04
	50	0.0035	0.8265	0.1700	5.01	5.65		50	0.1345	0.2435	0.6220	5.99	6.62
1.5 2.0							1.5 3.0						
	5	0.0640	0.0000	0.9370	8.03	-		5	0.5080	0.0000	0.4920	7.45	-
	10	0.4840	0.0000	0.5160	7.25	-		10	0.9730	0.0000	0.0270	5.32	-
	20	0.9250	0.0000	0.0750	5.43	-		20	1.0000	0.0000	0.0000	2.89	-
	50	1.0000	0.0000	0.0000	2.57	-		50	1.0000	0.0000	0.0000	1.32	-
2.0 1.0							2.0 1.2						
	5	0.0010	0.1140	0.8850	6.75	8.46		5	0.0075	0.0535	0.9390	7.38	8.63
	10	0.0025	0.5670	0.4305	7.44	6.85		10	0.0230	0.2915	0.6855	6.33	7.24
	20	0.0035	0.9140	0.0825	3.97	4.96		20	0.0515	0.6295	0.3190	5.40	5.78
	50	0.0030	0.9950	0.0020	2.40	2.52		50	0.0550	0.9020	0.0430	3.10	3.74
2.0 1.5							2.0 3.0						
	5	0.0440	0.0100	0.9460	6.93	8.65		5	0.8015	0.0000	0.1990	6.05	-
	10	0.1570	0.0890	0.7540	6.27	7.27		10	0.9905	0.0000	0.0095	3.92	-
	20	0.3180	0.1815	0.5005	5.52	5.98		20	1.0000	0.0000	0.0000	2.10	-
	50	0.5710	0.2720	0.1570	4.35	4.69		50	1.0000	0.0000	0.0000	1.02	-
3.0 1.0							3.0 1.2						
	5	0.0040	0.7865	0.2095	3.55	5.97		5	0.0210	0.6045	0.3745	5.19	6.46
	10	0.0035	0.9775	0.0190	3.67	4.02		10	0.0190	0.8895	0.0915	3.44	4.70
	20	0.0045	0.9955	0.0000	1.84	2.26		20	0.0150	0.9775	0.0075	3.37	2.94
	50	0.0020	0.9980	0.0000	0.50	1.05		50	0.0115	0.9885	0.0000	1.31	1.44
3.0 1.5							3.0 2.0						
	5	0.0705	0.3230	0.6065	5.50	6.76		5	0.3215	0.0955	0.5830	5.65	6.54
	10	0.1090	0.6320	0.2590	4.79	5.39		10	0.5075	0.1785	0.3140	3.88	5.17
	20	0.1270	0.8110	0.0620	3.33	4.05		20	0.6670	0.2365	0.0965	3.89	4.11
	50	0.1070	0.8915	0.0015	4.35	4.69		50	0.7770	0.2160	0.0070	2.28	2.45

Pa, Pb, PC=proportions of 2000 that crossed the upper boundary, lower boundary, vertical boundary, Ya, Yb= the average year when the upper, lower boundary was crossed conditional on having crossed the boundary. RI=relative incidence to be detected (design value), RI2=true relative incidence in the population. $\alpha = \beta = 0.01$ Type I and Type II errors.

6.8.6 True relative incidence=1 ($RI_2 = 1$)

We first investigated the performance of the surveillance system when the true relative incidence is 1. The quantity of interest is the Type *I* error, namely the probability of crossing the upper boundary (within 10 years) when the true relative incidence is 1. We see from the rows of Table 6.5 (rows $RI_2 = 1$, column Pa) that the empirical Type *I* error is smaller than that used to set the boundaries (0.01). This means that in most situations when the true relative incidence is 1, the surveillance system will not give a false alarm. As the design relative incidence increases and as the base line incidence increases, the system becomes more sensitive by signalling very quickly in favour of the null hypothesis (high proportions crossing the lower boundary 'Pb', see table 6.5). Figure 6.6 below shows the probability of correctly concluding that the relative incidence is 1, as indicated by the percentages of traces crossing the lower boundary. The probability of correctly concluding that the relative incidence is 1 was greater than 90% for design values RI in excess of 2, and baseline incidence of 20 or more.

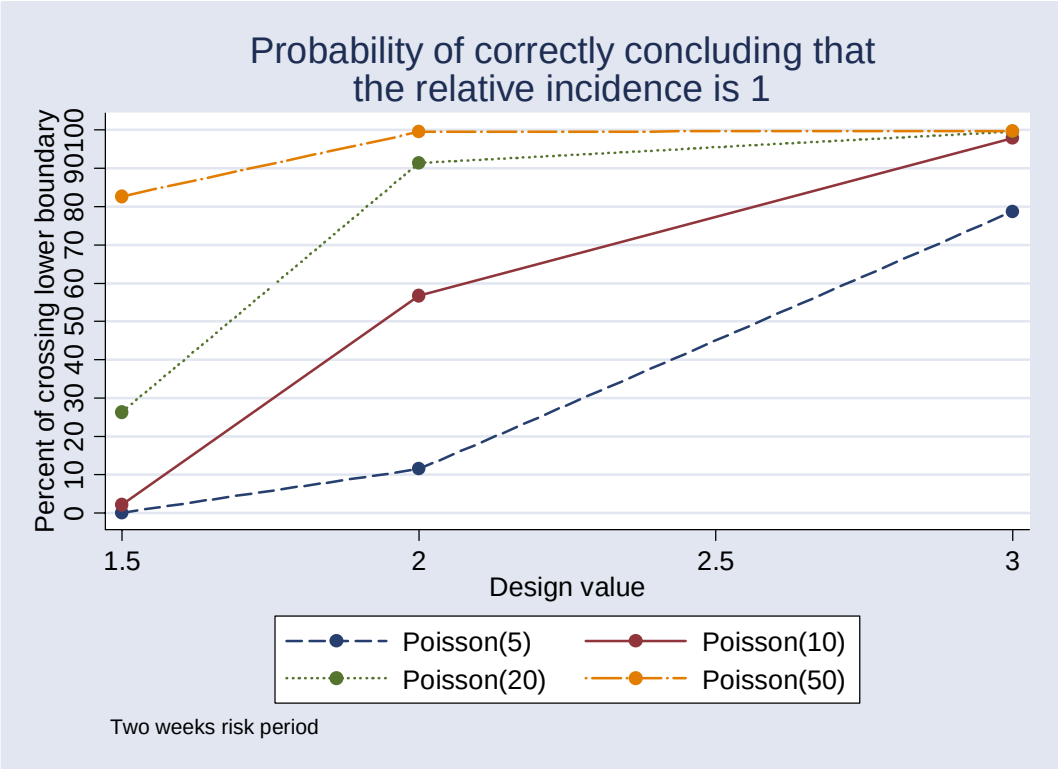


Figure 6.6. Probability of crossing lower boundary.

6.8.7 True relative incidence greater than 1, but not equal to the design value

Tables 6.5 also shows the results from the simulation with design value (RI) of the relative incidences of 1.5, 2, and 3 arising from simulations with true relative incidence (RI_2) 1.2, 1.5, 2, and 3. As one would expect, when the true relative incidence is larger than the design value (for example RI=1.5 and $RI_2 = 2$ or 3, $RI = 2$ and $RI_2 = 3$) the system very quickly signalled in favour of the alternative hypothesis. This is indicated by the decreasing average year when the upper boundary was crossed and by the high probabilities of crossing the upper boundary. In contrast, for the following pairs (RI, RI_2): (1.5, 1.2), (2, 1.2), (3, 1.2), (3, 1.5), the system signalled more frequently in favour of the null hypothesis than for the alternative. For the pairs (2, 1.5), (3, 2) the reverse was the case.

6.8.8 Average year to signal for simulations with different design values

In terms of how quickly the process was able to signal by way of either crossing the upper threshold or lower threshold, Table 6.5 show that the smaller the design value RI, the longer the process took to signal. The detection times (conditional on detection occurring) decrease as the true relative incidence RI_2 and the baseline incidence increase.

6.9 Conclusion

In this chapter, we have illustrated how the SPRT can be adapted for use with the self-controlled case series method. We have evaluated the performance of the SPRT by a simulation study of a possible surveillance system.

Overall we see from the simulation study that the performance of the surveillance system using the SPRT works broadly as intended. Ideally, we would like a system to be very quick to detect a true relative incidence greater than 1 and also if there is no problem we would like the process not to cross the upper boundary or ideally signal in favour of the null hypothesis. The simulation study showed that the system was able to achieve all these.

Using the SPRT with the self controlled cases series method has all the advantages of using the self-controlled case series method (see chapter 1). A further advantage compared to other methods [79-82] is the specification of Type *I* and Type *II* error probabilities which control against making wrong decisions. These error levels apply to the entire SPRT process, not to each specific monitoring time interval, and for this

reason the analysis takes account of multiple testing. The adjustment for multiple testing is not explicit as in the Bonferroni adjustment, but is incorporated into the SPRT in the way that the upper and lower boundaries are calculated. These boundaries preserve α and β until a final decision is reached as to whether the hypothesis should be accepted or rejected [92]. However, we note that the actual Type *I* and Type *II* error probabilities are lower than the nominal values $\alpha = \beta = 0.01$. Typically the actual values are less than half the nominal values. Thus the boundaries could be made less stringent without adversely affecting the actual error probabilities; nominal values $\alpha = \beta = 0.025$ might be appropriate to obtain actual values close to 0.01.

A possible limitation with the surveillance method is its inability to signal in real time since it is based on retrospective data. One possible solution to this would be to make the monitoring time interval shorter, though we have not investigated the implications of this other than through varying the baseline incidence λ. Finally, it should be stressed that such a surveillance system would rely on routinely collected data for signal detection. Such data have varying degrees of accuracy in diagnostic coding; it is for this reason that such a surveillance system should not be viewed as the final confirmatory epidemiologic investigation into potential vaccine adverse events. Such a surveillance system is also limited to conditions that develop relatively soon after vaccination, and would not be suitable for investigation of conditions with a longer induction period, for example an adverse outcome that manifests itself several years after exposure.

This chapter provides a ‘proof of principle’: the case series method can be used for focused surveillance using the SPRT. Further work is required to incorporate age effects, select optimal values of $\alpha = \beta$ and the best monitoring interval.

Chapter 7

Long-term surveillance using CUSUM charts with the self-controlled case series method

7.1 Introduction

In chapter six we explored various methods used in surveillance systems. We identified two possible methods that can be adapted for use with the self-controlled case series method. These were the sequential probability ratio test (SPRT) and the cumulative sum (CUSUM). We showed how the SPRT can be used with the case series method. We now show how the CUSUM can be used with this method. In chapter 6 the emphasis was on focused surveillance of a single vaccine and adverse event, as would be undertaken after licensure of a new vaccine. The situation we consider here is the second scenario described in section 6.2 of chapter 6, namely long-term surveillance of several vaccines or several adverse events. The presumption is that there is no problem, so that $\mathrm{RI}_2 = 1$ (where RI_2 is as defined in section 6.8.1 of chapter 6). The main differences with the earlier scenario are that there is no time limit (previously we had a vertical boundary at 10 years) and that we need to control the overall Type *I* error for several vaccines. We begin by a brief background of the genesis of the CUSUM charts in section 7.2, followed by some theory behind the CUSUM presented in section 7.3; a note on the control limit of the CUSUM is given in section 7.4. Section 7.5 describes the two sided tabular CUSUM. The application of the CUSUM to surveillance using the case series method is given in section 7.6, and

in section 7.7 we present results from a simulation study. We conclude the chapter by describing the overall findings in section 7.8.

7.2 Background on CUSUM

The CUSUM procedure is one of the most well-known monitoring methods for sequential data. There are two types of CUSUMs, the tabular (algorithmic, decision interval) CUSUM and the V-mask form. The tabular CUSUM was first introduced by Page [46]. It was developed from the Wald sequential test [45]. It was designed to detect changes in a process parameter of interest, for example in our case the relative incidence RI (where RI is as defined in section 6.8.1 of chapter 6). Later, Barnard [102] developed the V-mask form of the CUSUM. The idea behind the V-mask CUSUM was to enable combined detection of both an increase or a decrease of the parameter of interest. We restrict attention to the tabular CUSUMs which can easily be adapted for use with the self-controlled case series method.

The initial development of CUSUM by Page [46] was for use in industrial problems where monitoring of a production process is of interest. In these settings, the CUSUM charts have been shown to be ideally suited to detecting small persistent process changes[103]. Recently [86, 88, 92, 97, 104, 105], CUSUMs have been used in a medical context to monitor outbreaks of infectious disease or congenital malformations. Application of the CUSUM to monitoring surgical performance was first proposed by Williams et al [106]. The first application of a CUSUM chart to monitoring surgical performance is documented in De Leval et al [107] and Steiner et al [108] who considered the problem of monitoring outcomes in paediatrics cardiac

surgery. Rossi et al [109] used CUSUM charts to monitor respiratory and mortality in males in North Tuscany. Marshall et al [110] propose a CUSUM dealing with simultaneous surveillance of health outcomes over multiple units as well as time points.

7.3 The CUSUM

The CUSUM procedure involves plotting

$$Z_t = \max\left(0, Z_{t-1} + \Lambda_t\right), \quad t = 1, 2, 3, \ldots \qquad (7.1)$$

at the tth observation, where, as for the SPRT, $Z_0 = 0$ and Λ_t is the sample weight assigned to the tth subgroup as defined in an SPRT. For use with the self-controlled case series method, subgroups are a collection of cases taken from the surveillance system at fixed monitoring intervals. The CUSUM procedure differs from the SPRT because it has a holding barrier at zero rather than a lower absorbing barrier. The CUSUM sequentially tests the hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. The process is assumed to be in state H_0 as long as $Z_t < h$, and is deemed to have shifted to state H_1 if $Z_t \geq h$ at some time t. The constant h is called the control limit of the CUSUM. A CUSUM that exceeds the control limit is said to have 'signalled'. A signal means that the chart has accumulated enough evidence to conclude that the process (surveillance) parameter has changed. At this point, it is expected that monitoring will stop and remedial action will be taken. Notice that although individual scores (Λ_t) may be negative, the tabular CUSUM based on Z_t is restricted to values greater or equal to zero. This is mainly because the expression (7.1) is designed to detect an increase in

parameter θ_1. Later we will show how (7.1) can be rewritten to monitor decrease in parameter θ_1.

The hypothesis H_0 in the CUSUM can never be accepted, while H_1 will eventually be accepted with probability 1, thus $\alpha = 1$ and $\beta = 0$. Theoretically the CUSUM will always eventually signal, although the signal may be a false alarm. The run length of the CUSUM is defined as the time (or number of observations) required before the CUSUM first exceeds the control limit (i.e. signals). Good choices for the control limit h are based on the expected or average run length (ARL) of the CUSUM under H_0 and H_1. The ideal situation is to have a long ARL when the process is in state H_0 but a short ARL when the process has shifted to state H_1.

Whereas the performance of an SPRT is determined by its nominal error rates α and β, the efficiency of a CUSUM chart is quantified in terms of length of time before an alarm, false or true, is raised. The CUSUM's performance is assessed by the average run length to detection of an alarm. A useful review of some alternative measures that can be used to summarise the performance of Statistical Process Control (SPC) charts of which the CUSUM is one is provided by Frisén [111]. The most commonly used measure as reviewed by Frisén is the average run length. When the process is in state H_0, the average run length to detection is called the in-control ARL_0 and this is analogous to the Type *I* error of an SPRT, whereas the out-of-control ARL_1 is the average run length to detection when the process is in state H_1 which is analogous to 1 minus the Type *II* error (power) of an SPRT.

Brook et al [112] and also Grigg et al [92] showed that the distribution of in-control run lengths for a CUSUM scheme is approximately geometric, hence it possesses the memoryless property and because it is also discrete, it will usually remain close to zero. On the other hand, the out-of-control run length distribution is not geometric because the chart in this case will tend to move towards the out-of-control region rather than remaining at zero.

The CUSUM may be defined for weights Λ_t other than the log-likelihood ratio in contrast to the SPRT which is only defined with log-likelihood ratio weights. The log-likelihood ratio weights are the best to use in a CUSUM. Moustakides [113] showed that the log-likelihood ratio weights are optimal in the sense that, of all CUSUMs with the same ARL under the null hypothesis, the CUSUM with log-likelihood ratio weights has the shortest ARL under the alternative.

7.4 Determination of the limit h in a CUSUM

Choosing the control limit h should be based on the expected or average run length of the CUSUM under H_0 and H_1. Determining the average run length of a CUSUM is computationally intensive since it is based on all possible outcomes for a long series of observations of a monitoring process.

There are various ways of determining the ARL for a CUSUM. Some people use simulation, which is straight-forward but can be time consuming. Others have calculated the ARL using an integral equation approach [114]. In this approach, solutions are only possible via numerical methods. The method is only applicable to

charts with outcomes that follow a normal distribution. In fact, in some instances in this approach, the solution may not even be possible.

Steiner et al [74] have proposed an approach based on Markov chain methodology. In this method, properties of the run lengths distribution are used to determine the set of probabilities of moving from one point on a chart to another and then manipulating the resulting transition matrix [74]. Calculating the ARL using Markov chain methodology requires the state space to be discretised so that it is finite. Steiner et al do this by enlarging the weights and control limits by a factor (the multiplier) and rounding off to the nearest integer. Grigg et al [92] have argued that as the result of discretisation, the Markov chain methodology of calculating the ARL may induce some error, but the error settles very quickly as the process continues. In our case, we shall determine the average run length by simulation. We use this approach because it will allow us to explore the behaviour of the self-controlled case series CUSUM and it will enable us to explore ARLs for several CUSUMs.

7.5 Two-sided tabular CUSUM

The CUSUM described in the last few sections concentrated on observing a shift of one particular parameter of interest, say the relative incidence RI denoting an increase from the null value 1. In other circumstances one might also be interested in knowing whether a particular vaccine has developed some protective effect with respect to the adverse event resulting in a decrease of the relative incidence below 1. In such situations one could use a two-sided CUSUM. That is, one with the upper limit, denoting an increase (which in the case of the relative incidence represents a

deterioration) of the parameter from the expected, and the other, with the lower limit denoting a decrease (which represents protection) of the parameter from expected. Page [46] was the first to suggest a two sided CUSUM, that is, the combined use of two one-sided tabular CUSUMs, one to detect improvement and the other to detect deterioration. Two sided CUSUMs are now widely used [74, 92, 115] and calculations of the ARL are needed for both sides.

The CUSUM designed to detect the decrease in the parameter will accumulate negative values, hence the updating formula (7.1) can be modified as shown below:

$$Z_t = \min\left(0, Z_{t-1} - \Lambda_t\right), \quad t = 1, 2, 3, \ldots \qquad (7.2)$$

where $Z_0 = 0$ as before and Λ_t is still as defined in expression (6.9) of chapter 6. To enable plotting the CUSUM chart for the two sided on the same plot, the limit to detect a decrease in the parameter is usually assigned a negative value. We shall not explore the CUSUM based on a decrease in the parameter of interest since this is seldom of interest in a surveillance framework.

7.6 Use of the CUSUM for surveillance of adverse events

As outlined above, the CUSUM never results in 'acceptance' of the null hypothesis, and in this sense is well suited for long-term monitoring of established vaccines, the presumption being that such vaccines are safe. The aim of such monitoring could be to identify problems resulting from changes in vaccine production or delivery over time. Typically, one might expect several adverse events and several vaccines to be

monitored, thus increasing the chance of a false detection. For this reason, there is a danger that such a monitoring scheme will produce too many false warnings. Rather than ascribe precise detection limits, it is probably more sensible to use CUSUMs in a more informal manner, by plotting the updated values for the several conditions to be monitored, and inspecting them informally. In addition to a CUSUM signalling, two features might also be of interest:

a. Persistent increasing trends above baseline.
b. Persistent ranking in 'top' position of one CUSUM.

Either of these might suggest further, more formal, investigation, perhaps in the first instance using the SPRT, or by setting up a suitable epidemiological study. We shall concentrate on issues relating to a CUSUM signalling. In the next section, we investigate the performance of the self-controlled case series adapted CUSUM by determining the average run lengths in a simulation study.

7.7 Simulation study to evaluate the self-controlled case series CUSUM.

7.7.1 Simulation scenario

We considered two settings: surveillance of a single vaccine, and surveillance of several vaccines. The simulation study was carried out in a similar way as described in sections 6.6.1 and 6.8.1 of chapter 6. The main difference is that the lower and vertical boundaries were removed. We computed the average run length of the CUSUM for both in control and out of control processes. The other difference was that we looked at two approaches when simulating several vaccines. In the first

approach, when one vaccine signals we stop the whole process, and start again after resetting the CUSUM value of every vaccine to zero. In the second approach, when a signal is triggered, the CUSUM value for the signalling vaccine only is reset to zero. So as not to make the simulations too unwieldy, we used five vaccines to represent a situation corresponding to monitoring several vaccines. This represents a realistic choice in the light of childhood immunisation programmes. To determine the average run lengths, we simulated for long enough to be sure that the upper limit is eventually crossed. In practice we simulated for 100 years.

7.7.2 Average run length in control and out of control for one vaccine.

We begin by looking at the average run length for systems in control and out of control when monitoring a single vaccine. Finding the average run length in control is similar to investigating the Type *I* error in an SPRT (see section 6.8.6 of chapter 6). The parameters used in the simulation for the CUSUM were as follows.

Similar to the simulation in the SPRT, we used a monitoring interval of six months. We investigated various risk periods, but here we report only results for two weeks risk periods. Since in the present context the presumption is that the vaccine is safe, we are primarily interested in investigating relatively small changes in the relative incidence of adverse events. Accordingly we used design values for the relative incidence of 1.5, 2 and 3. We present values of 1, 2, 3, 4, for the control limits h because they gave ARLs that were realistic. The baseline incidences we investigated were same as in the SPRT, that is, Poisson mean of 5, 10, 20, and 50 per monitoring interval. So for one vaccine under investigation, we had a combination of 3 design

values, 4 control limits, 4 baseline incidences giving a total of 48 different simulations. The out-of-control data were simulated using design values, so $RI_2 = RI$.

In the case of a single vaccine, we simulated each scenario 2000 times and for each run of 2000 we found the average run length by calculating the average time at which the control limit was crossed. We also looked at the median time when the control limit was crossed, but only report the average because the distribution of crossing times was generally quite symmetric and the results based on means and medians were very similar.

Table 7.1 and Figure 7.1 below shows the average run lengths in years of a process in control and out of control for one vaccine. We can see from Figure 7.1 that the average run length in both situations decreases with decreasing control limit, increasing baseline incidence, and increasing design value. The decrease of the average run length with increasing design values may be explained as follows. The sample weight (6.9) in the CUSUM is $\Lambda_t = n_{.1t}\beta_1 - n_{..t}\log\left(1-r+e^{\beta_1}r\right)$. In this expression, for values of the relative incidence close to 1, for example relative incidences less than 5, and low values of r (here $r = \frac{2}{24}$) the values of $n_{.1t}\beta_1$ dominate the values of $n_{..t}\log\left(1-r+e^{\beta_1}r\right)$ hence as the relative incidence increases, the values of the CUSUM increases quickly such that it crosses the control limit sooner with larger design values (Figures 7.1). The average run length is shorter out of control than in control at each set of parameter values. The out of control ARL is at least 3 times the in control ARL at each control limit (see $\text{ratio} = \text{ARL}_0 / \text{ARL}_1$ in Table 7.1).

Table 7.1 Average run length for one vaccine in and out of control**.**

h	λ	ARL_0 for one vaccine in control. (years)			ARL_1 for one vaccine out of control. (years)			ratio $= \frac{ARL_0}{ARL_1}$		
		Design value			Design value			Design value		
		1.5	2	3	1.5	2	3	1.5	2	3
1	5	11.9	8.32	5.54	3.28	1.66	1.08	3.63	5.01	5.13
	10	7.48	6.05	4.27	2.05	1.20	0.83	3.65	5.04	5.15
	20	5.32	4.55	3.74	1.45	0.90	0.72	3.67	5.06	5.19
	50	4.63	3.66	3.13	0.98	0.72	0.60	4.72	5.08	5.22
2	5	31.4	21.9	13.0	8.58	4.35	1.98	3.66	5.03	6.57
	10	21.9	15.7	10.9	5.93	2.39	1.40	3.69	6.57	7.79
	20	17.0	10.8	7.65	3.60	1.62	0.98	4.72	6.67	7.81
	50	10.9	7.53	5.25	1.98	1.00	0.67	5.51	7.53	7.84
3	5	41.7	31.3	22.7	9.26	6.20	3.34	4.51	5.05	6.80
	10	35.3	28.4	19.7	7.21	4.63	2.51	4.90	6.14	7.85
	20	25.1	17.1	10.1	5.10	2.64	1.28	4.93	6.48	7.89
	50	18.6	11.4	9.67	3.35	1.43	0.75	5.55	7.97	12.9
4	5	45.8	36.1	28.9	9.99	7.10	4.24	4.58	5.09	6.82
	10	38.8	31.8	22.1	7.88	5.10	2.77	4.92	6.24	7.98
	20	29.5	21.8	15.7	5.92	3.35	1.46	4.98	6.51	10.8
	50	21.0	15.7	13.8	3.77	1.76	0.88	5.57	8.92	15.7

h is the control limit, λ is the baseline incidence with Poisson mean of 5, 10, 20 and 50.

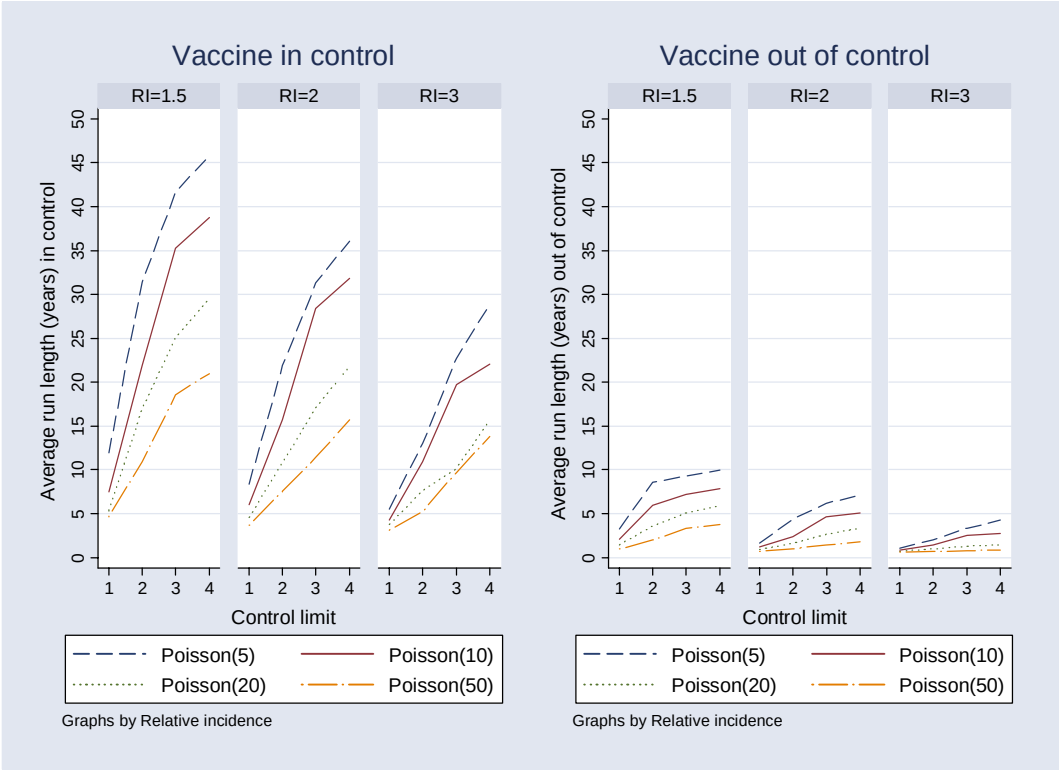


Figure 7.1 Average run lengths in and out of control for one vaccine

7.7.3 Average run length of a CUSUM in control and out of control for five vaccines: If one signals, correct it and reset all.

In the case of a surveillance system with several vaccines in which all CUSUMs are reset when one signal is triggered, we were interested in the average run length of the system as a whole. We investigated a system with 5 parallel CUSUMs. The overall in-control ARL is the average time to signal for any of the component CUSUMs. Thus it is the minimum of the five individual ARL_0s. We based each component CUSUM on the same parameters, and obtained 10 000 simulations. Then we calculated the minimum ARL_0s in groups of five. The overall ARL_0s is the average of the 2000 minima.

When monitoring five vaccines, it is unrealistic to expect all five vaccines to be out of control, hence we looked at a situation in which five vaccines are under surveillance and only one vaccine is out of control. The ARL_1 is the time to detection of the out of control vaccine, starting from 0. A possible realisation is given in Figure 7.2 below.

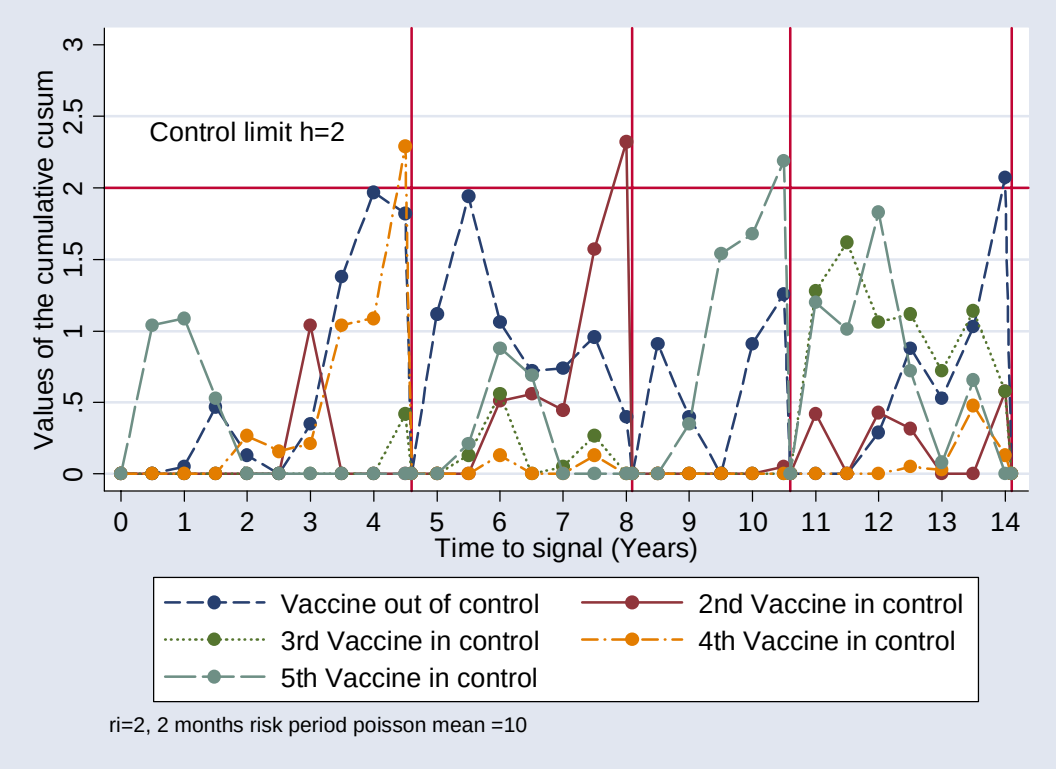


Figure 7.2 CUSUM chart monitoring possible 5 vaccines with one vaccine out of control.

Figure 7.2 shows a situation in which five vaccines are under surveillance and all but one are in control. The design value of the relative incidence is 2 and the expected baseline frequency of cases is 10 cases every six months. The control limit is 2. We can see that the system first signalled at about 4.5 years with the fourth vaccine (which was in control). The values of the CUSUM for all five vaccines were reset to zero and the process continued. After 8 years, the second vaccine signalled (again in control). Again all CUSUMs were reset to zero. After 10.5 years the fifth vaccine signalled; again all CUSUMs were reset. Finally in 14th year of observation the

vaccine out of control signalled and again all the vaccines were reset and the year of first signalling of the out-of-of control vaccine would have been recorded as 14. Note that this is somewhat unusual scenario: in most situations, it is the vaccine out of control that would signal first as can be seen by the closeness of its trace to the control limit when the in control vaccines signalled. This process was simulated 2000 times and the average time to signal was calculated. Table 7.2 below shows the results obtained from the simulation.

Overall Table 7.2 shows that the average run length for the system with one vaccine out of control was shorter than for the system with five vaccines in control. Figure 7.3 below illustrates that the average run length in either situation increased with increasing control limit, decreased with increasing baseline incidence and decreased with increasing design value. However, the values of the ratios ARL_0 / ARL_1 for five vaccines were consistently less than the corresponding values for a single vaccine (see Table 7.1 and Table 7.2).

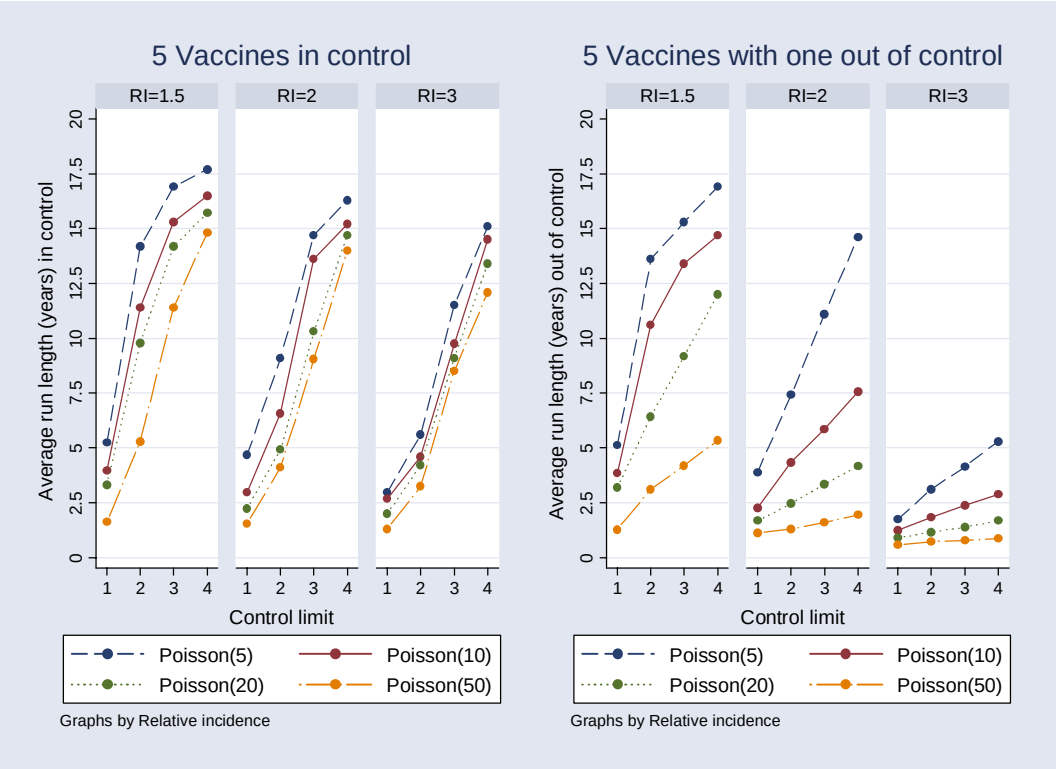


Figure 7.3 Average run length of a CUSUM in and out of control with 5 vaccines under surveillance.

Table 7.2 Average run length for in and out of control for 5 vaccines**.**

h	λ	ARL_0 for five vaccine in control (years)			ARL_1 for five vaccines and one out of control (years)			$\text{ratio} = \frac{ARL_0}{ARL_1}$		
		(Design value)			(Design value)			(Design value)		
		1.5	2	3	1.5	2	3	1.5	2	3
1	5	5.23	4.67	2.97	5.11	3.86	1.75	1.02	1.21	1.70
	10	3.95	2.96	2.66	3.85	2.26	1.23	1.03	1.31	2.16
	20	3.32	2.23	1.99	3.19	1.67	0.89	1.04	1.34	2.23
	50	1.62	1.54	1.31	1.27	1.12	0.58	1.28	1.37	2.25
2	5	14.2	9.09	5.61	13.6	7.42	3.12	1.04	1.23	1.80
	10	11.4	6.57	4.59	10.6	4.33	1.84	1.08	1.52	2.49
	20	9.79	4.93	4.22	6.44	2.46	1.16	1.52	2.00	3.64
	50	5.30	4.11	3.25	3.11	1.30	0.72	1.70	3.16	4.51
3	5	16.9	14.7	11.5	15.3	11.1	4.12	1.10	1.32	2.79
	10	15.3	13.6	9.76	13.4	5.84	2.38	1.14	2.33	**4.10**
	20	14.2	10.3	9.10	9.18	3.34	1.38	1.55	3.08	**6.59**
	50	11.4	9.05	8.51	4.19	1.59	0.78	2.72	5.69	**10.9**
4	5	17.7	16.3	15.1	16.9	14.6	5.30	1.05	1.12	2.85
	10	16.5	15.2	14.5	14.7	7.57	2.87	1.12	2.01	**5.05**
	20	15.7	14.7	13.4	12.0	4.16	1.68	1.31	3.53	**7.98**
	50	14.8	14.0	12.1	5.35	1.95	0.87	2.77	**7.18**	**13.9**

h is the control limit, λ is the baseline incidence with Poisson mean of 5, 10, 20 and 50. If one signals, correct it and reset all.

7.7.4 Average run length of a CUSUM in control and out of control for five vaccines: If one signals, correct and reset only the signalling vaccine.

We now present results from simulations in which when one vaccine of the five vaccines under surveillance signals, only the signalling vaccine is reset to zero. For this situation, we defined the system ARL_0 as the average time between signals in the long run when all vaccines are in control. So we left the process running until we had 2000 signals and calculated the average time interval between successive signals. Figure 7.4 below shows a possible realisation of such a surveillance system.

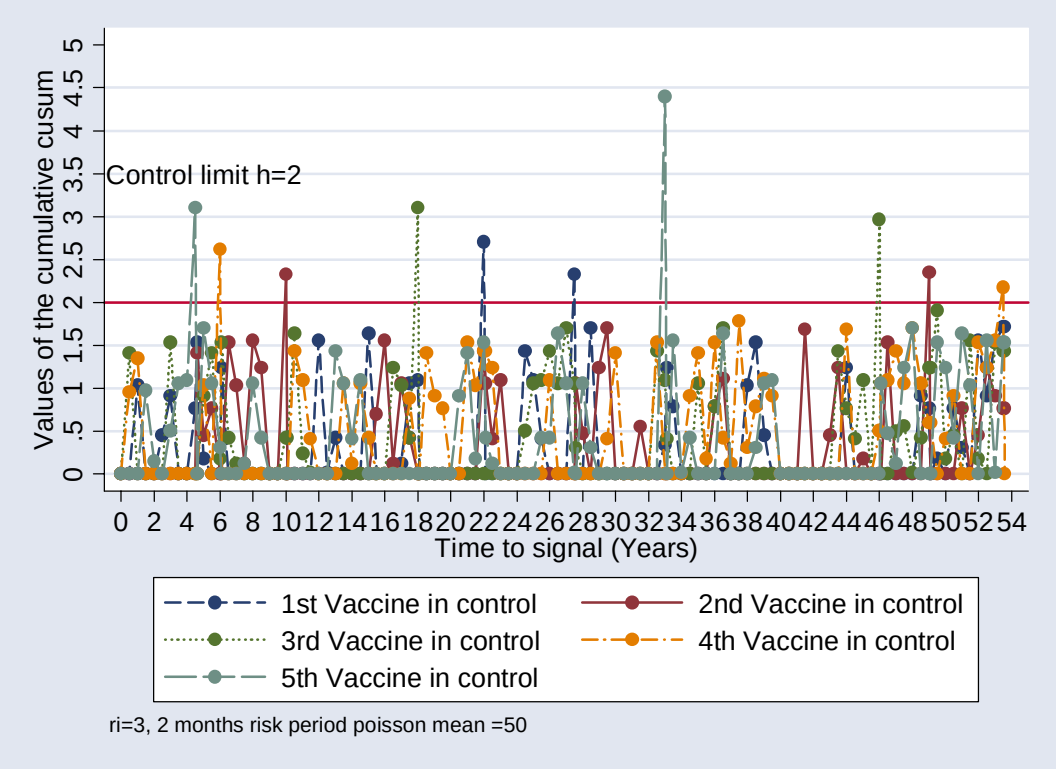


Figure 7.4 CUSUM chart monitoring possible 5 vaccines resetting only the vaccine out of control.

In the realisation shown in Figure 7.4, all five vaccines under surveillance are in control and the system is under surveillance for 54 years. The design value in this case was a relative incidence of 3, the baseline incidence was 50 cases in each monitoring

interval and the control limit was set at 2. We see that the fifth vaccine signalled first in the 5th year. This vaccine alone would then have been looked at, its CUSUM reset to zero and the surveillance would have continued. The second signal, by the fourth vaccine, was in the sixth year of observation and a similar action would have been taken for this vaccine alone, and so on. The ARL_0 of the system is therefore the average interval between signals when all vaccines are in control. The ARL_1 is the average time to detect an out-of-control vaccine starting from zero. Table 7.3 below shows the results from the simulations.

Table 7.3 Average run length for in and out of control for 5 vaccines.

h	λ	ARL_0 5 Vaccines in control. Average time interval between false signals (years)			ARL_1 5 vaccines with one out of control. Average time interval between signals (years)			ratio $= \frac{ARL_0}{ARL_1}$		
		(Design value)			(Design value)			(Design value)		
		1.5	2	3	1.5	2	3	1.5	2	3
1	5	4.35	2.36	1.78	3.26	1.67	1.10	1.33	1.41	1.62
	10	2.76	1.74	1.59	2.05	1.18	0.85	1.35	1.47	1.87
	20	2.03	1.51	1.49	1.46	0.92	0.70	1.39	1.64	2.13
	50	1.49	1.40	1.32	0.99	0.73	0.58	1.51	1.92	2.28
2	5	12.9	8.71	4.13	8.60	4.35	1.96	1.50	2.00	2.11
	10	10.9	5.92	4.39	6.00	2.38	1.38	1.82	2.49	3.18
	20	7.74	4.61	4.08	3.55	1.64	0.98	2.18	2.81	4.16
	50	4.62	3.84	2.92	1.98	1.04	0.65	2.33	3.69	4.49
3	5	15.1	13.4	9.60	9.20	6.20	3.37	1.64	2.16	2.85
	10	13.8	12.2	8.11	7.20	4.69	2.51	1.92	2.60	**3.23**
	20	11.8	9.10	7.90	5.12	2.66	1.27	2.30	3.42	**6.22**
	50	10.0	8.50	7.23	3.35	1.42	0.75	2.99	5.99	**9.64**
4	5	16.8	15.6	13.9	9.97	7.12	4.24	1.69	2.19	3.28
	10	15.4	14.3	10.3	7.90	5.14	2.76	1.95	2.78	**3.73**
	20	13.9	12.2	9.60	5.90	3.35	1.49	2.36	3.64	**6.44**
	50	11.9	10.7	8.26	3.77	1.72	0.85	3.16	**6.22**	**9.72**

h is the control limit, λ is the baseline incidence with Poisson mean of 5, 10, 20 and 50. If one signals, correct and reset only the signalling vaccine.

The average time intervals in control and out of control shows a similar trend as before, except that just resetting the problem vaccine means that the average

frequency of signals is greater. In particular, the ARL values are shorter than those found in Table 7.2. The other notable difference is that the $\frac{\text{ARL}_0}{\text{ARL}_1}$ ratios are generally larger when only the signalling vaccine is reset compared to the situation when all vaccines are reset (see Table 7.2, Tale 7.3 and the corresponding ratios $\frac{\text{ARL}_0}{\text{ARL}_1}$). There are exceptions, corresponding to high values of h and high design values, in which the ratios $\frac{\text{ARL}_0}{\text{ARL}_1}$ are smaller when only the signalling vaccine is reset. These values are indicated in bold in Tables 7.2 and 7.3.

7.8 Conclusion

In this chapter, we explored how the self-controlled case series sample weight $\Lambda_t = n_{.1t}\beta_1 - n_{..t}\log\left(1-r+e^{\beta_1}r\right)$ may be used in CUSUM charts. We have shown in different situations how such CUSUMs may be useful in long term surveillance of new vaccines. Unlike Marshall et al [110] who concentrated on false discovery rates (FDR) and successful discovery rates (SDR) in assessing the performance of the CUSUM, we assessed the performance of the case series CUSUM using average run lengths, suitably redefined for the surveillance of several vaccines. The method of Marshall et al [110] based on the false detection rate applies perhaps more appropriately to the surveillance of large number of units. However, it lacks the focus on detection times which is provided by the system ARL_0 and ARL_1 formulation. Our 'system' ARL_0 and ARL_1 are practically relevant parameters. The system ARL_0 measures the time interval between false signals when the system is in

control. A large value of ARL_0 is desirable. Surveillance of several vaccines drastically reduces the ARL_0 obtained for a single vaccine. In an effort to increase the ARL_0 we investigated a resetting scheme where all vaccines are reset, not just the signalling vaccine. This does indeed increase the ARL_0, but also affects the ARL_1.

The ARL_1 is an upper limit on the time interval between a problem occurring and when it is detected (it is an upper limit because the CUSUM will generally be greater than zero when the problem occurs). This time interval must be kept small; detection within 2 years might be a reasonable requirement. Thus using the ARL_1 may underestimate the speed at which problem vaccines are identified.

Choosing the control limit h should be based on the expected or average run length of the CUSUM under H_0 and H_1. Based on our simulations, if one vaccine is under observation, with a two week risk period, a six months observation period, and we are interested to detect a relative incidence of 3 based on the assumption that there are likely to be few cases arising in each six month monitoring interval (baseline incidence with mean of five), the simulation study (Table 7.1) shows that setting the control limit at 2 will have an average run length of 13.0 years when the system is in control and average run length of 1.98 years when it is out of control. These values appear reasonable.

However, the choice $h = 2$ with design value 3 is no longer adequate when 5 vaccines are involved. If all vaccines are reset upon signalling, then $ARL_0 = 5.61$ and $ARL_1 = 3.12$ (Table 7.2), whereas if just signalling vaccines are reset, then

$ARL_0 = 4.14$ and $ARL_1 = 1.96$ (Table 7.3). In both cases the ARL_0 values are rather too short. In this case, using $h = 3$ may be advisable.

In general when monitoring several vaccines, we note that the system ARL_0 is much shorter than the ARL_0 for a single vaccine. Also the ratios ARL_0 / ARL_1 were much close to 1 especially when trying to detect a smaller relative incidence (design value = 1.5) as apparent in Table 7.2. It is generally advisable to use large design values(2 and 3) and higher control limits. For most of the values we considered, it seems best to reset the signalling vaccine only rather than resetting all vaccines. However, this may not be the case for more frequent events (higher λ), in which case it could be best to reset all vaccines upon signalling.

It is not possible to suggest a single control limit to use when using the CUSUM based on the self-controlled case series method. This will depend on the risk period, the baseline incidences for the number of cases, the monitoring interval, the number of vaccines under observation, the way the surveillance is to be carried out especially when several vaccines are under observation, and the relative incidence to be detected. Our results show that a practical system may be possible, but requires careful choice of the parameters h and the design value, if we are to avoid swamping the system with false positive signals.

Note that even though the ARL is standard practice for evaluating the performance of a CUSUM [110], others [111, 116] have argued that the ARL is not ideal because the distribution of the run length may be skewed (though for the case series CUSUM, we found that median run lengths yielded similar results). Hence one has to be careful as

to what inferences can be drawn from an alarm signalled. Certainly an alarm signalled by a CUSUM does not constitute proof of causal association. However, among various methods for surveillance, it is argued [111, 113] that the ARL for a CUSUM is optimal to detect a change that occurs at the specific time.

Overall, a monitoring system using a combination of the SPRT and CUSUM based on the self-controlled case series method appears to be feasible, and could prove a very useful tool.

Chapter 8

Oral vaccines and intussusception

8.1 Introduction:

In this chapter, we describe a study conducted by GlaxoSmithKline (GSK) Biologicals to assess the incidence of intussusception in children less than 2 years of age in Latin America. We concentrate on an a-posteriori analysis of the data using the self-controlled case series method to assess whether intussusception is causally associated with oral polio vaccine (OPV). In section 8.2, we describe the background and rationale of the study. The objectives and study design are given in section 8.3, the study cohort and conduct of the study are described in section 8.4, the descriptive analysis is given in section 8.5, further statistical analysis is described in section 8.6, and conclusions are given in section 8.7.

8.2 Background and rationale

In August 1998, the first rotavirus vaccine, a tetravalent rhesus human reassortant rotavirus vaccine (RRV-TV) manufactured by Wyeth-Lederle (marketed as RotaShield^{TM}) was licensed in the United States of America (USA) and was recommended for routine immunisation of infants [68, 117]. The recommendations were suspended in July 1999 after the US Centres for Disease Control and Prevention (CDC) Adverse Events Reporting System identified 15 children who developed

intussusception after administration of the vaccine [69]. Additional epidemiological data lending support to a causal link was evident by October, 1999. Wyeth Lederle Vaccines and Pediatrics voluntarily withdrew RotaShield™ from the market, and CDC withdrew its recommendation for routine immunisation [118, 119]. Subsequent studies showed that RRV-TV is associated with increased risk of intussusception and the risk was shown to be highest between 3 to 7 days after the first vaccination dose [38, 120].

For most parents and paediatricians in the USA, the withdrawal of RotaShield™ was disappointing because it meant that the winter burden of severe rotavirus diarrhoea, which leads to an estimated 600 000 clinic visits, 50 000-60 000 hospital admissions, and 20-40 deaths, might continue for several years before another vaccine became available [119, 121]. The international medical community was disappointed because a vaccine that might have prevented 440 000 childhood deaths each year, or one in 20 deaths among children younger than 5 years, would remain a distant hope rather than an anticipated reality [122]. The rate of intussusception is not well known world wide, but a Cuban study [37] estimated a rate of about 45 per 100,000 live births which was similar to that found in the United States over a comparable period [123]. In some countries rates lower than those found in the United States have been observed [124].

Intussusception is a fairly uncommon type of acute intestinal obstruction. It occurs primarily in young children and is the most frequent cause of an acute abdominal emergency in the first 2 years of life. It rarely occurs in adults [125]. Most cases of intussusception are considered idiopathic. Children suffering from intussusception have problems, for example it is reported that about 5% to 10% of cases of

intussusception include an inverted appendiceal stump, Meckel's diverticulum (remnant of the embryonic yolk sac), intestinal polyps, lymphoid hyperplasia, hemangioma or lymphosarcoma. Twenty percent of the cases are noted to have upper respiratory tract infections [126]. Several other reports have indicated the presence of infectious agents in cases of intussusception, but the implications of these findings are unclear since most studies do not include a comparison group [127-133].

There is no clear evidence of association between natural rotavirus infection and intussusception [132, 134, 135]. Seasonality of rotavirus infection is well documented in the USA, and no seasonal variation in the occurrence of intussusception has been observed in most studies. Most studies have found that hospitalisation for intussusception was evenly distributed throughout the year while rotavirus disease peaked during the known season [136, 137]. However, in the Cuban study [37], cases showed a marked seasonality with cases peaking in December-May and low in June to May. The authors [37] argued that some of the observed seasonality was attributable to the seasonality of births in Cuba. In Nigeria, seasonality of intussusception has also been reported where most cases occur between October and April [138]. Generally, human rotavirus is not considered as a major etiological agent of intussusception in infants, though some studies have suggested that rotavirus and other viral epidemics may play a role in the aetiology of intussusception [139-141].

Following the withdrawal of the RotaShield^{TM} vaccine, there remained an urgent need for an effective vaccine because of the dramatic disease burden associated with rotavirus. The background incidence of intussusception in many countries is not known. The World Health Organisation (WHO) ethics workgroup [142]

recommended that such data be collected to help assess the risk/benefit ratio for use of rotavirus vaccines. Following the rotavirus vaccine experience, concerns arose as to whether oral polio vaccines might also be associated with intussusception. Two previous studies, both exploratory, had reported a significant increased risk of intussusception in the third or fourth weeks after doses of oral polio vaccine (OPV) administered at 4 months of age. Other studies have not confirmed these findings [36, 37, 143] albeit an increased risk after the third dose in the 14-27 days risk period was found in one case series study by Andrews et al [36] but this finding was thought to have been just a chance finding due to the number of risk periods examined. The authors [36] warned of the need for caution when looking at many risk periods without an *a priori* hypothesis . The Food and Drug Administration (FDA) and CDC had to tackle uncomfortable questions about how to detect such rare events before licensing future vaccines for rotavirus.

GlaxoSmithKline (GSK) Biologicals [144] have developed a new rotavirus vaccine based on human rotavirus strain and are currently performing clinical studies world-wide to evaluate this vaccine. Several studies are currently ongoing in Latin America (Argentina, Brazil, Chile, Costa Rica, Honduras, Mexico, Nicaragua, Panama, Peru, Dominican Republic, Columbia) to test GSK Biologicals' rotavirus vaccine in infants. In view of the recommendations to obtain intussusception data in different geographic settings, GSK Biologicals performed the GSK204 surveillance study described here. This was a hospital based multicentre study to assess the incidence of intussusception in children less than 2 years of age in Latin America.

8.3 Objectives and design of the study

The primary objective in the study was:

To estimate the incidence of intussusception in children less than 24 months of age in hospitals in Latin America.

For the purposes of the thesis, we undertook a-posteriori analyses following discussions with GSK as part of our collaboration. These analyses were not part of the initial objectives when the study was set up. But it was agreed that the data set collected may be suitable for the use of the self-controlled case series method to investigate the following question:

Is oral polio vaccine associated with an increase in intussusception in children less than 24 months of age?

This can be investigated using the self-controlled case series method taking OPV vaccination as the exposure. OPV has already been investigated in the UK for evidence of causal association with intussusception (Andrews et al [36]) and in Cuba [37]. The purpose of these a-posteriori analyses was to identify if there were any causal agents of intussusception other than rotavirus vaccine as found in the US [69].

8.4 The study

The study was designed as a hospital-based, multi-centre study. It was designed to enrol all cases of intussusception (definite, probable, possible or suspected) from children who received care at participating hospitals and listed on a Screening Sheet. Subjects were enrolled during a period of at least one year beginning at study start. For the enrolled subjects, the participation in the study consisted of an interview of the subjects' parents. The study was designed to be a self-contained study and the duration of the study was at least one year. Collection of data was by using hard copy Case Report Form (CRF).

8.4.1 Study cohort and conduct

The target population for enrolment was all subjects seen on an in or out-patient basis with confirmed diagnosis of intussusception during a one year period beginning at study start. All intussusception cases (definite, probable, possible, and suspected) seen at the participating hospitals were included in the study if they fulfilled the eligibility criteria. Only subjects whom the investigator believed had met the requirements of the protocol [144] were enrolled in the study. It was decided that for the purposes of the present analysis, only definite cases were to be analysed.

The inclusion criterion for the cases was as follows:

- A male or female infant aged less than twenty four months at the time of diagnosis of intussusception (patients became ineligible on the day of their second birthday).
- Subject was diagnosed with definite (radiographically, surgically or by post-mortem examination), probable, possible or suspected intussusception during the period of one year beginning at study start.
- He or she did not have a radiographically or surgically confirmed case of intussusception prior to the current episode.
- Written informed consent was obtained from the parent or guardian of the subject.

The study was conducted according to Good Clinical Practice, the Declaration of Helsinki (Protocol [144] Appendix AI) and the International Guidelines for Ethical Review of Epidemiological Studies (Protocol [144] Appendix AII) and logical rules and regulations of each participant country. The study was conducted in eleven Latin and Central American countries (Argentina, Brazil, Chile, Costa Rica, Honduras, Mexico, Nicaragua, Panama, Peru, Dominican Republic, and Columbia) between December 2002 and May 2005.

8.4.2 Case finding

Children admitted to or cared for at participating sites for definite, probable, possible, suspected intussusceptions were identified by daily reviews of admission logs, computerised hospital admission records, emergency department records, surgical records and radiology logs. Patients complaining of symptoms of intussusception

usually arrived at the admission and entry (A&E) department or outpatient paediatric clinics in the participating hospital. On preliminary query diagnosis by the paediatricians or A&E medical officers, the patients were admitted into the hospital and sent to Diagnostic Imaging for an ultrasound scan. On confirmation of intussusception by ultrasound, air enema (usually performed compared to barium) was carried out to confirm the diagnosis and reduce the intussusception. If attempts at reduction failed, the patient was sent for surgical reduction. If there were perforations or necrosis, then resections were carried out. Patients were then sent to the ward to recover. Data from each case were then keyed into the hospital computer under the ICD code for intussusception. Written informed consent was sought from child's parent or guardian if the child met the eligibility criteria.

8.4.3 Data collection

Data regarding the episode of intussusception including vaccination history, clinical symptoms noted on admission, diagnostic procedures, surgical and radiographic procedures performed, microbiology results and methods and outcome of admission were collected from hospital records, physician records, and vaccination booklets of all eligible subjects as well as interviews with parents or guardian.

8.5 Descriptive analyses of the GSK204 data

Overall, there were 531 cases in the GSK204 data set. Of these, 495 received polio vaccine, of which 492 had oral polio vaccine and 3 received injected vaccine. As

mentioned above, only definite cases that had oral polio vaccine were considered for analysis and these were 456. The following analyses are based on these 456 cases.

8.5.1 Distribution of cases by country

Table 8.1 below shows the number of cases per country and when the study started in each country.

Table 8.1 Distribution of cases by country

Country	Study Start	Study End	Number of cases
Argentina	02-Mar-03	02-Mar-05	40
Brazil	21-Mar-03	21-Dec-05	16
Chile	27-Jan-03	31-Jan-05	55
Costa Rica	17-Jan-02	31-Dec-03	24
Honduras	27-Jan-03	27-Jan-05	36
Mexico	06-Jan-03	20-Jan-05	120
Nicaragua	03-Mar-03	03-Mar-05	9
Panama	10-Jan-03	10-Jan-05	54
Peru	30-Sep-03	30-Sep-04	39
Dominican Republic	20-Jan-03	20-Jan-05	26
Columbia	02-May-03	02-May-05	37

There is much variation in the number of cases across the different countries. Mexico, Chile, and Panama seem to have had more cases than the other countries and Nicaragua had the fewest number of cases. The numbers of cases vary enormously, perhaps in part due to differences in case ascertainment.

8.5.2 Sex and age at diagnosis

Overall there were more boys (61%) than girls. A similar finding was made in the UK by Andrews et al [36]. Table 8.2 below summarises age at diagnosis in days by gender. The mean age at diagnosis did not vary substantially by gender, other than some indication that girls on average (mean=231.1 days) were slightly older at diagnosis compared to boys (mean=224 days). Age at diagnosis ranged from a minimum of 65 days to a maximum of 660 days for girls and from 66 days to 704 days for boys. Overall the distribution of age at diagnosis was positively skewed ranging from 65 days to 704 days with a mean age at diagnosis of 226.8 days and a median age of 196.5 days. Figure 8.1 below shows the distribution of age at diagnosis. The graph shows that most diagnoses were made between 100 days and 275 days.

Table 8.2 Distribution of age at diagnosis by gender

	Age at diagnosis days				
	Median	Mean	Std Dev	Minimum	Maximum
Female (176)	202.5	231.1	115.1	65.0	660.0
Male (280)	196.0	224.0	105.5	66.0	704.0

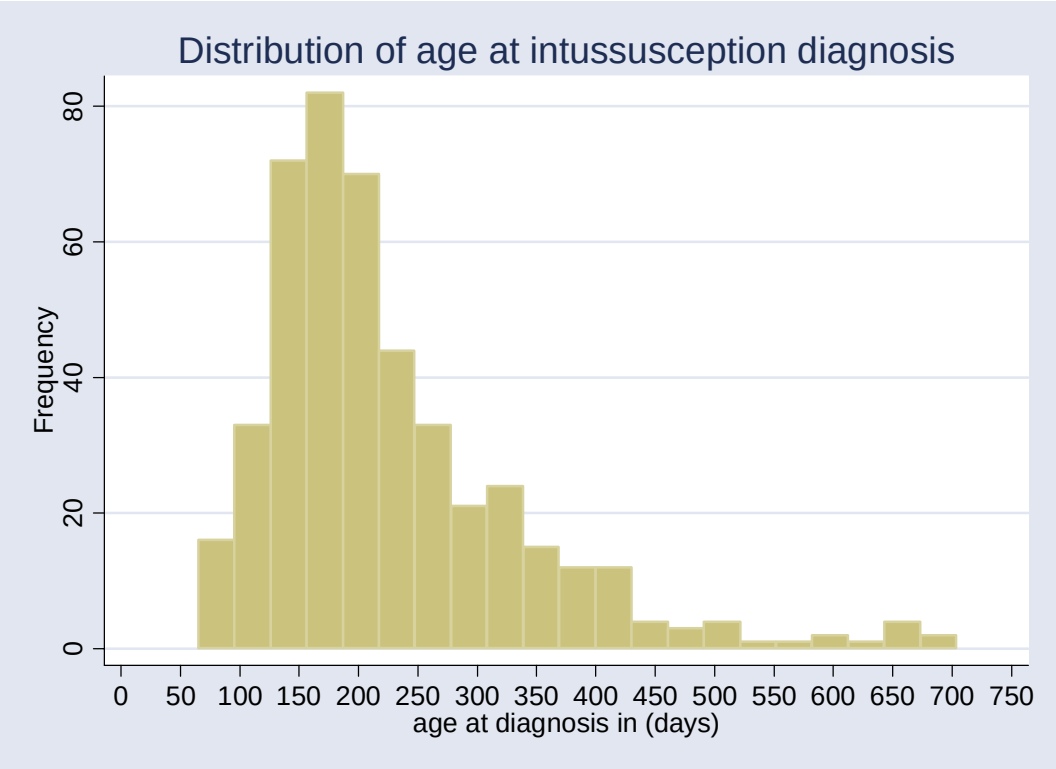


Figure 8.1. Distribution of age at diagnosis

8.5.3 Number of doses of OPV received

Children received a maximum of 5 doses of OPV. There was substantial variation in the number of doses of OPV received by different individuals. Most children received the first dose, the second dose and the third dose. A few children received a fourth, and only 12 had a fifth dose. Table 8.3 below shows the number of doses received by the children in the data set.

Table 8.3 Distribution of number of individuals who received OPV doses in the 204 data set

	Dose1	Dose2	Dose3	Dose4	Dose5	Dose12	Dose123	Dose1234	Dose12345
Yes	426	358	261	86	12	357	257	86	12
No	30	98	195	370	444	99	199	370	444
Total	456	456	456	456	456	456	456	456	456

Dose1 means number of individuals who received first dose, Dose2 means those who received second dose etc, and Dose12 means those who received dose1 and dose 2, Dose123 those who received dose1, dose2 and dose3 etc.

As one would expect, age at vaccination increases with the dose given. Table 8.4 below shows this relationship. Further, Figures 8.2 and 8.3 show the distribution of age at vaccination for each dose.

Table 8.4 Distribution of age at vaccination

Age at vaccination days				
Dose of OPV (Numbers)	Mean days	Std Dev days	Minimum days	Maximum days
First Dose (426)	41.1	36.9	0.0	236.0
Second Dose (358)	102.4	35.1	31.0	206.0
Third Dose (261)	161.0	42.8	57.0	387.0
Fourth Dose (86)	215.3	90.7	118.0	585.0
Fifth Dose (12)	314.4	153.0	179.0	619.0

The age distribution for doses 1, 2 and 3 are markedly bimodal, possibly reflecting different vaccination practices in different countries. Table 8.5 below shows the distribution of the interval between vaccination and diagnosis of intussusception at each dose. The interval between receipt of OPV and diagnosis of intussusception ranged from 3 days to 667 days for the first dose, for the second dose, the interval ranged from the day of vaccination to 602 days, for the third dose, six cases were diagnosed before receiving the third dose, and three cases were diagnosed before receiving the fourth dose, and the interval between diagnosis and receipt of 5^{th} dose ranged from 8 days to 193 days. This distribution reflects the way vaccination histories were collected, namely retrospectively from date of event (other than in the 9 cases with information on post-event vaccination).

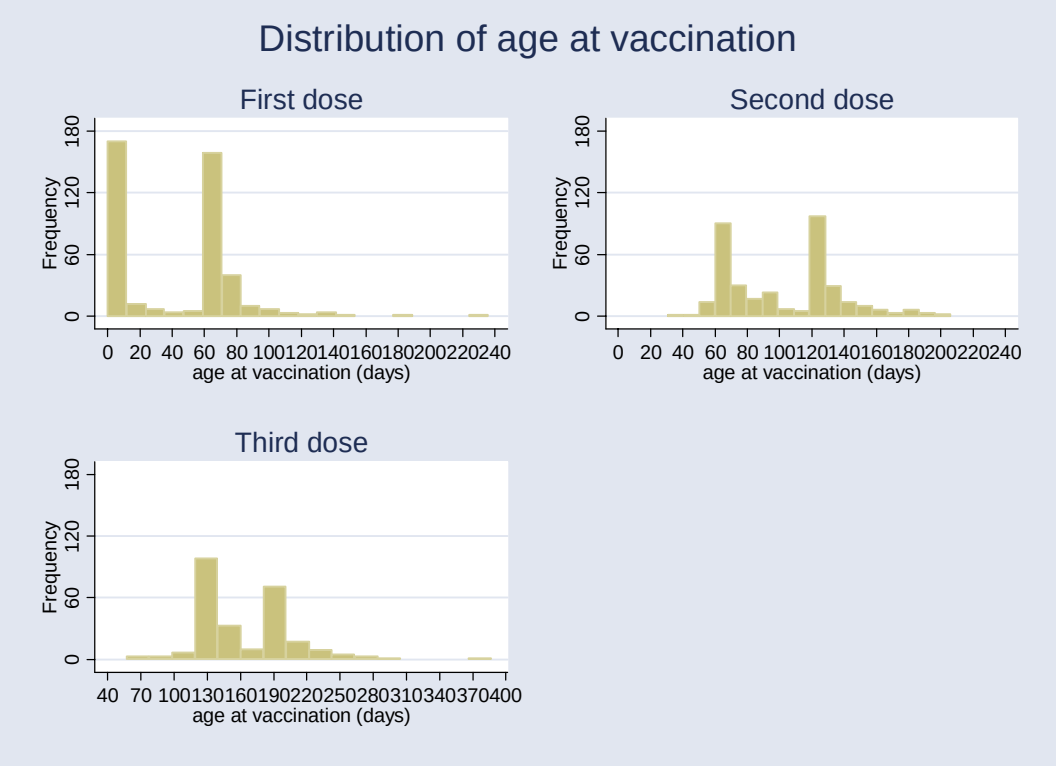


Figure 8.2 Distribution of age at vaccination

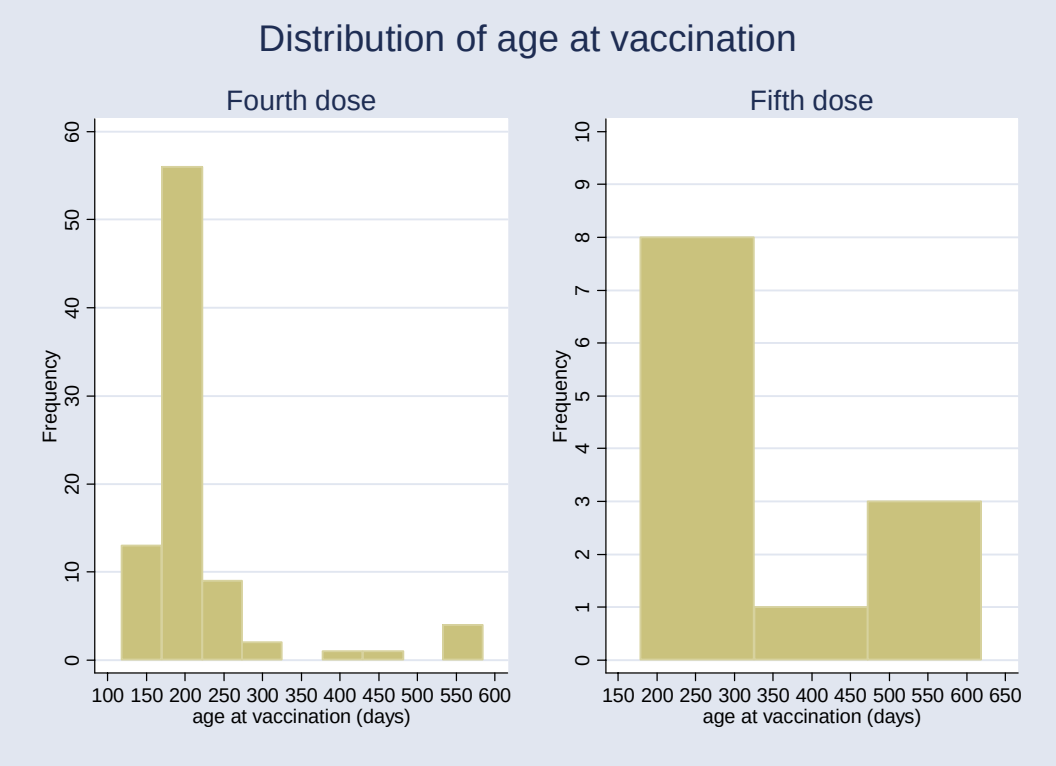


Figure 8.3 Distribution of age at vaccination continued

Table 8.5 Distribution of interval between vaccination and diagnosis of intussusception.

Interval between vaccination and diagnosis at each dose				
Dose of OPV (Numbers)	Mean days	Std Dev days	Minimum days	Maximum days
First Dose (426)	184.7	115.1	3.0	667.0
Second Dose (358)	139.4	111.6	0.0	602.0
Third Dose (261)	109.7	109.1	-48.0	545.0
Fourth Dose (86)	105.3	103.8	-54.0	482.0
Fifth Dose (12)	93.4	66.7	8.0	193.0

8.6 Statistical analysis

To assess the association between OPV and intussusception, the self-controlled case series method was used. We used risk periods of 31 days (0-30) after vaccination. We also split this into two risk periods, 0-15 and 16-30 days. These risk periods were chosen so as to compare with other studies[36, 37, 143]. For each analysis, we adjusted for age which was grouped into 24 different age categories of about 30 days each, except for the analysis of dose 5. Age was grouped in this way so as to take account of age in each month over the 2 years of the maximum possible observation period of each individual. For the analysis of dose 5, it was not sensible to have 24 different age groups as there were only 12 cases. To avoid unbounded estimates due to not having cases in some age groups, we grouped age in 5 longer age-groups as follows: 0-150 days, 150-300 days, 300-450 days, 450-600 days and 600-750 days.

One requirement of the self-controlled case series method is that the probability of exposure should not be affected by the occurrence of an outcome event [7]. In the

GSK204 study, the histories of the exposure was recorded when or soon after the cases were entered in the study. There was no follow-up after the outcome event and hence no information was collected on exposure after entering the study. In such situations when using the self-controlled case series method for single exposures, it is recommended to define the observation period starting from the exposure up to the end of the study (Farrington [2]). However, this study involved five different doses of vaccines, giving exposures at five different time points. The self-controlled case series method for censoring events has so far been used with one exposure. When there are several doses, one has to analyse the data starting with the latest dose first. If there is no significant association for this dose then one can analyse the previous dose and proceed iteratively in this way. The approach is necessary to avoid bias from unobserved exposures after the outcome events. Only if such later doses are not associated with the outcome can the current dose be evaluated in an unbiased way. Thus, we started by analysing data at dose 5, and only if we got a non significant result did we analyse dose 4, moving down the doses in this way. Thus, for the analysis of dose k, the observation period is $[v_k, b)$ where $v_k =$ age at dose k and $b =$ age at end of study, and we assume no effect for doses $k+1, k+2, ..., 5$. Hence the fact that later doses may be unobserved is immaterial.

Table 8.6 below shows the relative incidences obtained in each analysis for each dose and for each risk period. Included in the table is the number of events in each risk period at each dose. In all situations the relative incidence is not significantly different from 1, indicating no association between vaccination and OPV given at any dose, though the effect of dose 5 is very poorly estimated as there are so few cases.

Table 8.6 Relative incidence and number of events in risk periods after vaccination for each dose and 95% confidence intervals

Dose	0-15 days RI (95% CI)	No. of events	16-30 days RI (95% CI)	No. of events	0-30 days RI (95% CI)	No. of events
Opv5	2.79 (0.25, 30.9)	1	5.22 (0.73, 37.4)	1	4.04 (0.67, 24.4)	2
Opv4	1.01 (0.38, 2.72)	11	0.78 (0.30, 2.06)	7	0.88 (0.38, 2.04)	18
Opv3	0.71 (0.41, 1.24)	22	0.92 (0.58, 1.46)	28	0.84 (0.55, 1.28)	50
Opv2	0.86 (0.45, 1.61)	14	1.39 (0.89, 2.17)	28	1.19 (0.79, 1.80)	42
Opv1	1.34 (0.55, 3.24)	8	0.74 (0.30, 1.85)	7	0.97 (0.48, 1.95)	15

Opv1, Opv2, etc =Oral polio vaccine given at first dose, second dose etc, RI=relative incidence.

8.7 Conclusions

The aim of the analysis was to investigate whether incidence of intussusception in children less than 24 months of age in hospitals involved in the GSK204 study was associated with oral polio vaccine (OPV). The results shown in Table 8.6 do not support the hypothesis that OPV is causally related to intussusception. There is no evidence of causal relationship at any dose. With the exception of dose 5 the point estimates are generally close to unity, with narrow confidence intervals, indicating that the analyses have good power. For the dose 5 analysis, there were 12 cases, with two cases in the 30 day risk period. In this case, the confidence intervals are very wide as one would expect due to lack of power.

One of the advantages of the self-controlled case series method is that it implicitly controls for fixed factors, hence in these analyses, fixed factors such as social and economic factors, country, sex, and any bias due to individual level confounding, for example confounding due to vaccination and unmeasured risk factors for intussusception have been taken into account.

The rigour in the way the study was conducted and the way the cases were ascertained gives us confidence in these findings. Further our analyses were restricted to definite cases only. In addition, the cases of intussusception were ascertained independently of any perceived link of OPV with intussusception since the primary objective of the study was to assess the incidence of intussusception in children less than 24 months of age in hospitals involved in the study. The vaccination history of various vaccines in each case was then recorded without any focus on a particular vaccine, using data from vaccination booklets.

Our earlier findings relating to the properties of the self-controlled case series method in chapters 2 and 3, suggest that the self-controlled case series method gives effectively unbiased results when the number of cases is at least 20 for a relative incidence greater than one. It is only seriously biased for number of cases of about 10 if the ratio of the risk period to observation period takes extreme values. In this study, the ratio of risk period to observation period was approximately 0.04.The number of cases for each dose was all above 50 except for dose 5. Hence we expect estimates at each of these doses other than dose 5 not to be substantially biased.

In chapters 4 and 5 we saw that for a ratio of the risk period of about 0.05, with 100 or more cases (Table 5.2) one has power of 80% or more to detect a relative incidence of at least 3. In this data set, we had well over 100 cases for doses 1 to 3 hence the power appears adequate. The power for dose 5, however, is inadequate.

A limitation of the study was the censoring of post-event vaccination histories. This would have resulted in some difficulties completing the analysis of all doses had a

significant result been obtained at some point. But this was not an issue in these data as there was no significant result at any dose. A further problem we would have had to consider is the issue of multiple testing. Again here as there was no significant result, we did not need to worry unduly about this. Another limitation is the fact that there were only 12 cases with dose 5. This resulted in low power and imprecise confidence intervals for this dose. A possible solution to this is to obtain bootstrap confidence intervals and possibly using multiple imputations for the missing data so as to reanalyse the data with similar number of cases as those at first dose.

The study was undertaken with the primary objective of providing baseline information for the surveillance of a new rotavirus vaccine. Such surveillance could be undertaken using the methods described in chapters 6 and 7. Specifically, as surveillance would be focussed on a single new vaccine, the SPRT approach described in chapter 6 would be most appropriate.

Chapter 9

Conclusions

The first issue we considered in this thesis was to explore some further statistical properties of the self-controlled case series method, these were explored in chapter two. To this end we derived expressions to second order for the asymptotic bias and variance of the estimator of log relative incidence in a simplified setting. These enabled us to understand in qualitative terms the impact of quantities such as the length of the risk period and the relative incidence on the accuracy and precision of the estimates. We studied these effects graphically to examine how the bias, variance, and asymptotic mean square error vary with the ratio of the risk period to the observation period, and how they vary with the relative incidence at fixed sample sizes.

The main finding is that asymptotic bias and variance (and hence AMSE) are smallest when the expected number of events within the risk period and outside the risk period are equal. The greater the difference between these two expected frequencies, the greater the bias and variance. The asymptotic second order expression suggest that there is little bias with sample sizes in excess of 20 for the types of scenarios we might expect to encounter in practice.

All in all, the self-controlled case series model seems to perform very well. Asymptotically, the estimates obtained are not biased. The relative incidence

estimator is biased when trying to detect a relative incidence less than one (a situation which seldom arises in practice), when the risk period is short, for example 1 day risk period and when there are very few cases (for example n=10).

The main limitations of the findings from the expressions of the asymptotic bias, variance and AMSE is that they make no allowance for age effects. We did not take age into account because the calculations for the bias, variance and AMSE become unwieldy. Instead we explored the effect of age in the simulation study reported in chapter three.

The results from the simulations were presented starting with what we called the 'standard scenario' which is representative of many studies of paediatric vaccines. In the standard scenario we found that the estimates were substantially biased for sample sizes of 20 or less, when the true relative incidence was ≤ 1. However for relative incidence ≥ 1.5 the biases were moderate even with sample sizes of 10 cases, and very small when the number of cases was ≥ 50. Risk periods as short as 1 day and up to a maximum of 200 days (for a total observation period of 500 days) were investigated. In these situations, the estimates were biased for short risk periods. For example when the risk period was 1 day, the bias was large when the relative incidence was 0.5 even with 500 cases. Generally speaking, the longer the risk period in the range considered (up to 200 days), the less biased the estimates were.

Different age effects classified as weak symmetric, strong symmetric, weak monotone increasing, strong monotone increasing were explored as was the effect of different

distributions of age at exposure. There was little evidence that these affected the performance of the self-controlled case series model.

Indefinite risk periods were looked at. This was done to answer questions by some researchers [51, 52] who have argued that the self-controlled case series model may not be effective if one is looking at a situation were adverse events may manifest themselves a long time after exposure. We explored this issue by extending the risk periods to indefinite length. Results showed that overall there was little bias except for large relative incidences and distributions of age at event and age at exposure that induce confounding between exposure and age effects. This confounding and the bias it generates can be controlled by including unvaccinated cases.

In most situations explored in the simulation study, the coverage probabilities from ten thousand samples of different number of cases were in excess of their nominal values, even in the presence of substantial bias (see Tables 3.1 to 3.20 inclusive). This was not surprising, since when the expected number of event in the risk period is very small, the variance of $\hat{\beta} = \log(\hat{\rho})$ (where $\hat{\rho} = e^{\hat{\beta}}$ is the estimate of the relative incidence) is very large, as may be seen from the asymptotic calculations of chapter 2. Hence the confidence intervals will themselves be very wide. Confidence intervals based on profile likelihood methods may be preferable, but were not investigated.

In chapter 2 we found that when there are no age effects, the magnitude of the asymptotic bias depended largely on the imbalance of events in the risk and control periods, that is when the expected number of events in the risk period was less than that in the control period, the bias was negative, and vice versa. When the two

expectations were equal, the bias was zero. The simulation study explored more complex situations and finite samples. Qualitatively similar results emerged for a given sample size: the bias was greatest in magnitude when the expected number of events in the risk period was much smaller than the expected number in the control period. In practice, bias is only a real problem when the risk period is very short or relative incidence is low. In other circumstances, sample sizes in excess of 20 appear to give reliable results.

The overall conclusion from the analytical calculations of Chapter 2 and the simulation study of Chapter 3 is that estimates and confidence intervals based on asymptotic theory are reliable except in extreme scenarios (namely very small sample sizes, very small risk period, low relative incidence).

The next issue to be investigated was to improve the design of self-controlled case series studies by obtaining and validating sample size formulae. These were presented in chapters 4 and 5. We started off by investigating the first published sample size formula by Farrington et al [3] and found that this formula was not accurate. We then investigated other approaches. This led us to derive six other sample size formulae, one based on the distribution of the logarithm of the relative incidence, three based on the binomial distribution and two based on the signed root likelihood ratio statistic. Of the six sample size formulae derived, four were found to be accurate. Of the accurate sample size formulae, two were based on the binomial distribution namely that using a continuity correction and that using the arcsine variance stabilising transformation. The other two good formulae were based on the signed root likelihood ratio statistic with and without age adjustment. Our overall recommendation in designing a study

using the self-controlled case series method without age effects is to use either the formula based on the binomial distribution with arcsine variance stabilising transformation, or the formula based on the signed root likelihood ratio statistic. If age is to be taken into consideration, then there is only one sample size formula to use and this is the formula based on the signed root likelihood ratio statistic.

The third topic that we explored was to extend the self-controlled case series method to prospective surveillance. We had to find a way of incorporating the retrospective self-controlled case series method within a prospective surveillance system. We used the ideas of Wald [45] and Page [46] to derive a self-controlled case series based sequential probability ratio test (SPRT) and cumulative sum (CUSUM) for use in surveillance. Detailed findings are presented in chapters 6 and 7. We envisage using the SPRT for focused surveillance of a new vaccine, whereas the CUSUM can be used for routine surveillance of several established vaccines. A detailed simulation study showing how the SPRT and CUSUM can be applied was illustrated and results presented.

A possible limitation of an SPRT and CUSUM surveillance system using the case series method is the inability to produce a signal in real time since it is necessarily based on retrospective data. One possible solution to this is to make the monitoring time interval as short as possible. We used a 6-month monitoring interval. The monitoring time interval could be of any length depending on prior knowledge of the vaccine being investigated. The surveillance system would rely on routinely collected case data for signal detection. Such data have varying degrees of accuracy in diagnostic coding. It is for this reason that such a surveillance system should not be

viewed as the final confirmatory epidemiologic investigation into potential vaccine-associated adverse events. However, a system based on the case series method does provide stronger evidence of association than signal based solely on spontaneous reporting.

Overall, simulation studies showed that the performance of a focused surveillance system using the SPRT is as desired. Ideally, we would like a system to be very quick to detect a true relative incidence greater than 1 and also if there is no problem we would like the process not to cross the upper boundary or possibly signal in favour of the null hypothesis. The simulation study showed that the system was able to achieve both requirements as indicated both by the probability of crossing the upper boundary and also by the average time to detection of a signal.

Using the SPRT with the self-controlled cases series log-likelihood has all the advantages of using the self-controlled case series method (in particular control of confounders). The other advantage of using the self-controlled case series SPRT compared to other methods [79-82] that are frequently used in surveillance systems is the prior specification of the Type *I* and Type *II* errors. The Type *I* and Type *II* errors control against making wrong decisions. These error levels apply to the entire SPRT process, not to each specific monitoring time interval, and thus the analyses take account of multiple testing. The adjustment for multiple testing is not explicit as in the Bonferroni adjustment, but rather the adjustment is incorporated into the SPRT in the way that the upper and lower boundaries are calculated. These boundaries preserve the specific alpha and beta until a final decision is reached as to whether the hypothesis should be accepted or rejected. However, for the self-controlled case series

based SPRT, we found that the actual Type *I* and Type *II* error probabilities were lower than the nominal values.

Chapter seven explored how we can use the self-controlled case series method to construct CUSUM charts for the long term surveillance of several vaccines in routine use. We assessed the performance of the self-controlled case series CUSUM using the standard practice of using the average run length (ARL) to select chart thresholds and to summarise performance after adapting the definitions of ARL to the surveillance of several vaccines. We used this method in preference to the false discovery rate (FDR) and successful discovery rate of Marshall et al [110]. We found that our 'system average run length in control' denoted ARL_0 and 'system average run length out-of-control' denoted ARL_1 are practically relevant and interpretable parameters.

The system average run length in control of a self-controlled case series CUSUM measures the time interval between false signals when the system is in control. A large value of ARL_0 is desirable. We found that surveillance of several vaccines drastically reduces the ARL_0 obtained for a single vaccine. In an effort to increase the ARL_0 we investigated a resetting scheme where all vaccines are reset, not just the signalling vaccine. This does indeed increase the system average run length in-control, but also affects the system average run length out-of-control.

The system average run length out-of-control (ARL_1) is an upper limit on the time between a problem occurring and when it is detected (it is an upper limit because the CUSUM will generally be greater than zero when the problem occurs). Ideally this

must be kept small; for example detection within 2 years might be a reasonable requirement.

The method of Marshall et al [110] based on the false detection rate of a CUSUM applies perhaps more appropriately to the surveillance of large number of units. However, it lacks the focus on detection times which is provided by the system average run length in control and the system average run length out-of-control formulation. In general our findings showed that when monitoring several vaccines, the system ARL_0 of the self-controlled based CUSUM is much shorter than the ARL_0 when we are monitoring a single vaccine.

It was not possible to suggest a single control limit to use when using the CUSUM based on the self-controlled case series method. This depends on the risk period, the baseline incidence for the number of cases, the observation period (monitoring interval), the number of vaccines under observation, the way the surveillance will be carried out especially when several vaccines are under observation, and the relative incidence to be detected. Our results showed that a practical system is possible, but require careful choice of the parameters h and the design value, bearing in mind the need to avoid swamping the system with false positive signals.

Overall chapter six and seven showed that a monitoring system for a single new vaccine based on the self-controlled case series method with the SPRT is feasible, and could prove to be a very useful tool. Further work is required to incorporate age effects, select optimal values of α and β in the case of the SPRT and the best

monitoring interval. It is perhaps less clear that a CUSUM-based system for monitoring several vaccines would produce reliable results.

In chapter 8 we undertook an analysis of data on intussusception and oral polio vaccine. In line with other studies, we did not find strong evidence of association. One specific difficulty with these data was how to take account of censoring of exposure histories at different doses. We proposed a stepwise estimation procedure, starting at the last dose. In the latter part of the chapter we brought to bear the new insights obtained in this thesis on the issues of bias and surveillance of new rotavirus vaccines.

The work undertaken in this thesis suggests some avenues for further research. It would be interesting to obtain analytic expressions for the asymptotic bias and variance of the log relative incidence estimator, allowing for age effects. The sample size expression we obtained, taking into account age effects, works for short risk periods, but different methods are required for long risk periods. Perhaps most importantly, more work is required to design a practical surveillance system using SPRT or CUSUM, based on the case series method, taking account of age effects.

APPENDIX 1

This appendix contains tables of results from the simulation study to test the performance of the self-controlled case series method under different scenarios as described in each table. The terms used in the tables are explained below.

1. 90 CI Ninety percent confidence interval
2. 95 CI Ninety five percent confidence interval
3. 99 CI Ninety nine percent confidence interval
4. %*covered* Percentage of the 90%, 95% or 99% confidence intervals that contain the true relative incidence
5. %*low* Percentage of the 90%, 95% or 99% confidence intervals where the true value was less than the lower limit
6. %*hi* Percentage of the 90%, 95% or 99% confidence intervals where the true value was greater than the upper limit.
7. **RI** Relative incidence
8. **Log(RI)** Logarithm of the relative incidence
9. **Mean** Mean age at vaccination
10. **sd** Standard deviation of age at vaccination
11. **10 000 samples** Ten thousand samples of ten, twenty, fifty, hundred, five hundred, one thousand cases were simulated

12. **Median** The median value of the ten thousand estimates was used as a measure of central tendency of the estimates

13. **1, 5, 10 etc days risk period** These were risk periods selected for particular simulations, for example 1 day, 5 days, etc.

14. **Indefinite risk period** Sometimes indefinite risk periods were used instead of fixed number of days as in 13 above.

15. **Prop. Vacc.** Standing for proportion vaccinated in a particular simulation. The following proportions were explored, 1, $\frac{5}{6}, \frac{2}{3}, \frac{1}{2}$.

RI Log(RI)	10 000 samples of 20 cases **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 100 cases **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 500 cases **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*
0.500 -0.693	**0.000 −∞** *(5%, 95%, 0%)* *(2%, 98%, 0%)* *(2%, 98%, 0%)*	**0.459 -0.879** *(9%, 91%, 0%)* *(2%, 98%, 0%)* *(2%, 98%, 0%)*	**0.464 -0.767** *(5%, 93%, 2%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*
1.000 0.000	**0.000 −∞** *(7%, 92%, 1%)* *(6%, 94%, 0%)* *(1%, 99%, 0%)*	**0.947 -0.054** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.995 -0.005** *(5%, 90%, 5%)* *(2.9%, 95%, 2.1%)* *(0.6%, 99%, 0.4%)*
1.500 0.405	**0.000 −∞** *(7%, 93%, 0%)* *(3%, 97%, 0%)* *(2%, 98%, 0%)*	**1.404 0.339** *(5%, 95%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**1.504 0.396** *(5%, 91%, 4%)* *(3%, 95%, 2%)* *(0.8%, 99%, 0.2%)*
2.000 0.693	**2.070 0.727** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**1.899 0.641** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**1.978 0.682** *(5%, 91%, 4%)* *(3%, 96%, 1%)* *(0.7%, 99%, 0.3%)*
5.000 1.609	**4.891 1.587** *(4%, 96%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**4.932 1.596** *(6%, 91%, 3%)* *(3%, 96%, 1%)* *(6%, 91%, 3%)*	**5.000 1.609** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(0.6%, 99%, 0.4%)*
10.000 2.303	**10.033 2.306** *(6%, 94%, 0%)* *(4%, 96%, 2%)* *(1%, 99%, 0%)*	**9.996 2.302** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**9.997 2.302** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.6%, 99%, 0.4%)*

Table 3.5 Simulation results for 10 days risk period.

RI Log(RI)	10 000 samples of 20 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 100 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 500 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*
0.500 -0.693	**0.443 -0.813** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.491 -0.712** *(6%, 92%, 2%)* *(2%, 97%, 1%)* *(1%, 99%, 0%)*	**0.497 -0.699** *(5%, 91%, 4%)* *(3%, 95%, 2%)* *(0.6%, 99%, 0.4%)*
1.000 0.000	**0.944 -0.058** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.985 -0.015** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.997 -0.003** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.6%, 99%, 0.4%)*
1.500 0.405	**1.493 0.401** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**1.489 0.398** *(5%, 95%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**1.495 0.402** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.5%, 99%, 0.5%)*
2.000 0.693	**1.964 0.675** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**1.993 0.690** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**2.000 0.694** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.6%, 99%, 0.4%)*
5.000 1.609	**5.137 1.637** *(4%, 96%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**5.035 1.616** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.7%, 99%,0. 3%)*	**5.006 1.611** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.6%, 99%, 0.4%)*
10.000 2.303	**11.135 2.410** *(6%, 94%, 0%)* *(4%, 96%, 2%)* *(1%, 99%, 0%)*	**10.196 2.322** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**10.034 2.306** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.6%, 99%, 0.4%)*

Table 3.6 Simulation results for 50 days risk period.

RI Log(RI)	10 000 samples of 20 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 100 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 500 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*
0.500 -0.693	**0.454 -0.790** *(6%, 90%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.493 -0.712** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**0.498 -0.697** *(5%, 91%, 4%)* *(3%, 95%, 2%)* *(0.6%, 99%, 0.4%)*
1.000 0.000	**0.984 -0.016** *(6%, 92%, 2%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.996 -0.015** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**1.003 0.003** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
1.500 0.405	**1.510 0.412** *(6%, 91%, 3%)* *(2.6%, 97%, 0.4%)* *(1%, 99%, 0%)*	**1.505 0.409** *(5%, 95%, 0%)* *(2.5%, 95%, 2.5%)* *(0.7%, 99%, 0.3%)*	**1.499 0.405** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.5%, 99%, 0.5%)*
2.000 0.693	**2.031 0.709** *(6%, 91%, 4%)* *(3%, 96%, 1%)* *(1%, 99%, 0%)*	**2.013 0.700** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**2.001 0.694** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
5.000 1.609	**5.509 1.706** *(6%, 90%, 4%)* *(3%, 95%,2%)* *(0.5%, 99%, 0.5%)*	**5.071 1.623** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.7%, 99%,0. 3%)*	**5.006 1.611** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.6%, 99%, 0.4%)*
10.000 2.303	**11.440 2.437** *(6%, 94%, 0%)* *(4%, 96%, 2%)* *(1%, 99%, 0%)*	**10.327 2.335** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**10.058 2.308** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*

Table 3.7 Simulation results for 100 days risk period.

RI Log(RI)	10 000 samples of 20 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 100 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 500 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*
0.500 -0.693	**0.478 -0.737** *(5%, 92%, 3%)* *(1.5%, 97%, 1.5%)* *(1%, 99%, 0%)*	**0.493 -0.707** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**0.499 -0.695** *(5%, 91%, 4%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
1.000 0.000	**0.993 -0.007** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**0.998 -0.002** *(5%, 95%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**1.001 0.001** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
1.500 0.405	**1.529 0.425** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**1.505 0.409** *(5%, 95%, 0%)* *(2.5%, 95%, 2.5%)* *(0.7%, 99%, 0.3%)*	**1.502 0.407** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
2.000 0.693	**2.091 0.737** *(4%, 91%, 5%)* *(2%, 96%, 2%)* *(0.3%, 99%, 0.7%)*	**2.008 0.697** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**2.002 0.694** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
5.000 1.609	**5.589 1.721** *(6%, 90%, 4%)* *(3%, 95%,2%)* *(0.5%, 99%, 0.5%)*	**5.068 1.623** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.7%, 99%,0. 3%)*	**5.007 1.611** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
10.000 2.303	**12.048 2.489** *(6%, 94%, 0%)* *(4%, 96%, 2%)* *(1%, 99%, 0%)*	**10.286 2.331** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**10.038 2.306** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*

Table 3.8 Simulation results for 200 days risk period.

RI	Log(RI)	10 000 samples of 20 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 100 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 500 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	*Risk periods (days)*
1.000	**0.000**	**0.000 −∞** *(9%, 91%, 0%)* *(4%, 96%, 0%)* *(2%, 98%, 0%)*	**0.814 -0.206** *(5%, 95%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**0.983 -0.018** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.7%, 99%, 0.3%)*	***10***
2.000	**0.693**	**1.922 0.653** *(7%, 93%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**1.980 0.683** *(5%, 93%, 2%)* *(3%, 97%, 2%)* *(1%, 99%, 0%)*	**1.998 0.692** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	***10***
5.000	**1.609**	**4.533 1.511** *(7%, 93%, 4%)* *(4%, 96% ,0%)* *(1%, 99%, 0%)*	**4.980 1.605** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.8%, 99%, 0. 2%)*	**4.998 1.609** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	***10***
1.000	**0.000**	**0.822 -0.196** *(9%, 91%, 0%)* *(4%, 96%, 0%)* *(2%, 98%, 0%)*	**0.967 -0.033** *(5%, 95%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**0.997 -0.003** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.7%, 99%, 0.3%)*	***25***
2.000	**0.693**	**1.839 0.609** *(7%, 93%, 0%)* *(4%, 96% ,0%)* *(1%, 99%, 0%)*	**1.985 0.685** *(5%, 93%, 2%)* *(3%, 97%, 0%)* *(0.7%, 99%, 0.3%)*	**1.997 0.692** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	***25***
5.000	**1.609**	**5.281 1.664** *(7%, 93%, 4%)* *(3%, 96%, 1%)* *(1%, 99%, 0%)*	**5.049 1.619** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**5.003 1.610** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	***25***
1.000	**0.000**	**0.914 -0.090** *(9%, 91%, 0%)* *(4%, 96%, 0%)* *(2%, 98%, 0%)*	**0.991 -0.009** *(5%, 95%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**0.998 -0.002** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.7%, 99%, 0.3%)*	***50***
2.000	**0.693**	**2.015 0.701** *(7%, 93%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**2.009 0.698** *(5%, 93%, 2%)* *(3%, 97%, 0%)* *(0.7%, 99%, 0.3%)*	**2.002 0.694** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.7%, 99%, 0.3%)*	***50***
5.000	**1.609**	**5.462 1.698** *(7%, 93%, 4%)* *(3%, 96%, 1%)* *(1%, 99%, 0%)*	**5.090 1.627** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.32%)*	**5.018 1.613** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	***50***

Table 3.9 Simulation results for strong symmetric age effect .

RI	Log(RI)	10 000 samples of 20 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 100 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 500 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	*Risk periods (days)*
1.000	**0.000**	**0.000 −∞** *(9%, 91%, 0%)* *(4%, 96%, 0%)* *(2%, 98%, 0%)*	**0.974 -0.026** *(5%, 95%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**0.983 -0.018** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.7%, 99%, 0.3%)*	***10***
2.000	**0.693**	**2.286 0.827** *(7%, 93%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**1.973 0.679** *(5%, 93%, 2%)* *(3%, 97%, 2%)* *(0.7%, 99%, 0.3%)*	**1.985 0.686** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	***10***
5.000	**1.609**	**5.094 1.628** *(7%, 93%, 0%)* *(4%, 96% ,0%)* *(1%, 99%, 0%)*	**4.915 1.592** *(6%, 90%, 4%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**4.992 1.608** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	***10***
1.000	**0.000**	**0.958 -0.043** *(8%, 92%, 0%)* *(4%, 96%, 0%)* *(2%, 98%, 0%)*	**0.983 -0.017** *(6%, 95%, 1%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.991 -0.009** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.6%, 99%, 0.4%)*	***25***
2.000	**0.693**	**1.955 0.670** *(6%, 94%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**1.979 0.683** *(6%, 91%, 3%)* *(3%, 95%, 2%)* *(17%, 99%, 0%)*	**1.989 0.688** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	***25***
5.000	**1.609**	**4.789 1.566** *(6%, 94%, 0%)* *(3%, 97, 0%)* *(1%, 99%, 0%)*	**5.028 1.615** *(5%, 91%, 4%)* *(3%, 95%, 2%)* *(0.6%, 99%, 0.4%)*	**5.005 1.610** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	***25***
1.000	**0.000**	**0.958 -0.043** *(8%, 92%, 0%)* *(4%, 96%, 0%)* *(2%, 98%, 0%)*	**0.985 -0.016** *(6%, 95%, 1%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.998 -0.002** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.5%, 99%, 0.5%)*	***50***
2.000	**0.693**	**1.971 0.679** *(6%, 94%, 0%)* *(4%, 96% ,0%)* *(1%, 99%, 0%)*	**1.991 0.689** *(6%, 91%, 3%)* *(3%, 95%, 2%)* *(17%, 99%, 0%)*	**2.002 0.694** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	***50***
5.000	**1.609**	**5.231 1.655** *(6%, 90% ,4%)* *(3%, 96, 1%)* *(1%, 99%, 0%)*	**5.030 1.615** *(5%, 91%, 4%)* *(3%, 95%, 2%)* *(0.6%, 99%, 0.4%)*	**5.015 1.612** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	***50***

Table 3.10 Simulation results for weak monotone increasing age effect

RI	Log(RI)	10 000 samples of 20 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 100 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	10 000 samples of 500 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi*	*Risk periods (days)*
1.000	**0.000**	**0.000 −∞** *(8%, 92%, 0%)* *(4%, 96%, 0%)* *(2%, 98%, 0%)*	**0.969 -0.032** *(6%, 95%, 1%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.985 -0.015** *(6%, 90%, 3%)* *(3%, 96%, 1%)* *(1%, 99%, 0%)*	***10***
2.000	**0.693**	**2.197 0.787** *(6%, 94%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**1.987 0.686** *(6%, 94%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**1.979 0.683** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	***10***
5.000	**1.609**	**4.687 1.545** *(6%, 90%, 4%)* *(3%, 96, 1%)* *(1%, 99%, 0%)*	**4.957 1.601** *(6%, 90%, 4%)* *(3%, 96%, 1%)* *(1%, 99%, 0%)*	**5.000 1.609** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	***10***
1.000	**0.000**	**0.941 -0.061** *(8%, 92%, 0%)* *(4%, 96%, 0%)* *(2%, 98%, 0%)*	**0.972 -0.029** *(6%, 95%, 1%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.995 -0.005** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	***25***
2.000	**0.693**	**1.972 0.657** *(6%, 94%, 0%)* *(4%, 96%,0%)* *(1%, 99%, 0%)*	**1.981 0.684** *(6%, 94%, 0%)* *(4%, 96%,0%)* *(1%, 99%, 0%)*	**2.000 0.693** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(2.5%, 99%, 2.5%)*	***25***
5.000	**1.609**	**5.100 1.629** *(6%, 92%, 2%)* *(3%, 97, 0%)* *(1%, 99%, 0%)*	**5.031 1.616** *(6%, 90%, 4%)* *(3%, 96, 1%)* *(1%, 99%, 0%)*	**5.021 1.614** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	***25***
1.000	**0.000**	**0.927 -0.075** *(8%, 92%, 0%)* *(4%, 96%, 0%)* *(2%, 98%, 0%)*	**0.989 -0.011** *(6%, 95%, 1%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**1.000 0.000** *(4.5%, 91%, 4. 5%)* *(2.5%, 95%, 2.5%)* *(0.6%, 99%, 0.4%)*	***50***
2.000	**0.693**	**2.036 0.711** *(6%, 94%, 0%)* *(4%, 96% ,0%)* *(1%, 99%, 0%)*	**2.000 0.693** *(5%, 90%, 5%)* *(2.5%, 96%, 1.5%)* *(1%, 99%, 0%)*	**2.000 0.693** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(2.5%, 99%, 2.5%)*	***50***
5.000	**1.609**	**5.447 1.695** *(6%, 92% ,2%)* *(3%, 97, 0%)* *(1%, 99%, 0%)*	**5.096 1.628** *(6%, 90%, 4%)* *(3%, 96, 1%)* *(1%, 99%, 0%)*	**5.018 1.613** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	***50***

Table 3.11 Simulation results for strong monotone increasing age effect

RI	Log(RI)	10 000 samples of 100 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (10 days risk period)	10 000 samples of 100 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (25 days risk period)	10 000 samples of 100 cases Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (50 days risk period)	*Mean* *(standard deviation)*
1.000	0.000	0.890 -0.117 *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	0.970 -0.300 *(5%, 92%, 3%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	0.991 -0.093 *(5%, 91%, 4%)* *(3%, 96%, 1%)* *(0.8%, 99%, 0.2%)*	*250* *(50)*
2.000	0.693	1.911 0.648 *(6%, 93%, 1%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	1.971 0.678 *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	1.985 0.686 *(5%, 90%, 5%)* *(3%, 95% ,2%)* *(0.7%, 99%, 0.3%)*	*250* *(50)*
5.000	1.609	4.939 1.597 *(5%, 90%,5%)* *(2%, 96, 2%)* *(1%, 99%, 0%)*	5.032 1.616 *(5%, 90%,5%)* *(2.5%, 95, 2.5%)* *(0.6%, 99%, 0.2%)*	5.000 1.609 *(5%, 90%, 5%)* *(2.5%, 95, 2.5%)* *(0.6%, 99%, 0.3%)*	*250* *(50)*
1.000	0.000	0.984 -0.016 *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	0.980 -0.020 *(5%, 92%, 3%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	0.986 -0.014 *(5%, 91%, 4%)* *(3%, 96%, 1%)* *(0.8%, 99%, 0.2%)*	*125* *(100)*
2.000	0.693	1.963 0.674 *(6%, 93%, 1%)* *(3%, 97% ,0%)* *(1%, 99%, 0%)*	1.990 0.688 *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	1.996 0.691 *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	*125* *(100)*
5.000	1.609	4.948 1.599 *(6%, 91%, 3%)* *(3%, 96, 1%)* *(1%, 99%, 0%)*	5.007 1.611 *(6%, 90%, 4%)* *(3%, 95, 2%)* *(0.6%, 99%, 0.2%)*	5.053 1.620 *(5%, 90%, 5%)* *(2.5%, 95, 2.5%)* *(0.6%, 99%, 0.3%)*	*125* *(100)*
1.000	0.000	0.966 -0.035 *(6%, 94%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	0.970 -0.030 *(6%, 93%, 1%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	0.986 -0.014 *(6%, 90%, 4%)* *(3%, 96%, 1%)* *(0.6%, 99%, 0.3%)*	*125* *(50)*
2.000	0.693	1.930 0.657 *(6%, 93%, 1%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	1.985 0.686 *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	1.996 0.691 *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	*125* *(50)*
5.000	1.609	4.911 1.591 *(6%, 91%, 3%)* *(3%, 96, 1%)* *(1%, 99%, 0%)*	4.991 1.608 *(6%, 90% ,4%)* *(3%, 95, 2%)* *(0.6%, 99%, 0.2%)*	5.041 1.618 *(5%, 90%, 5%)* *(2.5%, 95, 2.5%)* *(0.6%, 99%, 0.3%)*	*125* *(50)*

Table 3.12 Simulation results for weak symmetric age effect

RI	Log(RI)	10 000 samples of 100 cases **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* **(10 days risk period)**	10 000 samples of 100 cases **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* **(25 days risk period)**	10 000 samples of 100 cases **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* **(50 days risk period)**	***Mean*** ***(standard deviation)***
1.000	**0.000**	**0.941 -0.061** *(6%, 94%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**0.970 -0.030** *(5%, 93%, 1%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.981 -0.019** *(5%, 91%, 4%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	***250*** ***(50)***
2.000	**0.693**	**1.933 0.659** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**1.984 0.686** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**1.997 0.692** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	***250*** ***(50)***
5.000	**1.609**	**4.937 1.597** *(6%, 91%, 3%)* *(3%, 96, 1%)* *(1%, 99%, 0%)*	**5.018 1.613** *(6%, 90%, 4%)* *(3%, 95, 2%)* *(0.6%, 99%, 0.2%)*	**5.085 1.626** *(5%, 90%, 5%)* *(2.5%, 95, 2.5%)* *(0.6%, 99%, 0.3%)*	***250*** ***(50)***
1.000	**0.000**	**0.789 -0.237** *(5%, 95%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**0.949 -0.052** *(6%, 94%, 0%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**0.974 -0.026** *(6%, 91%, 3%)* *(3%, 96%, 1%)* *(1%, 99%, 0%)*	***125*** ***(100)***
2.000	**0.693**	**1.799 0.587** *(6%, 93%, 1%)* *(3%, 97%, 0%)* *(1%, 99%, 0%)*	**1.966 0.676** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**2.000 0.693** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	***125*** ***(100)***
5.000	**1.609**	**4.927 1.595** *(6%, 91%, 3%)* *(3%, 96, 1%)* *(1%, 99%, 0%)*	**5.054 1.620** *(6%, 90%, 4%)* *(3%, 95, 2%)* *(0.6%, 99%, 0.2%)*	**5.095 1.628** *(5%, 90%, 5%)* *(2.5%, 95, 2.5%)* *(0.6%, 99%, 0.3%)*	***125*** ***(100)***
1.000	**0.000**	**0.823 -0.196** *(6%, 94%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**0.962 -0.039** *(6%, 94%, 0%)* *(4%, 96%, 0%)* *(1%, 99%, 0%)*	**0.983 -0.017** *(5%, 92%, 3%)* *(3%, 97%, 0%)* *(0.4%, 99%, 0.6%)*	***125*** ***(50)***
2.000	**0.693**	**1.896 0.640** *(6%, 93%, 1%)* *(3%, 97% , 0%)* *(1%, 99%, 0%)*	**1.977 0.686** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	**2.022 0.704** *(5%, 90%, 5%)* *(3%, 95%, 2%)* *(0.7%, 99%, 0.3%)*	***125*** ***(50)***
5.000	**1.609**	**4.949 1.599** *(6%, 91%, 3%)* *(3%, 96, 1%)* *(1%, 99%, 0%)*	**5.061 1.622** *(6%, 90%, 4%)* *(3%, 95, 2%)* *(0.6%, 99%, 0.2%)*	**5.097 1.629** *(5%, 90%, 5%)* *(2.5%, 95, 2.5%)* *(0.6%, 99%, 0.3%)*	***125*** ***(50)***

Table 3.13 Simulation results for strong monotone increasing age effect

Table 3.14 Simulations for indefinite risk period with weak symmetric age effect, 250 days mean age at exposure, with 100 days standard deviation.

RI **Log(RI)**	**10 000 samples of 100 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**100 cases, Prop. Vacc. 1**)	**10 000 samples of 120 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**120 cases, Prop. Vacc. 5/6**)	**10 000 samples of 150 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**150 cases, Prop. Vacc. 2/3**)	**10 000 samples of 200 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**200 cases, Prop. Vacc. 1/2**)
1.000 0.000	**1.007 0.007** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**1.000 0.000** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**0.999 -0.002** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**1.000 0.000** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
2.000 0.693	**2.167 0.704** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**2.009 0.698** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**2.004 0.695** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**2.004 0.695** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
5.000 1.609	**5.049 1.619** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**5.021 1.614** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**5.059 1.621** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**5.029 1.615** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*

Table 3.15 Simulations for indefinite risk period with weak symmetric age effect, 125 days mean age at exposure, with 50 days standard deviation.

RI **Log(RI)**	**10 000 samples of 100 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**100 cases, Prop. Vacc. 1**)	**10 000 samples of 120 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**120 cases, Prop. Vacc. 5/6**)	**10 000 samples of 150 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**150 cases, Prop. Vacc. 2/3**)	**10 000 samples of 200 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**200 cases, Prop. Vacc. 1/2**)
1.000 0.000	**1.009 0.009** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**0.999 - 0.001** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**0.999 -0.001** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**1.004 0.004** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
2.000 0.693	**1.993 0.690** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**1.993 0.690** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**2.016 0.701** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**2.004 0.695** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
5.000 1.609	**5.029 1.630** *(4.5%, 91%, 4.5%)* *(2%, 96, 2%)* *(0.1%, 99%, 0.9%)*	**5.160 1.641** *(4.5%, 91% ,4.5%)* *(1%, 96, 3%)* *(0%, 99%, 1%)*	**5.223 1.653** *(3%, 92%, 5%)* *(1%, 97, 2%)* *(0%, 99%, 1%)*	**5.155 1.640** *(3%, 91%, 6%)* *(2.5%, 96, 1.5%)* *(0%, 99%, 1%)*

Table 3.16 Simulations for indefinite risk period with strong monotone increasing age effect, 250 days mean age at exposure, with 100 days standard deviation.

RI **Log(RI)**	**10 000 samples of 100 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**100 cases, Prop. Vacc. 1**)	**10 000 samples of 120 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**120 cases, Prop. Vacc. 5/6**)	**10 000 samples of 150 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**150 cases, Prop. Vacc. 2/3**)	**10 000 samples of 200 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**200 cases, Prop. Vacc. 1/2**)
1.000 0.000	**1.002 0.002** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**0.999 0.001** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**0.996 -0.004** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**0.999 -0.001** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
2.000 0.693	**2.004 0.695** *5%, 90%, 5%)* *(2%, 96%, 2%)* *(0.5%, 99%, 0.5%)*	**2.011 0.699** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**1.997 0.691** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**2.011 0.699** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
5.000 1.609	**5.130 1.635** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**5.112 1.632** *(4.5%, 91%, 4.5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**5.131 1.635** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**5.077 1.625** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*

Table 3.17 Simulations for indefinite risk period with strong monotone increasing age effect, 125 days mean age at exposure, with 50 days standard deviation.

RI **Log(RI)**	**10 000 samples of 100 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**100 cases, Prop. Vacc. 1**)	**10 000 samples of 120 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**120 cases, Prop. Vacc. 5/6**)	**10 000 samples of 150 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**150 cases, Prop. Vacc. 2/3**)	**10 000 samples of 200 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**200 cases, Prop. Vacc. 1/2**)
1.000 0.000	**1.004 0.004** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*	**1.004 0.004** *(5%, 90%, 5%)* *(2%, 96%, 2%)* *(0.5%, 99%, 0.5%)*	**1.001 0.001** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.3%, 99%, 0.7%)*	**1.005 0.005** *(5%, 90%, 5%)* *(2.5%, 95%, 2.5%)* *(0.5%, 99%, 0.5%)*
2.000 0.693	**2.044 0.715** *(5%, 90%, 5%)* *(2%, 96%, 2%)* *(0.1%, 99%, 0.9%)*	**2.040 0.713** *(4%, 91%, 5%)* *(2%, 96, 2%)* *(0.1%, 99%, 0.9%)*	**2.039 0.712** *(3%, 91%, 6%)* *(1%, 96%, 3%)* *(0%, 99%, 1%)*	**2.039 0.713** *(3%, 91%, 6%)* *(1%, 97%, 2%)* *(0%, 99%, 1%))*
5.000 1.609	**5.450 1.696** *(1%, 93%, 6%)* *(2%, 96, 2%)* *(0%, 99%, 1%)*	**5.305 1.669** *(0.5%, 94%, 5.5%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**5.385 1.684** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**5.314 1.670** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*

Table 3.18 Simulations for strong monotone increasing age effect and indefinite risk period

RI **Log(RI)**	**10 000 samples of 100 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**100 cases, Prop. Vacc. 1**)	**10 000 samples of 120 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**120 cases, Prop. Vacc. 5/6**)	**10 000 samples of 150 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**150 cases, Prop. Vacc. 2/3**)	**10 000 samples of 200 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* **(200 cases, Prop. Vacc. 1/2**)	**Mean** **Sd**
1.000 **0.000**	**0.914 -0.090** *(2%, 96%, 2%)* *(0%, 100, 0%)* *(0%, 100%, 0%)*	**0.988 -0.012** *(5%, 91%, 4%)* *(2%, 96, 2%)* *(0%, 100%, 0%)*	**1.009 0.009** *(5%, 91%, 4%)* *(2%, 96, 2%)* *(0%, 100%, 0%)*	**1.003 0.003** *(1%, 95%, 4%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**125** **10**
1.000 **0.000**	**1.002 0.002** *(5%, 90%, 5%)* *(2%, 96, 2%)* *(0.5%, 99%, 0.5%)*	**0.992 -0.008** *(5%, 90%, 5%)* *(2.5%, 95, 2.5%)* *(0.5%, 99%, 0.5%)*	**0.988 - 0.012** *(5%, 90%, 5%)* *(2.5%, 95, 2.5%)* *(0.5%, 99%, 0.5%)*	**0.996 -0.004** *(5%, 90%, 5%)* *(2.5%, 95, 2.5%)* *(0%, 100%, 0%)*	**125** **20**
1.000 **0.000**	**1.003 0.003** *(5%, 95%, 5%)* *(2.5%, 95, 2.5%)* *(0.5%, 99%, 0.5%)*	**1.002 0.002** *(5%, 90%, 5%)* *(2.5%, 95, 2.5%)* *(0.5%, 99%, 0.5%)*	**1.001 0.001** *(5%, 90%, 5%)* *(2.5%, 95, 2.5%)* *(0.5%, 99%, 0.5%)*	**1.009 0.002** *(4%, 91%, 5%)* *(2%, 96, 2%)* *(0.5%, 99%, 0.5%)*	**125** **30**
1.000 **0.000**	**1.006 0.006** *(5%, 90%, 5%)* *(2.5%, 95, 2.5%)* *(0.5%, 99%, 0.5%)*	**0.999 0.001** *(5%, 90%, 5%)* *(2%, 95, 3%)* *(0%, 99%, 1%)*	**1.014 0.014** *(5%, 95%, 5%)* *(2.5%, 95, 2.5%)* *(0%, 100%,0%)*	**1.010 0.009** *(4%, 91%, 5%)* *(2%, 96, 2%)* *(0%, 99%, 1%)*	**125** **40**

Table 3.19 Simulation for strong monotone increasing age effect and indefinite risk period.

RI **Log(RI)**	**10 000 samples of 100 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**100 cases, Prop. Vacc. 1**)	**10 000 samples of 120 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**120 cases, Prop. Vacc. 5/6**)	**10 000 samples of 150 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**150 cases, Prop. Vacc. 2/3**)	**10 000 samples of 200 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**200 cases, Prop. Vacc. 1/2**)	**Mean** **Sd**
2.000 0.693	**1.850 0.615** *(0%, 96%, 4%)* *(0%, 98, 2%)* *(0%, 100%, 0%)*	**1.981 0.684** *(2%, 95%, 3%)* *(0%, 98, 2%)* *(0%, 100%, 0%)*	**2.046 0.716** *(4%, 91%, 5%)* *(1%, 96, 3%)* *(0%, 100%, 0%)*	**2.071 0.728** *(1%, 95%, 4%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**125 10**
2.000 0.693	**2.055 0.720** *(2%, 92%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**1.995 0.691** *(5%, 90%, 5%)* *(1%, 96, 3%)* *(0%, 100%, 0%)*	**2.051 0.719** *(4%, 91%, 5%)* *(1%, 96, 3%)* *(0%, 100%, 0%)*	**2.071 0.728** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**125 20**
2.000 0.693	**2.038 0.712** *(5%, 95%, 5%)* *(1%, 96, 3%)* *(0%, 99%, 1%)*	**2.047 0.764** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**2.058 0.722** *(4%, 92%, 4%)* *(1%, 97, 2%)* *(0%, 100%, 0%)*	**2.039 0.712** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**125 30**
2.000 0.693	**2.056 0.721** *(4%, 92%, 4%)* *(1%, 97, 2%)* *(0.5%, 99%, 0.5%)*	**2.024 0.705** *(5%, 90%, 5%)* *(2%, 95, 3%)* *(0%, 99%, 1%)*	**2.022 0.704** *(4%, 91%, 6%)* *(1%, 96, 3%)* *(0%, 100%, 0%)*	**2.071 0.728** *(3%, 92%, 5%)* *(1%, 97, 2%)* *(0%, 99%, 1%)*	**125 40**

Table 3.20 Simulations for strong monotone increasing age effect and indefinite risk period

RI **Log(RI)**	**10 000 samples of 100 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**100 cases, Prop. Vacc. 1**)	**10 000 samples of 120 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**120 cases, Prop. Vacc. 5/6**)	**10 000 samples of 150 cases** **Median** *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**150 cases, Prop. Vacc. 2/3**)	**10 000 samples of 200 cases** Median *90 CI %low %covered %hi* *95 CI %low %covered %hi* *99 CI %low %covered %hi* (**200 cases, Prop. Vacc. 1/2**)	**Mean** **Sd**
5.000 1.609	**7.824 2.824** *(0%, 94%, 6%)* *(0%, 98, 2%)* *(0%, 100%, 0%)*	**5.362 1.679** *(2%, 95%, 3%)* *(0%, 98, 2%)* *(0%, 100%, 0%)*	**5.253 1.659** *(1%, 95%, 4%)* *(0%, 98, 2%)* *(0%, 100%, 0%)*	**5.334 1.674** *(1%, 95%, 4%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**125 10**
5.000 1.609	**7.506 2.016** *(2%, 92%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**5.311 1.670** *(2%, 93%, 5%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**5.374 1.682** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**5.309 1.669** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**125 20**
5.000 1.609	**5.660 1.733** *(0%, 93%, 7%)* *(0%, 96, 4%)* *(0%, 99%, 1%)*	**5.221 2.445** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**5.378 1.682** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%,1%)*	**5.292 1.666** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**125 30**
5.000 1.609	**5.568 1.717** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**5.223 1.653** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%, 1%)*	**5.398 1.686** *(0%, 94%, 6%)* *(0%, 97, 3%)* *(0%, 99%,1%)*	**5.303 1.668** *(0%, 94%, 6%)* *(0%, 98, 2%)* *(0%, 99%, 1%)*	**125 40**

APPENDIX 2

The case series likelihood for the parameters β and α_j, $j = 0,...,J-1$ is

$$L(\beta,\alpha_0,...,\alpha_{J-1}) = \prod_{i=1}^{n}\prod_{j=0}^{J-1}\prod_{k=0,1}\left(\frac{\exp(\alpha_j+\beta k)e_{ijk}}{\sum_{s=0}^{J-1}\sum_{t=0,1}\exp(\alpha_s+\beta t)e_{ist}}\right)^{n_{ijk}}$$

where e_{ijk} is the observation time for event i in age group j and risk period k $(k=0,$ unexposed; $k=1,$ exposed$)$, and n_{ijk} is the number of events (0 or 1) occurring in this period. Note that in the formulation, independent multiple events within the same individual are represented as separate terms in the likelihood.

Suppose now that the α_j are regarded as known. The log likelihood ratio for β is

$$D(\beta) = 2\left[\sum_{i,j,k} n_{i,j,k}\beta k - \sum_{i=1}^{n} n_{i..}\log\left(\frac{\sum_{s,t}\exp(\alpha_s+\beta t)e_{ist}}{\sum_{s,t}\exp(\alpha_s)e_{ist}}\right)\right]$$

If the event i occurs in an unexposed individual, its contribution to $D(\beta)$ is zero. Otherwise, under the assumptions set out in section 5.6.1,

$$\sum_{s,t}\exp(\alpha_s)e_{ist} = \sum_{s=0}^{J-1}e^{\alpha_s}e_s$$

$$\sum_{s,t}\exp(\alpha_s+\beta t)e_{ist} = \sum_{s=0}^{J-1}e^{\alpha_s}e_s + \exp\left(\alpha_{s(i)}\right)\left(e^{\beta}-1\right)e^{*}$$

where $s(i)$ is the age group exposure. Thus

$$D(\beta) = 2\left[x\beta - \sum_{j=0}^{J-1} m_j \log\left(r_j e^{\beta}+1-r_j\right)\right]$$

where x is the total number of events occurring in a risk period, m_j is the total number of events occurring in individuals exposed at age j, and r_j is defined in

section 5.6.2. The log likelihood ratio reaches its minimum at the maximum likelihood estimator $\hat{\beta},$ which is the solution of

$$x = \sum_{j=0}^{J-1} m_j \frac{r_j e^{\hat{\beta}}}{r_j e^{\hat{\beta}} + 1 - r_j}$$

Substituting this expression for x in $D\left(\hat{\beta}\right)$ we obtain $D\left(\hat{\beta}\right).$ The test statistic upon which the sample size calculation is based is

$$T\left(\hat{\beta}\right) = \operatorname{sgn}\left(\hat{\beta}\right)\sqrt{D\left(\hat{\beta}\right)}$$

The asymptotic variance of $\hat{\beta}$ is

$$V\left(\hat{\beta}\right) = \frac{1}{\left[\sum_{j=0}^{J-1} m_j \pi_j \left(1-\pi_j\right)\right]}$$

where the π_j are defined in section 5.6.2. Expanding $T\left(\hat{\beta}\right)$ in a Taylor series around $\beta,$ and substituting $V\left(\hat{\beta}\right)$ we obtain, to first order in n,

$$E\left[T\left(\hat{\beta}\right)\right] \approx \operatorname{sgn}(\beta)\sqrt{\left\{2\sum_{j=0}^{J-1} m_j\left[\beta\pi_j - \log\left(r_j e^{\beta} + 1 - r_j\right)\right]\right\}}$$

$$V\left[T\left(\hat{\beta}\right)\right] \approx \frac{\beta^2}{\left\{E\left[T\left(\hat{\beta}\right)\right]\right\}^2}\sum_{j=0}^{J-1} m_j \pi_j \left(1-\pi_j\right)$$

Finally, replace m_j by nv_j, with v_j as defined in section 5.6.2. Thus $T\left(\hat{\beta}\right) \approx N\left(\operatorname{sgn}(\beta)\sqrt{nA}, B\right)$ where A and B are given in equation (5.7). Note that by expanding A and B to second order in β, it can be shown that $A \to 0$ and $B \to 1$ as $\beta \to 0,$ as expected.

List of acronyms

ADRs Adverse Drug Reactions

A&E Admission and Entry

AMSE Asymptotic Mean Square Error

$\boldsymbol{ARL}$ Average Run Length

$\boldsymbol{ARL}_0$ system Average Run Length in-control

$\boldsymbol{ARL}_1$ system Average Run Length out-control

BCPNN Bayesian Confidence Propagation Neural Network

CCR Coded Clinical Records

CDC Centres for Disease Control and prevention

CI Confidence Interval

CRF Case Report Form

CUSUM Cumulative Sum

DTP Diphteria Tetanus Pertussis

FDA Food and Drug Administration

FDR False Discovery Rate

GSK GlasxoSmithKline

GPRD General Practice Research Database

HES Hospital Episode Statistics

$\boldsymbol{H}_0$ Null hypothesis

$\boldsymbol{H}_1$ Alternative Hypothesis

ICD International Classification of Diseases

ITP Idiopathic Thrombocytpenic Purpura

MCSE Monte Carlo Standard Error

MHRA Medicines and Healthcare products Regulatory Agency

MMR Measles Mumps and Rubella

OPV Oral Polio Vaccine

PAS Patient Administration System

PRR Proportional Reporting Ratios

RI Relative Incidence

RI_0 Relative Incidence under the null hypothesis

RL_1 The design Relative Incidence value used in the SPRT

RL_2 The actual Relative Incidence value used to generate data

RRV-TV Tetravalent Rhesus Human Reassortant Rotavirux Vaccine

ROR Reporting Odds Ratios

SCCS Self-Controlled Case Series

SCCSM Self-Controlled Case Series Method

SDR Successful Discovery Rate

SPC Statistical Process Control

SPRT Sequential Probability Ratio Test

UK United Kingdom

USA United States America

US United States

VAERS Vaccine Adverse Event Reporting System

VSD Vaccine Safety Datalink

WHO World Health Organisation

APPENDIX 3

Papers published or submitted from the thesis

Sample sizes for self-controlled case series studies [44] (Covering chapter 4 and 5).

Self-controlled case series analyses: small sample performance[145] (Covering chapter 2 and 3).

Tutorial in Biostatistics: The self-controlled case series method [7] (Covering part of chapter 4).

References:

[1]. Philips EM and Pugh DS, How to get a PhD, a handbook for students and their supervisors. 3rd ed. 2000, Maidenhead, Philadelphia: *Open University Press.* pp50-51.

[2]. Farrington CP, Relative incidence estimation from case series for vaccine evaluation. *Biometrics*, 1995. **51**: p. 228-235.

[3]. Farrington CP, Nash J, and Miller E, Case series analysis of adverse reactions to vaccines: a comparative evaluation. *American Journal of Epidemiology*, 1996. **143**: p. 1165-1173 (Erratum 1998; **147**:93).

[4]. Hubbard R, Farrington CP, Smith C, Smeeth L, and Tattersfield A, Exposure to tricyclic and selective serotonin inhibitor antidepressants and the risk of hip fracture. *American Journal of Epidemiology*, 2003. **158**: p. 77-84.

[5]. Hocine M, Guillemot D, Tubert-Bitter P, and Moreau T, Testing independence between two poisson-generated multinomial variables in case series and cohort studies. *Statistics in Medicine*, 2005. **24**: p. 4035-4044.

[6]. Becker NG, Li Z, and Kelman CW, The effect of transient exposures on the risk of an acute illness with low hazard rate. *Biostastics*, 2004. **5**: p. 239-248.

[7]. Whitaker HJ, Farrington CP, Spiessens B, and Musonda P, Tutorial in Biostatistics: The self-controlled case series method. *Statistics in Medicine*, 2006. **25**: p. 1768-1797.

[8]. Farrington CP and Whitaker HJ, Semiparametric analysis of case series data. *Applied Statistics*, 2006. **55**(Part 5): p. 553-594.

[9]. Farrington CP, Pugh S, Colville A, Flower A, Nash J, Capner P, Rush M, and Miller E, A new method for active surveillance of adverse events from diphtheria/tetanus/pertussis and measles/mumps/rubella vaccines. *Lancet*, 1995. **345**: p. 567-5699.

[10]. Miller E, Goldacre M, Pugh S, Colville A, Farrington CP, Flower A, MacFarlane l, and Tettmar R, Risk of aseptic menigitis after measles, mumps and rubella vaccine in UK children. *Lancet*, 1993. **341**: p. 979-982.

[11]. Ray WA and Griffin MR, Use of Medicaid data for pharmaco-epidemiology. *American Journal of Epidemiology*, 1989. **129**: p. 837-849.

[12]. Altman DG, Practical Statistics for Medical Research. 1991, London: *Chapman & Hall.* pp93-94.

[13]. Fine PE and Chen RT, Confounding in studies of adverse reactions to vaccines. *American Journal of Epidemiology*, 1992. **136**: p. 121-135.

[14]. Smeeth L, Donnan PT, and Cook DG, The use of primary care databases: case-control and case-only designs. *Family Practice*, 2006. **23**: p. 597-604.

[15]. Aelen OO, Borgan O, Keiding N, and Thorman J, Interaction between life history events: nonparametric analysis for prospective and retrospective data in the presence of censoring. *Scandinavian Journal of Statistics*, 1980. **7**: p. 161-171.

[16]. Prentice RL, Vollmer WM, and Kalbfleisch JD, On the use of case series to identify disease risk factors. *Biometrics*, 1984. **40**: p. 445-458.

[17]. Maclure M, The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology*, 1991. **133**: p. 144-153.

[18]. Vines SK and Farrington CP, Within-subject exposure dependency in case-crossover studies. *Statistics in Medicine*, 2001. **20**: p. 3039-3049.
[19]. Greenland S, A unified approach to the analysis of case-distribution (case-only) studies. *Statistics in Medicine*, 1996. **18**: p. 1-15.
[20]. Maclure M and Mittleman MA, Should we use a case-cross-over design? *Annual Review of Public Health*, 2000. **21**: p. 193-221.
[21]. Feldmann U, Epidemiologic assessment of risks of adverse reactions associated with intermittent exposure. *Biometrics*, 1993. **49**: p. 419-428.
[22]. Marshall RJ and Jackson RT, Analysis of case-crossover designs. *Statistics in Medicine*, 1993. **12**: p. 2333-2341.
[23]. Andrews NJ, Statistical assessment of the association between vaccination and rare adverse events post licensure. *Vaccine*, 2002. **20**: p. S49-S53.
[24]. Farrington CP, Control without separate controls: evaluation of vaccine safety using case-only methods. *Vaccine*, 2004. **22**: p. 2064-2070.
[25]. Navidi W, Bidirectional case-cross-over designs for exposures with time trends. *Biometrics*, 1998. **54**(2): p. 596-605.
[26]. Lumley T and Levy D, Bias in the case-crossover design: implications for studies of air pollution. *Environmetrics*, 2000. **11**: p. 689-704.
[27]. Miller E, Waight P, Stowe J, and Taylar B, Idiopathic thrombocytopenic purpura and MMR vaccine. *Archives of Diseases in Childhood*, 2001. **84**: p. 227-229.
[28]. Dourado I, Cunha S, Teixeira MDG, Farrington CP, Melo A, Lucena R, and Barreto ML, An outbreak of aseptic meningitis associated with a Urabe-containing MMR mass vaccination campaign: implications for immunisation programs. *American Journal of Epidemiology*, 2000. **151**: p. 524-530.
[29]. Taylor B, Miller E, Farrington CP, Petropoulos M-C, Favot-Mayaud I, and Waight PA, Autism and measles, mumps and rubella vaccine: no epidemiological evidence for a causal association. *Lancet*, 1999. **353**: p. 2026-2029.
[30]. Farrington CP, Miller E, and Taylor B, MMR and autism: further evidence against a causal association. *Vaccine*, 2001. **19**: p. 3632-3635.
[31]. Miller E, Andrews N, Waight P, and Taylor B, Bacterial infections, immune overload, and MMR vaccine. *Archives of Diseases in Childhood*, 2003. **88**: p. 222-223.
[32]. Miller E, Andrews N, Grant A, Stowe J, and Taylor B, No evidence of an association between MMR vaccine and gait disturbance. *Archives of Diseases in Childhood*, 2005. **90**: p. 292-296.
[33]. Kramarz P, DeStafano F, Gargiullo PM, et al, Does influenza vaccination exacerbate asthma? Analysis of a large cohort of children with asthma. *Arch Farm Med.*, 2000. **9**: p. 617-623.
[34]. Tata LJ, West J, Harrison T, Farrington CP, Smith C, and Hubbard R, Does influenza vaccination increase consultations, corticosteroid prescriptions or exacerbations in people with asthma or chronic obstructive pulmonary disease? *Thorax*, 2003. **58**: p. 835-839.
[35]. Mutsch M, Zhou W, Rhodes P, Bopp M, Chen RT, Linder T, Spyr C, and Steffen R, Use of the inactivated intranasal influenza vaccine and the risk of bell's palsy in Switzerland. *New England Journal of Medicine*, 2004. **350**: p. 896-903.
[36]. Andrews N, Miller E, Waight P, Farrington CP, Crowcroft N, Stowe J, and Taylor B, Does oral polio vaccine cause intussusception in infants? Evidence

from a sequence of three self-controlled case series studies in the United Kingdom. *European Journal of Epidemiology*, 2001. **17**: p. 701-706.
[37]. Galindo Sardiñas MA, Zambrano Cárdenas A, Coutin Marie G, Santin Peña M, Aliño Santiago M, Valcárcel Sachez M, and Farrington CP, Lack of association between intussusception among infants given an oral polio vaccine in Cuban children. *European Journal of Epidemiology*, 2001. **17**: p. 783-787.
[38]. Murphy TV, Garguillo PM, Massoudi MS, Nelson DB, Jumaan AO, Okoro CA, Zanardi LR, Setia S, E. F, LeBaron CW, Wharton M, and Livingood JR, Intussusception among infants given an oral rotavirus vaccine. *New England Journal of Medicine*, 2001. **344**: p. 564-572.
[39]. Mullooly JP, Pearson J, Drew L, Sculer R, Maher J, Gargiullo P, De Stefano F, and Chen R, et al, Wheezing lower respiratory disease and vaccination of full-term infants. *Pharmacoepidemiology and drug safety*, 2002. **11**: p. 21-30.
[40]. Tata LJ, West J, Smith C, Farrington CP, Card T, Smeeth L, and Hubbard R, General population based study of the impact of tricyclic and selective serotonin reuptake inhibitor antidepressants on the risk of acture myocardial infarction. *Heart*, 2005. **91**: p. 465-471.
[41]. France EK, Glanz JM, Xu S, Davis RL, Black S, Shinefield H, Zangwill KM, Marcy SM, Mullooly JP, Jackson LA, and Chen R, Safety of the trivalent inactivated influenza vaccine among children-a population based study. *Archives of Pediatrics and Adolescent Medicine*, 2004. **158**: p. 1031-1036.
[42]. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington CP, and Vallance P, Risk of myocardial infarction and stroke after acute infection or vaccination. *New England Journal of Medicine*, 2004. **351**: p. 2611-2618.
[43]. Glanz JM, McClure DL, Xu S, Hambidge SJ, Lee M, Kolczak MS, Kleinman K, Mullooly JP, and France EK, Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations. *Journal of Clinical Epidemiology*, 2006. **59**: p. 808-818.
[44]. Musonda P, Farrington CP, and Whitaker HJ, Sample sizes for self-controlled case series studies. *Statistics in Medicine*, 2006. **25**: p. 2618-2631.
[45]. Wald A, Sequential analysis. 1947, New York: *Wiley*.
[46]. Page ES, Continuous inspection schemes. *Biometrika*, 1954. **41**: p. 100-115.
[47]. Griffin MR, Ray WA, and Livengood JR, et al, Risk of seizures and encephalopathy after immunization with the Diphtheria Tetanus-Pertussis vaccine. *JAMA*, 1990. **263**: p. 1641-1645.
[48]. McCullagh P and Nelder JA, Generalized Linear Models. second ed. 1989: *Chapman &Hall/CRC.* p 209.
[49]. Johnson NL, Kotz S, and Kemp AN, Univariate Discrete Distribution. 1993: *John Wiley & Sons Inc.* pp106-107.
[50]. Cox DR and Hinkley DV, Theoretical Statistics. 1979, London: *Chapman and Hall.* p252.
[51]. Holmes V, *A Possible Association Between MMR and Autism: An Examination of Case Series Analysis*, in *Department of Applied Statistics*. 2001, University of Reading: MSc Dissertation.
[52]. Spitzer WO, A sixty day war of words: is MMR linked to autism? *Adverse Drug Reaction Toxological Review*, 2001. **20**: p. 47-63.
[53]. Armitage P, Statistical Methods in Medical Research. (1971). *Blackwell scientific publications.* p96-97.
[54]. Fleiss JL, Levin B, and Paik MC, Statistical Methods for Rates and Proportions. **3rd** ed. 2003, New Jersey: *John Wiley & Sons.* p31-33.

[55]. Chernick MR and Liu CY, The Saw-Toothed Behaviour of Power Versus Sample Size and Software Solutions: Single Binomial Proportion Using Exact Methods. *The American Statistician*, 2002. **56**(p149-155).
[56]. Cesana BM, Reina G, and Marubin E, Sample size for testing a proportion in clinical trials. "Two-step" Procedure Combining Power and Confidence Interval Expected width. *The American Statistician*, 2001. **55**: p. 288-292.
[57]. Brown LD, Cai TT, and DasGupta A, Interval Estimation for a Binomial Proportion (with discussion). *Statistical Science*, 2001. **16**: p. 101-133.
[58]. Hoehler FK, Exact Power Calculations when the Dependent Variables is a Single Proportion and the Number of Events is Small. *Comput. Biol. Med.*, 1995. **5**: p. 447-449.
[59]. Thomas RG and Conlon M, Sample size determination based on Fisher's exact test for use in comparative trials with low event rates. *Controlled Clin. Trials*, 1992. **13**: p. 134-147.
[60]. Matthews JNS, An Introductions to Randomized Controlled Clinical Trials. 2000, *London*: *Arnold.* p30-31.
[61]. Chen RT, Rastogi SC, and Mullen JR, The Vaccine Adverse Event Reporting System (VAERS). *Vaccine*, 1994. **12**: p. 542-550.
[62]. Olsson S, The role of the WHO programme on International Drug Monitoring in coordinating worldwide drug safety efforts. *Drug safety*, 1998. **19**: p. 1-10.
[63]. Rosenthal S and Chen RT, The reporting sensitivities of two passive surveillance systems for vaccine adverse events. *American Journal of Public Health*, 1995. **85**: p. 1706-1709.
[64]. Zhou W, Pool V, De Stefano F, Iskander JK, Haber P, and Chen RT, A potential signal of Bell's palsy after parenteral inactivated influenza vaccines: reports to the Vaccine Adverse Event Reporting System (VAERS)-United states, 1991-2001. *Pharmacoepidemiology and drug safety*, 2004. **13**: p. 505-510.
[65]. Varicchio F, Iskander J, DeStefano F, et al, Understanding vaccine safety: information from the Vaccine Adverse Event Reporting System (VAERS). *Paediatric Infectious Disease Journal*, 2004. **23**: p. 1-8.
[66]. McMahon AW, Iskander J, Haber P, Chang S, Woo EJ, Braun MM, and Ball R, Adverse Events After Inactivated Influenza Vaccination Among Children Less Than 2 Years of Age: Analysis of Reports From Vaccine Adverse Event Reporting System, 1990-2003. *Pediatrics*, 2005. **115**: p. 453-460.
[67]. Murphy TV, Gargiullo PM, Massoudi MS, et al, Intussusception among infants given an oral rotavirus vaccine. *New England Journal of Medicine*, 2001. **344**: p. 564-572.
[68]. AAP, Prevention of rotavirus disease: guidelines for use of rotavirus vaccine. *Pediatrics*, 1998. **102**: p. 1483-1491.
[69]. CDC, Intussusception among recipients of rotavirus vaccine. *MMWR Morb Mortal Wkly Rep*, 1999. **48**: p. 577-581.
[70]. Spiegelhalter DJ, Grigg OA, Kinsman R, and Treasure T, Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery. *International Journal for Quality in Health Care*, 2003. **15**(1): p. 7-13.
[71]. Shewhart WA, The economic control of quality of manufactured product. 1931, New York: *D. Van Nostrand and Co.*

[72]. Deming WE, Quality, productivity, and competitive position. 1982, Cambridge, MA, USA: *Massachusetts Institute of Technology Center for Advanced Engineering Studies*.
[73]. Deming WE, On classification of the problems of statistical inference. *Journal of the American Statistics Association*, 1942. **37**: p. 173-185.
[74]. Steiner SH, Cook RJ, Farewell VT, and Treasure T, Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics*, 2000. **1**(4): p. 441-452.
[75]. Benneyan JC and Borgman AD, Risk-adjusted sequential probability ratio tests and longitudinal surveillance methods. *International Journal for Quality in Health Care*, 2003. **15**(1): p. 5-6.
[76]. Bates DW, The costs of adverse drug events in hospitalised patients. *Journal of Amrican Medical Association*, 1997. **277**: p. 307-311.
[77]. Bogner MS, Human error in Medicine. 1994, Hillside: *NJ: Erlbaum*.
[78]. Leape LL, Error in Medicine. *Journal of American Medical Association*, 1994. **272**: p. 1851-1857.
[79]. Evans SJW, Waller PC, and Davis S, Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and drug safety*, 2001. **10**: p. 483-486.
[80]. Rothman KJ, Lanes S, and Sacks ST, The reporting odds ratio and its advantages over the proprotional reporting ratio. *Pharmacoepidemiology and drug safety*, 2004. **13**: p. 519-523.
[81]. Waller PC, van Puijenbroek EP, Egberts ACG, and Evans SJW, The reporting odds ratio versus the proportional reporting ratio: 'deuce'. *Pharmacoepidemiology and drug safety*, 2004. **13**: p. 525-526.
[82]. van Puijenbroek EP, Bate A, Leufkens HGM, Linddquist M, Orre R, and Egberts ACG, A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiology and drug safety*, 2002. **11**: p. 3-10.
[83]. Linddquist M, Ståhl M, Bate A, Edwards IR, Fucik H, and Nunes AM, From association to alert-a revised approach to international signal analysis. *Pharmacoepidemiology and drug safety*, 1999. **8**: p. S15-S25.
[84]. Linddquist M, Ståhl M, Bate A, Edwards IR, and Meyboom RH, A retrospectiv evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database. *Drug safety*, 2000. **23**: p. 533-542.
[85]. Bate A, Linddquist, M, Edwards, IR, et al., A Bayesian neural network method for adverse drug reaction signal generation. *European Journal of clinical pharmacology*, 1998. **54**: p. 315-321.
[86]. Tillett HE and Spencer IL, Influenza surveillance in England and Wales using routine statistics. Development of 'cusum' graphs to compare 12 previous winters and to monitor the 1980/81 winter. *J.Hyg., Camb*, 1982. **88**: p. 83-94.
[87]. Choi K and Thacker SB, An evaluation of influenza mortality surveillance. Time series forecasts of expected pneumonia and influenza deaths. *American Journal of Epidemiology*, 1981. **113**: p. 215.
[88]. Hutwagner LC, Maloney EK, Bean NH, Slutsker L, and Martin SM, Using Laboratory-Based Surveillance Data for Prevention: An Algorithm for Detecting *Samonella* Outbreaks. *Emerging Infectious Diseases*, 1997. **3**(3): p. 395-400.

[89]. Benneyan JC, Number-between g-type statistical control charts for monitoring adverse events. *Health Care Management Science*, 2001. **4**: p. 305-318.
[90]. Bourke PD, Detecting a shift in fraction nonconforming using run-length control charts with 100% inspection. *Journal of Quality Technology*, 1991. **23**: p. 225-238.
[91]. Reynolds MR and Stoumbos ZG, The SPRT chart for monitoring a proportion. *IIE Transactions*, 1998. **30**: p. 545-561.
[92]. Grigg OA, Farewell VT, and Spiegelhalter DJ, Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. *Statistical Methods in Medical Research*, 2003. **12**: p. 147-170.
[93]. Wald A, Sequential tests of statistical hypotheses. *Ann Maths Statist*, 1945. **6**: p. 117-186.
[94]. Bartholomay AF, The sequential probability ratio test applied to the design of clinical experiments. *New England Journal of Medicine*, 1957. **256**: p. 498-505.
[95]. Armitage P, Sequential tests in prophylactic and therapeutic trials. *Q J Med*, 1954. **23**: p. 255-274.
[96]. Barnard GA, Sequential test in industrial statistics (with discussion). *J R Statist Soc*, 1946. **8(suppl.)**: p. 1-26.
[97]. Steiner SH, Geyer PL, and Wesolowsky GO, Grouped Data-Sequential Probability Ratio Tests and Cumulative Sum Control Charts. *Technometrics*, 1996. **38**(3): p. 230-237.
[98]. Davis RL, Kolczak M, Lewis E, Nodin J, Goodman M, Shay DK, Platt R, Black S, Shinefield H, and Chen RT, Active Surveillance of Vaccine Safety, A system to Detect Early Signs of Adverse Events. *Epidemiology*, 2005. **16**(3): p. 336-341.
[99]. Chen RT, Glasser, J.W, Rhodes, P.H, et al, Vaccine Safety Datalink Project: a new tool for improving vaccine safety monitoring in the United States. *Pediatrics*, 1997. **99**: p. 765-773.
[100]. Clayton D and Hills M, Statistical Models in Epidemiology. 1st ed. 1993, New Yook: *Oxford University Press Inc., New York*.
[101]. *SAS/STAT® User's Guide. Version 8.2 (2nd edn)*. 1999-2001, SAS Institute, Inc.: Cary, NC.
[102]. Barnard GA, Control Charts and stochastic processes. *Journal of the Royal Statistical Society, Series B*, 1959. **21**: p. 239-271.
[103]. Montgomery DC, Introduction to Statistical Quality Control. 2nd ed. 1991, New York: *John Wiley and Sons*.
[104]. Gallus G, Mandelli C, Marchi M, and Radeaelli G, On surveillance methods for congenital malformations. *Statistics in Medicine*, 1986. **5**: p. 565-571.
[105]. Nix AB, Rowlands RJ, and Kemp KW, Internal quality control in clinical chemistry: a teaching review. *Statistics in Medicine*, 1986. **6**: p. 425-440.
[106]. Williams SM, Parry BJ, and Schlup MM, Quality control: an application of the CUSUM. *British Medical Journal*, 1992. **304**: p. 1359-1361.
[107]. DeLeval, Marc R, François K, Bull C, Brawn WB, and Speigelhalter DJ, Analysis of a cluster of surgical failures. *The Journal of Thoracic and Cardiovascular Surgery*, 1994. **104**: p. 914-924.
[108]. Steiner SH, Cook RJ, and Farewell VT, Monitoring paired binary surgical outcomes using cumulative sum charts. *Statistics in Medicine*, 1999. **18**: p. 69-86.

[109]. Ross G, Lampugnani L, and Marchi M, An approximate CUSUM procedure for surveillance of health events. *Statistics in Medicine*, 1999. **18**: p. 2111-2122.
[110]. Marshall C, Best N, Bottle A, and Aylin P, Statistical issues in the prospective monitoring of health outcomes across multiple units. *Journal of Royal Statistical Society Series A*, 2004. **167**(Part 3): p. 541-559.
[111]. Frisén M, Evaluations of methods for statistical surveillance. *Statistics in Medicine*, 1992. **11**: p. 1489-1502.
[112]. Brook D and Evans DA, An approach to the probability distribution of cusum run length. *Biometrika*, 1972. **3**: p. 539-549.
[113]. Moustakides GV, Optimal stopping times for detecting changes in distributions. *Annals of Statistics*, 1986. **14**(4): p. 1379-1387.
[114]. Goel AL and Wu SM, Determination of ARL and contour nomogram for cusum charts to control normal mean. *Technometrics*, 1971: p. 221-230.
[115]. Khan RA, On cumulative sum procedures and the sprt with applications. *Journal of the Royal Statistical Society, Series B*, 1984. **46**(1): p. 79-85.
[116]. Woodall WH, The distribution of the run length of one-sided CUSUM procedures for continuous random variables. *Technometrics*, 1983. **25**: p. 295-301.
[117]. CDC, Rotavirus vaccine for the prevention of rotavirus gastroenteritis among children-recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm. Rep.*, 1999. **48**: p. 1-20.
[118]. CDC, Withdrawal of rotavirus vaccine recommendation. *MMWR Morb. Mortal. Wkly. Rep.*, 1999. **48**: p. 1007.
[119]. Glass RI, Bresee JS, Parashar UD, Jiang B, and J G, The future of rotavirus vaccines: a major setback leads to new opportunities. *Lancet*, 2004. **363**: p. 1547-1550.
[120]. Kramarz P and France EK, et al., Population-based study of rotavirus vaccination and intussusception. *Paediatric Infectious Disease Journal*, 2001. **20**: p. 410-416.
[121]. Tucker AW, Haddix A, Bresee JS, Holman RC, Parashar UD, and Glass RI, Cost-effectiveness analysis of a rotavirus immunization program for the United States. *JAMA*, 1998. **279**: p. 1371-1376.
[122]. Parashar UD, Hummelman EG, Bresee JS, Miller M, and Glass RI, Global illness and deaths caused by rotavirus disease in children. *Emerg Infect Dis*, 2003. **9**: p. 565-572.
[123]. Parashar UD, Holman, R.C, Cumming, K.C, et al, Trends in intussusception-associated hospitalizations and deaths among US infants. *Pediatrics*, 2000. **106**: p. 1413-1421.
[124]. Stringer MD, Pablot SM, and Brereton RJ, Pediatric intussusception. *Br J Surg*, 1992. **79**: p. 867-876.
[125]. Behrman RE, Kliegman RM, and Arvin AM, Wyllie R. Ileus, adhesions, intussusception and closed-loop obstructions.: *Nelsons Textbook of Pediatrics.* Chapter 29:1072-1074.
[126]. Hong-Yuan H, et al, Viro etiology of intussusception in Taiwanese childhood. *Paediatric Infectious Disease Journal*, 1998. **17**: p. 893-898.
[127]. Konno T, et al, HRV infection in infants and young children with intussusception. *Journal of Medical Virology*, 1978. **2**: p. 265-269.
[128]. Bell TM, et al, Virases in Lymph nodes of children with mesenteric adenitis and intussusception. *British Medical Journal*, 1962. **2**: p. 700-702.

[129]. Potter CW, et al, Adenovirus infection as an etiological factor in intussuception of infants and young children. *Journal of Pathology and Bacteriology*, 1964. **82**: p. 263-274.
[130]. Calico I, et al, Viral infections associated with intestinal invaginatioon. *Enferm Infec Microbiol Clin*, 1990. **8**: p. 406-410.
[131]. Nicolas JG, et al, A one-year virological survey of acute intussusception in chidhood. *Journal of Medical Virology*, 1982. **9**: p. 267-271.
[132]. Bhisitkul DM, et al, Adenovirus infection and childhood intussusception. *American Journal of disease control*, 1992. **146**: p. 1331-1333.
[133]. Gardner PS, et al, Virus infection and intussusception in childhood. *British Medical Journal*, 1962. **2**: p. 692-697.
[134]. Mulcahy DL, Kamath, K.R., et al, A two-part study of the aethiological role of rotavirus in intussusception. *Journal of Medical Virology*, 1982. **9**: p. 51-55.
[135]. Nakagomi T, Rotavirus infection and intussusception: a view from retrospect. *Microbiology and Immunology*, 2000. **44**(8): p. 619-628.
[136]. Rennels MB, Parashar UD, Holman RC, Le CT, Chang HG, and Glass RI, Lack of an apparent association between intussusception and wild or vaccine rotavirus infection. *Paediatric Infectious Disease Journal*, 1998. **17**: p. 924-925.
[137]. Parashar UD, Holman, R.C, Bresee, J.S, et al, Epidemiology of diarrhoea disease among children enrolled in four West Coast health maintenance organisations. *Paediatric Infectious Disease Journal*, 1998. **17**: p. 605-611.
[138]. Ugwu BT, Legbo JN, Dakum NK, Yiltok SJ, Mbah N, and Uba FA, Childhood intussusception: A 9-year review. *Ann Trp Paediatr*, 2000. **20**: p. 131-135.
[139]. Chang EJ, Zangwill, K, Lee,H, et al, Intussusception and rotavirus disease in children (Abstract 845). *Pediatr Res*, 2000. **47 (Suppl S)**: p. 144A.
[140]. Gay N, Ramsay M, and Waight P, Rotavirus vaccination and intussusception. *Lancet*, 1999. **354**(956).
[141]. Hellems MA, Waggoner-Fountain L, and Borowitz SM, Association between intussusception and rotavirus gastroenteritis (Abstract 1562). *Pediatr Res*, 2000. **47 (Suppl S)**: p. 265A.
[142]. WHO/V&B/00.23, *Report of the meeting on future directions for rotavirus vaccine research in developing countries*. 2000, Vaccines and Biologicals, Vaccine Development, World Health Organization: Geneva.
[143]. Anon, Oral poliovirus vaccine (OPV) and intussusception. *Weekly Epidemiological Record*, 2000. **75**: p. 345-347.
[144]. GlaxoSmithKline, *Personal Communication, Protocol 99910/204, GlaxoSmithKline Biologicals*. 2002: Rixensart (Belgium).
[145]. Musonda P, Mounia NH, Whitaker HJ, and Farrington CP, Self-controlled case series analyses:small sample performance. *Computational Statistics and Data Analysis*, 2006: p. Submitted.