Vaccine safety surveillance with self-controlled study designs

Tuomo A. Nieminen

Pro Gradu Thesis
University of Helsinki
Faculty of Social sciences, statistics
June 2017

For the National Institute for Health and Welfare, Finland (THL)

TAMA TULEE LOMAKEMUOTOON

- Tiedekunta: Valtiotieteellinen
- Laitos: Sosiaalitieteiden laitos
- Tekijä: Tuomo Nieminen
- Työn nimi: Vaccine safety surveillance with self-controlled study designs
- Oppiaine: Tilastotiede
- Työn laji: Pro gradu thesis
- Aika: Kesäkuu 2017
- Sivumäärä: 46
- Avainsanat: Drug safety, Vaccine safety surveillance, Sequential analysis, Self-controlled case-series, maxSPRT

Abstract:

All pharmaceutical products, including vaccines, can increase the risk of some undesired medical occurences (adverse events). Evaluating these risks post-licensure is essential for evaluating the safety of vaccines, since rare adverse events might go undetected in pre-licensure studies. Vaccine safety surveillance studies a suspected, biologically plausible causal relationships between a vaccine and an adverse event. Information regarding such relationships are called safety signals. This thesis introduces and applies a method for vaccine safety surveillance, suitable for monitoring the safety of vaccines in near real-time, utilizing electronic health care records.

Vaccine safety surveillance can be seen as an observational study for which different study designs could be used. The popularity of vaccination, self-selection and changes in diagnosis coding practises, along with other possible sources of bias, present challeges for commonly used cohort designs. Self-controlled study designs such as the self-controlled case series (SCCS), eliminate time-invariant confounders and are therefore often more suitable for evaluating vaccine risks. This thesis introduces both a simple and a more general version of SCCS and explicitly describes the assumptions related to the method.

A vaccine safety surveillance method involves a decision rule for generating safety signals. Natural goals of a safety surveillance method are to control the rates of false positive and false negative signals, as well as to generate a signal as soon as possible when an association between the vaccine and the adverse event exists. Statistical hypothesis testing can be used to derive the decision rules. This thesis describes the maximized sequential probability ratio test (maxSPRT), a hypothesis testing method designed for vaccine safety surveillance. The thesis then introduces the BmaxSPRT, a variant of maxSPRT based on a self-controlled study design such as the SCCS. The derivation of the decision rules for BmaxSPRT, including the computation of critical values, is described in detail both mathematically and algorithmically.

The BmaxSPRT method addresses hypotheses concerning the relative incidence of adverse events during specified risk and control periods. As a proof-of-concept BmaxSPRT is retrospectively applied to Finnish register data. The relationship between the incidence of febrile seizures and three childhood vaccines; Measles-

Mumps-Rubella (MMR), Pneumococcal (PCV) and the Rota virus vaccination (Rota) is studied. BmaxSPRT generated an expected safety signal related to MMR; the incidence rate of febrile seizures is higher during a period $0 - 13$ days following MMR vaccination compared to a period $14 - 41$ days following vaccination (relative rate $RR = 1.59$ at the time of signal). The experiment also highlights the need for more in depth analysis regarding PCV vaccinations and febrile seizures. The sensitivity of BmaxSPRT to the specifications of the risk and control periods is also studied in this thesis. The sensitivity analysis highlights the importance of careful consideration of the risk and control periods, by quantifying the loss of power due to poor choices.

Contents

1 Introduction

1.1 Safety surveillance

Pharmaceutical products (drugs) can have unintended side effects. An untoward medical occurrence in a patient administered a drug is called an adverse event. Rare adverse events related to drugs are often impossible to detect in pre-licensure studies and therefore there is an incentive to monitor the safety of a drug post-licensure [Kulldorff et al., 2011]. Monitoring the safety of drugs post-licensure is called safety surveillance.

The information related to a possible causal relationship between a drug and an adverse event is called a safety signal (signal). The interest in safety surveillance is in finding these signals. According to Nelson et al. [2015, p. 179], post-licensure safety aims related to drugs can be classified into three stages:

1. **Signal identification** considers a large number of events and involves detecting signals related to unexpected adverse events.
2. **Signal refinement** considers drug-event pairs suspected to have a causal relationship, based on biological plausibility or a previsouly identified signal. Addresses specified hypotheses related to the pairs.
3. **Signal confirmation** involves a one-time, more in depth study of a previously generated signal.

An assumption about the type of safety surveillance considered in this thesis is that a biologically plausible exposure-event pair has been previously identified. The method to be introduced requires that the relationship between the exposure and the adverse event can be characterized by defining a time interval of possibly increased risk. This places the safety surveillance considered in this work to the refinement stage.

In this thesis the drug of interest is a vaccine product (vaccine). The goal is to introduce and utilize a vaccine safety surveillance method to make decisions related to the generation of safety signals, utilizing electronic medical records.

1.2 Vaccine safety surveillance

In vaccine safety surveillance, the researcher or an automated system monitors the safety of vaccines. If an unexpectedly large number of adverse events are observed, a safety signal is generated. Further action can then be taken, for example a confirmation analysis can be performed.

The observations might not always correspond to the true state of association between the vaccine and the adverse event, leading to false positive or false negative signals. Natural goals of vaccine safety surveillance are to control the expected rates of false positive and false negative signals. It is also desirable to generate a signal as soon as possible, if an association between the vaccine and an adverse event exists.

Some statistical methods designed specifically for vaccine safety surveillance exist. This thesis will focus on the maximized sequential probability ratio test (maxSPRT), which is a statistical hypothesis testing method

designed for vaccine safety surveillance [Kulldorff et al., 2011]. Vaccine safety surveillance based on maxSPRT has been utilized in many countries such as in USA, UK, Taiwan and New Zealand. The main focus has been on studying the effects of influenza vaccines. [Leite et al., 2016]

Statistical hypothesis tests such as the maxSPRT aim to control the expected rate of errors when making decisions under uncertainty. In statistical theory the false positive signals are called type I errors, and the false negative signals are called type II errors. The expected true positive signal rate is called the power.

There are multiple variants of the maxSPRT method such as the Poisson maxSPRT (PmaxSPRT), binomial maxSPRT (BmaxSPRT) and conditional PmaxSPRT, along with further versions designed for handling grouped observations. According to Leite et al. [2016], the most popular variant of maxSPRT has been PmaxSPRT followed by BmaxSPRT. In this work the focus will be on the BmaxSPRT variant, a self-controlled safety surveillance method.

One situation when a safety surveillance method such as the BmaxSPRT could be applied, would be an introduction of a new vaccine to a national vaccination program, such as the introduction of chicken pox vaccination starting in Finland on September 1, 2017. If there are previously identified safety signals (any reasons to suspect a possible causal relation between the vaccine and some rare adverse event), safety surveillance could be initiated to refine these signals, utilizing accumulating data from electronic health records.

1.3 Elements of safety surveillance

In this work I will generally view a safety surveillance method as consisting of three elements: (1) data accumulation (2), study design and (3), decision rule for generating signals. Table 1.1 describes options for the three elements. A unique combination of the elements can be thought to define a unique safety surveillance method. The different variants of maxSPRT can therefore be viewed as unique safety surveillance methods which vary in terms of the underlying study design and assumptions related to data accumulation. They all use (sequential) hypothesis testing as the decision element.

Table 1.1: Elements of Vaccine safety surveillance and different options (separated by comma) related to them.

Element	Options
Data accumulation	Continuously over time, at discrete time points (grouped data)
Study design	Case-only design, cohort design, case-control design
Decision rule	Hypothesis testing

For any reasonably rare adverse events, daily collected data can be considered as continuous observations. In this work I will mainly consider situations where electronic health care data accumulates daily and therefore the focus will be on continuous observations. I will a utilize a case-only design – and specifically the self-controlled case series (SCCS) – as the design element and I will use maxSPRT sequential hypothesis testing as the decision element of the vaccine safety surveillance method. These choices (continuous observations, a case-only design and maxSPRT) lead to the continuous binomial variant of maxSPRT, BmaxSPRT.

1.4 The operationalization of exposure and outcome

The starting point of vaccine safety surveillance is the definition of one or several vaccine, adverse event pairs. In this thesis the biological condition of interest (the adverse event) is operationalized by a group of medical diagnoses related to hospital visits. The medical diagnoses are identified based on a diagnosis code recorded in an electronic medical record database (register). Vaccinations are operationalized similarily from medical records, identified using the vaccine's name or identification code.

Utilizing register data for vaccine safety research is both economical and powerful, because the data collection is practically free and the registers allow for near population level studies. However, it is important to note that registers are secondary data sources – the data are originally collected for other reasons than the research question at hand – and therefore the selection, quality and methods of the data collection are not controlled by the researcher [Sørensen et al., 1996]. The visit to a hospital resulting in a certain diagnosis code might not directly relate to the biological condition of interest.

In Finland, medical diagnoses data are available from electronic health record databases such as The National Hospital Discharge Register (HILMO) and The Register of Primary Health Care Visits (AvoHILMO). Vaccination data are available from the National Vaccination Register (NVR). AvoHILMO, HILMO and NVR all contain time-stamped patient-level information and the data from the registers can be linked using a personal identification code, unique to all individuals in Finland. Diagnoses and vaccinations can thus be linked on an individual level and the relative timing of vaccinations to medical diagnoses can be asserted.

There are two diagnosis classifications used in Finland: the International Classification of Diseases and Related Health Problems, 10th edition (ICD-10) and to lesser extent the International Classification of Primary Care, 2nd edition (ICPC-2). The ICD-10 is much broader than the ICPC-2, but the ICPC-2 can be used alongside the ICD-10 due to cross-tabulation provided by the World Health Organisation [THL, 2011].

Data from AvoHILMO and NVR are available as daily collections and at the time of writing this thesis (2017), data from HILMO are available three times a year. Near real-time vaccine safety surveillance considered in this thesis is a possibility in Finland, utilizing data from AvoHILMO and NVR.

1.5 Outline

This thesis introduces the vaccine safety surveillance method BmaxSPRT in detail by breaking it down to its elements. The study design and decision rules for signal generation are the main focus of this work and are given thorough treatment. BmaxSPRT is also applied to data from HILMO and the sensitivity of the method is studied by simulation. The outline of the thesis is as follows.

- **Chapter 2** discusses the design elements in vaccine safety surveillance, focusing on case-only study designs.
- **Chapter 3** discusses the decision elements in vaccine safety surveillance, focusing on sequential hypothesis testing. Data accumulation and the case of grouped observations is discussed in section 3.4.
- **Chapter 4** offers a proof of concept application of the BmaxSPRT method, utilizing three different vaccine exposures and a single outcome of interest. A simulation experiment studies the effect of design choices to the power of BmaxSPRT.

The thesis ends with conclusions.

2 Study designs for vaccine safety analyses

In vaccine safety surveillance the researcher (or a defined automated system) observes, but does not intervene, with the events that occur. Therefore safety surveillance falls under observational studies. The researcher observes accumulating data related to vaccinations and medical diagnoses and attempts to answer the question: has the rate of adverse events for a vaccinated individual increased due to the exposure.

A natural follow-up question is: increased as compared to what? To answer this question, the researcher needs a *study design* which defines the groups of individuals to be compared. Ideally, the defined groups would only differ with respect to their exposure status, so that observed changes in the rates of events could be directly assigned to the exposure. In practise this is difficult to achieve.

Electronic health care databases such as the hospital discharge register (HILMO) can be utilized as sources of data for medical diagnoses to operationalize the biological condition of interest (adverse event). However, the visit to a hospital resulting in a certain diagnosis code might not directly relate to the adverse event. This can introduce biases when comparing observations from different time periods, due to for example changes in diagnosis coding practises, which are then observed as changes in event rates. The chosen study design should be as robust to these changes as possible.

If the study design is poor and the chosen groups are not comparable, the studied sample can be biased: the observed effect of the exposure may be caused by some other differences between the groups. This means that the rate of false positive or false negative signals generated by the safety surveillance method may be inflated.

In their systematic review of vaccine safety surveillance applications, Leite et al. [2016] found that from 11 recently generated signals, only 3 were confirmed as true signals. The false positive signals were assigned to

1. Confounding factors that have not been considered.
2. Misclassification of the outcome.
3. Changes in the true incidence or coding practises.
4. Inappropriate comparison groups.
5. Uncertainty in background rates.
6. Type I errors. [Leite et al., 2016, p. 234].

All the above reasons except the 2nd one are affected by the study design.

In vaccine safety surveillance, the study design should make it unlikely to introduce biases while enabling the generation of a true signal as reliably as possible. In the following sections, I will first compare possible observational study designs for safety surveillance: Cohort designs, case-control designs and case-only designs. The rest of the chapter will then focus on the two most popular case-only study designs in pharmagovigilance (drug safety): the self-controlled case series (SCCS) and the case-crossover design (CCO).

2.1 Observational epidemiological study designs

Ultimately all epidemiological study designs share the same goal: to compare some group of individuals (cases or exposed) to another group of individuals (controls or unexposed). In vaccine safety surveillance, the question of interest is whether and how the exposure to the vaccine causes an increased risk of an adverse event. The main observations of interest relate to medical diagnoses which occur after exposure. I will consider the following three observational study designs for safety surveillance:

1. **Cohort design**: Time is viewed from the point of view of exposure. Groups of exposed and non-exposed individuals are followed forward in time. The incidences of events are compared between the groups.
2. **Case-control design**: Time is viewed from the point of view of the event. Individuals with events (i.e. cases) are matched to one or more individuals without events (i.e. controls). The proportions of those that have been exposed are compared between the cases and controls.
3. **Case-only designs**: Only individuals with events are sampled and are self-matched, using so called risk and control periods to define the events as either cases or controls. Event incidences or exposure rates are compared between the periods.

The designs have different strengths and weaknesses, some of which are listed in Table 2.1.

Table 2.1: Strengths and weaknesses of different epidemiological study designs.

Design	Strengths	Weaknesses
Cohort design	• Can provide an estimate of the baseline incidence • Utilizes all available data resulting in high power	• If the exposure rate is high, the unexposed group from the same time period will be small • Confounders can affect the distribution of exposure, possibly resulting in bias
Case-control design	• If matched by age and time , controls for time-varying confounders such as age and seasonality • Needs little data specially with rare events	• The unexposed group can be small • Finding suitable matches might be difficult • Confounders can affect the distribution of exposure, possibly resulting in bias
Case-only design	• Self-controlled: eliminates selection biases and confounding related to control subjects and time-invariant characteristics. • Does not need a baseline incidence estimate • Needs only data on cases	• The choice of the risk and control periods plays a crucial role • Time-dependent confounders such as age and seasonality must be explicitly included in the model

2.1.1 Cohort and case-control designs

The cohort design utilizes all available data and is therefore often a preferred choice if data from the whole population are available. However, the nature of the exposure-event pairs considered here poses problems to cohort designs. One problem is that the exposure to vaccination is optional, possibly resulting in a selection bias. A second problem relates to using medical diagnoses as proxies for the underlying biological condition

of interest. I will provide two examples to clarify.

One way to utilize a cohort or a case-control design would be to compare the rates of events between the vaccinated and unvaccinated populations of selected birth cohorts. This would provide a comparison which utilizes all available data from a single time period. The problem with this approach is that vaccination is both optional and usually very popular. Therefore, the group of unvaccinated individuals is often small and possibly biased due to selection. The latter may mean that the observed differences in the rates of events are due to the same reasons that cause individuals to become or not become vaccinated.

A second example of a cohort design would be a comparison of cohorts born before and after the introduction of a new vaccine to a vaccination program. The pre-vaccination era could be used to estimate the baseline incidence of adverse events and the incidence during the post era could then be compared to the baseline. This approach utilizes a great amount of data and provides a seemingly unbiased comparison, assuming that exposure to vaccination is the only significant difference between the cohorts. The problem with this approach is that diagnosis coding practises can change over time due to changes in recommendations or practises. The observed differences in event rates in electronic health care data bases can therefore be simply due to these changes.

2.1.2 Case-only designs

Case-only designs are designs which utilize data only from cases (individuals with the events of interest). In case-only designs, each individual with an event acts as his/her own control and hence the designs are also called self-controlled designs. The self-controlling property can be achieved by splitting the observation period of individuals into periods of "risk" and "control" and then labeling events as either "case" or "control", depending on which period they occur in.

Self-controlled designs have a very desirable property: comparisons are made within the individuals such that all time-invariant confounders are eliminated [Nordmann et al., 2012]. Since self-controlled designs utilize data only from cases, the smaller sample size could decrease power compared to other designs. However, simulation studies by both McClure et al. [2008] and Glanz et al. [2006] have found that self-controlled designs retain relatively high power compared to the cohort and case-control designs despite utilizing less data.

In their systematic review of case-only designs, Nordmann et al. [2012] found that the two most popular case-only designs utilized in pharmacovigilance are the case-crossover design (CCO) and the self-controlled case series design (SCCS). They also remark that "With the development of health care information technology ... these designs seem particularily appropriate to analyze pharmacovigilance data", assuming that the exposure is transient rather than chronic, such as vaccination.

2.1.3 CCO and SCCS

The case-crossover design (CCO) and the self-controlled case series design (SCCS) are two of the most popular case-only designs. Both designs implicitly or explicitly define risk and control periods during which rates of adverse events are of interest. A risk period is a time interval following exposure during which the incidence rate of the adverse event can be thought of being potentially related to the exposure. A control period consists of one or more time intervals during which the incidence rate is assumed to be according to the normal background (baseline) rate of the same individuals.

- **The case-crossover design** (CCO) is a case-only study design derived from a case-control design logic. Time is viewed from the point of view of the event. The random variables analysed are retrospectively ascertained exposures.
- **The self-controlled case series design** (SCCS) is a case-only study design derived from a cohort design logic. Time is viewed from the point of view of the exposure. The ages at vaccination are regarded as fixed, and the random variable of interest is age at event, conditionally on its occurrence within a pre-determined observation period. [Farrington, 2004, pp. 2066-2067].

In CCO and SCCS, events occuring during a risk period can be labeled as "cases" and events during the control period as "controls." Both designs aim at comparing the incidence rates of adverse events between the different periods, "within individuals". Individuals implicitly act as both cases and controls in the study because the person-time of the individuals is used to define both the cases and the controls.

From a statistical inference point of view, incidence rates can be compared with their ratio, called the *rate ratio*, denoted by RR. The rate ratio between two constant incidence rates λ_1 and λ_0 is

$$RR = \frac{\lambda_1}{\lambda_0}. \tag{1}$$

The rate ratio is the parameter of interest in both the CCO and SCCS. The SCCS approaches the problem by directly considering the rates of events during different periods of risk (e.g. a risk and a control period). CCO approaches the problem "backwards" by considering the rates of exposure instead.

CCO and SCCS are both suitable for studying rare adverse events. According to Farrington [2004], both CCO and SCCS can and have been used for vaccine safety analysis. According to Maclure et al. [2012], they are better than cohort designs for investigating transient effects of vaccines. What now follows is a detailed description of the SCCS method followed by a description of CCO.

2.2 The self-controlled case series design: SCCS

In the self-controlled case series design (SCCS) developed by Farrington [1995], the goal is to compare the relative rates of adverse events during disjoint risk and control time periods, specified in relation to the time

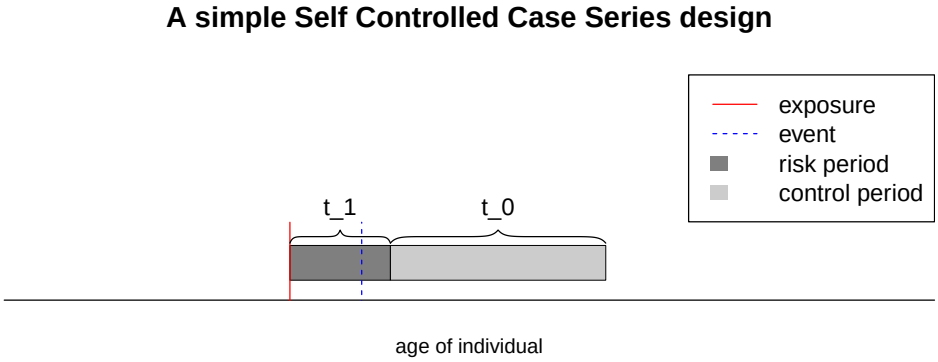


Figure 2.1: A simple example of a SCCS design where the period immediately following exposure is defined as the risk period with length t_1 and a time period following the risk period is defined as the control period with length t_0. In this example, the event occurs during the risk period

of vaccination. The effect of the vaccination is estimated via the rate ratio $RR = \lambda_1/\lambda_0$, where in a simple case

- λ_1: the rate of the event during a period when the exposure to vaccination is assumed to have an effect (risk period)
- λ_0: the rate of the event during a period when the vaccination is assumed to not have an effect (control period).

If there is a significant change in the incidence during the risk period, the rate ratio will differ from 1, with values greater than 1 indicating a positive association between the exposure and the event. Values less than 1 indicate a negative association and are usually not of interest in safety surveillance.

The risk period(s) are defined based on previous knowledge regarding the relationship between the vaccine and the adverse event. In a simple case a single risk period could be a period of two weeks immediately following vaccination. A control period can also be explicitly defined, or alternatively all other time during the observation period can be considered as the control period. See Figure 2.1 for an illustration of the former.

The general idea of SCCS is to describe the event incidence as a function of the (constant) individual baseline incidence, effects due to vaccine exposure and other possible time-dependent factors such as age. Statistical inference in SCCS is based on a conditional likelihood which removes the individual incidence as a nuisance parameter. Inference concerning relative incidences (rate ratios) can therefore be done "within individuals" so that the individual baseline incidences have no impact on the inference.

2.2.1 Observation period

In SCCS, the observation period for individual i is the time period during which an event could be sampled [Whitaker et al., 2006]. The observation periods could for example be based on a calendar time period and an age interval. This definition can lead to observation periods of different lengths, as the start or end of the

calendar time period might shorten the observation period of some individuals.

In this work I will use birth years of individuals to specify eligibility to the study. The observation period for the eligible individuals is based on an age interval during which events can be sampled. The smallest measurable unit for time is a day, since data in electronic health care data bases such as AvoHILMO are collected daily. Therefore, for each individual, I will index the observation period by age in days denoted by d. A single day has a time unit of 1. If not explicitly mentioned, the reader can assume that all individuals considered are eligible.

2.2.2 Statistical framework

In what follows is a mathematical description of SCCS largely inspired by the description of the general SCCS given in the case series tutorial by Whitaker et al. [2006, pp. 9-10] with influences from the description of SCCS by Bernardo et al. [2011, pp. 10-12]. I will first give a general framework for the SCCS method by introducing the concepts of a Poisson process and incidence rate function and then describe the conditional likelihood for the parameters of that function.

2.2.2.1 Poisson process

An underlying assumption in the SCCS method is that adverse events occur as a Poisson process. A Poisson process describes the probabilities for the numbers of events during some time interval(s). An important assumption related to the Poisson process is that the random variables corresponding to these numbers of events are independent between all disjoint time intervals. This implies that an occurrence of an adverse event should not influence the probabilities of future adverse events. This assumption is likely to be violated in practise, but if the event is rare (relative to the length of the observation period), then this violation is unlikely to have practical consequences.

Definition (Poisson process)**.** *A counting process [Ross, 1996]* $\{Y(t), t \geq 0\}$ *is a Poisson process with rate* λ *if*

1. $Y(0) = 0$

2. $Y(t)$ *has independent increments*

3. The probabilities for the number of events during an interval of length t *are given by the Poisson distribution with* $E[Y(t)] = \lambda t$*:*

$$y_t \sim Poisson(\lambda t) \Leftrightarrow P(Y(t) = y_t) = \frac{(\lambda t)^{y_t} \cdot e^{-\lambda t}}{y_t!}$$

2.2.2.2 Incidence rate function

An assumption is made that during the observation period, adverse events for individuals $i = 1, 2, ..., N$ arrive as a non-homogenous piecewise-constant Poisson process. The expected number of events (the incidence rate) for each individual is constant within each day, but can vary between days.

The incidence rate during day d for individual i is given by the incidence rate function λ_{id}, which depends on a constant individual effect ϕ_i and a time-dependent effect $\boldsymbol{\theta}$. The latter captures exposures to vaccination and can also capture factors such as age and seasonality. The effect of θ_d is assumed to be multiplicative, so that the incidence rate for individual i during day d is given by

$$\lambda_{id} = exp(\phi_i + \theta_d). \tag{2}$$

Due to the properties of the Poisson process the cumulative incidence over a number of days during the observation period is a sum of the daily incidences. If the incidence stays constant for a period of consequent, say, t_q days, the incidence during that period is given by the number of days times the daily incidence. Assuming that the observation period consists of such disjoint periods, the incidence rates are then given by

$$\lambda_{iq} = t_q \cdot exp(\phi_i + \theta_q), \tag{3}$$

where θ_q is constant for all d days in period q. For example in a simple case as in Figure 2.1, where $q \in \{\text{risk period}, \text{control period}\}$, the incidence rate (3) for any one individual takes on two different values.

2.2.2.3 Likelihood

Assume now that events of $i = 1, 2, ..., N$ individuals are sampled, with $n_i \geq 1$ events for each i for a total of n events. Assume that events during the observation period arrive as a Poisson process with an incidence rate as defined in (3). Denote the number of events for individual i during period q as y_{iq}. The probabilities for the numbers of events during the disjoint intervals q are given by the Poisson distribution: $y_{iq} \sim Poisson(\lambda_{iq})$. The individual contribution to the likelihood function, based on the obsevations, is therefore

$$L_i(\boldsymbol{\theta}) = P(\boldsymbol{y}_i \mid \phi_i, \boldsymbol{\theta}) = \prod_q P(y_{iq} \mid \phi_i, \theta_q) \propto \exp\{-\sum_q \lambda_{iq}\} \prod_q (\lambda_{iq})^{y_{iq}}, \tag{4}$$

where $\lambda_{iq} = t_q \cdot exp(\phi_i + \theta_q)$ as in (3). The likelihood contains the individual baseline incidence ϕ_i which is a nuisance parameter, since the real interest is in the effect of the vaccination and/or age, captured by $\boldsymbol{\theta}$.

The total number of events for the individual, $n_i = \sum_q y_{iq}$, turns out to be a sufficient statistic for ϕ_i, meaning that conditioning on n_i removes the dependency on ϕ_i. It is therefore convenient to operate with the conditional likelihood instead. Since the random variable corresponding to the total number of events is a sum of Poisson random variables, it is also Poisson distributed: $n_i \sim Poisson(\sum_q \lambda_{iq})$. The individual contribution to the conditional likelihood is

$$L_i(\boldsymbol{\theta} \mid n_i) = P(\mathbf{y}_i \mid n_i, \phi_i, \boldsymbol{\theta}) = \frac{P(\mathbf{y}_i \mid \phi_i, \boldsymbol{\theta})}{P(n_i \mid \phi_i, \boldsymbol{\theta})} \propto \frac{\prod_q \lambda_{iq}^{y_{iq}}}{\left(\sum_q \lambda_{iq}\right)^{\sum_{q'} y_{iq'}}} = \prod_q \left(\frac{\lambda_{iq}}{\sum_{q'} \lambda_{iq'}}\right)^{y_{iq}}. \tag{5}$$

Now, plugging in the incidence rate defined by (3), the conditional likelihood is

$$L_i(\boldsymbol{\theta} \mid n_i) \propto \prod_q \left(\frac{t_q \cdot e^{\phi_i} \cdot e^{\theta_q}}{\sum_{q'} t_{q'} \cdot e^{\phi_i} \cdot e^{\theta_{q'}}} \right)^{y_{iq}} = \prod_q \left(\frac{t_q \cdot e^{\theta_q}}{\sum_{q'} t_{q'} \cdot e^{\theta_{q'}}} \right)^{y_{iq}}, \tag{6}$$

which is a multinomial likelihood where the individual baseline incidences have been canceled out. This means that using the conditional likelihood, the inference concerning the parameter $\boldsymbol{\theta}$ cannot be influenced by the individual parameters ϕ_i. In that sense the SCCS design is self-controlled. However, it should be noted that all factors which depend on the time period q, such as age, remain. The effect of age can be separately included in the model in the general version of SCCS.

Next I will present the simple and the general SCCS which are both based on the assumptions of a piecewise-constant Poisson process and a conditional likelihood. The simple SCCS is particularily useful for the safety surveillance method BmaxSPRT, to be introduced later. The general SCCS is relevant for future research in developing safety surveillance methods.

2.2.3 Simple SCCS

Consider an SCCS with a simple incidence rate function where there are two periods of interest: a risk period and a reference control period. Denote these periods by the indicator $k \in \{0, 1\}$ where 1 indicates the risk period and 0 indicates the control period.

The risk and control periods are disjoint and of lengths t_1 and t_0 respectively, and are defined in relation to the time of exposure to vaccination. For example, a time period immediately following vaccination can be defined as a risk period and a time period following the risk period can be defined as a control period, as illustrated in Figure 2.1. The incidence rate function (3) takes the form

$$\lambda_{ik} = t_k \cdot exp(\phi_i + \psi_k), \quad k = 0, 1. \tag{7}$$

Assume that during the control period $\psi_0 = 0$ and the incidence is simply determined by the baseline incidence of the individual and by (7) is given by $t_0 \cdot e^{\phi_i}$. The incidence during the risk period is given by $t_1 \cdot e^{\phi_i} e^{\psi_1}$. The parameter of interest is ψ_1 which quantifies the multiplicative effect that exposure has on the incidence, as now $RR = \lambda_{i1}/\lambda_{i0} = e^{\psi_1}$ is the rate ratio of the incidence rates during the risk and control periods.

Now assume that events of $i = 1, 2, ..., N$ individuals are identified. Denote the number of events for individual i during the risk period as y_{i1} and during the control period as y_{i0}. Each individual has $n_i = y_{i1} + y_{i0}$ events during the periods of interest and the total number of events during the periods is $n = \sum_i n_i$. According to equation (6), the individual contribution to the conditional likelihood for the parameter ψ_1 is

$$L_i(\psi_1 \mid n_i) \propto \prod_{k=0,1} \left(\frac{t_k \cdot e^{\psi_k}}{\sum_{k'} t_{k'} \cdot e^{\psi_{k'}}} \right)^{y_{ik}} = \left(\frac{e^{\psi_1}}{e^{\psi_1} + t_0/t_1} \right)^{y_{i1}} \left(\frac{t_0/t_1}{e^{\psi_1} + t_0/t_1} \right)^{n_i - y_{i1}}. \tag{8}$$

which is the binomial likelihood, where the only unknown parameter is the common rate ratio $RR = e^{\psi_1}$. If one labels the events that occured during the risk period as "cases", and the events that occured during the control period as "controls", the probability of an event labeled as a case depends only on RR and the ratio of the lengths of the two periods $z = t_0/t_1$. The probabilities of a "case" ("success") and "control" ("failure") are given by

$$P(\text{"}case\text{"}) = p = \frac{RR}{RR+z}, \quad P(\text{"}control\text{"}) = 1 - p = \frac{z}{RR+z}. \tag{9}$$

Assuming that individuals are independent, the total conditional likelihood can be seen as independent binomial trials with a homogenous probability of success, $p = RR/(RR+z)$. Given that observations included $y_1 = \sum_i y_{i1}$ "cases", the total conditional likelihood for the parameter RR is

$$L(RR \mid n) = P(y_1 \mid RR, n) \propto \left(\frac{RR}{RR+z}\right)^{y_1} \left(\frac{z}{RR+z}\right)^{n-y_1}. \tag{10}$$

2.2.4 General SCCS

I will now present a more general version of SCCS, including an age effect parameter and multiple risk periods with varying risks. Varying risks can be motivated by for example the desire to include the effects of different doses of a vaccine. The inclusion of an age effect is motivated by the fact that many biological conditions are age-dependent.

Assume again that adverse events during the observation period arrive according to a piecewise-constant Poisson process with an incidence rate function similar to (3). The observation period for each individual is split into disjoint intervals by age groups indexed by j and disjoint risk periods indexed by k. The number of days that individual i spends in risk/control interval k while belonging to age group j is denoted by t_{ijk}. The incidence rate function for i can then be written as

$$\lambda_{ijk} = t_{ijk} \cdot exp(\phi_i + \gamma_j + \psi_k), \tag{11}$$

where for the reference age group $\gamma_0 = 0$ and for the reference risk period (control period) $\psi_0 = 0$. Now denote the number of events observed during interval (jk) by y_{ijk}. The individual contribution to the conditional likehood (5) is

$$L_i(\boldsymbol{\psi}, \boldsymbol{\gamma} \mid n_i) \propto \prod_{jk} \left(\frac{t_{ijk} \cdot e^{\gamma_j} \cdot e^{\psi_k}}{\sum_{j'k'} t_{ij'k'} \cdot e^{\gamma_{j'}} \cdot e^{\psi_{k'}}}\right)^{y_{ijk}}. \tag{12}$$

Assuming that individuals are independent, the total conditional likelihood is

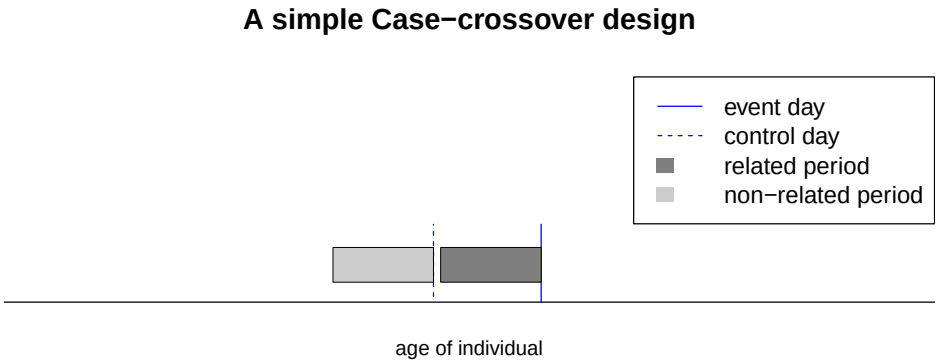


Figure 2.2: A graphical illustration of the CCO design. Exposures occuring immediately before the event day are considered as related to the event and exposures occuring immediately before the control day are considered as not related to the event.

$$L(\boldsymbol{\psi}, \boldsymbol{\gamma} \mid n) \propto \prod_i \prod_{jk} \left(\frac{t_{ijk} \cdot e^{\gamma_j} \cdot e^{\psi_k}}{\sum_{j'k'} t_{ij'k'} \cdot e^{\gamma_{j'}} \cdot e^{\psi_{k'}}} \right)^{y_{ijk}}, \tag{13}$$

which, again, is the multinomial likelihood where all the individual baseline incidences have been cancelled out.

2.3 The case-crossover design: CCO

The most popular case-only design utilised in pharmacovigilance is the case-crossover design (CCO). CCO is a modification of a case-control design, originally developed by Maclure [1991] to study the effect of transient exposures to acute events. Its use in drug safety studies has increased during 2000-2010 [Nordmann et al., 2012]. CCO has also been used for vaccine safety analysis [Farrington, 2004, p. 2066].

2.3.1 Description

Following Maclure et al. [2012, pp. 50-51], one way to describe the CCO design is as follows: Assume that rare events are observed for individuals $i = 1, .., N$. Each event occurs during some *event day* and a *control day* is defined as the day d time units (e.g. days) before the event day. For example, the control day could be two weeks (14 days) before the event day. For each sampled individual i, exposure status is retrospectively examined. Exposures during a time interval of length t immediately preceding the event day are considered as related to the event. Exposures occuring during a time interval of the same length (t) immediately preceding the control day are considered as not related to the event.

The periods where the exposure status is of interest are sometimes called risk and control periods in CCO literature. I will call these periods the *related period* and the *non-related period*, respectively. The idea of the CCO design is to compare the rates of exposure between the related and the non-related periods. This description of CCO is illustrated in Figure 2.2.

2.3.2 Prospective and retrospective arguments

The description of CCO given above follows the so-called retrospective conditional argument of case-control studies. In the retrospective probability model, the exposures are conditioned by the case or control status and the odds ratio of exposure between different event status ("case" or "control") is the parameter of interest. However, the estimation of disease-exposure relationship in case-control studies may also be approached using a prospective conditional argument. In the prospective probability model, the event status is conditioned on the exposure and the odds ratio of the event status between levels of exposure is estimated [Clayton and Hills, 1993, ch. 16].

From a prospective point of view, a time interval following exposure is called the *effect period* in CCO literature, during which the rate of events for the exposed individuals is possibly increased. The effect period corresponds to the related period and is of identical length. The effect period in CCO is comparable to the risk period in SCCS. I will therefore use the term risk period to refer to the effect period. Similarily, a control period of identical length to the non-related period can be defined.

For example, the risk period could be a period of two weeks following vaccination, in which case the related period is the period two weeks prior to the event day. A control period could be defined as the period of two weeks following the risk period, in which case the control day would be the day two weeks prior to the event day and the non-related period would be the period two weeks prior to the control day. See again Figure 2.2 for an illustration.

The question of interest in both CCO and SCSS is the same, i.e., to infer whether the rate of events is affected by exposure. Although the CCO method is based on the idea of exposures as the random variable, the parameter of interest can be defined as any of the prospective relative risk parameters: the odds ratio, relative risk or the incidence rate ratio [Nordmann et al., 2012, p.8].

2.3.3 CCO in the literature

The literature concerning the CCO approach may be confusing as the reader may have to interpret from the context whether the author has chosen to treat the design from a prospective or retrospective point of view and sometimes both views are used simultaneously. For example, Nordmann et al. [2012, p. 2] describe the method using phrasing *"With this design, the probability of exposure in the risk period is compared to the probability of exposure in control period(s)."*. They however then go on to state that *"same probability of event occurrence during case and control periods"* is a major assumption of the method.

When discussing CCO, Farrington [2004] instead writes that *"the underlying probability of vaccination should be the same in all intervals"*, which is in line wih treating exposure as the random variable of interest. Farrington [2004, p. 2066] then adds that this assumption *"is unlikely to hold for paediatric vaccines administered according to stricter schedules"*.

To estimate the relative risk in the CCO setting, Maclure [1991] suggests using the Mantel-Haenzel estimator

to approximate the rate ratio parameter. Further discussion of maximum likelihood estimation in the CCO setting is provided by Marshall and Jackson [1993].

2.4 Study design conclusions

Cohort designs are problematic for vaccine safety studies due to many possible sources of bias. Case-only designs are attractive because they implicitly control for time-invariant confounders.

Of the two most popular case only designs, SCCS compares somewhat favourably to CCO for vaccine safety analyses. Farrington [2004, p. 2066] has pointed out that in the case of paediatric vaccines there is a conceptual problem with the CCO design, where vaccination times are treated as random variables. It is more natural to view event times as the random variables of interest, since vaccinations often occur according to a strict schedule.

SCCS assumes that the vaccination time is fixed and the time of the adverse event is a random variable of interest. SCCS allows for direct inference concerning the population parameter of interest, which is the relative rate of adverse event incidences between specified risk and control time periods. The simple SCCS introduced in 2.2.3 offers a statistical model where the rate ratio RR is the only unknown parameter and fixed individual covariates have no effect on the inference. The likelihood function is binomial.

The following chapter continues discussion of safety surveillance from the point of view of the decision element. The simple SCCS is adopted as the design element.

3 Decision rules for vaccine safety surveillance

In vaccine safety surveillance, the main observations of interest are medical diagnoses used as proxies for the biological condition of interest (adverse event). Gradually accumulating diagnoses form the sample to be analyzed and decisions related to safety signal generation are made based on the sample. The goal is to stop observing and take some further action if an unexpectedly high number of adverse events have been observed. Dynamically determing what exactly is "unexpectedly high", is a problem that this chapter aims to solve.

One of the key elements of a safety surveillance method is a decision rule for generating safety signals. The available sample never perfectly represents the total study population of interest, which can be thought to include also unborn individuals, possibly experiencing adverse events in the future. Furthermore, there is random variation in the occurrence of adverse events in the population. The decision making process in safety surveillance involves uncertainty and inevitably some amount of error is expected.

A desirable quality of a safety surveillance method is to control the expected rate of false positive and false negative signals. When a true association between a vaccination and an adverse event exists, another desirable quality is to generate the signal as soon as possible. In other words, a natural goal is to minimize the number

of adverse events needed to generate the signal, given some fixed rates of false positive and false negative signals.

This chapter discusses sequential hypothesis testing, which is a statistical method that can be utilized for vaccine safety surveillance. Sequential hypothesis testing provides the means to both control the expected rates of false positive and false negative safety signals, as well as to minimize the expected time to signal generation.

3.1 Hypothesis tests in vaccine safety surveillance

Statistical hypothesis testing is a method of statistical inference for decision making under uncertainty. Uncertainty usually arises from the fact that available observations do not cover the whole population of interest, but rather are a sample from that population. In a regular hypothesis testing scenario, there is a single fixed sample and a single hypothesis test is performed to reach a conclusion. The setting in safety surveillance is different, since observations arrive sequentially.

In the following sections, I will discuss the use of hypothesis testing for deriving decision rules for safety signal generation. I will first describe hypothesis testing in general and then discuss testing with sequential observations.

3.1.1 Statistical hypotheses

Statistical inference is based on on a family of probability models $P(Y|\boldsymbol{\theta})$ for the observations Y, indexed by the parameters $\boldsymbol{\theta}$. A statistical hypothesis is a proposition which assigns restrictions for the parameter of the statistical model. Usually there are two hypotheses: the null hypothesis H_0 and the alternative hypothesis H_1. These are expressed by

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad H_1 : \boldsymbol{\theta} \in \Theta_1,$$

where Θ_0 and Θ_1 are disjoint and usually $\Theta_0 \cup \Theta_1 = \Theta$, so that the two hypothesis together cover the whole parameter space (all possible values of the parameter). For short I will denote $\boldsymbol{\theta} \in \Theta_0$ as $\boldsymbol{\theta}_0$ and similarily for $\boldsymbol{\theta}_1$.

A hypothesis is called simple if it addresses only a single point in the parameter space. It is common that the null hypothesis is simple, i.e. $H_0 : \theta = \theta_0$. In many applications such as in vaccine safety, the most interesting alternative hypothesis is of the form $H_1 : \theta > \theta_0$. This type of hypothesis – which addresses more than a single point in the parameter space – is called a composite hypothesis.

When the objective is to find evidence of an association between an exposure and an event (for example that the rate ratio for exposed and not exposed is positive), the null proposition (hypothesis) is a state of

no association. The alternative hypothesis is composite: some positive association. Using the rate ratio parameter RR, this can be stated as two competing hypotheses as follows:

$$H_0 : RR = 1 \quad H_1 : RR > 1 \tag{14}$$

3.1.2 Hypothesis testing

In statistical hypothesis testing, two statistical models $P(\mathbf{Y} \mid \boldsymbol{\theta}_0)$ and $P(\mathbf{Y} \mid \boldsymbol{\theta}_1)$, defined by the competing hypotheses H_0 and H_1, are compared. Usually the comparison is done by defining a test statistic $T = T(\mathbf{y})$ for which high values are critical to the null hypothesis and the probability distribution of T is known under H_0. If the observed value for T is very unlikely when H_0 were true, then one chooses to reject H_0.

The test statistic can be for example the likelihood ratio:

$$LR = \frac{L(\boldsymbol{\theta}_1; \mathbf{y})}{L(\boldsymbol{\theta}_0; \mathbf{y})} = \frac{P(\mathbf{y} \mid \boldsymbol{\theta}_1)}{P(\mathbf{y} \mid \boldsymbol{\theta}_0)}. \tag{15}$$

Any value of the likelihood ratio greater than 1 means that the observations are more likely under the alternative hypothesis ($P(\mathbf{y} \mid \boldsymbol{\theta}_1) > P(\mathbf{y} \mid \boldsymbol{\theta}_0)$).

3.1.3 Errors and power

When a decision is made between two competing hypothesis, two types of errors can be made:

- Type I error: H_1 is chosen (H_0 rejected) when the H_0 is true.
- Type II error: H_0 is chosen (H_0 accepted) when the H_1 is true.

The probabilities of type I and II errors are usually denoted by α and β, respectively. Terminology related to hypothesis testing is displayed in Table 3.1.

The complement of β is called the power of the hypothesis test: the probability of choosing H_1 (rejecting H_0) when H_1 is true. If the hypotheses are as in (14) and the parameter of interest is RR, the power of a hypothesis test is usually a function of RR, meaning that higher values of RR make it more likely for the test to reject the null hypothesis. This also means that the actual type II error rate is a decreasing function of RR.

In the context of vaccine safety surveillance the above means that one would be more likely to conclude that there is a difference in the rate of events between two groups, the bigger that difference is. This is of course a positive observation and a desirable quality. But it is also important because it leads to the following: knowing the type II error rate or the power of a test before collecting observations is impossible without knowledge of the actual RR.

Table 3.1: Terminology related to statistical hypothesis testing

Term	Description
H_0	The null hypothesis ($H_0 : \boldsymbol{\theta} \in \Theta_0$)
H_1	The alternative hypothesis ($H_1 : \boldsymbol{\theta} \in \Theta_1$)
Composite hypothesis	A hypothesis which adresses more than a single point in the parameter space. For example $H_1 : \theta > 1$
α	P(reject $H_0 \mid H_0$). The type I error rate (false positive rate)
β	P(accept $H_0 \mid H_1$). The type II error rate (false negative rate)
Power	$1 - \beta$

3.2 Sequential analysis

From a statistical inference point of view, a situtation where observations accumulate gradually is different from a more common situation with a fixed number of observations. With a fixed bumber of observations, one can perform a single statistical hypothesis test and make a single decision. With accumulating data, multiple such tests can be done. Sequential analysis focuses on situations where there is a need to perform an analysis whenever new observations arrive. Sequential hypothesis tests can be used to derive decision rules for each new observation.

A naive approach to sequential analysis would be to perform a standard hypothesis test on the accumulating data set each time new observations become available. Hypothesis testing is based on the general idea that if observations are unlikely under a statistical model defined by H_0, then some assumptions of that model can be questioned. However, if an experiment is repeated multiple times, then even very unlikely outcomes of the experiment are more likely to be observed during at least one of the trials. Repeated analysis of accumulating data creates a problem, since repetition introduces multiple chances to reject the H_0. The naive approach needs adjustment: in a situation of accumulating data, methods designed specifically for sequential analysis are needed.

Sequential analysis, developed by Wald [1945] during the second world war, adresses the problem of hypothesis testing in a situation where observations arrive sequentially. Wald defined a sequential test of a statistical hypothesis as a test procedure which gives a rule of making one of three possible decisions at a single trial of the experiment:

1. Accept the null hypothesis.
2. Reject the null hypothesis.
3. Continue the experiment by making an additional observation.

Wald [1945] developed the sequential probability ratio test (SPRT), which is the optimal sequential hypothesis test between two simple hypotheses [Wald and Wolfowitz, 1948].

SPRT has since been extended to adress composite hypotheses with so called sequential generalized probability

ratio tests. Kulldorff et al. [2011] introduced a version of such a test, called the maxSPRT, designed for vaccine safety surveillance. I will now briefly introduce SPRT and then describe in detail the maxSPRT and its self-controlled binomial variant BmaxSPRT.

3.2.1 Sequential probability ratio test: SPRT

Wald's [1945] sequential probability ratio test (SPRT) is a sequential hypothesis test designed for testing two simple hypotheses. SPRT is based on the likelihood ratio:

$$LR_n = \frac{L(\boldsymbol{\theta}_1; \mathbf{y}_n)}{L(\boldsymbol{\theta}_0; \mathbf{y}_n)} = \frac{P(\mathbf{y}_n \mid \boldsymbol{\theta}_1)}{P(\mathbf{y}_n \mid \boldsymbol{\theta}_0)}, \tag{16}$$

where $\mathbf{y}_n$ are the current observations for $n = 1, 2, ...$ The SPRT procedure is described in Algorithm 3.2.1. Even though there is no defined upper limit to the number of observations, Wald [1945, p. 128] proved that the SPRT experiment will eventually terminate with probability 1. Wald and Wolfowitz [1948] also showed that SPRT is the optimal sequential test for testing a simple null hypothesis against a simple alternative, in the sense that it has the lowest expected sample size among tests of equal power.

Algorithm 3.2.1: SPRT

Input: Desired Type I and Type II error rates α and β, test statistic function LR_n as in (16)

1. Compute critical upper and lower boundaries $A = (1 - \beta)/\alpha$ and $B = (1 - \alpha)/\beta$.
2. After observation n = 1, 2, ... do

- if $LR_n \geq A$ stop, reject H_0
- if $LR_n \leq B$ stop, reject H_1
- otherwise draw an additional observation.

3.2.2 Critical values of SPRT

The SPRT test has two critical regions: $Q_0 = \{LR_n \mid LR_n \geq A, B < LR_{n-1} < A\}$ for rejecting H_0 and $Q_1 = \{LR_n \mid LR_n \leq B, B < LR_{n-1} < A\}$ for rejecting H_1 (accepting H_0), defined by the two critical values A and B. These regions define the stopping criteria for the test for all observations $n = 1, 2, ...$ The probability α of rejecting the null hypothesis when it is true (type I error) is given by $P(Q_0 \mid H_0)$ and the probability β of rejecting the alternative hypothesis when it is true (type II error) is given by $P(Q_1 \mid H_1)$. The SPRT procedure and the critical regions are illustrated in Figure 3.1.

The critical values A and B should be chosen to match desired (low) error rates (probabilities). In order to determine the values A and B, Wald considered the entire sample space consisting of all possible realisations in the sequential test. He showed that A and B can be approximated by functions of the desired α and β regardless of the statistical model by setting $A = (1 - \beta)/\alpha$ and $B = (1 - \alpha)/\beta$. He also remarked that this

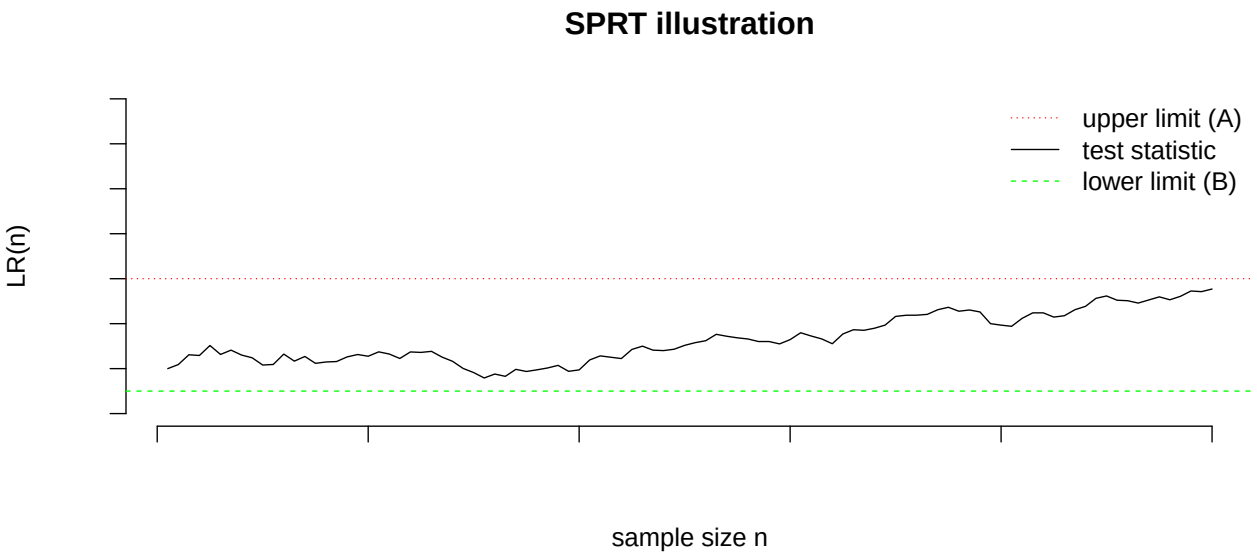


Figure 3.1: A graphical illustration of the critical boundaries of the sequential probability ratio test (SPRT) where the test statistic is the likelihood ratio (LR). If the numerator of LR is the likelihood according to the alternative hypothesis, then high values of LR mean that the alternative model fits the data better and are thus critical to the null hypothesis. The upper limit A (dashed red line) then defines the critical region for the null hypothesis. Similarily, the lower limit B (dashed green line) defines the critical region for the alternative hypothesis.

procedure will guarantee that the actual type I and II errors will not exceed α and β and will only differ from them slightly [Wald, 1945, 127-133].

3.2.3 SPRT and a composite alternative hypothesis

SPRT was designed for testing two simple hypotheses. Wald did however propose a solution to deal with composite alternative hypotheses. In this section I will introduce this solution and then illustrate a weakness related to it.

Wald [1945, p. 158] remarked that in common statistical models the power of the SPRT test is an increasing function of the parameter of interest θ. He therefore suggested dealing with a composite alternative hypothesis by simply defining a value θ_1 such that the difference $\theta_1 - \theta_0$ would be of significant interest in the application and then setting a point alternative hypothesis $H_1 : \theta = \theta_1$. Then one could simply utilize SPRT as described in algorithm 3.2.1 and test a simple null hypothesis against a simple alternative.

When the parameter of interest is the rate ratio, one example of the strategy above would be to view rate ratios $1 \leq RR < 1.2$ as of no interest and therefore for example set $H_1 : RR = 1.2$. Kulldorff et al. [2011] remarked that an unfortunate relation between the choice of RR_1 and the actual (i.e. true) RR_a can either

1. delay the rejection of H_0 when the actual rate of events is higher than the rate suggested by H_1 (delayed signal generation)
2. increase the type II error when the actual rate of events is closer to the rate suggested by H_0 than the

rate suggested by H_1 (decreased power).

In other words, scenario (1) can happen when $RR_0 < RR_1 << RR_a$ and scenario (2) when $RR_0 < RR_a << RR_1$.

The intuition leading to (1) is that if RR_a is far from both RR_0 and RR_1, then both models, defined by H_0 and H_1, are inappropriate and therefore $P(\mathbf{y} \mid RR_1)$ and $P(\mathbf{y} \mid RR_0)$ will on average be close to each other. For example, if one specifies $RR_0 = 1$ and $RR_1 = 1.2$ when in reality $RR_a = 6$, then both H_0 and H_1 specify inappropriate models. In such a case the high number of adverse events expected to be observed would be given low probability by both models, as illustrated in the left panel of Figure 3.2.

In the example described above the two likelihoods would remain close to each other and the likelihood ratio would remain close to one. It might take a large number of samples to reach a point where $LR \geq A$ and H_0 is rejected. This is clearly undesirable especially with serious adverse events, where it is desirable that an unexpectedly high number of adverse events would lead to a quick decision to reject H_0.

To see the intuition behind scenario (2), assume again that we are interested to find if $RR > 1$. Following Wald's suggestion one might for example choose $RR_0 = 1$ and $RR_1 = 2$, when $RR_a = 1.4$. In this case the model specified by H_0 is closer to the real value of RR and it is thus expected that $P(\mathbf{y} \mid RR_0) > P(\mathbf{y} \mid RR_1)$, making it more likely that $LR \leq B$ (H_1 rejected). Thus the type II error is increased. See the right panel of Figure 3.2 for an illustration.

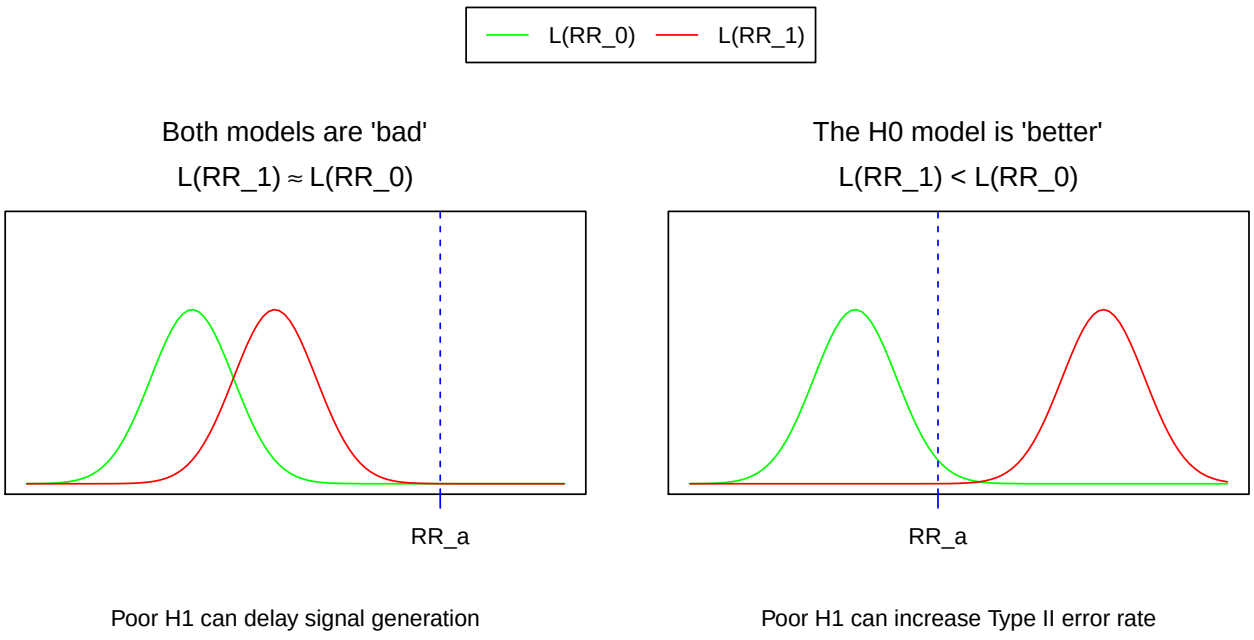


Figure 3.2: A graphical illustration of a weakness of Wald's SPRT. The blue line describes the actual relative incidence rate (RR_a) and the green and red lines describe the likelihood under the null and alternative hypothesis, respectively. Left: A choice of point alternative hypothesis (RR_1) close to the rate specified by the null hypothesis (RR_0) and far from RR_a can delay the rejection of H0, because a high relative rate is given low likelihood under both hypotheses. Right: A choice of point alternative hypothesis that is 'too agressive' compared to the actual relative rate can increase the type II error rate, because the null hypothesis gives higher likelihood to the actual relative rate.

3.2.4 Maximized sequential probability ratio test: maxSPRT

A solution suggested by Kulldorff et al. [2011] to remove the weakness of SPRT as described above is to modify the test in two ways:

1. Maximize the likelihood ratio in the space of the alternative hypotheses Θ_1.
2. Instead of setting a lower bound B to reject H_1, define a maximum number of observations N and reject H_1 if $n \geq N$.

The modified sequential test is called the maximized sequential probability ratio test (maxSPRT). The maxSPRT is a general sequential hypothesis testing method, which can be used on any statistical model.

Kulldorff et al. [2011] introduced two versions of the maxSPRT method: one based on Poisson likelihood and the other on binomial likelihood. The binomial model arises when the study design is a simple self-controlled design such as the simple SCCS introduced in section 2.2.3.

When the parameter of interest is the rate ratio RR, the maxSPRT test statistic is

$$LR_n = \max_{RR_1} \frac{L(RR_1; \mathbf{y}_n)}{L(RR_0; \mathbf{y}_n)} = \max_{RR_1} \frac{P(\mathbf{y}_n \mid RR_1)}{P(\mathbf{y}_n \mid RR_0)}. \tag{17}$$

The maxSPRT procedure is described in Algorithm 3.2.4. The procedure uses a test statistic such as (17) and requires that the desired type I error rate and the maximum number of observations are chosen before the experiment is carried out.

Algorithm 3.2.4: MaxSPRT

Input: Desired type I error rate α', upper boundary for the sample size N, test statistic function LR_n as in (17)

1. Compute the critical value c of the test.
2. After observation $n = 1, .., N-1$ do

- if $LR_n \geq c$ stop, reject H_0
- otherwise continue

3. After observation N do

- if $LR_N \geq c$ stop, reject H_0
- otherwise reject H_1

3.2.5 Critical values of maxSPRT

The values of the test statistic (17) are easy to compute. What then remains is the definition of the critical region of the test: which values of the test statistic should lead to the decision of rejecting the null hypothesis.

Since higher values of the test statistic are always more critical to the H_0, it is sufficient to determine a single critical value, which defines the boundary of the critical region.

The first step of the maxSPRT procedure – as described in Algorithm 3.2.4 – is to compute the critical value c corresponding to the desired type I error α'. Values of (17) higher than c then lead to rejection of H_0. The critical region of maxSPRT thus is

$$Q_c = \{LR_n \mid LR_n \geq c, LR_{n-1} < c\}, \quad \text{for all} \quad n = 1, 2, ..., N. \tag{18}$$

The actual type I error of the test is given by $P(Q_c \mid H_0)$. If this probability can be computed, then c can be found iteratively. The computation of $P(Q_c \mid H_0)$ depends on the statistical model for the observations (i.e. the likelihood function). Kulldorff et al. [2011, pp. 65-67, p. 72] describe how to determine c for the Binomial and Poisson likelihoods. In what follows I will present the binomial case in detail, where a Markov chain probability model can be utilized to determine c.

3.3 Binomial maxSPRT: BmaxSPRT

Let us now adopt the maxSPRT method in the setting of the simple SCCS. Assume that for the sequence of observations $(y_n, n), n = 0, 1, ...N$, where y_n denotes the number of "cases" out of n events, the probability distribution for y_n is given by the binomial distribution as described in section 2.2.3. The probability of "success" (i.e. adverse event clasified as a "case"), depends on the unknown rate ratio parameter RR and the known ratio between the lengths of the control and risk periods, z and the conditional likelihood (10).

Let the two hypotheses be as in (14). Using equation (17), the maxSPRT test statistic is given by

$$LR_n = \underset{RR>1}{max} \frac{P(y_n \mid RR)}{P(y_n \mid RR = 1)} = \underset{RR>1}{max} \frac{(\frac{RR}{z+RR})^{y_n}(\frac{z}{z+RR})^{n-y_n}}{(\frac{1}{z+1})^{y_n}(\frac{z}{z+1})^{n-y_n}}. \tag{19}$$

Kulldorff et al. [2011] call a sequential test based on the test statistics (19) the binomial maxSPRT (BmaxSPRT). A simple SCCS design combined with the maxSPRT method is one way to arrive at the BmaxSPRT, which shows that the BmaxSPRT method is self-controlled.

Computation of (19) requires maximization. It is easy to see that maximization in terms of RR depends only on the numerator, which is a likelihood function. Maximizing a likelihood function is a common task in statistics and the value that maximizes the likelihood in terms of the parameter RR is called the maximum likelihood estimate (MLE) for RR, denoted by $\hat{RR}$.

For computational reasons, it is usually convenient to operate with the log likelihood ratio instead. Since the logarithm is a strictly increasing function, maximizing the log likelihood ratio is equivalent to maximizing the likelihood ratio. The (log) likelihood ratio is maximized by finding the MLE for RR, which is easily seen to be $(y_n \cdot z)/(n - y_n)$. Since we are not interested in situations where $RR < 1$, one should use $\hat{RR} = max\{1, \frac{z \cdot y_n}{n-y_n}\}$.

Then the test statistic becomes

$$LLR_n = log(LR_n) = log\left(\frac{(\frac{\hat{RR}}{z+\hat{RR}})^{y_n} \cdot (\frac{z}{z+\hat{RR}})^{n-y_n}}{(\frac{1}{z+1})^{y_n} \cdot (\frac{z}{z+1})^{n-y_n}}\right). \tag{20}$$

A simplified form of (20) is given by Kulldorff et al. [2011, p. 71]. The BmaxSPRT experiment proceeds as described in Algorithm 3.2.4, utilizing the test statistic (20).

3.3.1 BmaxSPRT as a Markov chain

In the discrete binomial case, the critical value c of the maxSPRT procedure can be found iteratively by utilizing a Markov chain probability model to compute the type I error probability $P(Q_c \mid H_0)$. In what follows I will formulate BmaxSPRT as a Markov chain by defining the state space and transition probabilities of the experiment.

3.3.1.1 State space

In the BmaxSPRT sequential test, the possible states of the experiment comprise all possible combinations of "trials" (n) and "successes" (y_n) during the experiment, bounded by the maximum number of observations N. Therefore the state space S is defined as follows:

$$S = (n, y_n), \quad \text{where} \quad n = 0, 1, ..., N \quad \text{and} \quad y_n = 0, 1, .., n. \tag{21}$$

The experiment always starts at the state $(0, 0)$. Clearly, there are a finite number of states. In fact, there are $M = \sum_{n=0}^{N}(n+1) = (N+1)(N+2)/2$ possible states.

3.3.1.2 Transitions and critical region

In the BmaxSPRT experiment, adverse events are classified as a "case" ("success", $Y = 1$) or a "control" ("failure", $Y = 0$), depending on the outcome of a Bernoulli random variable Y (See section 2.2.3). Possible transitions in the state space S are therefore as follows:

$$\begin{aligned} "success" &: (n, y_n) \to (n+1, y_n+1) \\ "failure" &: (n, y_n) \to (n+1, y_n). \end{aligned} \tag{22}$$

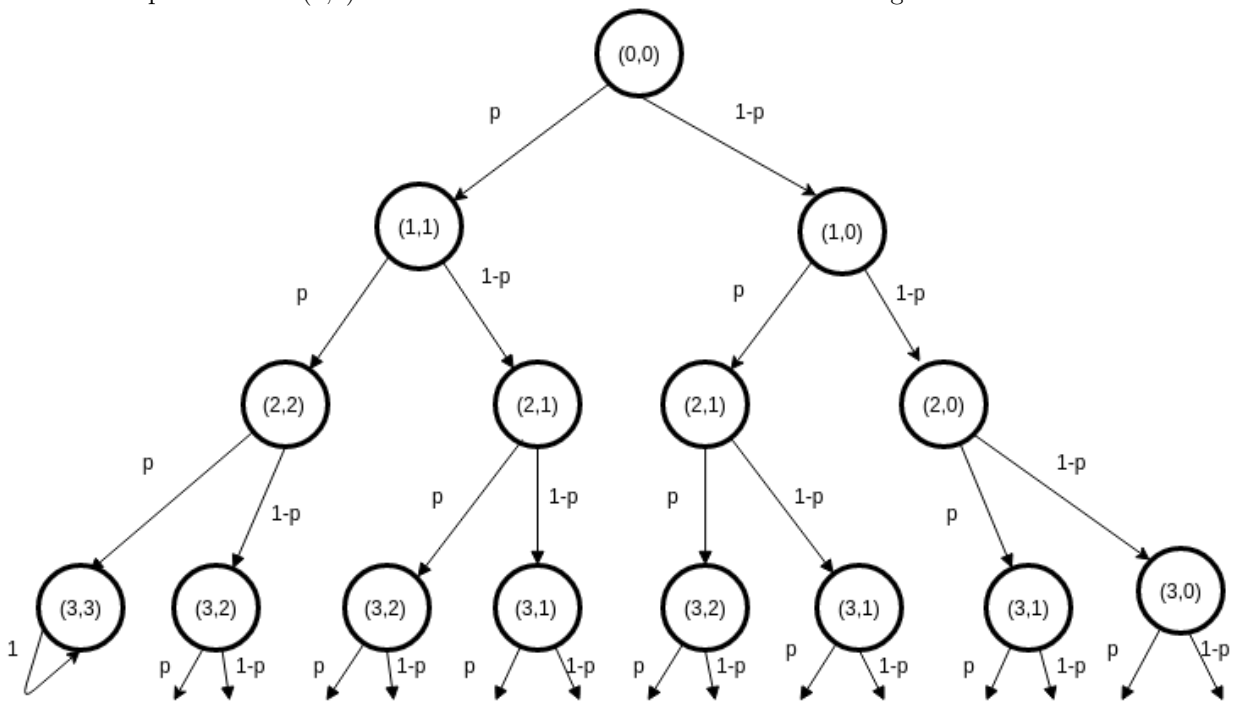
If the experiment is stopped at some state $s \in S$, then that state is called absorbing: it is impossible to leave the state. Otherwise, a state is called transient. In BmaxSPRT, the states for which $n = N$, are absorbing. Also, the states for which the value of the test statistic reaches the critical value c, are absorbing.

Let $LLR(s)$ denote the value of the test statistic (20) for state $s \in S$. The critical region of the test can then be defined as follows:

$$Q = \{s \in S \mid LLR(s) \geq c\}. \tag{23}$$

It should be noted that by this definition Q may contain states which are impossible to reach (because of previous absorbing states). However, this does not cause theoretical or practical problems.

Figure 3.3: Visualization of a BmaxSPRT type Markov Chain. The chain starts from the node (0,0) and a trial with the result "success" or "failure" is performed at each node, with p being the probability for "success". In this example the state (3,3) with 3 successes out of 3 trials is an absorbing state.



3.3.1.3 Transition probabilities

The BmaxSPRT experiment starts at the state $(0, 0)$ with probability 1. Therefore a vector of initial probabilities $\mathbf{v}$ for being in each of the (M number of) states at the beginning of the experiment is given by

$$\underset{1\times M}{\mathbf{v}} = (1, 0, ..., 0). \tag{24}$$

A Bernoulli random variable Y determines the transitions. The transition probabilities for the transient states, i.e., for those for which $LLR(n, y_n) < c$ are given by:

$$
\begin{array}{c}
 \\
(0,0) \\
(1,0) \\
(1,1) \\
(2,0) \\
(2,1) \\
(2,2) \\
(3,0) \\
(3,1) \\
. \\
. \\
(N,0) \\
. \\
. \\
(N,N)
\end{array}
\begin{array}{c}
\begin{array}{ccccccccccc}
(0,0) & (1,0) & (1,1) & (2,0) & (2,1) & (2,2) & (3,0) & (3,1) & .. & (N,0) & .. & (N,N)
\end{array} \\
\left(\begin{array}{ccccccccccc}
0 & 1-p & p & 0 & 0 & 0 & 0 & 0 & .. & 0 & .. & 0 \\
0 & 0 & 0 & 1-p & p & 0 & 0 & 0 & .. & 0 & .. & 0 \\
0 & 0 & 0 & 0 & 1-p & p & 0 & 0 & .. & 0 & .. & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1-p & p & .. & 0 & .. & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-p & .. & 0 & .. & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & .. & 0 & .. & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .. & 0 & .. & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .. & 0 & .. & 0 \\
. & . & . & . & . & . & . & . & .. & . & .. & . \\
. & . & . & . & . & . & . & . & .. & . & .. & . \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .. & 1 & .. & 0 \\
. & . & . & . & . & . & . & . & .. & . & .. & . \\
. & . & . & . & . & . & . & . & .. & . & .. & . \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .. & 0 & .. & 1
\end{array}\right)
\end{array}
$$

Figure 3.4: An example of a transition matrix describing the transition probabilities between the states in the BmaxSPRT experiment. The rows and columns of the matrix correspond to the possible states and $\mathbf{P}_{ij}$ gives the probability of the transition $s_i \to s_j$. For example under the null hypothesis of the probability of the transition $(0,0) \to (1,0)$ ("failure") is given by $1-p = z/(1+z)$ and the probability of the transition $(0,0) \to (1,1)$ ("success") is given by $p = 1/(1+z)$. In this example the state $(2,2)$ is an absorbing state and a transition can happen only onto itself. The states where the maximum number of observations is reached are all absorbing states.

$$
\begin{aligned}
P\{(n+1, y_{n+1}) \mid (n, y_n)\} &= P(Y=1) = p \\
P\{(n+1, y_n) \mid (n, y_n)\} &= P(Y=0) = 1-p
\end{aligned}
\tag{25}
$$

where p is as in (9). Under the null hypothesis ($RR = 1$) $p = 1/(1+z)$. The probabilities of other transitions from the transient states are zero. The transition probabilities for the absorbing states, i.e. for those for which $LLR(n, y_n) \geq c$ or $n = N$, are:

$$
P\{(n, y_n) \mid (n, y_n)\} = 1, \tag{26}
$$

meaning that if such a state is reached, it is never left. These transition probabilities show that the BmaxSPRT experiment has the Markov property: the transition probabilities only depend on whether the previous state was absorbing or not.

The transition probabilities are gathered in the transition matrix $\mathbf{P}$, which defines the probability distribution over all the transitions from state to state in the state space S. Each row and column of $\mathbf{P}$ corresponds to one of the states $s \in S$ and $\mathbf{P}_{ij}$ gives the probability of the transition from s_i to s_j. The size of $\mathbf{P}$ depends on N and the values of $\mathbf{P}$ depend on the values of p and c, of which c defines the absorbing states.

The probabilities of being in each of the states after N transitions in the state space are given by

$$\underset{1\times M}{\mathbf{p}} = \mathbf{v}\mathbf{P}^{(N)}. \tag{27}$$

3.3.2 Type I error

The critical value c of the BmaxSPRT experiment defines the states in the critical region Q which then defines the type I error α. To compute the value of α for a given c, the overall probability of reaching the critical region can be computed by summing over the probabilities of being in each of the states $s \in Q$ after N transitions in the state space S. Using the result introduced in (27), the type I error probability for any BmaxSPRT critical region Q is given by

$$\alpha_{(BmaxSPRT)} = P(Q \mid H_0) = \sum_{s\in Q} \mathbf{v}\mathbf{P}_0^{(N)}, \tag{28}$$

where $\mathbf{P}_0$ is the transition matrix, where $p = 1/(1+z)$ is according to H_0.

3.3.3 Critical values

Any critical value candidate c' defines the states in the critical region Q, corresponding to some type I error α. For a desired type I error α', the goal is to find a c such that $P(Q \mid H_0) \approx \alpha'$, meaning that the overall probability of rejecting the null hypothesis when it is true is (at least approximately) α'. Utilizing equation (28), it is possible to try out possible values c' until the c is found which best matches the desired α'. One way to execute this procedure is given in algorithm 3.3.3.

It should be noted that for a desired α', the algorithm finds the value c which corresponds to a type I error α for which the inequality $\alpha \leq \alpha'$ holds. It is unlikely that $\alpha = \alpha'$ exactly. This is due to the discrete nature of the binomial distribution.

Algorithm 3.3.3: BmaxSPRT critical

Input: States S, initial probabilities $\mathbf{v}$, transition matrix $\mathbf{P_0}$, test statistic function LLR_n as in (20)

1. Compute all possible test statistic values $L = \{LLR(s) \mid s \in S\}$ and sort L from lowest to highest.
2. Choose $c' = min\{L\}$ to be the critical value. The absorbing states are $Q = \{s \in S \mid LLR(s) \geq c\}$.
3. Compute $P(Q \mid H_0) = \pi'$

- if $\pi' \leq \alpha'$, stop, choose c' to be the critical value c for the type I error rate $\pi' = \alpha$
- otherwise remove c' from L and go to 2.

3.3.4 Power

The power of the BmaxSPRT experiment depends on the true value of the rate ratio. When the true rate ratio is RR_a, the probability of an event classified as a "case" is given by $p_a = RR_a/(RR_a + z)$. The true transition matrix of the experiment $\mathbf{P}_a$ is then defined by these probabilities.

From this observation it is easy to see that the power of the BmaxSPRT experiment can be computed for any chosen fixed value of RR, say $RR_1 = 1.5$. One simply needs to define the transition matrix $\mathbf{P}_1$, where the probabilities are given by $p_1 = RR_1/(RR_1 + z)$. This value only corresponds to the actual power of the experiment in the case where $RR_1 = RR_a$, which is of course unlikely. However, this procedure can provide a conservative estimate for the power of the experiment for a reasonable (small) choice of fixed RR_1.

Denoting the critical region corresponding to a desired α' as Q_0, the power of the BmaxSPRT experiment is given by the overall probability of reaching the critical region:

$$\text{Power}_{(BmaxSPRT)} = P(Q_0|H_1) = \sum_{s \in Q_0} \mathbf{v}\mathbf{P}_1^{(N)}, \tag{29}$$

where $\mathbf{v}$ are the initial probabilities of the states given in equation (24) and $\mathbf{P}_1$ is the transition matrix, where $p = RR_1/(RR_1 + z)$ for some fixed choice of RR_1.

3.4 Grouped observations

The sequential hypothesis testing methods discussed so far in this chapter (SPRT, maxSPRT, BmaxSPRT) assume that the value of the test statistic is evaluated whenever a new observation (a medical diagnosis) arrives. This applies to situations where data are available in near real time and observations arrive individually. This type of analysis is called continuous sequential analysis. Continuous sequential analysis is applicaple in situtations where the adverse events of interest are reasonably rare and medical diagnoses are collected reasonably often (for example daily).

Due to administrative reasons, sometimes medical diagnoses become available for analysis in groups. For example, the National Health and Welfare Institute (Finland) receives data from the HILMO register three times a year (2016). Unless the adverse events of interest are extremely rare, it can be expected that more than a single observation becomes available at the same time. The methods applicable in these situations are called group sequential methods [Silva and Kulldorff, 2015].

According to Silva and Kulldorff [2015] group sequential and continuous sequential analyses can be formally defined as follows. These definitions are given here only to aid discussion and in an effort to distinguish between group sequential and continuous sequential analyses.

Let X_t be a non-negative integer valued stochastic process describing the number of adverse events that occur during a $[0, t]$ time window.

Definition (Group sequential analysis)**.** *For a set of constants $A_1, ..., A_K$ and a sequence of $\{t_i\}_{i=1}^K$ of times, a group sequential analysis design is any procedure that rejects the null hypothesis if $X_{t_i} \geq A_i$ for some $i \in \{1, ..., K\}$*

Definition (Continuous sequential analysis)**.** *For a function $B(t)$, a continuous sequential analysis design is any procedure that rejects the null hypothesis if $X_t \geq B(t)$ for some $0 < t \leq L$.*

In the coming sections I will discuss why the grouped nature of the data should affect a sequential hypothesis test and its critical values. I will then discuss possible solutions.

3.4.1 Adjusting for grouped observations

Silva and Kulldorff [2015] discuss the differences between group sequential and continuous sequential analyses. I will present a short overview of that discussion. I will show that the following statements are true.

1. Continuous sequential analysis should not be used if data become available in groups.
2. Any post-market safety surveillance system should attempt to obtain data as frequently as possible.

To see the intuition for statement 1, assume a special case of group sequential analysis with only a single group of observations. The two choices are then to either perform a retrospective continuous sequential analysis or do a single hypothesis test (using the single group of data). Assume that the same test statistic is used in both cases. For the continuous sequential analysis, the critical value of the test statistic must adjust for multiple chances to reject the H_0. This means that it must be more difficult to reject the H_0 when H_1 is true. Continuous sequential analysis would therefore always have lower power than group sequential analysis if a group of data are already available.

The reason for statement 2 is that a continuous sequential method is always superior to a group sequential method if observations can be made separately. To see the intuition for this, assume a group sequential design that rejects H_0 when the total number of events (after a group of observations) is at least y_c. Now assume a continuous sequential design that checks the observations separately and rejects H_0 as soon as there are y_c events. The error rates of the two designs are identical because the number of events is non-decreasing (both designs reject H_0 iff at some point there are y_c events). However, the continuous method is superior because it can reject the H_0 sooner (with a smaller sample size). Therefore for every group sequential design there is a superior continuous sequential design.

A conclusion can be made that not adjusting for the availability of the data can result in either loss of power or an increase in the expected sample size. Analyses should always be performed as soon as possible, using all the available data.

3.4.2 Group sequential methods

The same statistical model and test statistic that would be utilized in a continuous sequential analysis can be applied in group sequential analysis. However, when analyses are done for groups of data, there are less

opportunities to reject the H_0 and therefore for a given critical value the type I error of a group sequential analysis is smaller. The critical values should threfore be chosen differently. Unfortunately, if the number of observations per group is unkown, the computation of the critical values becomes difficult. Two assumptions could be made to simplify this computation, but the validity of the assumptions is questionable.

The problem of determining the critical values of a group sequential test would simplify if the number of observations per group were assumed to be fixed (fixed group sizes). Silva and Kulldorff [2017] have implemented this solution for computing critical values of a group sequential BmaxSPRT test in their Sequential R package, available in the Comprehensive R Archive Network (CRAN). This assumption makes the problem very similar to the situation discussed in section 3.3.1. The intution for the similarity is that continuous BmaxSPRT can be seen as a special case of a group sequential hypothesis test with group sizes fixed at 1.

It is quite obvious that if adverse events are assumed to arrive as a random process in time but are collected during infrequent time intervals, the number of observations per group is by assumption a random variable. Therefore the solution of fixed group sizes does not seem satisfactory if the grouped nature of the data were due to administrative reasons.

The problem would also simplify if the maximum number of groups was fixed. However, assuming that the group sizes are unkown, fixing the number of groups will not fix the number of observations at the end of surveillance. Rememeber that the goal of safety surveillance is to keep collecting and analyzing data until there is sufficient information regarding a possible association between the exposure and the event. Naturally, the amount of information depends on the amount of observations. Therefore it would clearly be better to fix the number of observations than the number of groups.

A possible solution which needs neither simplifying assumption is to use an error spending approach, as described by Jennison and Turnbull [1999] in their book focusing on group sequential methods in clinical trials. Next, I will introduce the concept of error spending functions and maximum information trials, which could be useful approaches for future research regarding vaccine safety surveillance with grouped observations.

3.4.3 Error spending

One solution to group sequential analysis for random group sizes is to use an error spending approach (also known as alpha spending) [Jennison and Turnbull, 1999, ch. 7]. The idea of error spending is that for K groups, the type I error α is partioned into probabilities $\pi_1, ..., \pi_K$ which sum to α. For the test statistics Z_k, critical values c_k are calculated so that

$$P(|Z_1| < c_1, ..., |Z_{k-1}| < c_{k-1}, |Z_k| \geq c_k) = \pi_k \tag{30}$$

Intuitively this means that for each group k, only a proportion of the desired error probability α is spent. A practical problem remaining is: how should each π_k be chosen? To solve this problem, Jennison and Turnbull

[1999, pp. 148-150] introduce families of error spending functions and compare their properties.

An error spending function is a non-decreasing function which partitions the desired type I error rate α and for which $f(0) = 0$ and $f(t) = \alpha$ for $t \geq 1$. In the paradigm of maximum information trials (discussed next), the value $f(t)$ indicates the cumulative type I error to be spent when a fraction t of the maximum anticipated information I_{max} has been obtained. Jennison and Turnbull [1999, p. 148] suggest a family of error spending functions defined by

$$f(t) = min\{\alpha \cdot t^p, \alpha\} \tag{31}$$

where the choices of $p \in \{1, 3\}$ yield similar results to more classical approaches suggested by Pocock [1977] and O'Brien and Fleming [1979], which belong to the Wang & Tsiatis family of error spending functions.

3.4.4 Maximum information trials

Utilizing the concept of error spending, Jennison and Turnbull [1999, pp. 146-148] discuss decision rules in the case of two-sided alternative hypotheses, where the stopping rule for accepting H_0 is defined by a target maximum information level, denoted by I_{max}. The information level for group k is defined as

$$I_k = \{var(\hat{\theta}^{(k)})^{-1}\}, k = 1, 2, .. \tag{32}$$

where $\hat{\theta}$ is the estimator for the parameter of interest θ. A maximum information trial uses an error spending function $f(t)$ as described above. The type I errors allocated to each analyses are

$$\begin{aligned} \pi_1 &= f(I_1/I_{max}) \\ \pi_k &= f(I_k/I_{max}) - f(I_{k-1}/I_{max}). \end{aligned} \tag{33}$$

A decision rule for a maximum information trial is described in Algorithm 3.4.4 [Jennison and Turnbull, 1999, p. 54].

Algorithm 3.4.4: A maximum information trial

Input: target information I_{max}, sequence of groups k, test statistic function Z.

1. Define K as the smallest value k for which the information reaches the target information: $K = min\{k \mid I_k \geq I_{max}\}$

2. After group $k = 1, ..., K-1$
- if $|Z_k| \geq c_k$ stop, reject H_0
- otherwise continue to group k + 1

3. After group K
- if $|Z_k| \geq c_k$ stop, reject H_0
- otherwise stop, accept H_0

The target information I_{max} is a similar idea to the maximum sample size in maxSPRT in that it too defines a boundary related to the sample size after which the experiment ends and H_0 is accepted. In their example utilizing the maximum information trial approach, Jennison and Turnbull [1999, pp. 150-153] assume that the test statistic has a normal distribution and the alternative hypothesis is two-sided. In maxSPRT type surveillance, the alternative hypothesis is one-sided and the test statistic is not assumed to be normal.

Maximum information trials provide a promising approach for grouped observation in vaccine safety surveillance, but further research is needed to adopt the method in the maxSPRT setting.

4 Childhood vaccinations and febrile seizures: application of BmaxSPRT

Fevers can cause a child to experience spasms or jerky movements called seizures. Seizures caused by a fever are called febrile seizures. Febrile seizures usually last for less than one or two minutes and they do not cause any permanent harm or have any lasting effects. In this application I use the BmaxSPRT vaccine safety surveillance method to study the relationship between febrile seizures and three different childhood vaccinations: Measles-Mumps-Rubella vaccination (MMR), Pneumococcal vaccination (PCV) and the Rota virus vaccination (Rota). MMR, PCV and Rota are all included in the vaccination programme in Finland and are free for all children. The three vaccines are interesting because:

- For MMR there exists a known relationship to increased risk of febrile seizures
- For PCV there exists a suspected relationship to increased risk of febrile seizures
- For Rota there exist a suspected relationship to decreased risk of febrile seizures

It is established that there is an increased risk of acute fever seizures during the $5-12$ days following the first

dose vaccination with MMR in children aged from 6 months to 2 years. This relationship has been previously shown by both international studies and using Finnish register data. A study by the Centers for Disease Control and Prevention (CDC) has also indicated that PCV vaccination together with a flu vaccination could be associated with an increased risk of febrile seizures during the 24 hours following exposure [CDC]. A PCV vaccination alone might also cause a small increased risk of seizures. There is no evidence of Rota vaccination having similar effects and Sheridan et al. [2016] have found that a Rota vaccination can prevent febrile seizures.

In this proof-of-concept application, BmaxSPRT vaccine safety surveillance is retrospectively applied to data from the Finnish HILMO register. The main interests are to apply the BmaxSPRT method on real data, study if and when the method generates a signal for the MMR and febrile seizure pair, and also to see what the results are for the two other vaccines.

4.1 Data

In a real application of BmaxSPRT surveillance, data would be collected and analyzed until a decision to reject or accept the null hypothesis can be made. In this proof-of-concept application, the goal is to utilize a limited amount of already existing data.

Finnish children under 2 years of age, born during $2010 - 2014$ were considered eligible for this study. For MMR, the 2014 birth cohort was not considered, as due to the higher expected first dose vaccination age, no sufficient data for the 2014 cohort were yet available at the time of the study. Vaccination data of all the three vaccines (Rota, PCV, MMR) were collected from the National Vaccination Register maintained by THL. Health care data regarding the febrile seizures were collected from the HILMO register.

4.1.1 Vaccinations

One challenge with analyzing the effects of vaccines in an observational study is that multiple vaccinations can be given during the same day. For example, the second dose of Rota is recommended to be taken at the same age as the first dose of PCV. The recommended schedule for each of the three vaccines of interest is given in Table 4.1. In an effort to make the vaccines (i.e. exposures) more identifiable, only the first doses of each of the vaccines were considered in this study.

Sometimes information about the same vaccination is collected multiple times so that there are duplicate records in the vaccination register. Duplicate records identified with the same personal identification code, same vaccine identifier and the same date as another record were excluded from analyses.

The three vaccines were analyzed separately by defining respective vaccination groups. A child belongs to a specific vaccination group if he/she was vaccinated with the respective vaccine during the observation period. A child can belong to multiple vaccination groups.

4.1.2 Hospitalizations

Febrile seizures were operationalized utilizing medical diagnoses related to hospitalizations. Febrile seizures were identified with the following ICD-10 diagnosis codes:

- A858: *Other specified viral encephalitis*, A86: *Unspecified viral encephalitis*, A87: *Viral meningitis*, A88: *Other viral infections of central nervous system, not elsewhere classified*
- G038: *Meningitis due to other specified causes*, G039: *Meningitis, unspecified*, G04: *Encephalitis, myelitis and encephalomyelitis*, G05: *Encephalitis, myelitis and encephalomyelitis in diseases classified elsewhere*
- R291: *Meningismus*, R55: *Syncope and collapse*, R560: *Febrile convulsions*, R568: *Other and unspecified convulsions*

Hospitalization associated with any one of these diagnosis codes was classified as a febrile seizure event. The date of the seizure was taken to be the date of hospitalization. Any seizure within 7 days from the previous one for the same child was considered to represent the same event and were thus excluded from the analysis.

4.2 Method

The method used in this study is the binomial variant of maxSPRT, BmaxSPRT. In order to deploy the BmaxSPRT, one needs to specify

1. The observation period.
2. The risk and control periods.
3. The desired type I error rate.
4. The maximum number of observations.

The chosen maximum number of observations affects the power of BmaxSPRT. Instead of directly specifying the number of observations, a desired power given a specified rate ratio could be chosen and the maximum number of events correspoding to that power could then be computed. In this study the maximum number of observations was directly chosen.

The choices $1-4$ were used to compute the critical value of BmaxSPRT. In what follows I will briefly describe the BmaxSPRT method and the parameters for this study. For details on the method please refer to earlier chapters.

4.2.1 Observation period

For each individual born during the chosen years, the observation period in BmaxSPRT is an age interval during which an event can be sampled. To define the age intervals of interest, the realised age distribution of the first dose was analyzed for each of the three vaccines (vaccination groups). The observation period for

each vaccination group was then defined so that the chosen age interval covers most first dose vaccinations (>95%). The observation periods as ages in days for each of the vaccination groups are displayed in Table 4.1.

Table 4.1: The recommended vaccination schedule and observation ages by vaccination group.

Exposure	Schedule	Observation.ages
MPR	12 months	250-650 days
Rota	2, 3 and 5 months	30-140 days
PCV	3, 5, and 12 months	60-200 days

The analysis is performed with the assumption that new data may become available each day, starting from 2010-01-01. It should be noted that in reality the first observation is expected to arrive much later, since a person born during 2010 will first have to be vaccinated, then hospitalized and then pass through the risk and control periods in order to contribute to the study.

4.2.2 Risk and control periods

From an epidemiological design standpoint, the study design in BmaxSPRT is a self-controlled, case-only design, such as the simple SCCS design introduced in section 2.2.3. In the simple SCCS, the observation period for each child is divided into risk and control periods following exposure. A febrile seizure during the risk period is classified as a case and a febrile seizure during the control period is classified as a control.

In this application, the risk period was chosen to be $0 - 13$ days following vaccination and the control period $14 - 41$ days following vaccination, for each of the vaccines. This risk period simultaneously covers the suspected period of increased risk for both PCV (24 hours following exposure) and MMR ($5 - 12$ days following exposure). The chosen risk and control periods are displayed in Table 4.3.

It should be noted that the chosen risk period is likely to be too long especially for PCV. This could affect the power of BmaxSPRT for the PCV group. In a real safety surveillance application, specific knowledge of the actual period of increased risk might not be available, which is why a choice of a conservative common risk period was used in this proof-of-concept application. See section 4.4 for a simulation study of how the choice of the risk period affects the power of BmaxSPRT.

4.2.3 Maximum number of observations

BmaxSPRT surveillance lasts until a determined maximum number of adverse events is observed or a safety signal is generated. There is no predetermined calendar time (or last included birth cohort) for the end of surveillance. In this retrospective proof-of-concept application however, it was desirable to reach a conclusion utilizing the available data. Measures were taken to make this more likely, by utilizing information on the

incidence of febrile seizures available in the data. This information would not be available nor would it be needed in a real application of BmaxSPRT.

For each of the age intervals of interest, the incidence of febrile seizures during the observation period, but excluding the risk period, was estimated to aid determining the maximum number of observations. I will call these the baseline incidences. The baseline incidences for each vaccination group are displayed in Table 4.2. For the oldest age group the baseline incidence of febrile seizures is higher than for the two younger groups.

The size of the vaccinated population was estimated using available birth data and vaccination coverage estimates. The expected number of seizures during surveillance, during which each vaccinated individual contributes a total of 42 person days to the study, was then computed. These expected numbers of events were then used as the maximum number of observations.

Table 4.2: The included birth cohorts, size of the study population and estimated vaccination coverage for each of the observed age groups, along with the baseline event incidence estimates given by 1000 person years. The incidence estimates were used to calculate the expected number of events during surveillance, during which vaccinated individuals contribute 42 person days to the study.

Age	Cohorts	Population	Vaccination.coverage	Baseline.incidence	Expected.Events
250-650 days	2010-2013	238 568	95%	14.6	380
30-140 days	2010-2014	295 800	93%	8.1	259
60-200 days	2010-2014	295 800	94%	7.9	252

It should be noted that alternative strategies could have been used. For this retrospective study, it would have also been possible to simply count the total number of events during the risk and control periods in the available data for each vaccination group, instead of estimating the expected number of events. One could then choose the smallest of those as the mutual maximum number of observations. It is likely that other feasible approaches exist as well.

4.2.4 Hypotheses and decision rule

For each vaccination group, the hypotheses are

- $H_0 : RR = 1$
- $H_1 : RR > 1$,

where $RR = \lambda_1/\lambda_0$ is the rate ratio of the incidence of febrile seizures during the risk period (λ_1) and control period (λ_0).

Inference is based on a conditional likelihood (10), which can be thought of as independent binomial trials,

where the probability of an adverse event being classified as a "case" is a function of the rate ratio parameter RR. Higher values of RR correspond to a higher probability that an event occured during the risk period instead of the control period. The test statistic is the maximized log-likelihood ratio as in (20).

The study proceeds as described in algorithm 3.2.4. For each day of surveillance, the value of the test statistic is computed and compared to the critical value to make a decision. If at any day during surveillance the value of the test statistic reaches the critical value, surveillance is stopped for the vaccination group and H_0 is rejected. If there are no new observations, the value of the test statistic will remain the same and the decision made during the previous day will not change. If the maximum number of observations is reached without rejecting H_0, surveillance ends and H_0 is accepted.

The desired type I error probability was chosen to be $\alpha' = 0.05$. Critical values were computed with the Sequential R package by Silva and Kulldorff [2017], which implements a similar algorithm as described in section 3.3.3. The actual type I error probabilities matching the critical values were approximately the desired probabilities (< 0.0001 difference for all three groups). The power was calculated for RR set at 1.5.

The parameters defining the stopping conditions for each vaccination group are displayed in Table 4.3, along with power estimates. It can be seen that the power is reasonable for each group. However, the actual power depends on the actual relative incidence rate and the actual risk period.

Table 4.3: Parameters of the BmaxSPRT surveillance. z is the ratio of the lengths of the control and risk periods, p is the expected proportion of cases under the null hypothesis (p = 1 / (z + 1)), N is the maximum number of observations. alpha is the type I error. The power was computed for a rate ratio of 1.5.

	risk	ctrl	z	p	N	c	alpha	power
MMR	0-13	14-41	2	1/3	380	3.78	0.05	0.92
ROTA	0-13	14-41	2	1/3	259	3.69	0.05	0.80
PCV	0-13	14-41	2	1/3	252	3.68	0.05	0.79

4.3 Results

All analyses were carried out using the R program for statistical computing [R Core Team, 2017]. Functions from the Sequential R package were used for parts of the computations [Silva and Kulldorff, 2017], such as computing the critical values and power. In addition to the main results of BmaxSPRT surveillance for the three vaccines, the method was studied in the case where surveillance continued even after a decision was reached (MMR and PCV vaccines).

4.3.1 Surveillance results

The BmaxSPRT method generated a signal on day 693 of surveillance for PCV and on day 1041 for MMR. In both cases, the null hypothesis was rejected. No signal was generated for Rota and the maximum number of observations was reached without rejecting H_0. The values of the test statistic and the rate ratio estimates for each day of surveillance are illustrated in Figure 4.1.

According to the choices of the risk and control periods, the probability that an event is classified as a "case" is 1/3 under the H_0. For MMR there were 66 cases out of 149 events ($\approx 44\%$) and for PCV there were 29 cases out of 56 events ($\approx 52\%$). For Rota, the proportion of cases ($\approx 34\%$) closely matched the expected proportion. The rate ratio estimes at the time of signal generation were 1.59 for MMR and 2.15 for PCV. For Rota, the rate ratio estimate at the end of surveillance was 1.01. These results are displayed in Table 4.4.

Table 4.4: Summary of the results of the BmaxSPRT surveillance for the three vaccination groups. A signal was generated for MMR and PCV.

	signal.day	LLR	cases	controls	n	prop.cases	RR
MMR	1041	3.85	66	83	149	44%	1.59
Rota	-	0.00	87	172	259	34%	1.01
PCV	693	4.03	29	27	56	52%	2.15

4.3.2 More results and conclusions

In a real application of BmaxSPRT, the surveillance immediately ends once evidence of the association between the vaccine and the adverse event has been found (a signal is generated) or if the maximum number of observations is reached. In this proof-of-concept application however, it is possible to satisfy one's curiosity and see what would happen if surveillance was instead continued for the two vaccines for which signal was generated and early stopping occured (MMR, PCV).

The values of the test statistic and the rate ratio estimates for each day of the continued surveillance are illustrated in Figure 4.2. Further surveillance shows that when using all the available data, the estimate of the rate ratio for the PCV group goes down and the value of the test statistic stays below the critical value. The evidence of an association to febrile seizures for the PCV group therefore seems inconclusive. For MMR the evidence of an association becomes stronger with more observations.

It should be noted at this point that there are two reasons for why in this proof-of-concept application the results of the sequential hypothesis test should not be considered as conclusive results, especially for the PCV group. First, since a group of data is already available, the continuous maxSPRT is not the optimal hypothesis test. A regular hypothesis test would have greater power, as discussed in section 3.4. Second,

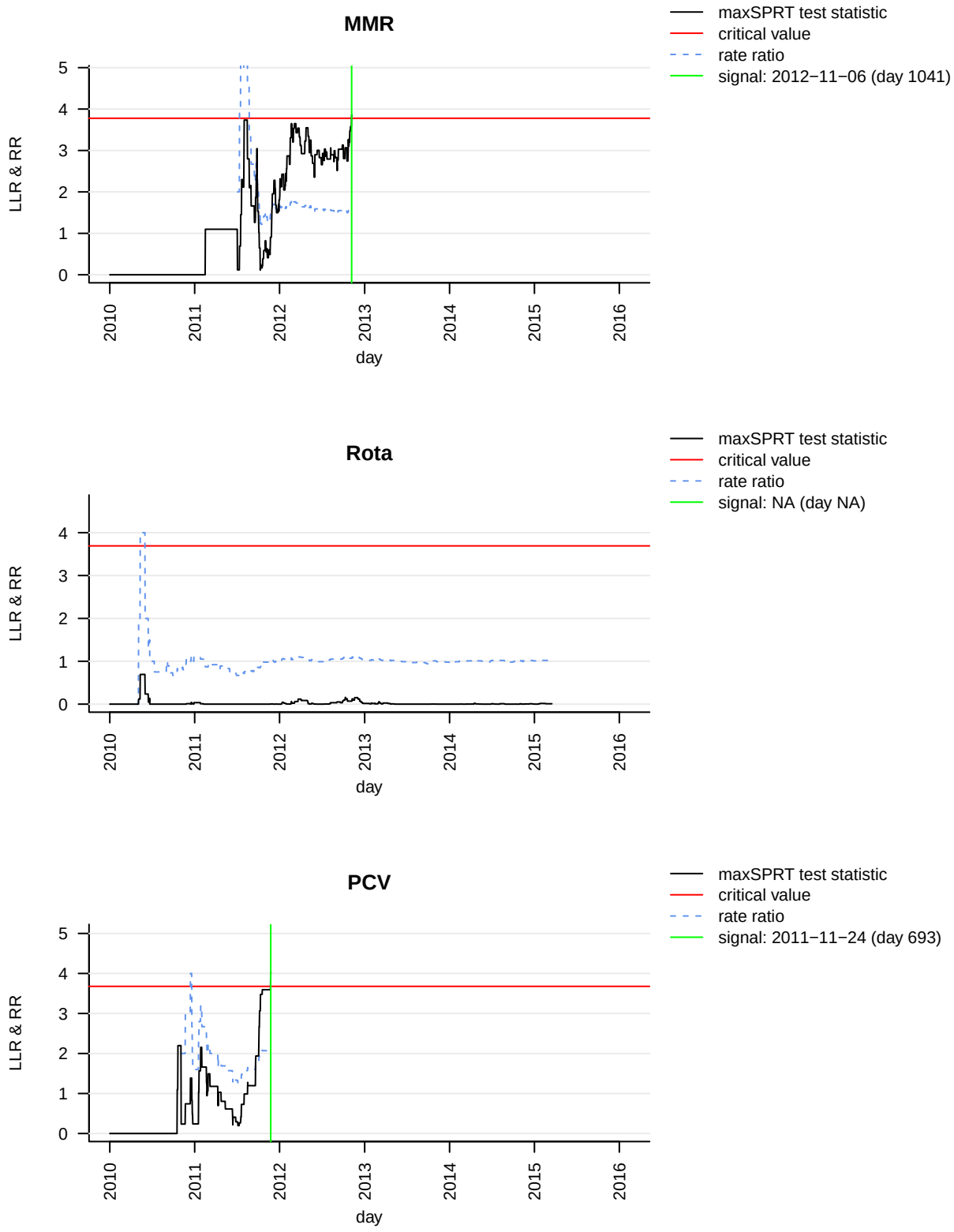


Figure 4.1: Results of BmaxSPRT surveillance for MMR, Rota and PCV. A signal was generated on day 1041 for MMR and on day 694 for PCV as the value of the test statistic (solid black) reached the critical value (solid red). The maximum number of observations (259) was reached for Rota without generating a signal. The running maximum likelihood estimate for the rate ratio is also displayed in the picture (dashed blue).

the chosen risk period (0-13 days from vaccination) is possibly too long for PCV, where according to the literature the suspected increased risk of a febrile seizure is within a 24 hour period following vaccination.

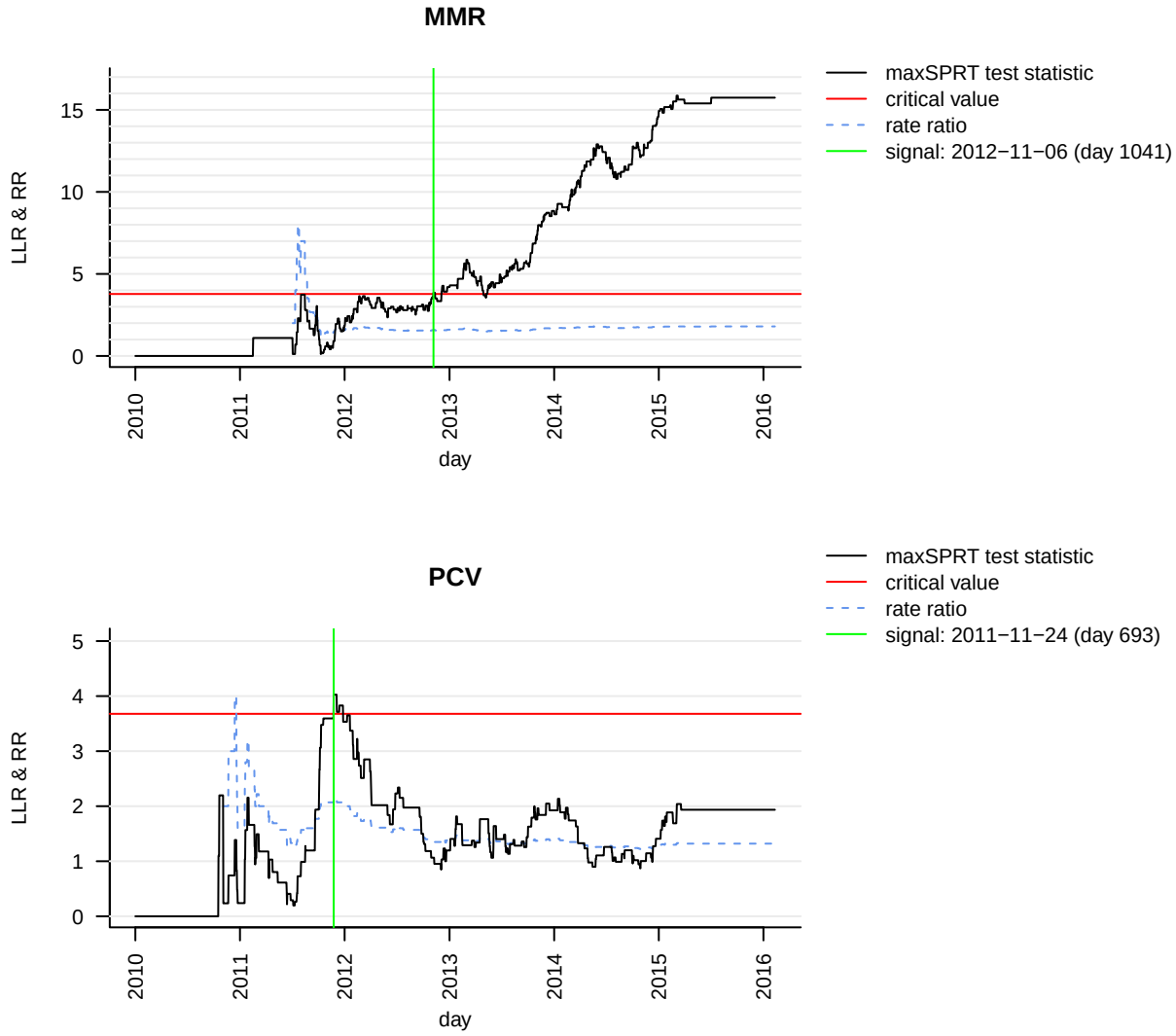


Figure 4.2: Results for continued surveillance for the PCV and MMR groups, both of which a safety signal was generated for. Continued surveillance shows that the evidence of an association to febrile seizures for MMR becomes stronger when utilizing the complete data set: the value of the test statistic (solid black) grows higher, i.e. further away from the critical value (solid red). The estimate for the rate ratio (dashed blue) remains above 1. Results for PCV are inconclusive, as the evidence of an association weakens with more observations.

4.4 Sensitivity analysis of BmaxSPRT

It is clear that the results of BmaxSPRT surveillance depend on the succesful choice of the risk period. Intuitively, if the actual time period of increased risk is shorter or longer than the chosen risk period, that should have a decreasing effect on the power of the test. An interesting question is therefore to quantify this effect for different choices of risk and control periods.

I used simulation to study how the power of BmaxSPRT is affected by poor choices of the risk and control periods. The analysis was carried out by repeatedly simulating adverse events data and applying the BmaxSPRT method to the simulated data with different choices for the risk and control periods used in the analysis. Below I will describe the assumptions of the simulation and then present the results of the analysis.

4.4.1 The studied risk and control periods

Six risk and control periods which differed from the actual risk and baseline time periods of the simulation were chosen, along with a seventh pair of reference choices which matched the actual risk and baseline periods perfectly. Figure 4.3 illustrates the chosen case and control periods. All the chosen periods have a common ratio between the lengths of the risk and control periods, as shown in Table 4.5.

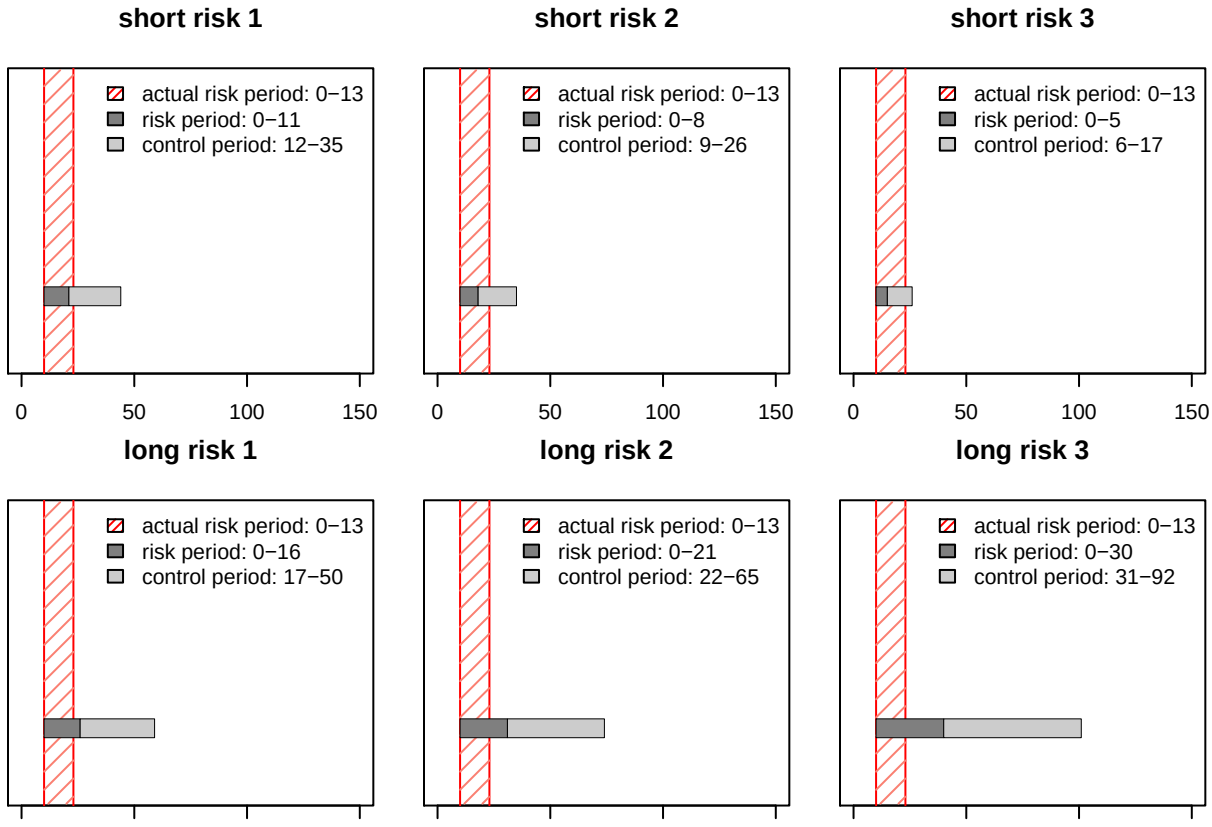


Figure 4.3: The six risk and control periods studied in the experiment. In the top row the risk period is too short and the control period overlaps with the risk period. In the bottom row the risk period is too long and overlaps with the control period.

Table 4.5: The chosen risk and control periods of the simulation experiment. The reference group matches the actual risk period used in the simulation of adverse event data. The relative length (z) of all the risk and control periods is identical.

	risk	control	risk.length	control.length	z
reference	0-13	14-41	14	28	2
short risk 1	0-11	12-35	12	24	2
short risk 2	0-8	9-26	9	18	2
short risk 3	0-5	6-17	6	12	2
long risk 1	0-16	17-50	17	34	2
long risk 2	0-21	22-65	22	44	2
long risk 3	0-30	31-92	31	62	2

4.4.2 Assumptions of the simulation

The following assumptions were made when simulating the adverse events data:

- The baseline rate of an event is 10 events per thousand person years
- The incidence rates are equal (homogenous) for all individuals
- The actual risk period is always 0-13 days after exposure
- For each exposed individual, the rate of events is multiplied by 1.5 during the risk period ($RR_a = 1.5$)
- The number of exposed individuals is 200 000 x 6 during a 6 year period (6 birth cohorts)
- The type I error of the sequential test is 0.05
- When the risk and control periods are chosen correctly, the power of the sequential test is 0.9

Adverse events were simulated for the 6 birth cohorts 10 000 times. For each of these iterations, multiple maxSPRT surveillances (with differing risk and control periods) were carried through. The date of signal generation was saved (if there was a signal).

4.4.3 Results of the simulation

Table 4.6 shows the power for each risk and control period pair in the simulation. The power is the proportion of simulation iterations where the null hypothesis was rejected. When the risk and control periods were chosen perfectly as was done with the reference group, the power in the simulation was 0.9. As expected, the power is lower the more the choice for the risk and control periods differ from the reference periods. For example, if the chosen risk period is approximately 2.5 times the size of the actual risk period ($0 - 30$ compared to $0 - 13$), the power decreases from 0.9 to 0.35.

If the chosen risk period is shorter than the actual risk period and the control period simultaneously overlaps

with the actual risk period, the situation is even worse. If the actual risk period is 2.5 the size of the chosen risk period and the control period overlaps, the power decreases from 0.9 to 0.16.

Table 4.6: Results of the simulation. The desired power in the experiment was 0.9 and the actual period of risk was 0-13 days after exposure. The power is lower the more the choices for the risk period differs from the reference period.

	risk	control	power
reference	0-13	14-41	0.90
short risk 1	0-11	12-35	0.82
short risk 2	0-8	9-26	0.58
short risk 3	0-5	6-17	0.16
long risk 1	0-16	17-50	0.78
long risk 2	0-21	22-65	0.57
long risk 3	0-30	31-92	0.35

5 Conclusions

All pharmaceutical products, including vaccines, can increase the risk of some undesired medical occurrences (adverse events). Evaluating these risks post-licensure is essential for evaluating the safety of the vaccine, since rare adverse events might go undetected in pre-licensure studies.

This thesis has presented vaccine safety surveillance as an observational study where the goal is to evaluate the safety of a vaccine (or vaccines) in real-time. The goal of a safety surveillance method is to generate a safety signal as soon as possible, when an association between the vaccine and the adverse events exists.

Electronic health care data bases (registers) can be utilized as sources of data for safety surveillance. Registers such as AvoHILMO in Finland provide daily data and thus enable near real-time surveillance. Adverse events can be operationalized by one or several medical diagnoses, identified by appropriate ICD codes.

An important consideration in vaccine safety studies is that sources of bias such as self-selection and changes in diagnosis coding practises present challenges to which some observational study designs such as the cohort design, are vulnerable. Case-only designs such as SCCS and CCO eliminate all time invariant confounders and can therefore be considered more suitable for drug safety analysis.

5.1 Contribution to existing literature

This thesis has considered vaccine safety surveillance from the general perspectives of drug safety and epidemiological studies. The type of safety surveillance considered is placed in the refinement stage of drug

safety studies, where the interest is in refining a previously identified, biologically plausible causal relationship between a vaccine and an adverse event. The refinement stage is preceeded by an identification stage and possibly followed by a confirmation stage.

This thesis has focused on the BmaxSPRT safety surveillance method introduced by Kulldorff et al. [2011]. A key contribution of this work is in expanding the theoretical treatment of the method. The self-controlled aspect of BmaxSPRT and the connection to the self-controlled case series (SCCS) is made explicit by deriving the BmaxSPRT from SCCS and maxSPRT. Additionally I have, both mathematically and algorithmically, presented the computation of the critical values of BmaxSPRT, which are essential elements of the decision rule of BmaxSPRT, also presented in detail in this work.

I have studied the relationship of three childhood vaccinations to febrile seizures by applying the BmaxSPRT method to Finnish register data. I have also analyzed the sensitivity of BmaxSPRT to the choice of risk and control periods by simulation, showing how the power of BmaxSPRT is affected by the choices. The sensitivity analysis highlights the importance of careful selection of the risk and control periods as the loss of power due to poor choices can be considerable.

5.2 Discussion of application results

As a proof-of-concept, I applied the BmaxSPRT method retrospectively to Finnish register data to show how the method performs with real data. The outcome of interest was febrile seizures. Three vaccines, MMR, PCV and Rota, each with a different type of known or suspected relationship to febrile seizures, were the exposures of interest.

The effects of each of the vaccines were studied during a risk period of $0 - 13$ days following vaccination. The results of the experiment were as expected for MMR and Rota, as a safety signal was generated for the former and not generated for the latter. A safety signal was also generated for PCV.

The result of the experiment for PCV raises further questions. Utilizing all the available data, the point estimate for the rate ratio between the chosen risk and control periods stayed consistently, but not considerably, above 1, indicating some increased risk during the risk period. Given that the chosen risk period was significantly longer than the risk period suspected by CDC for PCV (24 hours following exposure), there are at least two plausible interpretations for this outcome: (i) The results of the study are valid. The actual risk period for PCV is longer than the period suspected by CDC, and the risk of febrile seizures increases only slightly. (ii) The risk period that was used in the study is too long compared to the actual risk period. The risk of febrile seizures increases substantially, but during a short time period.

A confirmation analysis studying the relationship between febrile seizures and PCV vaccinations should be performed.

5.3 Further reasearch

The thesis provides starting points for further reseach in developing the BmaxSPRT method. The general version of SCCS can provide a framework for a more general version of a self-controlled maxSPRT, a multinomial maxSPRT if you will. Maximum information trials provide a promising approach for dealing with grouped observations.

This thesis has argued that the BmaxSPRT method belongs to the refinement stage of drug safety studies and requires that a biologically plausible vaccine-adverse event pair has been previously identified. Electronic medical records could however also be used to identify safety signals. According to Nelson et al. [2015], signal identification has primarily been conducted using spontaneous report databases, which are not considered in this work. Dumouchel [1999] introduces Bayesien data mining methods for detecting adverse drug reactions from such databases.

References

Childhood vaccines and febrile seizures. URL https://www.cdc.gov/vaccinesafety/concerns/febrile-seizures.html. Accessed: 2017-05-19.

JM Bernardo, MJ Bayarri, JO Berger, AP Dawid, D Heckerman, AFM Smith, and M West. Bayesian methods in pharmacovigilance. *Oxford University Press*, 23:29, 2011.

David Clayton and Michael Hills. *Statistical methods in epidemiology*. Oxford University Press, 1993.

William Dumouchel. Bayesian data mining in large frequency tables, with an application to the fda spontaneous reporting system. *The American Statistician*, 53(3):177–190, 1999. doi: 10.1080/00031305.1999.10474456. URL http://www.tandfonline.com/doi/abs/10.1080/00031305.1999.10474456.

CP Farrington. Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics*, pages 228–235, 1995.

CP Farrington. Control without separate controls: evaluation of vaccine safety using case-only methods. *Vaccine*, 22(15):2064–2070, 2004.

Jason M Glanz, David L McClure, Stanley Xu, Simon J Hambidge, Martin Lee, Margarette S Kolczak, Ken Kleinman, John P Mullooly, and Eric K France. Four different study designs to evaluate vaccine safety were equally validated with contrasting limitations. *Journal of clinical epidemiology*, 59(8):808–818, 2006.

Christopher Jennison and Bruce W Turnbull. *Group sequential methods with applications to clinical trials*. CRC Press, 1999.

Martin Kulldorff, Robert L Davis, Margarette Kolczak, Edwin Lewis, Tracy Lieu, and Richard Platt. A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Sequential analysis*, 30 (1):58–78, 2011.

Andreia Leite, Nick J Andrews, and Sara L Thomas. Near real-time vaccine safety surveillance using electronic health records - a systematic review of the application of statistical methods. *Pharmacoepidemiology and drug safety*, 25(3):225–237, 2016.

Malcolm Maclure. The case-crossover design: a method for studying transient effects on the risk of acute events. *American journal of epidemiology*, 133(2):144–153, 1991.

Malcolm Maclure, Bruce Fireman, Jennifer C. Nelson, Wei Hua, Azadeh Shoaibi, Antonio Paredes, and David Madigan. When should case-only designs be used for safety monitoring of medical products? *Pharmacoepidemiology and Drug Safety*, 21:50–61, 2012. ISSN 1099-1557. doi: 10.1002/pds.2330. URL http://dx.doi.org/10.1002/pds.2330.

Roger J Marshall and Rodney T Jackson. Analysis of case-crossover designs. *Statistics in medicine*, 12(24): 2333–2341, 1993.

David L McClure, Jason M Glanz, Stanley Xu, Simon J Hambidge, John P Mullooly, and James Baggs. Comparison of epidemiologic methods for active surveillance of vaccine safety. *Vaccine*, 26(26):3341–3345, 2008.

Jennifer C Nelson, Andrea J Cook, Onchee Yu, Shanshan Zhao, Lisa A Jackson, and Bruce M Psaty. Methods for observational post-licensure medical product safety surveillance. *Statistical methods in medical research*, 24(2):177–193, 2015.

Sandra Nordmann, Lucie Biard, Philippe Ravaud, Marina Esposito-Farèse, and Florence Tubach. Case-only designs in pharmacoepidemiology: a systematic review. *PLoS One*, 7(11):e49444, 2012.

Peter C. O'Brien and Thomas R. Fleming. A multipe testing procedure for clinical trials. *Biometrics*, pages 549–556, 1979.

Stuart J. Pocock. Group sequential methods in the design and analysis of clinical trials. *Biometrica, Vol 64, No. 2*, pages 191–199, 1977.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL https://www.R-project.org/.

Sheldon M. Ross. *Stochastic processes 2nd ed.* New York: Wiley, 1996.

Sarah L. Sheridan, Robert S. Ware, Keith Grimwood, and Stephen B. Lambert. Febrile seizures in the era of rotavirus vaccine. *Journal of the Pediatric Infectious Diseases Society*, 5(2):206, 2016. doi: 10.1093/jpids/piu097. URL +http://dx.doi.org/10.1093/jpids/piu097.

Ivair R Silva and Martin Kulldorff. Continuous versus group sequential analysis for post-market drug and vaccine safety surveillance. *Biometrics*, 71(3):851–858, 2015.

Ivair Ramos Silva and Martin Kulldorff. *Sequential: Exact Sequential Analysis for Poisson and Binomial Data*, 2017. URL https://CRAN.R-project.org/package=Sequential. R package version 2.3.1.

Henrik Toft Sørensen, Svend Sabroe, and Jørn Olsen. A framework for evaluation of secondary data sources for epidemiological research. *International journal of epidemiology*, 25(2):435–442, 1996.

THL. Tautiluokitus icd-10. *Luokitukset, termistöt ja tilasto-ohjeet / Terveyden ja hyvinvoinnin laitos (THL)*, pages 3–4, 2011.

A. Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.

A. Wald and J. Wolfowitz. Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 19(3):326–339, 1948. ISSN 00034851. URL http://www.jstor.org/stable/2235638.

Heather J Whitaker, C Paddy Farrington, Bart Spiessens, and Patrick Musonda. Tutorial in biostatistics: the self-controlled case series method. *Statistics in medicine*, 25(10):1768–1797, 2006.