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)

TÄMÄ 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:

Tiivistelmä:

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]. In this work, monitoring the safety of drugs 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). Safety surveillance is interested in finding and generating these signals. According to Nelson et al. [2015], 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 biologically plausible drug-event pairs suspected to have a causal relationship. Addresses multiple specified hypotheses related to the pairs.
3. **Signal confirmation**: Involves a one-time, more in depth study of a previously generated signal.

Safety surveillance can relate to stages one or two, with the aim of either finding new associations (identification) or validating the evidence of suspected associations (refinement).

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

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

1.2 Vaccine safety surveillance

In vaccine safety surveillance the researcher or an automated system monitors the safety of vaccines. Natural goals of vaccine safety surveillance are to control the expected rate 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. After a signal has been generated, further action can be taken, for example a confirmation analysis can be performed.

Statistical methods designed specifically for vaccine safety surveillance exist. This thesis will largely 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 survaillance based on maxSPRT has been utilized between 2005 - 2015 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]

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 major focus will be on the BmaxSPRT variant.

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 1.9.2017 in Finland. 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 initialized 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 elements. A unique combination of the three elements can be thought to define a unique safety surveillance method. The different variants of maxSPRT can 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.

For any reasonably rare adverse events, daily accumulating observations can be considered as continuous observations. In this work I will mainly consider situations where observations (electronic health care data) accumulates daily. Therefore the focus will be on continuous observations. I will present a case for utilizing case-only designs – and specifically the self-controlled case series (SCCS) – as the design element of a vaccine safety surveillance method. I will utilize maxSPRT sequential hypothesis testing as the decision element.

These choices (continuous observations, a case-only design and maxSPRT) lead to the continuous binomial variant of maxSPRT, BmaxSPRT.

Table 1.1: Elements of Vaccine safety surveillance and options related to them.

Element	Options
Data accumulation	Continuous observations, grouped observations
Study design	Case-only design, cohort design, case-control design
Decision rule	Hypothesis testing

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 data base. Vaccinations are operationalized similarily from medical records, identified using the vaccines name or an 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 data bases 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 data from HILMO is available three times a year (2016). 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 will introduce the vaccine safety surveillance method BmaxSPRT in detail by breaking it down to it's elements. The study design and decision rules for signal generation are the main focus of this work and are given thorough treatment. The outline of the thesis is as follows.

Chapter 2 offers discussion related to the design elements in vaccine safety surveillance, focusing on case-only study designs.
Chapter 3 offers discussion related to the decision elements in vaccine safety surveillance and focuses on sequential hypothesis testing. Dealing with grouped observations (data accumulation) is briefly 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.

The thesis ends with conclusions and discussion.

2 Study designs in vaccine safety surveillance

In vaccine safety surveillance the researcher (or a defined automated system) observes, but does not alter, 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 increased due to an 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 as related to the exposure. In practise this is difficult to achieve.

Electronic health care data bases such as the hospital discharge register (HILMO) can be utilized as sources of data for the medical diagnoses used 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, then the studied sample can be biased: the observed effect of the exposure can 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 can increase.

In their systematic review of vaccine safety surveillance applications, Leite et al. 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].

Reasons 1,3,4,5 are all affected by the study design.

In the following sections, I will first compare possible observational study designs for safety survaillance: 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 (CCO) desings.

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 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. This terminology will be useful in the coming sections.

2.1 Observational epidemiological study designs

Ultimately all epidemiological study designs share the same goal: 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 if 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. The following three observational study designs can be considered 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. 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). Proportions of having been exposed are compared between the cases and controls.
3. **Case-only designs**: Only individuals with events are sampled and are self-matched, using for example risk and control intervals to define the events as either cases or controls. Event incidences or exposure rates are compared.

The designs have different strengths and weaknesses, 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 the most 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

Design	Strengths	Weaknesses
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 little data (only 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 the most data and is therefore often a preferred choice if data from the whole population is available. However, the nature of the exposure-event pairs considered here poses a problem to cohort designs. One problem is that the exposure to vaccination is optional, possibly resulting in selection bias. The 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 the not vaccinated populations of selected birth cohorts. This would provide a comparison which can potentionally utilize a lot of 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 not vaccinated individuals is often small and possibly biased due to selection. The observed differences in the rates of events could be 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 full birth cohorts from pre- and post eras of introduction of a new vaccine to a vaccination program. The pre era could be used to estimate the baseline incidence and the incidence during the post era could then be compared against the baseline. This approach

utilizes a lot of data and provides a seemingly unbiased comparison, assuming that exposure to vaccination is the only significant change between the different birth 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 (persons with the events of interest). In case-only designs, each individual with an event acts as their own control and hence the designs are also called self-controlled designs. The designs considered here are based on comparison of event or exposure rates between defined risk and nonrisk periods.

In self-controlled designs, comparisons are made within the individuals such that all time-invariant confounders are eliminated [Nordmann et al., 2012]. Simulation studies by both McClure et al. [2008] and Glanz et al. [2006] have found that the 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.

According to Farrington [2004], both CCO and SCCS can and have been used for vaccine safety analysis and according to Maclure et al. [2012], they are better than cohort designs for investigating transient effects of vaccines.

In the following sections, I will present the CCO and SCCS in detail. CCO and SCCS are suitable for studying rare adverse events for which a risk interval can be defined based on previous knowledge.

2.1.3 CCO and SCCS

The case-crossover design (CCO) and the self-controlled case series design (SCCS) are tow of the most popular case-only designs. Both designs implicitly or explicitly define risk and control periods during which rates of events are of interest. The risk period is a time interval following exposure during which the incidence rate of the adverse event can be thought of as 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 rate of the individuals

- **The case-crossover design** (CCO): 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) A case-only study design derived from a cohort design logic. Time is viewed from the point of view of the exposure. Ages at vaccination are regarded as fixed, and the random variable of interest is the age at event, conditionally on its occurrence within a pre-determined observation period. [Farrington, 2004]

In CCO and SCCS, events occuring during the risk period can be labeled as "cases" and events during the control period as "controls." Individuals 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. Both designs aim at comparing the incidence rates of the adverse events between the risk and control periods.

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 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 designs. The SCCS aproaches the problem by directly considering the rates of events during the different periods of risk. CCO approaches the problem "backwards" by considering the rates of exposure instead. In SCCS the rate ratio can be directly estimated and in CCO it can be approximated with the odds ratio of exposure.

What follows is a detailed description of the SCCS method followed by a description of the probability model of the CCO, where I will further discuss the relationship of the rate ratio, risk ratio and exposure odds ratio in CCO and case-control studies in general.

2.2 The self-controlled case-series design (SCCS)

I will now present the self-controlled case series design (SCCS) developed by Farrington [1995]. In SCCS the goal is to compare the relative rates of adverse events during disjoint risk and control time periods, specified in relation to the time 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: event incidence during a period when the exposure to vaccination is assumed to have an effect (risk period)
- λ_0: event incidence 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 significantly differ from 1, with values greater than 1 indicating a positive association between the exposure and the event. The risk period could for example 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.

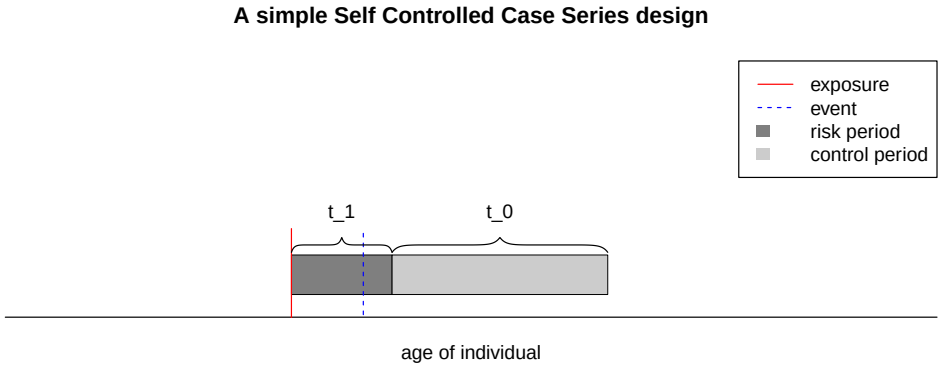


Figure 2.1: A simple example of a SCCS design where the period immediately following exposure is defined as the risk period with length t_1 and a time period following the risk period is defined as the control period with length t_0.

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 nuissance parameter. Inference concerning relative incidences (rate ratios) can therefore be done within individuals such that the individual baseline incidences have no effect on the inference.

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 concept of a Poisson incidence function and the conditional likelihood for the parameters of that function. I will then proceed to present a simple version of the SCCS design, which I will call the simple SCCS, where interest is to quantify the relative incidence of adverse events during a specified risk and control period. I will then describe the general SCCS, which allows for inference regarding multiple risk periods and age.

2.2.1 Poisson process

An underlying assumption in the SCCS method is that adverse events arrive 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. This implies that an occurance 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 consequence.

Definition (Poisson process)**.** *Let D be the time period of interest and d disjoint subsets of D. Let Y_d denote the number of events during d with $E[Y_d] = \lambda_d$. A Poisson process is defined by the following assumptions:*

1. The random variables Y_d are independent

2. The point probabilities for the number of events are given by the Poisson distribution:

$$y_d \sim Poi(\lambda_d) \Leftrightarrow P(Y_d = y_d) = \frac{\lambda_d^{y_d} \cdot e^{-\lambda_d}}{y_d!}$$

2.2.2 Incidence function

The observation period in SCCS is the calendar period and ages in days during which individuals can be sampled as cases. An assumption is made that during this observation period, adverse events for individuals $i = 1, 2, ..., N$ arrive as a non-homogenous piecewise-constant Poisson process. What this means is that while the intensity of the process can vary, it partly remains constant. For each individual, I will index the observation period by days denoted by d. In this work the smallest measurable unit for time is a day, since data in electronic health care data bases such as AvoHILMO is collected daily.

The rate of the adverse events during each day for each individual depends on the idividual intensity of the Poisson process during that day. The incidence is assumed to be constant within each day, but can vary between days. The intensity during day d for individual i is given by the incidence 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 changes in age. The effect of each θ_d is assumed to be multiplicative, so that the incidence function 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 total 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 function is then given by

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

For example in a simple case as in figure (2.1), where $q \in \{\text{risk period}, \text{control period}\}$, the incidence function takes on two different values.

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 the observation period is as described above and that events arrive as a Poisson process with an incidence function 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: $y_{iq} \sim Poi(\lambda_{iq})$. The individual contribution to the likelihood for the observations 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 nuissance 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 corresponsing to the total number of events is a sum of Poisson random variables, it is also Poisson distributed $n_i \sim Poi(\sum_q \lambda_{iq})$. 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, pluggin in the actual incidence function (3), the conditional likelihood is

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

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 incidences 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 SCCS 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.

2.2.4 Simple SCCS

Consider a SCCS with a simple incidence 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 a risk period and 0 indicates a reference control period. The parameter of interest is ψ_k which quantifies the multiplicative effect that the risk period has on the incidence.

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 function is then

$$\lambda_{ik} = t_k \cdot exp(\phi_i + \psi_k) \tag{7}$$

Assume that during the control period $\psi_0 = 0$ and the incidence is simply determined by the baseline incidence of the individual $t_0 \cdot e^{\phi_i}$ and during the risk period the incidence is given by $t_1 \cdot e^{\phi_i} e^{\psi_1}$. Then, $RR = \lambda_{i1}/\lambda_{i0} = e^{\psi_1}$ is the common rate ratio of the incidence between the risk and conrol periods.

Now assume that the observation period is as described earlier and assume again 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. Denote the number of events for individual i during the risk period as y_{i1} and during the control period as y_{i0}. The individual contribution to the conditional likelihood (5) for the parameter $\boldsymbol{\psi}$ is

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

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 interval as "cases", and the events that occured during the control interval 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

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

2.2.5 General SCCS

I will now present a more general version of SCCS, including an age effect parameter and multiple risk intervals 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.

Concider an observation period as described earlier and assume that adverse events arrive as a piecewise-constant Poisson process with an incidence function similar to (3). The observation period for each individual is split into disjoint intervals by age groups denoted by γ_j and disjoint risk periods denoted by ψ_k. The number of days in an interval where i is in age group j and risk period k is denoted by t_{ijk}. The incidence 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 each interval 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

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

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. It's use in drug safety studies has increased during 2000-2010 [Nordmann et al., 2012].

Following the description of Maclure et al. [2012], 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*. A *control day* is defined as the day d number of 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, but I will call these periods the *related period* and the *non-related period*, respectively, in order to avoid confusion with the risk and control periods of the SCCS design, where the risk period is a time period

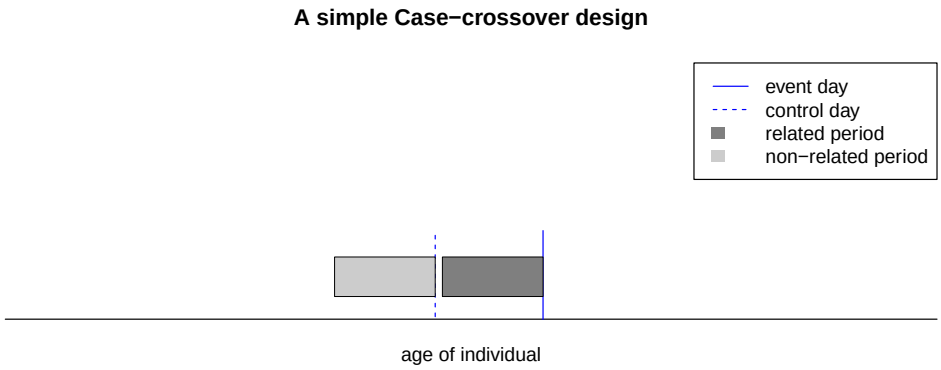


Figure 2.2: A graphical illustration of the CCO design. Exposures occuring immediately before the event day are considered as related to the event and exposures occuring immediately before the control day are considered as not related to the event.

during which the rate of events (not exposures) is of interest. The idea of the CCO method 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

The description given above follows a 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. The estimation of disease-exposure relationship in case-control studies may however also be approached using a prospective conditional argument. In a 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 $[t_0, t_1]$ 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 is comparable to the risk period of the SCCS design and thus I will use the term risk period to refer to the effect period. I will also refer to the time outside the risk period as the control period, as in SCCS. The individuals in CCO can be thought of as sampled as either "cases" or "controls", where individuals for which vaccination occured during the related period are sampled as "cases" and individuals for which vaccination occured during the non-related period are sampled as "controls".

CCO is interested in the same question as SCCS, which is to infer if the rate of events is affected by exposure. Even though the method is based on the idea of exposures as the random variable, the parameter of interest can be defined as either the odds ratio of exposure, relative risk or the incidence rate ratio [Nordmann et al., 2012, p.8]. In what follows, I will introduce the probability model and assumptions of the CCO design in more detail by treating the desing in the general framework for case-control studies following the presentation of Clayton and Hills [1993, ch. 16].

2.3.1 The probability model for CCO

Assume that we have sampled individuals who each experienced at least one of the events of interest. The date of the events is labeled as the event day. Define a control day as the day d days prior to the event day. The control day could for example be two weeks and one day before the event day. Also define the related and the non-related periods, or equivalently the risk and the nonrisk risk periods, as each being of length t days. The risk period could for example be a period of two weeks following vaccination, in which case the related period is the period two weeks prior to the event day. Then, the non-related period is the period two weeks prior to the control day. See again figure 2.2 for an illustration.

To set up the notation, denote the probabilities related to the occurance of an adverse event for individual i during the risk period as $P(AE_1) = \pi_{i1}$ and during the control period as $P(AE_0) = \pi_{i0}$. Denote the random variable corresponding to being vaccinated during the related period by V_{i1} and during the non-related period by V_{i0}.

Now assume that in the underlying cohort (the study base), exposure to vaccination for each individual is equally likely during both the related and non-related periods. The random variables $V_{i\cdot}$ are then Bernoulli random variables with equal probabilities of success $P(V_{i1}) = P(V_{i0}) = p_i$. Note, however the distinction to the conditional probabilities $P(V_{i\cdot} \mid AE)$ which are not assumed to be equal.

To solve the probability that an individual who experienced an event was vaccinated during the related period, notice that this sequence of events requires that an individual was first vaccinated and then experienced an event during the risk period. Therefore, the probability is given by $p_i \cdot \pi_{i1}$. Similarily, the probability that an individual with an event was vaccinated during the non-related period is given by $p_i \cdot \pi i0$. This is illustrated in figure 2.3. I will call these probabilities the probability model for the CCO.

Using the probability model for the CCO, I will next derive equations for the odds ratio of exposure between the related and non-related periods and discuss it's relationship to the risk ratio and rate ratio.

2.3.2 The odds ratio and risk ratio

Continue from the probability model for the CCO and define O_1 as the odds of vaccination during the related period and O_0 as the odds of vaccination during the non-related period.

$$O_1 = \frac{p_i \cdot \pi_{i1}}{(1 - p_i) \cdot \pi_{i0}}, \quad O_0 = \frac{p_i \cdot \pi_{i0}}{(1 - p_i) \cdot \pi_{i0}}. \tag{14}$$

Then, the odds ratio (OR) of vaccination between the related and non-related periods is

$$OR_i = \frac{O_1}{O_0} = \frac{p_i \cdot \pi_{i1}}{(1 - p_i) \cdot \pi_{i0}} \cdot \frac{(1 - p_i) \cdot \pi_{i0}}{p_i \cdot \pi_{i0}} = \frac{\pi_{i1}}{\pi_{i0}}. \tag{15}$$

The underlying probability of exposure is canceled out and the odds ratio is seen to be equivalent to the risk

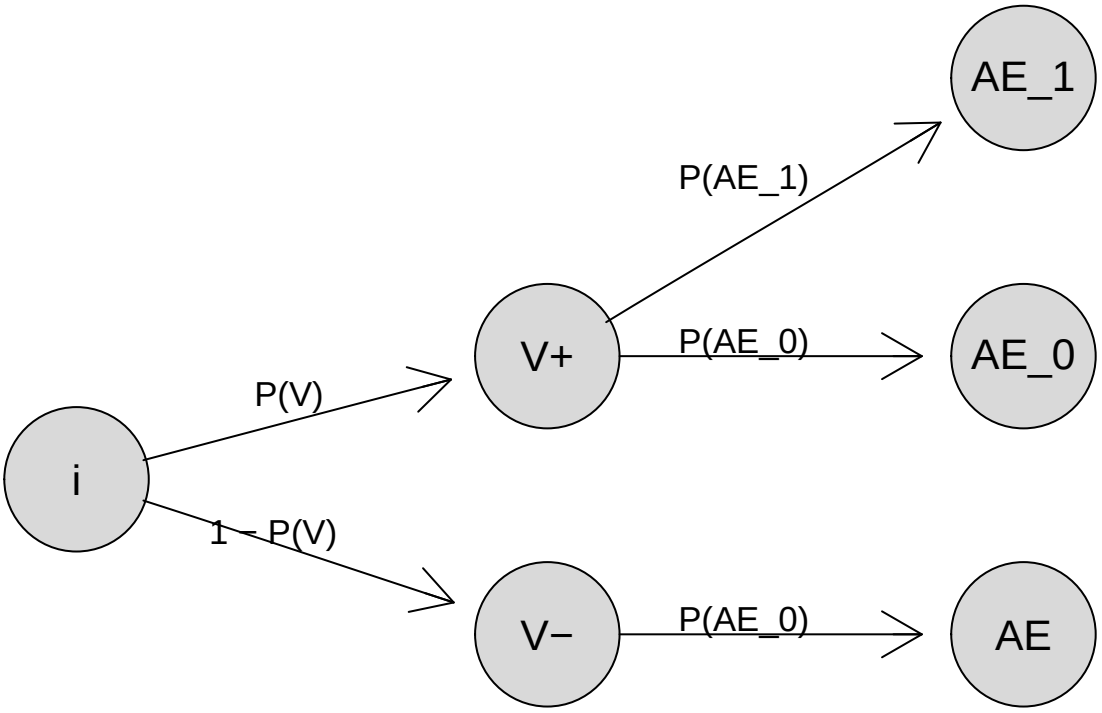


Figure 2.3: The probability model in the study base for the case-crossover design, where only individuals with events (AE) are considered. An individual is first either vaccinated (V+) with probability P(V) or not vaccinated (V-). If she was vaccinated, the AE she experienced happened during a risk period with probability P(AE_1) or a control period with probability P(AE_0). If she was not vaccinated the AE happened with probability P(AE_0)

ratio of the event occurrance between the risk and control periods, defined as π_{i1}/π_{i0}. When the probability of the event occurance is small (rare event), this ratio is also approximately equal to the event rate ratio [Clayton and Hills, 1993, ch. 16].

There is no analytical solution to the likelihood equations of the CCO design. In his original paper, Maclure 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 [Marshall and Jackson, 1993].

2.3.3 CCO in the literature

The literature concerning the CCO approach can be confusing, since it must be interpret from context if 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, Norman et al. 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. [Nordmann et al., 2012, p. 2]

When discussing CCO, Farrington [2004] writes that *"the underlying probability of vaccination should be the same in all intervals"*, which from equation 15 can be seen is a necessary condition for the fact that the odds ratio of exposure and the odds ratio of event status are equal. Farrington [2004, p. 2066] then adds that this assumption *"is unlikely to hold for paediatric vaccines administered according to stricter schedules"*, which implies that the related and non-related periods would be defined in relation to age or calendar time. It is my understanding that this however need not be true, since both periods can be defined in relation to the event time, which clearly must not necessarily obey a strict schedule.

2.4 Study design conclusions

Case-only designs are attractive for vaccine safety surveillance mainly because they implicitly control for time-invariant confounders. However, it remains important to control for time-varying confounders such as age and seasonality [Farrington, 2004, *Introduction*].

Of the two most popular case only designs, the SCCS compares somewhat favourably to CCO. The SCCS allows for direct inference concerning the population parameter of interest, which is the rate ratio of adverse events between specified risk and control time periods. Farrington [2004, p. 2066] has also pointed out that in the case of paediatric vaccines there is at least a conceptual problem with the CCO design, where the idea is to treat vaccination times as random variables. It is more natural to view the event times as the random variables of interest as in SCCS, since vaccinations often occur according to a strict schedule.

The simple SCCS introduced in 2.2.4 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 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 safety surveillance

One of the key elements of a safety surveillance method is a decision rule for generating safety signals. In safety surveillance, the 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 analysed and decisions related to safety signal generation are made based on the sample. The sample never represents the total study population of interest, which can be thought to also include unborn individuals, possibly experiencing adverse events in the future. Furthermore, there is presumably random variation in the occurrance of adverse events in the population.

The decision making process in safety surveillance involves uncertainty. In this chapter, the focus will be on statistical hypothesis testing, which is a well known method for decision making under uncertainty.

3.1 Hypothesis tests in vaccine safety surveillance

Statistical hypothesis testing is a method in statistical inference, used to make decisions 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.

When an association between a vaccination and an adverse event exists, the goal of a safety surveillance method is to generate a safety signal as soon as possible. This goal can be expressed as a problem of minimizing the expected sample size until signal generation, for some fixed rates of false positive and false negative signals. In other words, the goal is to minimize the number of observed adverse events needed to generate a signal.

In the following sections, I will discuss the use of sequential hypothesis testing for deriving decision rules for vaccine safety signal generation. I will first describe hypothesis testing in general and then discuss testing with sequential observations. I will introduce the sequential probability ratio test (SPRT), which was the first hypothesis testing method developed for sequential analysis. I will then describe in detail the maximized sequential probability ratio test (maxSPRT), which is a sequential hypothesis testing method designed for vaccine safety surveillance.

3.1.1 Statistical hypotheses

Statistical inference is based on a probability model $P(Y|\boldsymbol{\theta})$ for the observations Y, under some parameters $\boldsymbol{\theta}$. A statistical hypothesis is a proposition which assigns restrictions for the parameter of a 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 test can only find evidence *against* a defined hypothesis, by showing that the observations are unlikely under the statistical model defined by the hypothesis. Therefore, evidence for H_1 is obtained by finding evidence against H_0.

3.1.2 The hypotheses in safety surveillance

A hypothesis is called simple, if it addresses only a single point in the parameter space. 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 non-exposed is positive), the null proposition (hypothesis) is a state of no association. The aim is to find evindence against the null proposition. The alternative hypothesis is composite: some positive association. Using the rate ratio parameter RR, this can be stated as two competing hypothesis as follows:

$$H_0 : RR = 1 \quad H_1 : RR > 1 \tag{16}$$

3.1.3 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 if H_0 were true, then one can choose to reject H_0.

The test statistic can be for example the likelihood ratio:

$$LR = \frac{L(\boldsymbol{\theta}_1; \mathbf{y})}{L(\boldsymbol{\theta}_0; \mathbf{y})} = \frac{P(\mathbf{y} \mid \boldsymbol{\theta}_1)}{P(\mathbf{y} \mid \boldsymbol{\theta}_0)}, \tag{17}$$

Now, if the hypotheses are as in (16), and the parameter of interest is RR, the power (see table 3.1) 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. In the context of vaccine safety surveillance this 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 interesting because it leads to the following observation: when the alternative hypothesis is composite, calculating the power of a test before collecting observations is impossible without fixing the value of RR.

Next, I will introduce sequential hypothesis testing and Walds sequential probability ratio test (SPRT), which was designed for testing two simple hypotheses. Wald's suggestion for computing the critical values of the SPRT test requires a known power (or Type II error, β).

Table 3.1: Terminology related to statistical hypothesis testing

Term	Description
H_0	The null hypothesis ($H_0 : \theta = \theta_0$)
H_1	The alternative hypothesis ($H_1 : \theta = \theta_1$)
Composite hypothesis	A hypothesis which adresses more than a single point in the parameter space. For example $H_1 : \theta > 1$
α	$P(\text{reject } H_0 \mid H_0)$. The type I error (false positive rate)
β	$P(\text{accept } H_0 \mid H_1)$. The type II error (false negative rate)
$Power$	$1 - \beta$

3.2 Sequential analysis

The goal in safety surveillance is to stop observing and take some further action if an unexpectedly high number of adverse events are observed. Dynamically determing what exactly is an "unexpectedly high number" for each point during the surveillance, is the problem that this chapter aims to solve by using methods of statistical hypothesis testing.

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 could 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 regular hypothesis test on the accumulated data set each time new observations become available. As discussed earlier, hypothesis testing is based on

the general idea that if observations are very 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 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 was developed by Wald [1945] during the second world war and it 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 most powerful sequential hypothesis test between two simple hypotheses in the sense that for given error rates, SPRT minimizes the expected sample size until a decision can be made.

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. In the following sections I will first briefly introduce SPRT and then describe the maxSPRT and it's self-controlled binomial variant BmaxSPRT in detail.

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:

$$LR_n = \frac{L(\boldsymbol{\theta}_1; \mathbf{y}_n)}{L(\boldsymbol{\theta}_0; \mathbf{y}_n)} = \frac{P(\mathbf{y}_n \mid \boldsymbol{\theta}_1)}{P(\mathbf{y}_n \mid \boldsymbol{\theta}_0)}, \tag{18}$$

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. He also later 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 [Wald and Wolfowitz, 1948].

Algorithm 3.2.1: SPRT

Input: Desired Type I and Type II error rates α and β, test statistic function LR_n as in (18)

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, defined by the two critical values A and B. These regions define the stopping criterias for the test for all observations $n = 1, 2, ...$ The SPRT procedure and the critical regions are illustrated in 3.1.

In order to determine the values A and B, Wald [1945, pp. 127-133] considered the entire sample space consisting of all possible realisations in the sequential test. He showed that the critical values 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 procedure will guarantee that the actual type I and II errors will not exceed α and β and will only differ from them slightly.

The above is a beautiful and simple result, but the solution requires a fixed α and β. The latter, as discussed before, in turn would require fixing the parameter of interest (RR), which can be considered a weakness. In what follows I will introduce Wald's suggested solution for dealing with a composite alternative hypothesis and then illustrate the weakness related to the solution.

3.2.3 SPRT and a composite alternative hypothesis

SPRT was designed for testing two simple hypotheses. Wald did however propose a solution to a situation of a composite alternative hypothesis. He 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. [Wald, 1945, p. 158]

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 $H_1 : RR = RR_1$ and the actual (i.e. true) RR_a can either

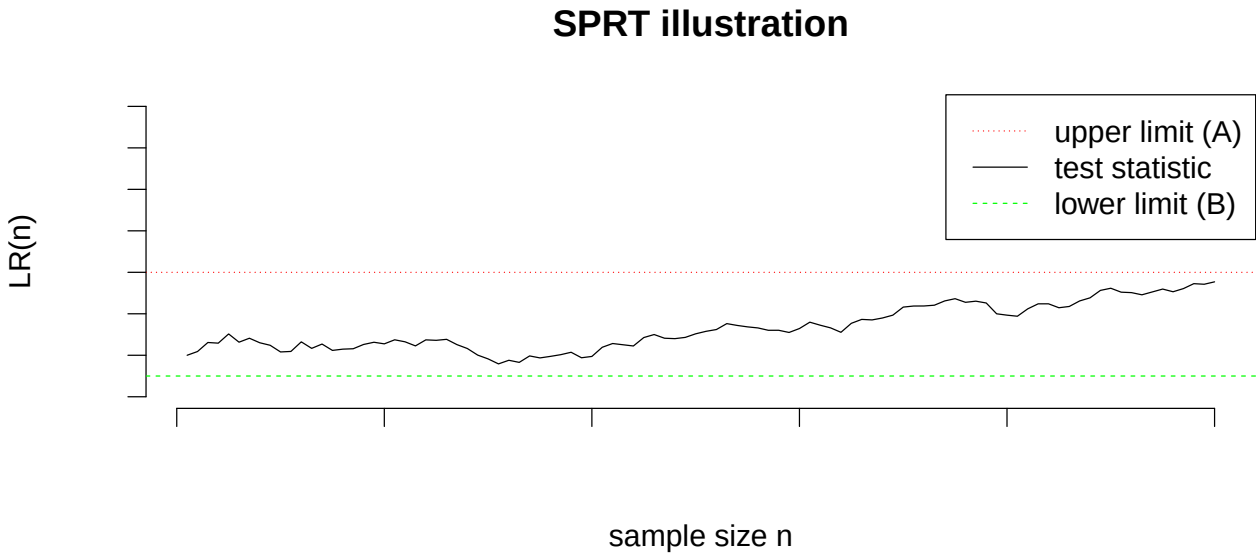


Figure 3.1: A graphical illustration of the critical boundaries of the sequential probability ratio test (SPRT) where the test statistic is the likelihood ratio (LR). If the nominator of LR is the likelihood according to the alternative hypothesis, then high values of LR mean that the alternative model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

1. delay the rejection of H_0 when the observed rate of events is higher than the specified alternative hypothesis would expect (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 the models defined by H_0 and H_1 are bad models 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 bad models. In such a case the high number of adverse events expected to be observed would be given low probability by both models. The 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.

The above 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 figure (3.2) for a

graphical illustration.

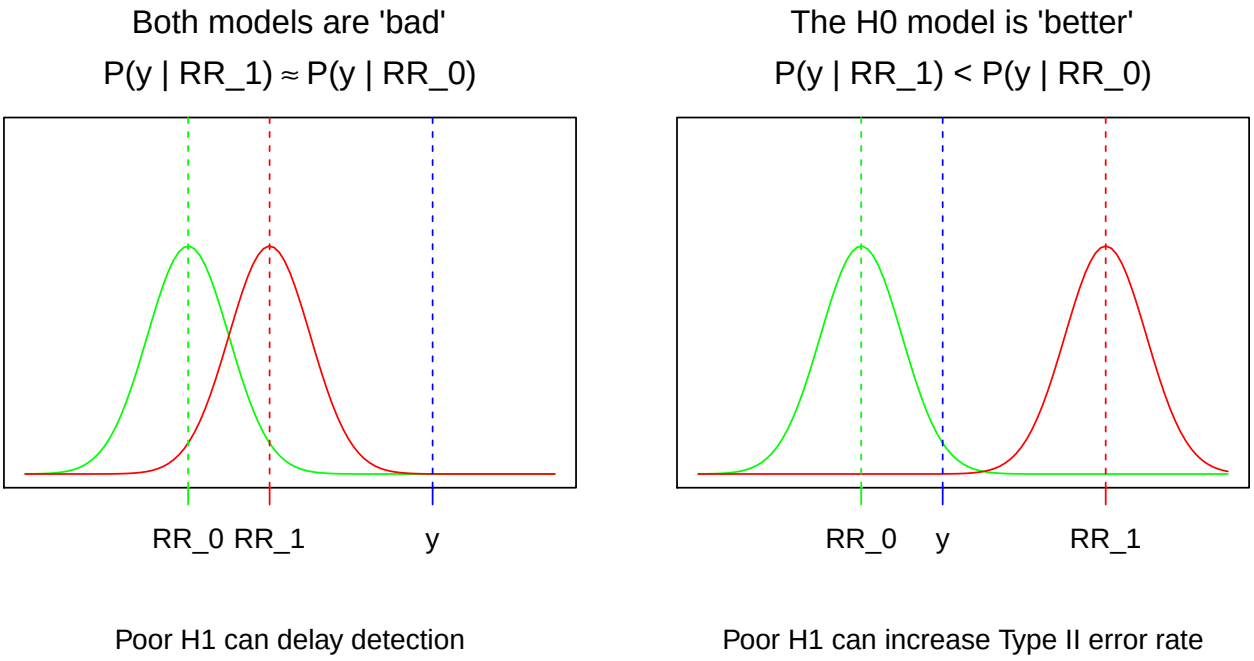


Figure 3.2: A graphical illustration of a weakness of Wald's SPRT. The blue line describes the number of observations and the green and red lines describe the statistical model for the observations under the null and alternative hypothesis, respectively. Left: A choice of point alternative hypothesis close to the null hypothesis and far from the actual observations can delay the rejection of H0, because the (high number of) observations are unlikely under both hypotheses. Right: A choice of point alternative hypothesis that is 'too agressive' compared to the actual observations can increase the type II error rate, because the data is more likely under the null hypothesis.

3.3 Maximized sequential probability ratio test: maxSPRT

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

1. Maximize the likelihood ratio in the space of the alternative hypothesis 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] introduce 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 2.2.4.

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

$$LR_n = \max_{RR_1} \frac{L(RR_1; \mathbf{y}_n)}{L(RR_0; \mathbf{y}_n)} = \max_{RR_1} \frac{P(\mathbf{y}_n \mid RR_1)}{P(\mathbf{y}_n \mid RR_0)}. \tag{19}$$

The maxSPRT procedure is described in algorithm 3.3.

Algorithm 3.3: MaxSPRT

Input: Desired type I error rate α', upper boundary for the sample size N, test statistic function LR_n as in (19)

1. Compute the critical value of the test c.
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

3.3.1 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 2.2.4. The probability of "success" (i.e. adverse event clasified as a "case"), depends on the unkown rate ratio parameter RR and the known ratio between the lengths of the nonrisk and risk periods, z. The conditional likelihood is as in (10).

Let the two hypotheses be as in 16. Using the equation (19), the maxSPRT test statistic is given by

$$LR_n = \max_{RR>1} \frac{P(y_n \mid RR)}{P(y_n \mid RR = 1)} = \max_{RR>1} \frac{(\frac{RR}{z+RR})^{y_n}(\frac{z}{z+RR})^{n-y_n}}{(\frac{1}{z+1})^{y_n}(\frac{z}{z+1})^{n-y_n}}. \tag{20}$$

Using this test statistic is what Kulldorff et al. [2011] call the binomial maxSPRT (BmaxSPRT) method. A simple SCCS design combined with the the maxSPRT method is one way to arrive at the BmaxSPRT, which shows that the BmaxSPRT method is self-controlled.

Computation of (20) requires maximization. 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. It is easy to see that maximization in terms of RR depends only on the nominator, 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 $\hat{RR}$. The (log) likelihood ratio is thus 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

$$LLR_n = log(LR_n) = log\left(\frac{(\frac{\hat{RR}}{z+\hat{RR}})^{y_n} \cdot (\frac{z}{z+\hat{RR}})^{n-y_n}}{(\frac{1}{z+1})^{y_n} \cdot (\frac{z}{z+1})^{n-y_n}}\right). \tag{21}$$

3.3.2 Computing the critical values of BmaxSPRT

The values of the test statistic are easy to compute. What then remains is the definition of the critical region Q of the test: which values of the test statistic should lead to a decision to reject 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 c, which defines the boundary of the critical region. Values of (21) higher than c then lead to rejection of H_0. The critical region of BmaxSPRT is

$$Q = \{LLR_n \mid LLR_n \geq c, LLR_{n-1} < c\}, \quad \text{for all} \quad n = 1, 2, ..., N,$$

where N is the maximum number of observations (adverse events). In order to compute the critical value, one must first choose a desired Type I error rate α' and an upper boundary for the sample size N. The goal is to then define c so that when H_0 is true, Q is only expected to be reached α' proportion of the time.

The upper boundary of the sample size makes it possible to compute the critical value of the maxSPRT test to any desired precision. Kulldorff et al. [2011, pp. 65-67, p. 72] describe how to do this for the Binomial and Poisson likelihoods. In the discrete binomial case the critical value can be computed by using a Markov chain approach.

Next, I will introduce the concept of a Markov chain in the setting of the BmaxSPRT experiment. I will define the state space, the initial probabilities and the transition probabilities, and show that the BmaxSPRT experiment has the Markov property. I will then proceed to present an algorithm which can be used to compute the critical values of BmaxSPRT.

3.3.2.1 State space

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

$$S = (n, y_n), \quad \text{where} \quad n = 0, 1, ..., N \quad \text{and} \quad y_n = 0, 1, .., n. \tag{22}$$

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

In the BmaxSPRT experiment, a new observation is either a "success" ($Y = 1$) or a "failure" ($Y = 0$), depending on the outcome of a Bernoulli random variable Y. Possible transitions in the state space S are therefore as follows.

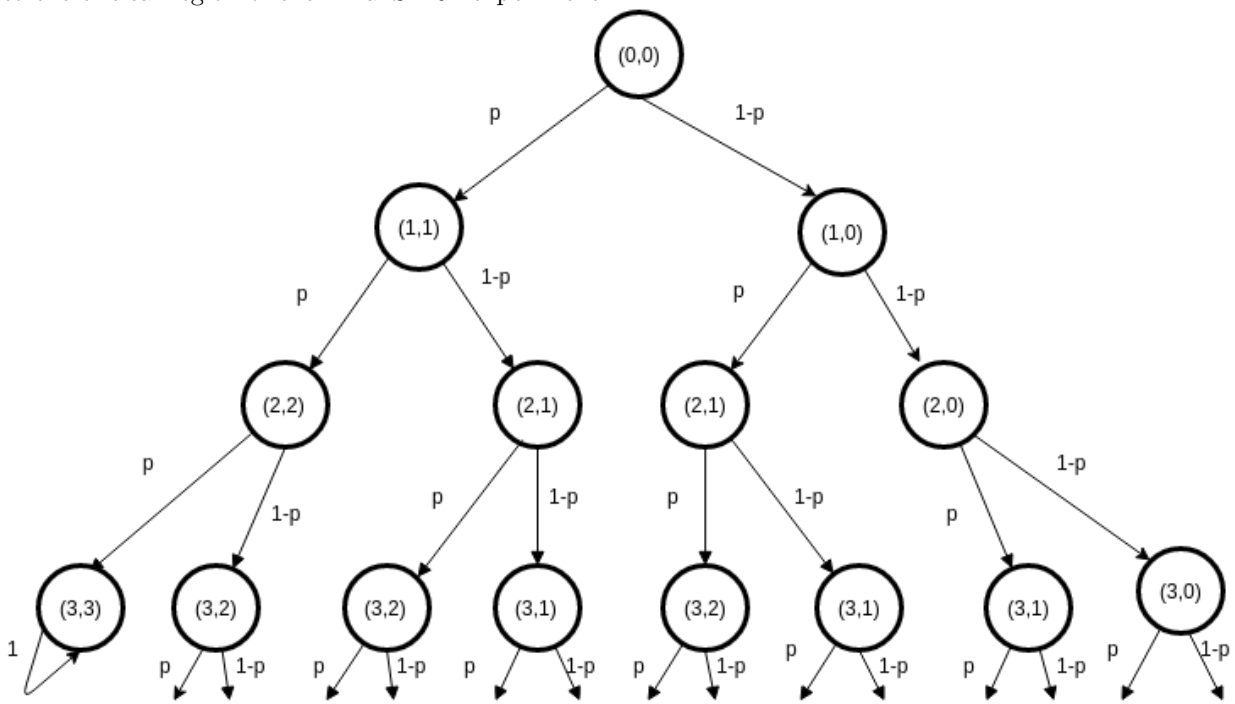
$$\begin{aligned} "success" &: (n, y_n) \to (n+1, y_n+1) \\ "failure" &: (n, y_n) \to (n+1, y_n). \end{aligned} \tag{23}$$

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 set of absorbing states in the experiment are the states for which the value of the test statistic reaches the critical value c and the experiment ends. I will denote the value of the test statistic (21) for the state s by $LLR(s)$. The set of absorbing states in the experiment are then

$$Q = \{s \in S \mid LLR(s) \geq c\}, \tag{24}$$

which is the critical region of the test.

Figure 3.3: Visualization of the BmaxSPRT 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, corresponding to belonging to the critical region of the BmaxSPRT experiment.



3.3.2.2 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

$$\underset{1\times M}{\mathbf{v}} = (1, 0, ..., 0). \tag{25}$$

In the BmaxSPRT experiment the random variable Y which corresponds to a single labeling of an adverse event as a "case" or "control" ("success" or "failure") determines the transitions and has a Bernoulli distribution. The transition probabilities for the transient states are given by

$$\begin{aligned} P\{(n, y_n) \to (n+1, y_n+1) \mid LLR(n, y_n) < c\} = P(Y = 1) = p \\ P\{(n, y_n) \to (n+1, y_n) \mid LLR(n, y_n) < c \ \} = P(Y = 0) = 1 - p, \end{aligned} \tag{26}$$

where p is as in (9) and thus 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 are

$$P\{(n, y_n) \to (n, y_n) \mid LLR(n, y_n) \geq c \ \} = 1, \tag{27}$$

meaning that if such a state is reached, it is never left.

The matrix $\mathbf{P}$ in (28) is a transition matrix which gathers the transition probabilities and defines the probability distribution over all the transitions from state to state in the state space S as in (22). 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 $s_i -> s_j$.

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

$$\underset{1\times M}{\mathbf{p}} = \mathbf{v}\mathbf{P}^{(N)} \tag{29}$$

3.3.2.3 Markov property

The Markov property means that the probability of transitioning to the next state depends only on the current state. It is easy to see that the BmaxSPRT experiment has a Markov property. Since the transitions are defined by independent Bernoulli trials, the probability of moving to the next state depends only on whether the current state is absorbing or not, as is shown in (30).

$$
\mathbf{P} = \begin{array}{c} \\ (0,0) \\ (1,0) \\ (1,1) \\ (2,0) \\ (2,1) \\ (2,2) \\ . \\ . \\ (N,0) \\ (N,1) \\ . \\ . \\ (N,N) \end{array}
\begin{array}{c}
\begin{array}{cccccccccccc} (0,0) & (1,0) & (1,1) & (2,0) & (2,1) & (2,2) & ... & (N,0) & (N,1) & ... & (N,N) \end{array} \\
\begin{pmatrix}
0 & 1-p & p & 0 & 0 & 0 & ... & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 1-p & p & 0 & ... & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & 0 & 1-p & ... & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & ... & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & ... & 0 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & ... & 0 & 0 & ... & 0 \\
. & . & . & . & . & . & ... & . & . & ... & . \\
. & . & . & . & . & . & ... & . & . & ... & . \\
0 & 0 & 0 & 0 & 0 & 0 & ... & 1 & 0 & ... & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & ... & 0 & 1 & ... & 0 \\
. & . & . & . & . & . & ... & . & . & ... & . \\
. & . & . & . & . & . & ... & . & . & ... & . \\
0 & 0 & 0 & 0 & 0 & 0 & ... & 0 & 0 & ... & 1
\end{pmatrix}
\end{array}
\tag{28}
$$

Figure 3.4: A transition matrix describing the transition probabilities between the states in the state space S of the BmaxSPRT experiment. The rows and columns of the matrix are the possible states and $\mathbf{P}_{ij}$ gives the probability of the transition $s_i -> s_j$. For example under the null hypothesis of the BmaxSPRT experiment the probability of the transition $(0,0) -> (1,0)$ ("failure") is given by $1-p = z/(1+z)$ and the probability of the transition $(0,0) -> (1,1)$ ("success") is given by $p = 1/(1+z)$. The states where the maximum number of observations is reached are all absorbing states.

$$
P(S_{i+1} = s_{i+1} \mid S_1, .., S_i) = \begin{cases} P(Y_{i+1} = y_{i+1} \mid Y_1, .., Y_i) \overset{!!}{=} p, & \text{when } LR(s_i) < c, \\ 0, & \text{when } LR(s_i) \geq c. \end{cases} \tag{30}
$$

3.3.3 Algorithm for BmaxSPRT critical values

Let N now be the chosen maximum number of observations (adverse events) and α' the desired type I error rate in the BmaxSPRT experiment. Using the notation for the absorbing states defined in (24), the goal is to define c such that $P(Q \mid H_0) = \alpha'$, meaning that the overall probability of rejecting the null hypothesis when it is true is α'.

For any given critical value c, the overall probability of reaching the critical region of the BmaxSPRT experiment can be computed by summing over the probabilities of being in each of the absorbing states $s \in Q$ after N transitions in the state space S. Using the result introduced in (29), this probability is given by

$$
P(Q \mid H_0) = \sum_{s \in Q} \mathbf{v}\mathbf{P}_0^{(N)}, \tag{31}
$$

where $\mathbf{P}_0$ is the transition matrix (28), where $p = 1/(1+z)$ is according to H_0. Using this result, it is then possible to try out possible values of c until the one 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 (21)

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 of BmaxSPRT

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 (low) 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, as in equation 32, where $\mathbf{v}$ are the initial probabilities of the states and $\mathbf{P}_1$ is the transition matrix (28), where $p = RR_1/(RR_1 + z)$ for some fixed choice of RR_1.

$$P(Q_0|H_1) = \sum_{s \in Q_0} \mathbf{v}\mathbf{P}_1^{(N)}, \tag{32}$$

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 is expected that in this case sometimes more than a single observation arrives simultaneously. 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.

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 requires adaptation and 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 becomes 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 'harder' 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) are at least y_c. Now assume a continuous sequential design that looks at the observations separately and rejects H_0 as soon as there are y_c events. The error rates of the designs are identical because the number of events are non-decreasing. 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

When data is collected during unfrequent calendar times (such as three times a year) and observations arrive in groups, group sequential analysis should be used. The same statistical model and test statistic that would be utilized in a continuous sequential analysis can be applied for group sequential analysis. However, when analyses are done for groups of data there are less chances to reject the H_0 and therefore the critical values of the test should be affected. Unfortunately, if the number of observations per group is unkown, the computation of the critical values becomes difficult.

There are two simplifying assumptions which could be made to make the computation of the critical values of a group sequential hypothesis test easier.

1. Fix the number of observations per group.
2. Fix the number of groups.

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. Silva and Kulldorff [2017] have implemented this solution for computing the 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 3.3.2. 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 and are then collected during some rare time intervals, the number of observations per group is by assumption a random variable. Therefore assumption 1 does not seem realistic if the grouped nature of the data were due to administrative reasons.

The problem of computing the critical values of a group sequential test would also simplify if the maximum number of groups was fixed. However, if the group sizes (i.e. the number of observations per group) are unknown, fixing the number of groups will not fix the number of observations at the end of surveillance.

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 and a decision to stop the experiment can be made. 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.

Since both simplifying assumptions seem unsatisfactory, a possible solution is to use and error spending approach, as introduced by Jennison and Turnbull [1999] in their book which focuses on group sequential methods in clinical trials. Next, I will introduce the concepts 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 functions

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

$$P(|Z_1| < c_1, ..., |Z_{k-1}| < c_{k-1}, |Z_k| \geq c_k) = \pi_k \tag{33}$$

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 $f(t)$ is a non-decreasing function which partions the desired type I error rate. $f(t)$ satisfies $f(0) = 0$ and $f(t) = \alpha$ for $t \geq 1$. Jennison and Turnbull suggest a family of error spending functions defined by

$$f(t) = min\{\alpha \cdot t^p, \alpha\} \tag{34}$$

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

3.4.4 Maximum information trials

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

$$I_k = \{var(\hat{\theta}^{(k)})^{-1}\}, k = 1, 2, .. \tag{35}$$

where $\hat{\theta}$ is the estimator for the parameter of interest θ. A maximum information trial uses an error spending function such as (34), where the value of $f(t)$ indicates the cumulative type I error to be spent when a fraction t of the maximum anticipated information has been obtained. The type I errors allocated to each analyses are

$$\begin{aligned} \pi_1 &= f(I_1/I_{max}) \\ \pi_k &= f(I_k/I_{max}) - f(I_{k-1}/I_{max}). \end{aligned} \tag{36}$$

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

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

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 seemingly similar idea to the maximum sample size in maxSPRT. In their example utilizing the maximum information trial approach however, 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 safety surveillance method to Finnish register data

In this proof-of-concept application, BmaxSPRT vaccine safety surveillance was retrospectively applied to data from the Finnish HILMO register. The exposures of interest are three different childhood vaccinations: Measles-Mumps-Rubella Vaccine (MMR), Pneumococcal vaccination (PCV) and the Rota virus vaccination (Rota). The outcome of interest are febrile seizures. The three vaccines are interesting because:

- For MMR there exists a known relationship to febrile seizures
- For PCV there exists a suspected relationship to febrile seizures
- For Rota there is no evidence of a relationship to febrile seizures

MMR, PCV and Rota are all included in the vaccination program in Finland and are free for all children. The main interest of this application was to see if and when the BmaxSPRT method would generate a signal for the MMR-febrile seizure pair, and also to see what the results would be for the other pairs.

4.1 Febrile seizures and vaccinations

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.

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 for childred of ages 6 months - 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 shot is associated with an increased risk of febrile seizures during the 24 hours following exposure. A PCV vaccination alone might also cause a small increased risk of seizures. There is no evidence of Rota vaccination having similar effects. [CDC]

4.2 Study population and data sources

Finnish children under 2 years of age, born during 2010 - 2014 were considered eligible for this study. Vaccination data of all the three vaccines (Rota, PCV, MMR) was collected from the National Vaccination Register maintained by THL. Health care data regarding the febrile seizures was collected from the HILMO register.

4.2.1 Vaccinations and vaccination groups

One challenge with analysing the effects of vaccines in an observational study is that multiple vaccinations can be given during the same day. The recommended schedule for each of the three vaccinations of interest is given in table 4.1. For example, the second dose of Rota is recommended to be taken dring the same age as the first dose of PCV. In an effort to make the vaccines more identifiable, only first doses of each of the vaccines were considered in the experiment.

The vaccinations of interest are the first doses of MPR, PCV and Rota. Only unique vaccination records in the vaccination register were considered. A non-unique vaccination record was defined as a record with the same personal identification code and the same date as another record.

The three vaccines were analysed separately by defining vaccination groups. A child belongs to a vaccination group if he/she was vaccinated with that vaccine during the observation period. A child can belong to multiple vaccine groups. For each of the vaccines, age distributions at first dose were analysed and the observation periods for each vaccination group were defined so that the observation period covers most first dose vaccinations (>95%). The observation periods for each vaccine group as ages in days are displayed in table 4.1, along with information regarding vaccination coverage and the underlying study population size.

For the MPR group, the 2014 birth cohort was not considered in the study, as due to the higher expected first dose vaccination age, no sufficient data for the 2014 cohort was yet available at the time of the study.

Table 4.1: The recommended vaccination schedule, observation ages, included birth cohorts, size of the study population and the vaccination coverage by vaccination group.

Group	Schedule	Observation.ages	Cohorts	Population	Vaccination.coverage
MPR	12 months	250-650 days	2010-2013	238 568	95%
Rota	2, 3 and 5 months	30-140 days	2010-2014	295 800	93%
PCV	3, 5, and 12 months	60-200 days	2010-2014	295 800	94%

4.2.2 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. A seizure within 7 days from the previous one for the same child was considered to represent the same event and thus the latter event was not included in the analysis.

For each of the vaccine age groups, the event incidences during the non-risk periods were estimated to aid with determining the maximum number of observations for each vaccine group, which is needed as a parameter for the BmaxSPRT experiment. The non-risk incidences were estimated by dividing the number of hospitalizations during the non-risk parts of the observation period by the non-risk person years. The size of the vaccinated population was estimated using available birth data and vaccination coverages (see table 4.1). Table 4.2 displays these incidences as the number of events per 1000 person years.

Table 4.2: Event incidence estimates for the vaccinated populations during the non-risk time periods for each vaccination group. Non-risk time is the time during the observation period excluding the defined risk period of 0-13 days following vaccination. Incidence estimates are given by 1000 person years.

Group	Observation.ages	Nonrisk.events	Nonrisk.incidence
MPR	250-650 days	3588	14.6
Rota	30-140 days	640	8.1
PCV	60-200 days	815	7.9

4.3 Method

The method used in this experiment is the binomial variant of maxSPRT, BmaxSPRT. The analysis was performed with the assumption that new data might 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 be hospitalized, and then pass through both the risk and control periods in order to contribute to the study.

In what follows I will briefly describe the design- and decision elements of the BmaxSPRT method. For details please refer to earlier chapters.

4.3.1 Study design

From an epidemiological design standpoint, the study design in BmaxSPRT is a case-only design, such as the simple SCCS design introduced in section 2.2.4. 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. The children who are not vaccinated have no contribution to the likelihood and so only children who were both vaccinated and were hospitalized during the observation period contribute to the study.

In this application, the risk interval was chosen to be 0 - 13 days following vaccination and the control interval 14 - 41 days following vaccination, for each of the vaccines. This risk interval simultaneously covers the suspected risk intervals for both PCV (24 hours following exposure) and MMR (5 - 12 days following exposure).

It should be noted that the chosen risk interval is likely to be too long especially for PCV, which is suspected to be related to febrile seizures within 24 hours from vaccination. This could affect the power of the BmaxSPRT experiment for the PCV group. However, in a real safety surveillance application, specific knowledge of the actual risk period for the vaccine and adverse event pair might not be available, which is why in this

proof-of-concept application a choice of a conservative common risk period was used. See section 4.5 for a simulation study of how the choice of the risk interval affects the power of the experiment.

4.3.2 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 incidences of adverse events during the risk period (λ_1) and control period (λ_0).

Inference in the experiment is based on a conditional likelihood as in (10), which can be thought of as independent binomial trials, were 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 of an event occuring during the risk period instead of the control period (and thus classified as "case").

The test statistic is the maximized log-likelihood ratio as in (21). The experiment proceeds as described in algorithm 3.3. For each vaccination group a maximum number of observations N is first defined, a desired type I error level α' is chosen and a critical value c is computed. Then for each day of surveillance, the value of the test statistic is computed and compared to the critical value to make a decision. In the BmaxSPRT experiment, 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 at any point during surveillance the value of the test statistic reaches the critical value, the experiment is stopped for the vaccination group and the H_0 is rejected. If this does not happen and the maximum number of observations is reached without rejecting H_0, the experiment ends and H_0 is accepted.

4.3.3 Parameters of the experiment

In order to deploy the BmaxSPRT, one needs to specify

1. The risk interval
2. The control interval
3. The desired maximum type I error rate
4. The maximum number of observations

Alternatively to directly specifying 4, the desired power of the test corresponding to the choices made for 1-3 and a specified rate ratio could be chosen instead, which then determines the maximum number of observations. In this experiment the maximum number of observations was directly defined. An estimate for the expected number of events during surveillance under the null hypothesis (no increased risk during the

risk period) was used as the maximum number of observations. This was done for practical reasons for each vaccination group to help ensure that the experiment would reach a conclusion.

The parameters defining the stopping conditions of the maxSPRT experiment for each vaccination group are displayed in table 4.3. The risk and control intervals are as specified earlier in section 4.3.1 and the desired type I error rate was chosen to be $\alpha' = 0.05$. The power of the sequential hypothesis test was computed for a choice of $RR = 1.5$.

Table 4.3: Parameters of the BmaxSPRT experiment. 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. 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

4.4 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]. In some cases the source code was modified for efficiency.

4.4.1 Surveillance results

The BmaxSPRT method generated a signal on day 693 of surveillance for PCV and on day 1041 for MMR. 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).

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. According to the choices of the risk and control intervals, the expected proportion of cases under the H_0 is approximately 33%. 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. These results are gathered 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	RR	prop.cases
MMR	TRUE	1041	3.85	66	83	149	1.59	44%
Rota	FALSE	1903	0.00	87	172	259	1.01	34%
PCV	TRUE	693	4.03	29	27	56	2.15	52%

4.4.2 More results and conclusions

In a real application of BmaxSPRT, the surveillance immidiately ends once evidence of an association between the vaccine and the adverse event has been found (a signal is generated) or if the maximum number of observation is reached. In this proof-of-concept application however, it is possible to satisfy ones curiosity and see what would happen if surveillance was instead continued for the two vaccination groups 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 for MMR and PCV are illustrated in figure (4.2). Further surveillance shows that when using all the available data for the cohorts under study, 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, when utilizing the whole data set. For MMR the evidence of an association gets 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, the chosen risk intervals (0-13 days from vaccination) are possibly too long especially for PCV, where according to the literature the suspected increased risk of a febrile seizure is within a 24 hour period following vaccination. The sensitivity of the power of the BmaxSPRT experiment to the choice of the risk period is an important question, which I will consider next.

4.5 Sensitivity analysis of BmaxSPRT

It is clear that the results of the BmaxSPRT experiment depend on the choice of the risk interval. Intuitively, if the actual time interval of increased risk is shorter or longer than the chosen risk interval, that should have a decreasing effect on the power of the test. An interesting question is therefore to quantify this effect for

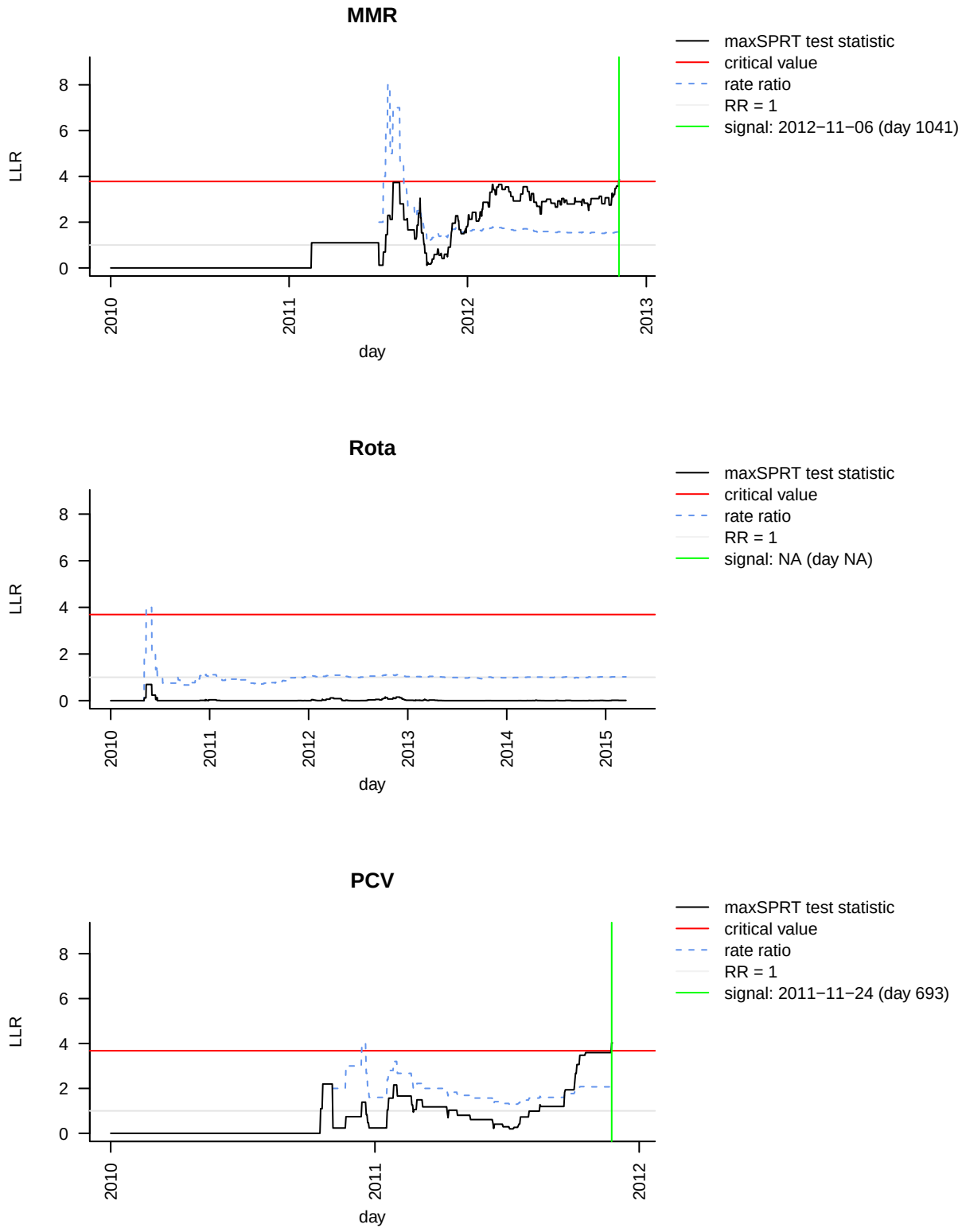


Figure 4.1: Results of BmaxSPRT surveillance for MMR, Rota and PCV. A signal was generated on day 1041 for MMR and on day 694 for PCV as the value of the test statistic (solid black) reached the critical value (solid red). The maximum number of observations was reached for Rota without generating a signal. The running maximum likelihood estimate for the rate ratio is also displayed in the picture (dashed blue) along with a reference line for the value RR = 1 (solid grey).

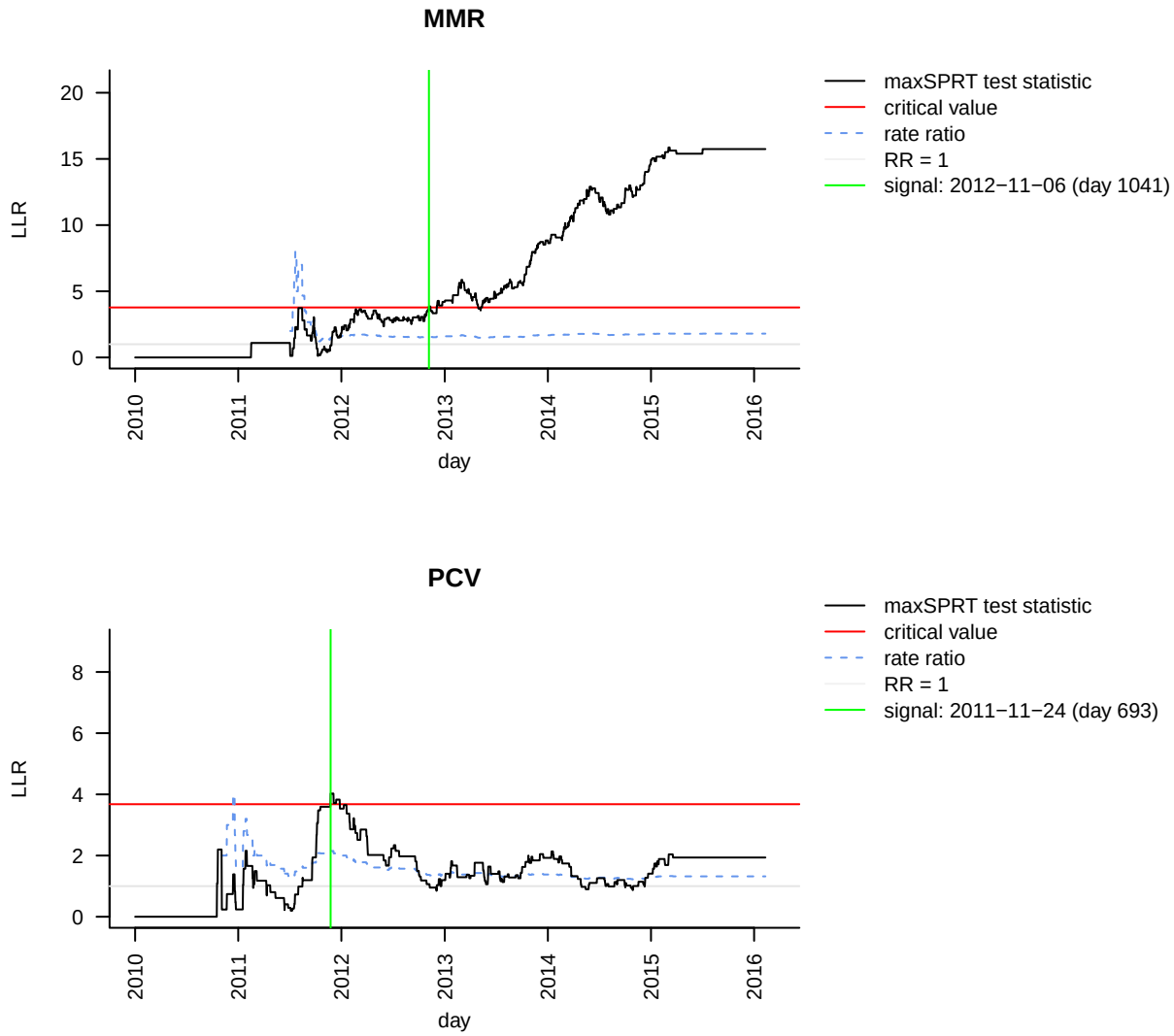


Figure 4.2: Results for continued surveillance for the PCV and MMR groups, both of which a safety signal was generated. Continued surveillance shows that the evidence of an association to febrile seizures for MMR gets stronger when utilizing the complete data set: The value of the test statistic (solid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

different choices of risk and control intervals. I used simulation to study how the power of the BmaxSPRT method is affected by the (poor) choices of risk and control intervals.

The analysis was carried out by repeadetly simulating adverse events data and applying the BmaxSPRT method to the simulated data with different choices for the risk and nonrisk intervals. Below I will describe the assumptions of the simulation and then present the results of the analysis.

4.5.1 The studied risk and control intervals

Six types of choices for the risk and nonrisk intervals which differ from the actual risk and non-risk periods were chosen for the simulation, along with a seventh pair of reference intervals choices which matched the actual risk (and therefore nonrisk) period perfectly. The figure (4.3) illustrates the chosen case and control intervals. All the chosen intervals have a common ratio between the risk and control periods, as shown in table 4.5.

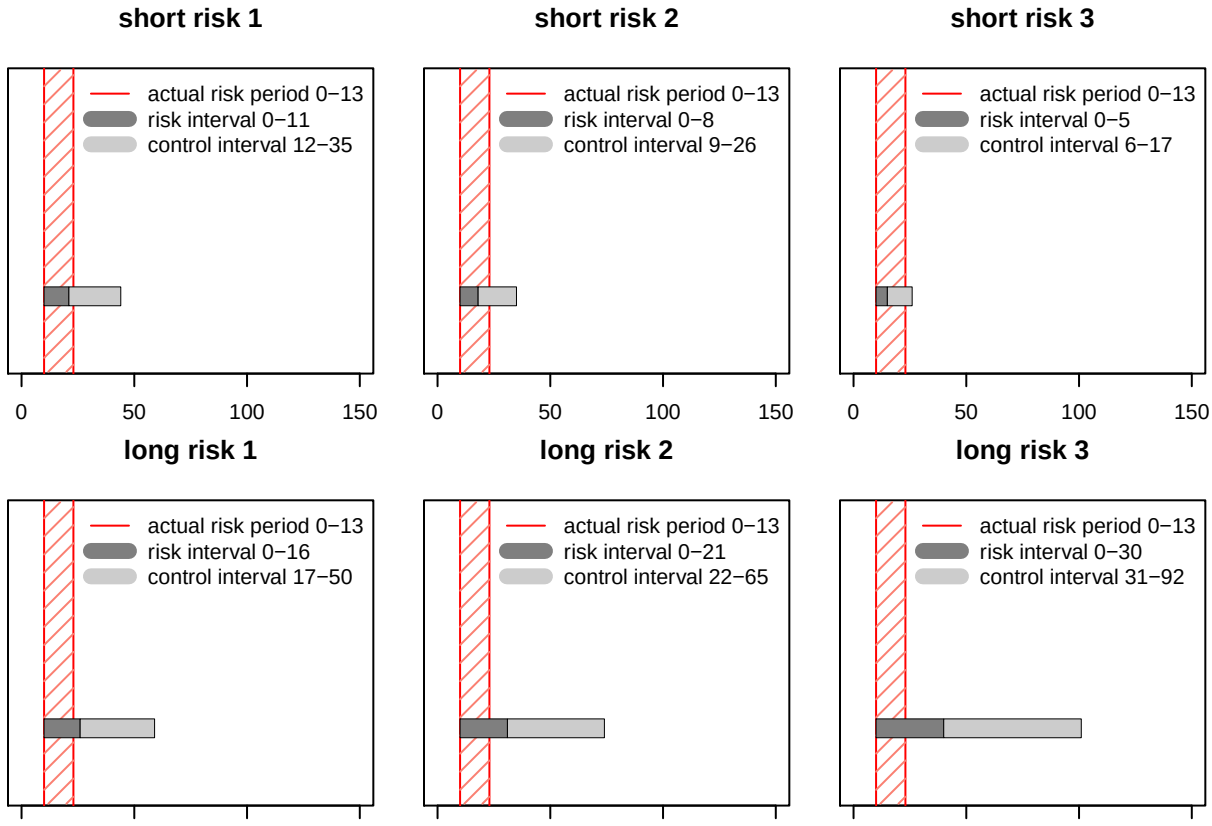


Figure 4.3: The six risk and control interval choices studied in the experiment. In the top row the risk interval is too short and the control interval overlaps with the risk interval. In the bottom row the risk interval is too long and overlaps with the control interval.

Table 4.5: Risk and control interval choices for the simulation experiment. The reference group matches the actual risk interval used in the simulation.

	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

4.5.2 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 interval is always 0-13 days after exposure and the chosen risk and control intervals vary
- For each exposed individual, the rate of events is multiplied by 1.5 during the risk interval ($RR_a = 1.5$)
- 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 intervals are chosen correctly, the power of the 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).

4.5.3 Results of the simulation

Table 4.6 shows the detection rates for each risk and control interval pair in the simulation. Detection rate is the proportion of simulation iterations where the null hypothesis was rejected and it corresponds to the power of the test. If the risk and control intervals are chosen perfectly as was done with the reference group, the expected detection rate in the simulation was 0.9.

Table 4.6: Results of the simulation. The signal rate for the reference group matches the desired power of the experiment. As expected, the signal rates are lower the more the choices for the risk and control intervals differ from the reference.

	risk	control	signal.rate
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 Discussion

UNDER CONSTRUCTION - strengths and weaknesses of self-controlled designs - dealing with confounders (general SCCS) - data-mining methods in signal indentification

References

Childhood vaccines and febrile seizures. 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.

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.

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/.

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.