Implementation of a framework to apply Sequential Analysis for Vaccine Safety Surveillance on SAEFVIC Dataset

Md Samiullah · Jim Buttery · Hazel J. Clothier · Jiying Yin · John Mallard · Jeremiah Munakabayo · Gonzalo Sepulveda Kattan · Gerardo Luis Dimaguila

Abstract Timely detection of rare adverse events following immunisation (AEFI) is critical for vaccine safety. For post-market surveillance, sequential analysis techniques that were initially created for clinical trials (Gordon Lan and DeMets, 1983) have been modified to allow for near real-time signal detection while strictly controlling Type I (Silva, 2018) and Type II errors. The conventional Sequential Probability Ratio Test (SPRT), however fundamental, has limitations due to its dependence on predetermined alternatives and sensitivity to predefined relative risks. These drawbacks are addressed by the Maximised SPRT (MaxSPRT), which supports both Poisson and Binomial models for a variety of surveillance designs and permits composite alternatives (e.g., Relative Risk, $RR > 1$). By conditioning on observed events, the Conditional MaxSPRT (CMaxSPRT) goes one step further and maintains error validity even when there is a lack of historical data. Alpha spending functions (Silva et al, 2020), especially when optimised by linear programming techniques, provide better control over error rates to handle repeated testing, allowing for more effective designs with lower expected time to signal (ETS) and expected sample size (ESS).

Despite these methodological breakthroughs, sequential analysis of spontaneous surveillance systems are still difficult to apply in the real world and lack standardisation. This paper proposes a practical implementation of sequential analysis in real-world settings, encompassing everything from data wrangling to signal detection. By synthesising existing literature, we identify crucial contextual factors that influence design choices, such as data structure, surveillance objectives, and vaccine-AEFI characteristics. With SAEFVIC (a system for Surveillance of Adverse Events Following Vaccination in the Community), Victoria's jurisdictional vaccine safety system, as a case study, we outline implementation steps, talk about common issues, and provide solutions. This work acts as a guide for organisations intending to start from scratch when integrating sequential analysis into vaccine pharmacovigilance operations.

Keywords Sequential Analysis, Continuous and Group Surveillance, Adverse events, Vaccine pharmacovigilance, Sequential Probability Ratio, Signal detection, reporting delay, real-world datasets

M Samiullah · J Buttery · H J Clothier · J Mallard · J Munakabayo · G Kattan · GL Dimaguila
Murdoch Children's Research Institute
Royal Children's Hospital, 50 Flemington Rd, Parkville, VIC - 3052, Australia
E-mail: {md.samiullah, jim.buttery, hazel.clothier, john.mallard, jeremiah.munakabayo, gonzalo.sepulveda, gerardoluis.dimaguil}@mcri.edu.au

J Yin
Murdoch Children's Research Institute, Australia
E-mail: Jiying.Yin@afsa.gov.au

Key Points

Sequential analysis especially MaxSPRT is a powerful method for ongoing vaccine surveillance, offering flexibility to adapt to various situations and data limitations. However, the offered flexibility and variations as well as the lack of step-by-step guidance can lead to challenges in implementation. Our paper simplifies the MaxSPRT method with clear explanations and step-by-step guidance from scratch with illustrative computation, addressing potential issues and proposing solutions to improve its use in monitoring vaccine and drug

safety. Furthermore, we offer an extendable and generic tool to implement such analysis from scratch that includes data wrangling and preprocessing.

1 Introduction

Vaccines keep us safe against harmful diseases by increasing the natural defence of our body and inducing immunity to various infections (Murphy et al, 2023). Every licensed vaccine goes through rigorous testing over numerous trial phases before being approved for use by regulators. Nonetheless, adverse events following immunisation could occur, although they are typically modest and transient, like a painful arm or a low-grade fever. Though relatively rare, more severe side effects may occur. Following approval, vaccine safety surveillance services continue to scan and analyse data from a variety of sources for any signs that a vaccine might have unexpected harmful consequences on health, or an increase in incidence or severity of known reactions, known as post-licensure surveillance.

In order to perform post-licensure vaccine safety signal detection, researchers propose diverse approaches that involve various statistical methods (Lee et al, 2020; Mesfin et al, 2021), techniques, and measures in original, modified, or combined forms. Research and vaccine surveillance groups may also impose necessary modifications to the approaches during implementation, addressing their organisational strengths and limitations, availability of data, and resources. Note, unlike the works performed so far in the literature, we are targeting a unique aspect of the implementation such as strengths, variations and limitations of data due to organisational factors.

Key statistical measures and techniques used in post-licensure vaccine surveillance include Proportional Reporting Ratio (PRR) (Clothier et al, 2019), Cumulative Sum (CUSUM) (Mesfin et al, 2021), Reporting Odds Ratio (ROR) (Yoon et al, 2020), Empirical Bayesian Geometric Mean (EBGM) (Lee et al, 2020), Chi-Square, and MaxSPRT (Kulldorff et al, 2011)-a specialised variation of Sequential Probability Ratio Test (SPRT) (Wald, 1992) proposed by Wald.

Sequential analysis for post-market surveillance evolved to enable early detection of rare adverse events while controlling error rates. An important aspect of vaccine surveillance is the continuous monitoring of adverse events that vaccinees may experience when vaccines are administered, through a continuous sequential analysis that detects a signal when adverse events exceed the probability that they occur by chance (Kulldorff et al, 2011). Sequential Probability Ratio Test (SPRT) (Wald, 1992) was proposed by Wald to detect

adverse events through hypothesis testing. It assumes fixed alternatives and known expected event rates and is appropriate for sequentially independent and identically distributed (iid) data, which implies that the data comes from the same distribution. SPRT is used for surveillance to test if the null hypothesis (no vaccine safety signal or the vaccine is safe), is more likely than the alternative hypothesis (there is a vaccine safety signal). However, SPRT is highly sensitive to the relative risk chosen in the alternative hypothesis construction, which makes it difficult to use for ongoing, routine surveillance.

To overcome SPRT's rigidity and sensitivity, the Maximized SPRT (MaxSPRT) (Kulldorff et al, 2011) was introduced, allowing for a composite alternative (e.g., $RR > 1$) and implemented for both Poisson and Binomial models. MaxSPRT facilitated real-time safety monitoring by enabling signals to occur even after a single event, depending on its timing. Poisson MaxSPRT is best for rate-based monitoring having time-to-event data or counts over exposure, while Binomial MaxSPRT fits self-controlled or matched cohort designs where outcomes are binary (event or no event).

However, when expected event rates are estimated from limited historical data, MaxSPRT can inflate Type I error. To adjust for such issues, Li and Kulldorff proposed the Conditional MaxSPRT (CMaxSPRT), which incorporates variability in both historical and surveillance data by conditioning on observed events, thus preserving Type I error validity under small-sample conditions or when historical data is sparse.

To further improve control over false positives across multiple looks at data, alpha spending functions were introduced. These functions (e.g., power-type, Pocock (Pocock, 1977), O'Brien-Fleming (O'Brien and Fleming, 1979)) spread the allowable Type I error across time, with concave forms ($\rho < 1$) minimizing expected time to signal (ETS) and convex ones ($\rho > 1$) minimizing expected sample size (ESS). Exact optimization of alpha spending using linear programming (Silva et al., 2020) enables designs that minimize ETS while controlling Type I and Type II errors and sample sizes, and maximizing statistical power, outperforming classical methods.

Note that it is evident from (Silva et al, 2021) that alpha spending is implicitly defined by sequential test plans, regardless of the test statistic scale that is selected. This can be readily illustrated with the R Sequential (Silva and Kulldorff, 2024) package. However, since identical alpha spending produces equivalent results across all scales, the practical lesson is not the test statistic selected but rather the careful selection of a near-optimal alpha spending plan. The package inte-

grates Silva and Kulldorff's (Silva and Kulldorff, 2015) methodological advancements.

A significant improvement is the replacement of the linear interpolation method suggested by Kulldorff et al. (Kulldorff et al, 2011) with a bisection procedure to precisely compute exact critical values (CVs). This method may result in infinite loops depending on the fixed expected number of events. Moreover, Silva and Kulldorff (Silva and Kulldorff, 2015) showed that, in contrast to depending on the continuous critical values that (Kulldorff et al, 2011) had previously tabulated, employing group sequential tests with unpredictable sample sizes through alpha spending functions greatly improves statistical power.

Silva et al.'s extensive guide (Silva et al, 2021) on planning sample sizes and the alpha spending approach in sequential analysis plays a crucial role in maintaining the desired statistical power. They present detailed methodological insights and introduce optimized alpha spending plans to minimize the expected time to signal (ETS) and expected sample size (ESS). However, while Silva et al.'s work expertly addresses statistical optimizations, building on it, our paper takes a more practical, step-by-step approach tailored for real-world implementation, particularly highlighting common challenges such as data wrangling and preprocessing that arise in operational vaccine safety surveillance systems like SAEFVIC.

Moreover, Leite et al. (Leite et al, 2016) present a comprehensive systematic review of real-world applications of statistical methods in vaccine safety surveillance, presenting a variety of statistical tools and approaches. Their work outlines practical surveillance methods using electronic health records and provides foundational insights into the surveillance landscape. In contrast, our manuscript uniquely integrates these surveillance methods into a hands-on, comprehensive framework, clearly detailing implementation steps, parameter selection, and strategies for addressing common real-world issues like reporting delays and missing data.

Our paper addresses important practical gaps not fully explored in the literature, including: − Explicit procedures for data preprocessing and handling of missing values and noisy data. − Detailed computation and interpretation of real-time measures such as the observed exposed person-time (L_{stiwa}) and the proportion of data completeness ($p(s,t)$), essential for accurate real-time signal detection. − Flexible options for adapting to various surveillance scenarios, such as age stratification and dynamic binning, not extensively detailed in existing literature. − Clear illustration of calculating the expected number of events, addressing and calculating

reporting delays and proportion of data complete, leveraging practical heuristics and real-world constraints, thereby providing explicit operational guidance beyond theoretical formulations.

By explicitly addressing these pragmatic considerations and providing a comprehensive, illustrative approach to sequential analysis for vaccine safety, our manuscript offers complementary value to existing tutorials and reviews. Thus, our paper serves not as a replacement but rather as an essential operational extension to the seminal methodological contributions of Silva et al. (Silva et al, 2021) and the foundational review by Leite et al. Leite et al (2016). It is worth noting that the R Sequential package (Silva and Kulldorff, 2024) is central to the sequential surveillance system developed to enhance analysis of SAEFVIC data; all data analyses, experimental results, visualizations, and related research presented in this paper were conducted using this package.

2 Methods

In addition to the introductory developmental trajectory provided in Section 1, this section describes the evolution of sequential analysis focused on likelihood ratio tests such as SPRT and MaxSPRT as reported in the literature. It forms the foundation for the framework proposed in this paper for the implementation of the variations of the sequential surveillance technique for vaccine safety. We look at how variations in data types, distribution features, parameter availability, and organisational constraints—such as concerns about data timeliness and completeness—all affect the form and application of sequential analysis methods, leading to variations in implementation strategies and surveillance results.

We start with the difference in the calculation of the parameters for SPRT and MaxSPRT. In both Wald's SPRT (Wald, 1992), and Kulldorff's MaxSPRT (Kulldorff et al, 2011), the number of adverse events is considered to be random, while the cumulative person-time or the cumulative number of vaccinations, if the follow-up time per vaccination is uniform, is considered to be fixed. The predicted number of adverse events under the null hypothesis is then considered to be a known function of the total person-time and certain possible confounding variables like age, sex, and location. Figure 1 denotes this variation by the first branching that occurs for sequential analysis between SPRT and MaxSPRT. Note, another factor that signifies the difference is, SPRT is highly sensitive to the relative risk chosen in the alternative hypothesis and cannot deal with composite alternatives.
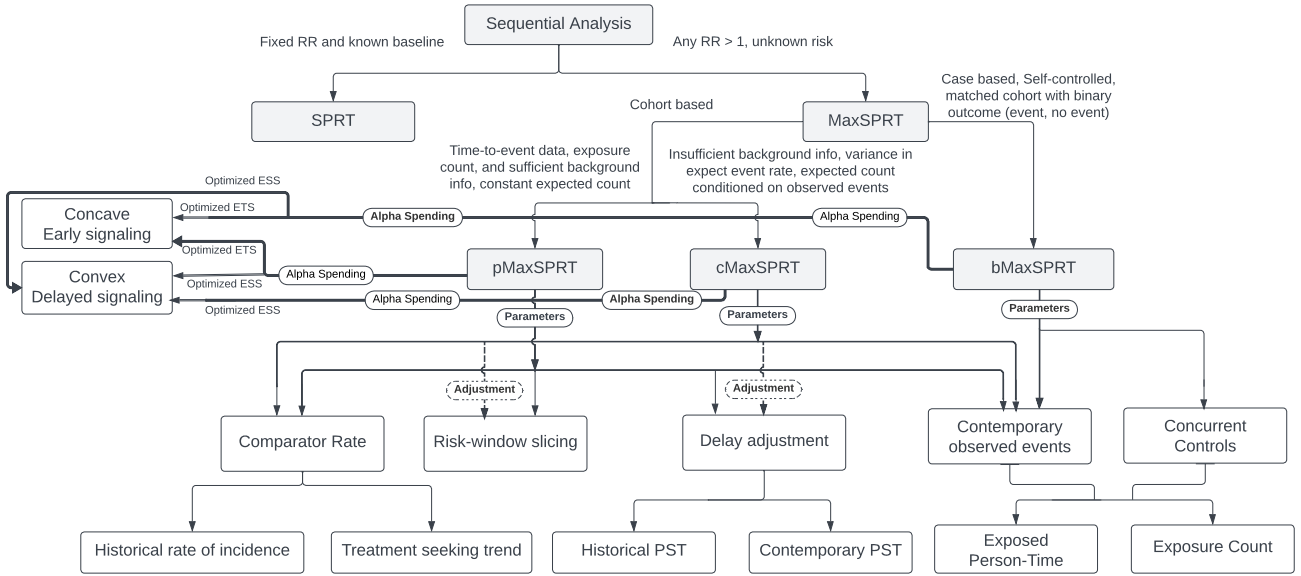
Fig. 1 Sequential Analysis: Comparative View Across Parameters and Conditions.

Based on the mode of analysis—whether cohort-based or case-based—MaxSPRT can be applied in two primary forms: Poisson MaxSPRT for cohort-based analyses and Binomial MaxSPRT for case-based or self-controlled designs. The Binomial MaxSPRT (bMaxSPRT) is suitable for self-controlled approaches in which the comparator is calculated from contemporary rather than historical data and the analysis compares the presence or absence of an event (binary outcome).

Furthermore, the dependence and sufficiency of the comparator data lead to another branching: (i) Conditional MaxSPRT (cMaxSPRT) and (ii) unconditional/-conventional MaxSPRT (we refer to it as pMaxSPRT for the sake of simplicity). For the scenario when the background information is insufficient and conditional on an event, Li et al. developed conditional MaxSPRT (cMaxSPRT) (Li and Kulldorff, 2010), where it factors in the number of adverse events in the historical data and the surveillance population and computes the cumulative person-time it took to observe so many events as the random variable.

In addition to these methodological variations, Silva et al.'s introduction of optimised alpha spending plans adds a further dimension to MaxSPRT design. By distributing the Type I error across interim analyses using tailored spending functions, this approach enhances control over both Type I and Type II errors, increases statistical power and significance, and effectively reduces the expected time to signal (ETS) by allowing early stopping, while also optimising the expected sample size (ESS) to increase statistical power. This optimisation, often achieved through linear programming, enables more efficient and robust sequential surveillance compared to traditional fixed-threshold designs. Here, ETS, ESS and CV are calculated dynamically based on the choice of alpha spending functions. There are generally three categories of alpha expenditure functions: (i) Linear, (ii) Convex, and (iii) Concave. Silva et al. further demonstrated that the analytical objective should guide the choice of alpha spending function among the three types. Concave functions, which use a larger portion of the Type I error budget at the start, are preferable for early signal detection. Although this lowers ETS, it comes at the cost of reduced statistical power. Convex expenditure functions, on the other hand, preserve statistical power in longer surveillance windows by conserving alpha for subsequent analyses, providing improved control of ESS and more stable estimation when previous data is ambiguous.

The mode, outcome and processing steps of surveillance vary for some factors and parameters that need to be considered while performing the analysis where some of them are also important to track which variation of MaxSPRT is to be used. The parameters, hyperparameters, and conditions are as follows: (i) the grouping of the reports based on age; (ii) the decision to use any comparator rate, sources of the comparator rate, the process of calculating the comparator rate; (iii) the length of the risk window and the decision to slice or not slice risk windows; (iv) using self-controls or other controls; (v) choosing the unvaccinated population in the concurrent data as comparator; (vi) the decision to use the length of exposure and the method to compute

it, e.g., using l_{stiwa}, as a proxy for the calculation of exposed person time, and the variations in calculating the exposed person-time; (vii) the choice of using a test margin (Kulldorff et al, 2011) in the calculation; and (viii) the decision to use a delay adjustment, as well as the ways to calculate the delay adjustment.

Each combination of the aforementioned factors makes a different variation of sequential analysis, therefore, the factors and hyperparameters are introduced briefly as follows and described in greater detail with relevant technical information later.

- **Age grouping:** Sometimes, age-based stratification of the whole MaxSPRT analysis is required. In such cases, all the calculations are performed at a more granular level, and all the hyperparameters, such as the historical comparator rate, are taken based on age grouping. For example, age strata can be 5-year or 10-year age groups, or bespoke grouping according to vaccine schedule. More small-ranged grouping can also be performed to address the risk of different AESIs for different age ranges. Moreover, unlike the static binning detailed above, a dynamic binning approach of risk and AEFI-based age grouping is also possible, where all groups may not have the same interval. This could provide an option for sequential analysis. This kind of analysis has not been tried yet in the literature, to the best of our knowledge.

- **Comparator group:** One of the hyperparameters is the historical comparator group. This is required as a reference value in calculating the likelihood of an event. This value sometimes needs scaling down based on the length of the observed period. There are several ways of calculating this rate, and the information available at the time of analysis also plays an important role. In the case where more than one option is available, the most relevant and reasonable one needs to be used. Two such potential options, as per (Kulldorff et al, 2011), are:
  - *Retrospective rate of incidence:* This approach calculates the rate of incidence of the same AEFI in previous years. This rate can be calculated from the historical data, reflecting the ratio of reports in the previous years.
  - *Treatment-seeking trend:* In this approach, the rate is calculated from General Practitioners' (GP) data, which is a term used in Australia that refers to primary care in general, clinics', and hospitals' data and health records, by finding the trend of treatment-seeking behaviour for a particular AEFI.

- **Risk-window slicing:** In MaxSPRT, while calculating various metrics, we can take the elapsed time of an observation period into consideration. This

helps mitigate the gap of overestimating the risk or control period of a vaccine. It also allows slicing the risk window into smaller periods so that an observation period that is shorter than the risk window can scale down the required metrics as per the length of the observation period. In the sliced version, we further stratify the data, especially reports for different intervals of the risk window. This provides a more granular level of calculation and tuning. In (Greene et al, 2011), authors show the efficacy of using a sliced risk window.

- **Exposure length (l_{stiwa}) calculation:** the measure that counts the length of exposure by a particular population under observation and consideration. It is required to estimate the expected risk ratio if the dose distribution information is unknown or insufficient.
  - *Exposed person-time:* The Centers for Disease Control and Prevention, USA (CDC) in (Center for Biologics Evaluation and Research, 2021) proposed a technique and a formula to calculate the length of exposure in a person-time unit by capturing the trend of overlapping exposed time for the vaccinees within the risk window.
  - *Other alternative estimations:* If the CDC's approach can be replaced by some more reasonable approach that can capture the length of exposure more realistically and effectively.

- **Delay adjustment:** AESIs are not reported in real-time for various reasons. This factor must be addressed in any version of sequential analysis to adjust the observed number of events within the observation period. This adjustment to the observed number of events is required to help sequential analysis find appropriate values for CVs to decide on the likelihood of signals in the study.

There are several ways of calculating and adjusting the delay, especially estimating the delay from retrospective analysis. Based on the availability of the data or reports in previous years for the same AEFI and for the same vaccines, we have divided the ways into two high-level classes.
  - *Historical $p(s,t)$:* If the analyser has access into the reports for the AEFI and the vaccine under consideration, then CDC's approach (Center for Biologics Evaluation and Research, 2021) as explained in later sections (Section 3.2.3) can be followed to calculate $p(s,t)$ for dealing with delay adjustment.
  - *Concurrent $p(s,t)$:* Where the analyser does not have historical information on the trend of delays in reporting, such as in the case of a new AEFI or a new vaccine, the delay adjustment must be

Table 1 Notations

Terms	Notations
Study week	s
Any week within the study period	t
Exposure weeks	w
Time to onset	TTO
Slices in Risk window	δ
Expected number of events	μ_s
Observed number of events	Y_s
Historical comparator rate	θ
Per day comparator incidence rate of age-group a	θ_a
Proportion of data complete	$p(s,t)$
Observed exposed person-time	L_{stiwa}
Log-likelihood Ratio	LLR

done by some predictive or approximation approach using the so far obtained concurrent data. We also explain this and propose a method to deal with this situation in Section 3.2.3.

3 Results

After analysing the relevant literature, we developed a taxonomy of variations of sequential analysis as shown in Figure 1.

The structure and organisation of the taxonomy depict the factors that establish the implementation variations in a hierarchical manner. In Section 2, we explained how the variation starts with the availability of the base-line data and the nature of Risk ratio.

As explained in Section 2, SPRT differs from MaxSPRT due to the nature of relative risk; Poisson MaxSPRT and Binomial MaxSPRT (bMaxSPRT) are the results of choosing between cohort-based and case-based analysis, respectively. The usage of either conditional (cMaxSPRT) or unconditional (pMaxSPRT) models is further determined by the availability and dependability of background data. Another layer of diversity is added by the choice to use alpha spending and the choice of suitable spending functions depending on the goal, such as optimising anticipated time to signal (ETS) or expected sample size (ESS). All of these factors are included in the taxonomy.

To aid readability and implementation, we list the terminologies used in the literature to construct the narrative of the paper clearly. Table 1 enlists the terms, their notations, and symbols that are commonly used throughout the paper. We refer to Appendix A for the definitions, examples, and explanations of the terms.

3.1 Properties of the Dataset

SAEFVIC (a Surveillance of Adverse Events Following Vaccination in the Community) data (Clothier et al, 2017) is a set of timely passive surveillance data consisting of AEFI reports submitted by vaccinees, their relatives, GPs, nurses, or other healthcare professionals. It aligns to the minimum data for AEFI surveillance as per WHO (World Health Organization and others, 2014). Each row of the data contains a report converted to a transactional database format, where each report is identified by a unique identifier named $Vaccinee\_ID$. Corresponding to each unique identifier, there are the age and sex information of the vaccinee, the time stamps of the dose administration, reporting time (when was it posted), and report submission. It also contains the time to onset expressed in the number of days for the adverse event, followed by the description of the event, reaction sign/symptom or clinical term of the event name, and vaccine brand, as well as the dose received. There are other data that are neither relevant nor required for MaxSPRT analysis, and we filtered them in the preprocessing step.

Like other real-world datasets, our data includes missing values, too. This needs special attention before the data can be used for analysis. Data are collected through an online interactive form that is used by end-users (reporters). Hence, it is prone to errors like human errors that include misspellings and typographical errors. This implies that we need to use some mechanism to deal with unclean data.

Table 2 shows the headers of a sample, a synthetic dataset developed from real SAEFVIC data after altering it so that MaxSPRT analysis can detect a signal for demonstration purposes. We used this sample data for the analysis with selected columns, the corresponding data types, and the domains of the columns.

3.2 Steps and processes of implementation

Here, we explain the step-by-step calculation of each measure, factor, and parameter required to calculate the LLR (Log-likelihood Ratio), which is used to detect a signal in Sequential analysis and MaxSPRT. Throughout the paper, we use a sample SAEFVIC dataset for a particular vaccine and "Abdominal pain" as an AEFI as a running example to demonstrate the calculations.

Refer to Appendix B (Implementation) for full details of each step, particularly during a MaxSPRT implementation in the SAEFVIC setting.

Table 2 SAEFVIC data set used in this paper with their attributes, types, and domains of the attributes, with pseudo sample data.

Attributes	VacID	Vaccine Name	Age	Sex	Time of Vaccination	TTO	Reporting Time	Submission Time	Reactions
Types	String	String	Number	Binary	Time stamp	Number	Time Stamp	Time Stamp	String
Domains	Report ID	Name	Years	M/F	Time Format: dd/mm/yyyy HH:MM:SS	Days	Time Format: dd/mm/yyyy HH:MM:SS	Time Format: dd/mm/yyyy HH:MM:SS	Set of names Comma Separated
Semantics	Unique Identifier	Name of the vaccine brand	Age of the Vaccinee at vaccination	Gender of the Vaccinee	Time of dose administration	Time to On setting the AEFI	When the AEFI was reported	When the report was recorded in SAEFVIC	The Adverse Event

3.2.1 Computing Cumulative Observed Number of events

We have taken all the reports into account after pre-processing (refer to Appendix B.1, Preprocessing) the data. For each week $1 \leq t \leq s$, and each age group a, we counted the number of reports, say $Y_{(t,a)}$. Summing all the counts for all age groups for a particular week, t, gives the number of observed events for the week. We added a column Y_a for each age group a's report count and two more columns to the data, one for keeping track of the weekly observed report count Y_t and the other one for keeping the cumulative counts of observed reports from week 1 up to a corresponding week t, $Y_{1,t}$.

As an illustration, Table 3 shows the weekly count of observed events with a breakdown based on age strata. The last column shows the cumulative count of events.

3.2.2 Length of exposure calculation

The length of exposure is roughly the cumulative count of exposed time per vaccinee after a dose is administered and until the reaction is set on them in the person-time unit. Definition 7 (Appendix A, Terminology) expresses a way of capturing the overall length of exposure for all vaccinee under consideration for a MaxSPRT analysis when dose administration information is not available at all, or at least not during the analysis in real-time.

3.2.3 Proportion of data complete calculation

AEFIs are not reported in real-time for various reasons. Even after being reported, AEFI recording may also be delayed. This phenomenon is known as observation delay. Due to such delays, at the time of analysis, the amount of the observed data for a particular study week is likely to be less than the actual amount. This leads to an overestimation of the expected number of events, which imposes a potential risk of missing a signal. To address the issue, we can adopt an approximation technique, and this is known as calculating the proportion of data complete (the ratio of the observed

and expected amount of data). We defined the term in Definition 14.

As we discussed in Appendix A, CDC defined $p(s,t)$ in (Center for Biologics Evaluation and Research, 2021) to capture the proportion of data complete. However, the process of calculation or estimation of the term $p(i)$ was not clearly stated, and neither was it illustrated clearly in the paper. Although it looks quite straightforward to calculate this value from the ratios of reports obtained and expected, when we tried to apply the method we ended up with several possible ways of calculating the value. This adds more factors and hence variations of MaxSPRT calculation.

We also propose a new but similar way of calculating the proportion of data complete. We describe an equation to estimate the proportion of data complete that can be used to adjust the factor of delays in reporting incidents in Section 3.2.6 using concurrent data.

3.2.4 Computing comparator rate, θ

There are several ways of calculating the comparator rate, and two of the plausible ways are described in Appendix A. In our current implementation, we calculated this value retrospectively from the previous years' data (Phillips et al, 2023). We have utilised our available data from previous years of the same reaction for calculating the background rates and used the formula mentioned in Definition 11 and Definition 12. These background rates were extracted and analysed by our epidemiology team.

3.2.5 Estimating the Expected number of events

The expected number of events is the number of events expected to be observed without rejecting the null hypothesis. Calculating expected number of events is crucial and can be achieved in various ways. It depends on the available information, objective and type of the sequential analysis being performed. There is a significant difference in calculating or estimating this value based on the available information. An approach to estimate this measure is proposed in (Center for Biologics Evaluation and Research, 2021) and defined in Defini-

Table 3 Demonstration of computing observed report count for MaxSPRT analysis for Abdominal Pain

						Computing Observed Report count for MaxSPRT analysis for Abdominal Pain					
#Week	Report count of age 0-9	Report count of age 10-19	Report count of age 20-29	Report count of age 30-39	Report count of age 40-49	Report count of age 50-59	Report count of age 60-69	Report count of age 70-79	Report count of age 80+	Weekly Total Report	Cumulative Report Count, Y_t
1	0	0	2	2	3	2	0	0	0	9	9
2	0	0	4	2	3	3	2	1	0	15	24
3	0	0	7	12	7	3	2	1	1	33	57
4	0	0	8	13	11	8	5	5	1	51	168
5	0	0	17	16	18	14	7	5	1	78	246
6	0	0	0	1	0	4	6	6	2	19	265
7	0	0	0	2	1	7	6	7	1	24	289
8	0	0	1	1	1	10	14	3	1	31	320
9	0	0	0	1	0	20	18	2	0	41	361
10	0	0	0	0	1	4	1	2	0	8	369

Table 4 Demonstration of computing l_{stiwa} for MaxSPRT analysis for Abdominal Pain

					Computing l_{stiwa} for MaxSPRT analysis for Abdominal Pain					
#Week	l_{stiwa} (0-9)	l_{stiwa} (10-19)	l_{stiwa} (20-29)	l_{stiwa} (30-39)	l_{stiwa} (40-49)	l_{stiwa} (50-59)	l_{stiwa} (60-69)	l_{stiwa} (70-79)	l_{stiwa} (80+)	**Weekly** l_{stiwa}
1	0	0	2	2	5	1	0	0	0	10
2	0	0	18	5	13	9	2	0	0	47
3	0	0	24	18	37	10	8	4	3	104
4	0	0	51	57	51	38	14	28	15	254
5	0	0	76	79	84	52	23	25	7	346
6	0	0	10	12	22	7	17	26	2	96
7	0	0	0	3	6	11	30	14	1	65
8	0	0	6	7	0	33	24	16	1	87
9	0	0	7	8	1	91	53	12	0	172
10	0	0	7	7	4	41	13	10	0	82

Table 5 Demonstration of Reporting delay distribution & computing $p(s,t)$ for Abdominal Pain

					Reporting delay distribution & computing $p(s,t)$ for Abdominal Pain					
AEFI Occuring Week	Reported on Week 3	Reported on Week 4	Reported on Week 5	Reported on Week 6	Reported on Week 7	Reported on Week 8	Reported on Week 9	Reported on Week 10	Reported after Week 10	Total Reports	Total Reports up to week s [s=10]
3	1	0	2	2	1	0	0	0	0	6	6
4		5	10	5	2	0	1	1	1	25	24
5			11	33	10	2	2	0	10	68	58
6				23	27	9	1	2	10	72	62
7					31	38	9	0	10	88	78
8						7	10	2	2	21	19
9							3	6	16	25	9
10								9	18	27	9

tion 8 as "Approximate Expected number of events" in Appendix A.

In passive surveillance or reporting systems, such as SAEFVIC, we do not have dose administration information in real time. In such systems, only the numerator that is the report count are available. Therefore, in this paper and in our current implementation, we focused on l_{stiwa} (observed exposed person-time, see Appendix A, Terminology) based estimation of the denominator. Once all the required metrics for calculating the expected number of events (μ_s) are found, we compute the expected count using Definition 8 (Appendix A, Terminology). While processing the data, another column is added to it to keep the weekly cumulative expected number of event counts that is going to be utilised later to compute *LLR*.

3.2.6 Calculation of Data Complete

We propose a new but similar way of calculating the proportion of data complete. We describe an equation to estimate the proportion of data complete that can be used to adjust the factor of delays in reporting incidents. We specified the way of calculating the ratio. The denominator of the ratio calculation is subject to the choice of dataset and timeline we choose.

$$p(s,t) = \sum_{i=t}^{s} p(i)$$

where, $p(i) = \frac{c_t(i)}{c_t}$, and c_t is the total number of reports for week t, $c_t(i)$ is the total number of reports for week t at week i.

Using this factor is optional for sequential analysis and can be ignored without the risk of impacting the outcome significantly. However, using this parameter

ensures better approximation of the likelihood of a signal. In case we decide to use this factor, the calculation of the proportion of observed data can be performed in two ways. Firstly, similar to the existing works (Center for Biologics Evaluation and Research, 2021; Greene et al, 2011), we can use historical data for the same AEFI and the same vaccine or a similar vaccine. Secondly, where no historical data is available, a special approach should be adopted to calculate the $p(i)$ using the concurrent data by estimating the proportion of later weeks of exposure from the earlier weeks of exposure.

In order to estimate the proportion of data complete from concurrent data, we can use the trend of delays in reporting for the weeks up to the study week. As an illustration of the reporting delay distribution, see Table 5. This distribution is for a MaxSPRT analysis on week 10 observed in an analysis performed to check for a safety signal for Abdominal pain. Note that the count of reports for each week with corresponding report submission weeks. The table contains the distribution of AEFI reporting with certain delays. As an example, it is expected that there will be 6 reports for week 3 and those were reported in weeks 3, 5, 6, and 7. Similarly, 25 reports of weeks 4, and 5 were reported in week 4, 10 on week 5, 5 on week 6, 2 on week 7, and 1 on each of week 9, 10, and after 10 weeks.

Refer to Appendix 3.2.8 to check the detailed calculation of the ratio to get $p(s,t)$. This metric is used in the MaxSPRT analysis and demo calculations.

3.2.7 Demonstration of l_{stiwa} calculation

In Table 9, we provide a visual demonstration of the calculation of l_{stiwa} and how it is calculated for each person. As an example, vaccinee V_1 got a dose on the 1st day of week-1. Their TTO was 5 days. Hence, the person was exposed from day 2 to day 6 of week-1. Hence, l_{stiwa} of the person is denoted as $l_{sti(w=0)a}$ and calculated as 5. Also, vaccinee V_5 got a dose on the 5th day of week-1. Their TTO was 11 days. Hence, the person was exposed from day 6 of week-1 to day 2 of week-3. Hence, the l_{stiwa} of the person for weeks 1, 2, and 3 are denoted and calculated as $l_{sti(w=0)a} = 2$, $l_{sti(w=1)a} = 7$ and $l_{sti(w=2)a} = 2$, respectively. At the end, the last row of the table shows the weekly l_{stiwa}.

Similar to the observed event count, we have added additional columns for l_{stiwa} for each age group a, and for weekly l_{stiwa} by adding all age-group based l_{stiwa} for a particular week t. As an illustration, Table 4 shows the computed weekly l_{stiwa} and age-strata based administration of l_{stiwa}.

3.2.8 Variations in $p(s,t)$ calculation

To calculate the ratio for calculating $p(s,t)$, we need to know the total number of reports for the week. This ratio can be estimated from historical data. That means, it is quite challenging to get the ratio in case the historical information is not available or insufficient. Estimating the ratio without a proper or exact denominator of the ratio from past data needs an approximation technique.

One of the approximation techniques would be to take the total reports found so far as a denominator to compute the ratio. However, this technique would not be sufficient to approximate the ratio properly. Because, as shown in Table 5, it is quite possible that we have more information about earlier weeks and less information regarding later weeks. However, the ratio indicates a wrong proportion of data complete, that is, more for later weeks and less for earlier weeks, unlike a real scenario. In such cases, earlier week ratios will be more accurate than the later weeks. We can estimate the ratios of later weeks from the earlier weeks to deal with this challenge.

Our idea is to estimate a certain number of later weeks from the remaining weeks' $p(s,t)$. As an example, we can compute the last 50% of weeks' $p(s,t)$ from all previous weeks' $p(s,t)$. There is scope to take lesser or greater than 50%. We propose the following equation to calculate the proportion with a certain number of weeks' $p(s,t)$ to be estimated.

$$p(s,t) = \frac{1}{\lceil \frac{s}{d} \rceil} \sum_{w=1}^{\lceil \frac{s}{d} \rceil} p(s, t-w)$$

Here we estimate the $p(s,t)$ for week t by taking the average of the $p(s,t')$ where t' is the first half of the study weeks taking the division or partition number, $d = 2$. We can substitute this equation with some other more appropriate estimator, as well as we can tune the equation by changing the values of d. As an example, if $d = 3$, the equation will take the average of the 1st one-third weeks' $p(s,t)$.

Table 11 shows the possible ways of calculating $p(s,t)$ for the delay distribution of the reports shown in Table 5. All five different approaches shown in the table for computing $p(s,t)$ do not comply with CDC's conclusions that the earlier weeks will have more data complete than the later weeks, but they also have different logic and merits. To choose the best and most effective one, further empirical assessment is required.

Table 6 Computing required parameter values for MaxSPRT analysis for Abdominal Pain using the formula in Definition 8

#Week	l_{stiwa}	Θ	$l_{stiwa} \times \Theta$	$p(s,t)$	μ_t	Cumu. μ_t
1	10		2.35	0.9	2.12	2.12
2	47		11.05	1	11.05	13.17
3	104		24.44	0.95	23.22	36.39
4	254		59.69	0.93	55.51	91.9
5	346	0.235	81.31	0.93	75.62	167.52
6	96		22.56	0.9	20.30	187.82
7	65		15.28	0.75	11.46	199.28
8	87		20.45	0.89	18.20	217.48
9	172		40.42	0.75	30.32	247.8
10	82		19.27	0.25	4.82	252.6

3.2.9 Calculating LLR and detecting signals

We used cumulative expected number of events, μ_s (a snapshot of calculation of all components/measures of μ is shown in Table 6) and Y_s up to week s, to calculate the LLR using the formula in Definition 17. Either, the LLR or Y_s is used to detect signals by comparing against the critical value, CV. Note, there are two types of critical values: (i) based on LLR of the corresponding week's cumulative expected number of events, we denote it as $CV_{LLR_{\mu_s}}$ and (ii) based on Y (Silva and Kulldorff, 2015), the corresponding week's observed event count, we denote it as CV_{Y_s}. In other words, say in week t, if $LLR_t \geq CV_{LLR_{\mu_t}}$ or $Y_t \geq CV_{Y_t}$, then H_0 can be rejected and a signal can be announced.

Table 7 Demonstration of sequential signal detection using an alpha spending approach for Abdominal Pain (null hypothesis, H_0: no signal; a signal is indicated when H_0 is rejected based on time-varying critical values)

Week#, t	Test#	μ_t	Y_t	Cumu. μ_t	Cumu. Y_t	RR_t	LLR_t	CV	Reject H_0?
1	1	2.12	9	2.12	9	4.25	6.13	10	No
2	2	11.05	15	13.17	24	1.82	3.57	27	No
3	3	23.22	33	36.39	57	1.57	4.97	58	No
4	4	55.51	51	91.9	108	1.18	1.33	124	No
5	5	75.62	78	167.52	186	1.11	0.98	209	No
6	6	20.3	19	187.82	205	1.09	0.76	233	No
7	7	11.46	24	199.28	229	1.15	2.11	247	No
8	8	18.2	31	217.48	260	1.2	3.91	267	No
9	9	30.32	41	247.8	301	1.21	5.34	301	Yes
10	10	4.82	8	252.6	309	1.22	NA	NA	NA

In this study, we choose the alpha spending method to demonstrate the sequential analysis, as this is the most vibrant method, providing more controls to analysts and, through additional challenges at the same time, including decisions to choose the right plan and understanding their consequences. It has a significant difference from plain MaxSPRT without any alpha spending plan. Although both approaches are used for sequential signal detection, they differ fundamentally in how statistical significance is evaluated over time. MaxSPRT employs a fixed critical value (CV) derived
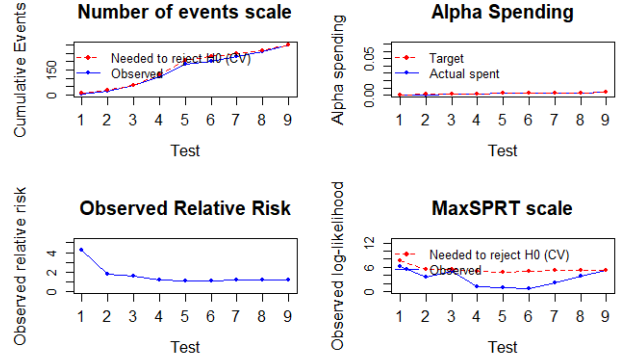
Fig. 2 Visualising signal detection using a sequential testing approach with alpha spending

from a likelihood ratio test to control the overall Type I error rate, resulting in high efficiency but a rigid implementation. In contrast, the alpha spending approach distributes the overall alpha (Type I error budget) across interim analyses, using a time-varying critical value determined by a pre-specified spending function.

Figure 2 illustrates an alpha spending sequential analysis. The top-left panel ("Number of events scale") compares the observed cumulative events with the threshold needed to reject H_0 at each test; the top-right panel tracks the target versus actual alpha spending over time. The bottom-left panel shows the observed relative risk decreasing and stabilizing, while the bottom-right panel ("MaxSPRT scale") presents the log-likelihood ratio (LLR) compared to the time-varying critical value—again, a feature of alpha spending, not MaxSPRT. The accompanying table reinforces this interpretation by showing that the CV changes across test periods, which would not occur in a true MaxSPRT design. Specifically, the null hypothesis was rejected only at Test 9, when the observed LLR (5.34) exceeded the then-current CV (3.01), confirming a statistically significant signal. Thus, both the figure and the table represent an application of alpha spending methodology. The boundaries and critical values are obtained from Poisson MaxSPRT analysis using R Sequential package (Silva and Kulldorff, 2024) with a sample size = 500, total α = 0.05 with 10 installments for 10 tests, each test having α = 0.005, ρ = 1, the minimum number of events = 1. We explained about R Sequential package briefly later in the paper.

Note in the above example, to perform MaxSPRT, especially using the Sequential package (Silva and Kulldorff, 2024) of R language, we need to provide some values as parameters: Sample size, Statistical significance (α) and alpha spending plan, alpha spending function (ρ), statistical power, the minimum number of events,

the observed number of events, and the expected number of events. Calculation of the expected number of events is highlighted in the paper. To aid the readers, we briefly introduced those terms in Appendix 3.3.

3.3 Parameters for Sequential Analysis

This section introduces some parameters required to perform MaxSPRT analysis for continuous surveillance using the Sequential package (Silva and Kulldorff, 2024). Most of these parameters are explained in the manual (Silva and Kulldorff, 2024) with example scenarios and values. These parameters play a very important role in the analysis.

Expected Sample size (ESS): A non-zero upper limit of the expected number of events or cumulative information (e.g., person-time or sample size) required before a decision is made in a sequential analysis under the null hypothesis under the null hypothesis is referred to as Expected Sample Size (ESS), assuming no true signal exists.

It shows how much information would, on average, be gathered before the surveillance was stopped, either because of a signal or because it was completed, if the null hypothesis were correct. Numerous techniques are available in the Sequential R package to optimise ESS by choosing suitable Type I error rates, alpha spending functions, and relative risk values while weighing the trade-off between statistical power and surveillance length.

Expected Time to Signal (ETS): The expected time (or information accrual point) expressed in terms of the number of events at which a safety signal is detected, assuming the alternative hypothesis is true (i.e., there is a real increase in risk), is referred to as ETS.

It shows how quickly a surveillance system can detect a true safety signal. The factors that affect ETS are: the frequency of monitoring, the alpha spending function selected, and variables such as relative risk and incident. When focusing on early detection, especially in vaccine safety monitoring, minimising ETS is frequently a design objective.

Statistical significance assesses whether the observed risk is not occurring due to random variation. The significance level α focuses on controlling Type-I errors and influences statistical significance in the analysis. MaxSPRT allows the defining or controlling of the alpha spending plan by the analyser. There are parameterisable methods to conduct an analysis using MaxSPRT where a specific plan to spend α, such as, uniform in every analysis period, or conservative or generous at the beginning and changing exponentially or

logarithmically in spending α) can be executed. As per the plan, a certain amount of alpha will be spent in each analysis period (where period means weeks if the data is stratified into weeks and analysed weekly). ρ is a positive number used for the power-type alpha spending function.

Statistical power measures the probability of identifying the risk if it is truly present. It focuses on controlling Type-II errors and is influenced by sample size and significance level.

Relative risk: The relative risk RR is the ratio of incidence rate in exposed and unexposed populations. In other words, this is the ratio of observed and expected rates of events. It is required that $RR > 1$, because, with $RR = 1$ means there is no difference between observed and expected incidences and hence no risk due to the vaccine. The higher the value of RR the more the risk is, and the smaller will be the expected sample size if statistical significance and alpha remain constant.

Minimum number of events: In vaccine safety surveillance, normally a minimum number of events need to be observed before rejecting the null hypothesis and marking the vaccine unsafe. This minimum number of events must be positive and recommended to be at least 4 (Kulldorff and Silva, 2017).

3.4 Incorporating variations

MaxSPRT exhibits numerous variations that we explained in Section 1. In this section, we explain what variations are allowed in our current implementation. From the beginning, our focus was to make the implementation as flexible as possible so that the end user could utilise different variations of sequential analysis by tuning the values of the relevant parameters.

The main goal of this research was to provide support for vaccine safety analysis even in case sufficient baseline information is not handy. For this purpose, we follow the calculation of the expected number of events proposed by CDC (Center for Biologics Evaluation and Research, 2021). Their calculation of the expected number of events is termed in this paper as Approximate Expected Number of Events as defined in Definition 8 (Appendix A, Terminology). The equation in the definition takes into account several parameters, such as l_{stiwa}, $p(s,t)$, α, and M. Each of these parameters introduces different variations of sequential analysis, as discussed in previous sections (such as Sections 1 and Appendix B), at least because of differ-

ent approaches available and adapted by various research groups in their implementations of the parameters and MaxSPRT. Moreover, if we follow the approach of (Greene et al, 2011), we can add more variations. From our point of view, the following variations can be applied, and we provide the flexibility to use all the variations in our current implementation.

1. $\mu_s = \sum_{t=1}^{s} \sum_{i=1}^{n_t} \sum_{w=0}^{T_i} \sum_{a=1}^{A} l_{stiwa} \times \theta_a \times P(s,t)$

2. $\mu_s = \sum_{t=1}^{s} \sum_{i=1}^{n_t} \sum_{w=0}^{T_i} l_{stiw} \times \theta \times P(s,t)$

3. $\mu_s = \sum_{t=1}^{s} \sum_{i=1}^{n_t} \sum_{w=0}^{T_i} \sum_{a=1}^{A} l_{stiwa} \times \theta_a$

4. $\mu_s = \sum_{t=1}^{s} \sum_{i=1}^{n_t} \sum_{w=0}^{T_i} l_{stiw} \times \theta$

Note that in Options 1 to 4, we have not used the test margin, unlike CDC's original calculation (Center for Biologics Evaluation and Research, 2021). We can also vary the calculation by not using any age stratification for θ and thus for other parameters, as we did in Option 2 and Option 4. We also omitted the use of $p(s,t)$ in Option 3 and Option 4. More such variations could be introduced by taking the combination of all four parameters used in Definition 8 (Appendix A, Terminology).

In our future work, we plan to investigate the performance of our MaxSPRT implementation and compare it with other statistical measures for vaccine safety surveillance. There, we plan to incorporate these variations in the performance analysis and describe our findings in detail following (Greene et al, 2011).

3.5 Flowchart

Figure 3, shows a flow diagram of various sequential analyses. It describes the steps from taking input through to signal detection, including possible decision points to get a variation of sequential analysis utilising the available information and selecting from possible options. The diagram, however, shows a general flow of processes for most of the possible implementations and variations.

4 Discussion

In the previous sections, we introduced the variations of sequential analysis implementation, the reasons and the factors behind the variations, the challenges associated with particular variations, and how to deal with them. We also explain how we implemented it with maximum flexibility to offer as many variations as possible while minimising editing to the underlying code. Our current implementation, however, does not include as yet all possible variations of sequential analysis, however, our Object-Oriented code can be easily extended with new variations, functionalities, and parameters without breaking any working code or module. This extensibility offers the opportunity to add new variations and devise more effective and adaptable signal detection. Some additional potential variations that can be added are as discussed in Appendix D.

We also plan to extend our current implementation in various dimensions, such as applicability in surveillance and in other domains, to increase reusability, and enhance capability; towards wider accessibility, performance assessment and comparison, providing soundness in signal detection, and adaptability. We can extend our current implementation and the work on MaxSPRT in the following ways:

– **Applicability in surveillance:** Integrate MaxSPRT in a Bayesian Network (BN) (Koller, Daphne and Friedman, Nir, 2009; Jensen, Finn V and Nielsen, Thomas Dyhre, 2007) model that is being developed for signal detection within our organisation. Our plan is to build a BN model containing widely used statistical surveillance techniques as integrated nodes. Finally, the BN is expected to perform probabilistic inference (Samiullah et al, 2023; Flores et al, 2004) and find signals more effectively.

– **Performance benchmarking:** We plan to compare the performance of different variations of MaxSPRT and different statistical surveillance methods and techniques, such as proportional reporting ratio (PRR) (Clothier et al, 2019), Empirical Bayes (Harpaz et al, 2013), CUSUM (Musonda et al, 2008), for safety signal detection. This requires some standards of comparison, and our initial plan is to apply all the variations of sequential analysis and all other statistical methods under consideration on the same data sample. We can then compare sequential analysis variations in their capacity to detect signals under particular circumstances. Among the methods that can detect signals, we can rank them based on the time required to detect a signal. We can also check if the generated signal is a false alarm or not, by checking the results among the methods. This develops another way of evaluating the capability of the methods in signal detection. We can then perform the analysis of effectiveness among the methods by calculating precision and recall.
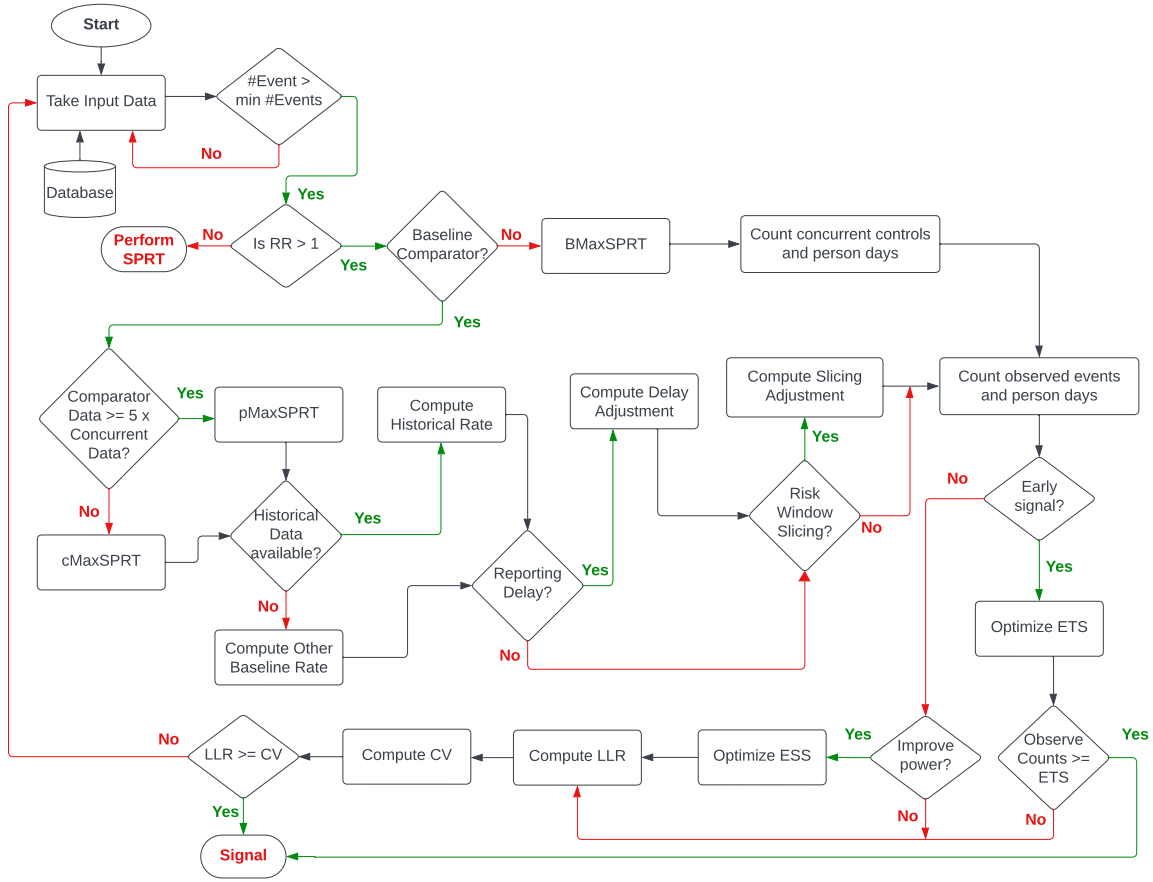
Fig. 3 Flowchart of MaxSPRT-based Sequential Analysis

– **Providing wider accessibility:** Difficulties faced in the implementation can occur due to challenges associated with the range of potential variations and no definite answer to some critical issues; and a lack of top-to-bottom detailed illustration of use cases. Another set of challenges was the verification of the correctness and soundness of the implementation. This motivated us to write this paper and further plan to develop an Application Programming Interface (API) that can be used by researchers to detect signals using their data. Using an open-access API can also assist in verifying implementation by comparing implementation results with the output and scores generated by our API on the same data.

– **Adaptability and extensibility:** Our framework is a generic framework that adapts to the variation in the structure and schema of data for analysis. Moreover, it has the flexibility to accommodate new methods and ways of calculating parameter values required for the sequential analysis.

4.1 Challenges and Lessons Learnt

SPRT and MaxSPRT were introduced to conduct continuous (Leite et al, 2016) and group (Jennison and Turnbull, 1999) sequential analysis that detects potential signals when adverse events exceed the probability that they occur by chance, and allows for continuous monitoring of adverse events that may occur in temporal association following immunisation. However, Continuous surveillance for signal detection through MaxSPRT is challenging. The measure itself is quite complex and has been used and implemented in various ways. The challenges can be divided into three stages: 1. pre-implementation, 2. implementation, and 3. post-implementation.

Before we started the implementation of MaxSPRT, we faced a number of challenges. The first challenge was establishing a clear understanding of the procedure, terms, and underlying mathematics due to its many variations. The variations depend on the availability, nature, and properties of the data in hand. In addition, there is no hard and fast rule or specified model for the implementation. Organisations have different capabili-

ties and limitations that direct implementers towards different variations, as explained in Section 1.

The absence of thorough, detailed instructions for doing sequential analysis from the bottom up was one of the biggest obstacles we faced during implementation. A number of studies, including (Center for Biologics Evaluation and Research, 2021; Mesfin et al, 2019; Silva et al, 2021), provide insightful information, however they mostly concentrate on methodological or high-level analytical issues. There is still a dearth of thorough documentation on the actual implementation procedure, especially when it comes to people with little experience with epidemiology or vaccination. This emphasises the necessity of repeatable procedures that connect statistical theory to practical implementation in vaccination safety monitoring. As an example, the definition of $p(s,t)$ is insufficiently defined to reliably reproduce (as discussed in Section 3.2.3). The data source and technique to calculate the "proportion of data complete" that captures and addresses the observation delay in reporting reactions add several variations in MaxSPRT, and dealing with the calculation of this term was challenging.

Calculating the expected number of events was the most onerous. Its significance in the whole surveillance process increases the difficulty because any significant mistake or error in the calculation of this term will mislead the surveillance. It has several components such as $p(s,t)$, l_{stiwa}, θ and each adds different variations (discussed in Section 1 and Section 3.4), especially when we do not have any prior information about the admissibility and correctness of the approaches and heuristics we have adopted.

Deciding whether to use age stratification or not, and if used, choosing between static grouping or dynamic binning technique adds to the challenge. Moreover, in the case of static grouping, deciding which grouping would give us optimal and more indicative results is not known and is subject to empirical analysis. In the case of either type of grouping, calculating the background rates for each age group also adds another layer of difficulty.

Interestingly, the data source and technique to calculate the background rates add some variations and thus, additional challenges are introduced. Data properties and characteristics, such as the distribution of the data, also increase complexity in the implementation. One of the significant inherent challenges in any process involving data is missing values. In the case of MaxSPRT analysis, missing information in the data needs special consideration. We handled the missing data with great care, explained in Appendix B.1.

We also had to deal with some challenges regarding the selection of the development and implementation tool and setting up the environment. We explained the issue in detail with our workaround in Appendix B.3.

Moreover, it is also challenging to validate or assess if the implementation of MaxSPRT was correctly performed; to the best of our knowledge, there is no common and publicly available data to be used as a benchmark and standard outcome to compare the results due to the privacy, security, and ethics policies of the data used in the domain of public health. Choosing and using the right version of sequential analysis adds to the complexity. To deal with all of these challenges, we implemented the framework in a way that allows the end user to exhibit different variations of sequential analysis for analysis using some arguments, and adjustment or changes to their values. We listed the available variations in Section 3.4.

5 Conclusions

MaxSPRT is a complex but powerful method for sequential group and continuous surveillance. It provides flexibility for implementing in different situations and conducting surveillance with various data limitations. However, its flexibility comes with implementation and operational trade-offs. The associated challenges can make it difficult to implement in some contexts or organisations, which can prevent benefits from being realised.

In this paper, we explained the variations; factors behind the variations; parameters associated with different variations; potential ways of calculating parameters; challenges in calculating parameters; and consequences of using and not using various parameters. We also provided a step-by-step implementation description; demonstrating the corresponding output of each step. The step-by-step discussion also included a detailed description of our data, handling and preprocessing of the data, and our working environments.

This paper explained the MaxSPRT surveillance method in a new, different, clear, and intuitive way, aiming to support its reproduction by providing mathematical and scientific notations and definitions of the relevant terms that mitigate the gaps between concept and reality. We included a flowchart of the overall implementation method for different variations of sequential analysis; and also added a hierarchy of the variations to demonstrate the evolution of the methods.

Furthermore, we proposed new measures to assist in the calculation of some parameters. In case there are any special attributes or some missing attributes in our

data, we showed in a detailed manner how to deal with the issues and proposed appropriate workarounds as a guideline for reproduction.

We discussed further variations of sequential analysis that may better capture a signal and our plan to extend the work, aiming to provide greater support to the vaccine safety surveillance community and to the public health surveillance community more broadly. We intend to further the work with the empirical outcomes of the variations of sequential analysis and other surveillance techniques.

Our paper provides a clear road map for future researchers and pharmacovigilance practitioners to implement MaxSPRT not only for vaccine safety surveillance but also for broader adverse drug event sequential analysis and safety monitoring.

Declarations

Conflicts of Interest/Competing Interests None to mention.

Ethics Approval This paper does not require ethics approval.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Availability of Data and Material Not applicable.

Code Availability Not applicable.

Authors' Contributions All authors contributed to the study conception. The frst draft of the manuscript was written by Samiullah, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding acquisition: Buttery

Investigation: Samiullah, Clothier, Yin, Munakabayo, Kattan, Mallard

Methodology: Samiullah

Project administration: Samiullah, Dimaguila

Supervision: Clothier, Dimaguila, Buttery

References

Center for Biologics Evaluation and Research (2021) COVID-19 Vaccine Safety Surveillance: Active Monitoring Master Protocol. Tech. rep., Food and Drug Administration, USA

Clothier HJ, Crawford NW, Kempe A, Buttery JP (2011) Surveillance of adverse events following immunisation: the model of SAEFVIC, Victoria. Communicable diseases intelligence quarterly report 35(4):294–298

Clothier HJ, Crawford NW, Russell M, Kelly H, Buttery JP (2017) Evaluation of 'SAEFVIC', a pharmacovigilance surveillance scheme for the spontaneous reporting of adverse events following immunisation in Victoria, Australia. Drug safety 40:483–495

Clothier HJ, Lawrie J, Russell MA, Kelly H, Buttery JP (2019) Early signal detection of adverse events following influenza vaccination using proportional reporting ratio, Victoria, Australia. Plos one 14(11):e0224702

Flores J, Gamez JA, Olesen KG (2004) Incremental compilation of Bayesian networks in practice. In: Intelligent Systems Design and Applications (ISDA 2004), < Forlag uden navn>

Gordon Lan K, DeMets DL (1983) Discrete sequential boundaries for clinical trials. Biometrika 70(3):659–663

Greene SK, Kulldorff M, Yin R, Yih WK, Lieu TA, Weintraub ES, Lee GM (2011) Near real-time vaccine safety surveillance with partially accrued data. Pharmacoepidemiology and drug safety 20(6):583–590

Harpaz R, DuMouchel W, LePendu P, Shah NH (2013) Empirical Bayes model to combine signals of adverse drug reactions. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 1339–1347

Jennison C, Turnbull BW (1999) Group sequential methods with applications to clinical trials. CRC Press

Jensen, Finn V and Nielsen, Thomas Dyhre (2007) Bayesian networks and decision graphs. Springer

Koller, Daphne and Friedman, Nir (2009) Probabilistic graphical models: principles and techniques. MIT press

Kulldorff M, Silva IR (2017) Continuous post-market sequential safety surveillance with minimum events to signal. Revstat statistical journal 15(3):373

Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R (2011) A maximized sequential probability ratio test for drug and vaccine safety surveillance. Sequential analysis 30(1):58–78

Lee H, Kim JH, Choe YJ, Shin JY (2020) Safety surveillance of pneumococcal vaccine using three algorithms: Disproportionality methods, empirical bayes geometric mean, and tree-based scan statistic. Vaccines 8(2):242

Leite A, Andrews NJ, Thomas SL (2016) 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

Li L, Kulldorff M (2010) A conditional maximized sequential probability ratio test for pharmacovigilance. Statistics in medicine 29(2):284–295

Mesfin YM, Cheng A, Lawrie J, Buttery J (2019) Use of routinely collected electronic healthcare data for postlicensure vaccine safety signal detection: a systematic review. BMJ global health 4(4):e001065

Mesfin YM, Cheng AC, Enticott J, Lawrie J, Buttery J (2021) Post-vaccination healthcare attendance rate as a proxy measure for syndromic surveillance of adverse events following immunisation. Australian and New Zealand journal of public health 45(2):101–107

Murphy DM, Cox DJ, Connolly SA, Breen EP, Brugman AA, Phelan JJ, Keane J, Basdeo SA, et al (2023) Trained immunity is induced in humans after immunization with an adenoviral vector COVID-19 vaccine. The Journal of clinical investigation 133(2)

Musonda P, Hocine MN, Andrews NJ, Tubert-Bitter P, Farrington CP (2008) Monitoring vaccine safety using case series cumulative sum charts. Vaccine 26(42):5358–5367

O'Brien PC, Fleming TR (1979) A multiple testing procedure for clinical trials. Biometrics pp 549–556

Phillips A, Jiang Y, Walsh D, Andrews N, Artama M, Clothier H, Cullen L, Deng L, Escolano S, Gentile A, et al (2023) Background rates of adverse events of special interest for COVID-19 vaccines: A multinational Global Vaccine Data Network (GVDN) analy-sis. Vaccine 41(42):6227–6238

Pocock SJ (1977) Group sequential methods in the design and analysis of clinical trials. Biometrika 64(2):191–199

Samiullah M, Nicholson A, Albrecht D (2023) Shareable and inheritable incremental compilation in iOOBN. In: Pacific Rim International Conference on Artificial Intelligence, Springer, pp 91–103

Silva IR (2018) Type I error probability spending for post-market drug and vaccine safety surveillance with Poisson data. Methodology and computing in applied probability 20(2):739–750

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

Silva IR, Kulldorff M (2024) Sequential-package: Analysis Support, Critical Values, Power, Time to Signal and Sample Size for Sequential Analysis with Poisson and Binomial Data. URL https://www.rdocumentation.org/packages/Sequential/versions/4.3.3

Silva IR, Lopes WM, Dias P, Yih WK (2019) Alpha spending for historical versus surveillance Poisson data with CMaxSPRT. Statistics in medicine 38(12):2126–2138

Silva IR, Kulldorff M, Katherine Yih W (2020) Optimal alpha spending for sequential analysis with binomial data. Journal of the Royal Statistical Society Series B: Statistical Methodology 82(4):1141–1164

Silva IR, Maro J, Kulldorff M (2021) Exact sequential test for clinical trials and post-market drug and vaccine safety surveillance with Poisson and binary data. Statistics in Medicine 40(22):4890–4913

Wald A (1992) Sequential tests of statistical hypotheses. In: Breakthroughs in statistics: Foundations and basic theory, Springer, pp 256–298

World Health Organization and others (2014) Global manual on surveillance of adverse events following immunization. Tech. rep.

Yoon D, Kim JH, Lee H, Shin JY (2020) Updates on vaccine safety and post-licensure surveillance for adverse events following immunization in South Korea, 2005–2017. Yonsei Medical Journal 61(7):623

Appendix

A Terminology

We refer the interested reader to Silva and Kulldorff (2015) and Silva et al (2019) for the definitions of "group" and "continuous" sequential analysis, concave and convex type alpha spending functions, ETS, and ESS. Some of the other related and important terms are defined here.

Definition 1 Study Week: Let s be the week at which the MaxSPRT analysis is being performed to find any potential signal for an AEFI on a particular dataset. This week, s, is known as study week.

Definition 2 Study period: Let s be the study week of surveillance through MaxSPRT analysis. The time between the beginning of an observation or surveillance and the study week s is known as the study period, i.e., from the 1st week to week s, and denoted as $[1:s]$.

As an example, if an epidemiologist wants to perform a MaxSPRT signal detection analysis for fever post-influenza vaccination as an AEFI at week 10 of the observation period, then s=10. In this case, the study period is Week 1 to 10.

Definition 3 Intermediate week: Let s be the study week of surveillance through MaxSPRT analysis. Any week, ranging from the start week (time/week of vaccine administration starts) up to the study week s, is called an intermediate week and denoted as t where $1 \le t \le s$.

For MaxSPRT analysis, we need to calculate several values in a weekly fashion, from week 1 of the observation week up to the observation/study week s. These weeks are known as intermediate weeks and are denoted by t. So, t=1, 2, ..., s.

Definition 4 Exposed weeks: Let s be the study week of a MaxSPRT analysis of a vaccine for an AEFI that has a risk window of length RW days. Then, for any vaccinee who was administered a dose of the vaccine on week t and having TTO days of time-to-onset is in the exposed state for $min(TTO, RW)$ days starting from week t just after the day of the dose administration. The weeks from t to $t + \lceil min(TTO, RW)/7 \rceil$ are known as exposed weeks for the vaccinee and denoted as w.

In other words, exposed weeks of a vaccinee are the weeks after a dose is administered to the person until the reaction/AEFI is observed on them.

As an example, if someone gets a dose on week-2 of the study period, and the risk window is 4 weeks or 28 days, then 2nd week of the study period (t=2) is considered as the administered week and denoted as w=0 at week $t = 2$, week $t = 3$ is the 1st exposed week and denoted as w=1, and so on. This continues up to w=4 during week t=6 of the study period.

Definition 5 Time to onset: Let a vaccinee V be administered a dose on day d (calendar date D) of the week t and a MaxSPRT analysis being performed in week s where $t \le s$. If the AEFI is observed on that vaccinee in day d' (calendar date D') of week t' where $t \le t' \le s$, then the time required for the reaction to start/set is the difference between the two days D and D' is known as the time to onset and denoted as TTO where $TTO = interval(D' - D)$.

TTO can be expressed in hours or in days in the SAE-FVIC environment/system. As an example, if a vaccinee reported that on the 15th day after the dose the AEFI occurred, then TTO=15 days. This is rounded up to a day if TTO is given in hours.

Definition 6 Risk Window: Let a vaccinee V be administered a dose on day d of the week t. Say, the AEFI is observed on that vaccinee on day d' of week t' where $t \le t' \le s$ and TTO be the time-to-onset of the AEFI for the vaccinee V. The maximum value of (TTO, RW) for which the AEFI be considered as an effect of the vaccine is known as the Risk window and is denoted as RW.

In other words, the Risk window is the maximum number of days/weeks a vaccinee is at risk of the AEFI after a dose. As examples, some AESIs with relevant risk windows are shown in the following Table 8.

Table 8 Example of risk windows for potential AESIs (Center for Biologics Evaluation and Research, 2021)

AEFI	Risk Window
Guillain-Barré syndrome	1-42 days
Bell's palsy	1-42 days
Anaphylaxis	0-1 day
Encephalomyelitis/Encephalitis	1-42 days
Narcolepsy	1-42 days
Appendicitis	1-42 days
Non-hemorrhagic stroke	1-28 days
Hemorrhagic stroke	1-28 days
Acute myocardial infarction	1-28 days
Myocarditis/Pericarditis	1-42 days
Deep vein thrombosis (DVT)	1-28 days
Pulmonary embolism (PE)	1-28 days
Disseminated intravascular coagulation (DIC)	1-28 days
Immune thrombocytopenia (ITP)	1-42 days
Transverse myelitis	1-42 days
Multisystem inflammatory syndrome	1-42 days

Definition 7 Observed exposed person-time: Let s be the study week of a MaxSPRT analysis and a vaccinee i of age group a is administered a dose on week t and $1 \le t \le s$. The number of days of a week w ($t \le t + w \le s$) after vaccination and until the reaction/AEFI is set on the vaccinee is known as exposed person-time for the vaccinee i in wth week and denoted as l_{stiwa}.

It is required for calculating the expected number of events, μ_s, and represents the length of exposed time for a vaccinee after a dose administration in a person-time unit.

Note, l_{stiwa} has five terms incorporated in it, namely, s, t, i, w, and a that corresponds to observation week s, dose week t, i^{th} vaccinee, wth exposed week and a age group in which the vaccinee i belongs to. It is defined as the number of exposed days or length of exposure of a person i of age group a who got a dose on week t where t^{th} week is a week within the study week s. Note that for any vaccinee, $w = 0$ means the week the vaccinee i was administered a dose and $w = 1$ means 1st exposure week of the person.

l_{stiwa} ranges from 0 to 7. If the dose is administered on the selected last day of a week t, l_{stiwa} for the week t will be 0 and denoted as $l_{sti(w=0)a} = 0$. More specifically, if the person was 7th in the list of people reporting for that week of their AEFI and got the dose on the third week of the study period of 10 weeks, then the length of exposure for the first week of the person (or third week of the study for the person) is 0 and denoted as $l_{(s=10)(t=3)(i=7)(w=0)(a=50-59)} = 0$, where the person is categorised in the age range between 50 and 59. If the risk window of the AEFI is greater than or equal to 7 days, then for the same person the l_{stiwa} for the next week is $l_{(s=10)(t=4)(i=7)(w=1)(a=50-59)} = 7$.

Say, for a risk window of 14 days, if the dose was administered on day 2 of a week, then $l_{sti(w=0)a} = 5$, $l_{sti(w=1)a} = 7$ and $l_{sti(w=2)a} = 2$. On the other hand, if the dose was administered on day 4 of a week, then $l_{sti(w=0)a} = 3$, $l_{sti(w=1)a} = 7$ and $l_{sti(w=2)a} = 4$.

Definition 8 Approximate Expected number of events: Following the definition of (Center for Biologics Evaluation and Research, 2021), let s be the study week of a MaxSPRT analysis and t be any intermediate week and $1 \le t \le s$. The approximate number of events that are to occur in a week t under the null hypothesis, is known as the approximate expected number of events. It is denoted as μ_s'' and defined as:

$$\mu_s'' = \sum_{t=1}^{s} \sum_{i=1}^{n_t} \sum_{w=0}^{T_i} \sum_{a=1}^{A} l_{stiwa} \times \theta_a \times P(s,t) \times M$$

where,

- s is a number representing the week when the signal detection analysis is being performed,
- t is a number within $[1, s]$ denoting any particular week from week 1 up to the study week,
- n_t is the number of vaccinee got a dose on week t,
- i is any particular individual or vaccinee,
- T_i is the exposed weeks for vaccinee i ranging from $[1, RW]$,
- RW is the risk window length, expressed in the number of weeks,
- w is a value in the range $[0, T_i]$ where $t + w$ represents a week under risk where $w = 0$ represents $t + 0$ week of the surveillance which is the dose week of vaccinee i and $w = 1$ is the $t + 1$ week in the study and 1st exposed week for the vaccinee,
- A is the set of age group strata,
- a is the strata of the vaccinee i,
- l_{stiwa} is the observed exposed person-time, and
- θ_a is the per-day comparator rate of age group a.
- M is the test margin.

Let RW be the risk window for the AEFI under consideration for MaxSPRT analysis. If the analysis being performed is in δ slices, then all the calculations for MaxSPRT are performed for each slice of the window. For some variations of MaxSPRT, we can split the risk window into several parts. Say, the risk window is 42 days, if $\delta = 6$, then for each slice, the number of days = 7.

Definition 9 Expected number of events: Let s be the study week of a MaxSPRT analysis and t be any intermediate week and $1 \le t \le s$. The number of events that are likely to occur in week t under the null hypothesis, is known as the expected number of events and denoted as $\mu_{s,t}$.

It is one of the two parameters required to calculate LLR. Calculating $\mu_{s,t}$ is tricky and complicated. Estimating this value requires approximation in person-time. There are different approaches that can be adopted based on available data and information to estimate the value of $\mu_{s,t}$.

Definition 10 Observed number of events: The number of events that occurred under the risk window in week t and reported as well as observed by week s, is known as the observed number of events and denoted as $Y_{s,t}$.

This is another parameter for calculating LLR. It counts the weekly events within the risk window and is reported and observed by the study period.

Definition 11 Historical comparator rate: Let a MaxSPRT analysis be performed for an AEFI due to a vaccine. Say, in previous y years, n people were observed to be affected by the AEFI out of N people, then the historical comparator rate is defined as the ratio of the observed number of affected people by the AEFI, i.e., $\frac{n}{N \times y}$ and denoted as θ.

It is the value in person-time unit required to calculate the expected number of events, μ_s, representing the number of people per year that suffer from an AEFI. It is calculated retrospectively from the previous years' data.

Definition 12 Per-day comparator rate: Let θ be the historical comparator rate in previous y years obtained by observing n people to be affected by the AEFI out of N people, then the per-day comparator rate is defined as the ratio of the observed number of affected people by the AEFI in a day, i.e., $\frac{n}{N \times y \times 365}$ and denoted as θ'.

In the case of age grouping stratified MaxSPRT analysis, the historical comparator rate and per-day comparator incidence rate may also vary. In order to address that, these rates for any age group a are denoted as θ_a and θ_a', respectively. There are several ways of calculating the rate where two of the plausible ways are:

- the rate of the AEFI within the general population
- healthcare-seeking behaviour for AEFI following vaccination

Definition 13 Observation delay: Let s be the study week when a MaxSPRT analysis is being performed, $V = \{v_1, v_2, ..., v_n\}$ be the set of vaccinees reporting the AEFI, $T = \{t_1, t_2, ..., t_n\}$ and $T' = \{t_1', t_2', ..., t_n'\}$ be the set of corresponding AEFI observation weeks and AEFI reporting weeks of the vaccinees in V, respectively, where $1 \le t_i \le t_i' \le s, 1 \le i \le n$. Then the time elapsed between the dose administration and AEFI reporting is known as observation delay and denoted as, $\delta = interval(T', T)$.

AESIs may not be promptly reported in real time due to various factors. Even when reported, there can be delays in recording AEFI cases, a phenomenon referred to as observation delay. These delays result in a reduced amount of observed data compared to the actual amount during analysis, potentially leading to an underestimation of expected event numbers and a risk of missing signals. To mitigate this issue, cMaxSPRT incorporates an approximation technique involving the calculation of the proportion of complete data (the ratio of observed to expected data). This concept is defined in the following Definition 14.

Definition 14 Proportion of data complete: Let s be the study week when a MaxSPRT analysis is being performed and δ be the observation delay in the reports. Then, the proportion of data complete at week s is the ratio of reports of week $1 \le t \le s$ that are available by week s due to observational delay and total expected reports for week t. It is denoted as $p(s,t)$.

As an example, at week $s=10$, we observe 6 reports of week $t = 5$, and it is expected that 10 reports to be noted for week $t = 5$. That means, only 60% of week $t=5$ data or report is available. This is denoted as $p(10, 5) = 0.6$.

There are several ways or heuristics to calculate this value. A retrospective analysis of historical data can be an option to compute this metric.

At the time of analysis, we may not obtain all the reports in time due to various reasons. Addressing this factor while estimating the expected number of events is crucial. Otherwise, we may overestimate the expected number of events. To address this issue and adjust the expected number of events, CDC (Center for Biologics Evaluation and Research, 2021) proposed the following equation.

$$p(s,t) = \sum_{i=0}^{s-t} p(i)$$

We explained the calculation techniques and challenges of this metric in the paper in Section 3.2.3.

Definition 15 Cumulative expected number of events: Let s be the study week when a MaxSPRT analysis is being performed and $\mu_{s,t}$ be the expected number of events (reports) to occur in weeks $1 \leq t \leq s$. Then the cumulative expected number of events is denoted by μ_s and defined as:

$$\mu_s = \sum_{t=1}^{s} \mu_{s,t}$$

Definition 16 Cumulative observed number of events: Let s be the study week when a MaxSPRT analysis is being performed and $Y_{s,t}$ be the observed number of events (reports) occurring in weeks $1 \leq t \leq s$. Then the cumulative observed number of events is denoted by Y_s and defined as:

$$Y_s = \sum_{t=1}^{s} Y_{s,t}$$

It represents the number of AESIs occurring within the exposed risk window. It can be easily calculated by counting the cumulative sum of AEFI counts occurring weekly straightway. In order to find the number of observed events up to week s, we have to simply accumulate the number of events occurring weekly.

Definition 17 Log-likelihood ratio: Let s be the study week of a MaxSPRT analysis where t be any intermediate week and hence $1 \leq t \leq s$. Also, let μ_t and Y_t be the cumulative approximate expected number of events and cumulative observed number of events in the study, respectively, up to week t. Now, according to (Kulldorff et al, 2011), the Log-likelihood ratio of week t be denoted as LLR_t and defined as follows:

$$LLR_t = \begin{cases} (\mu_t - Y_t) + \mu_t log\left(\frac{Y_t}{\mu_t}\right) & Y_t \geq \mu_t \\ 0 & Y_t < \mu_t \end{cases}$$

Definition 18 Critical value: Let s be the study week of a MaxSPRT analysis and LLR_t be the Log-likelihood ratio for any intermediate week $1 \leq t \leq s$. According to (Kulldorff et al, 2011), the boundary of the value of LLR_t (or μ_t if the analyser chooses this over LLR_t) that helps decide if the alternative hypothesis is true is defined as a critical value and denoted as CV.

Depending on whether an alpha spending plan is employed, there are two techniques to establish critical values (CVs) in sequential analysis. The CV stays constant during the whole surveeillance period in the absence of an alpha spending plans, providing a steady threshold for signal identification. Using an alpha spending plan, on the other hand, makes the CV dynamic, changing at every analysis point according to the amount of Type I error spent. More flexible and effective monitoring is made possible by this method, which preserves overall Type I error control across several data analyses while allowing for earlier signal discovery.

There are two ways of detecting a signal. Firstly, if the cumulative number of events observed at any time is beyond a prespecified number aka ETS, then there is no need to further continue the analysis and the null hypothesis can be rejected straightaway. Otherwise, at any week t, if the LLR_t (or μ_t) is beyond the corresponding critical value CV expressed in likelihood (or, expressed in event count), then there is a signal, that is, $LLR_t \geq CV_{LLR_t}$, where CV_{LLR_t} or $\mu_t \geq CV_{\mu_t}$,

where CV_{LLR_t} or CV_{μ_t} is the critical value for μ_t being the cumulative expected number of events for week t.

Kulldorff et. al. (Kulldorff et al, 2011) proposed the technique, calculation, and uses of CVs for signal detection. Silva et al. developed an R Sequential package (Silva and Kulldorff, 2024) with all functionality native to sequential analysis calculation. If someone uses R, they do not need to reinvent the wheel for computing CV and other parameters required for MaxSPRT analysis. Moreover, in their paper, they have given CVs for up to 1000 expected number of events, for the reason that if someone wants to use a CV for signal detection, they do not need to compute CVs again. The R Sequential package provides comprehensive functionality for sequential analysis, supporting multiple test statistic scales (e.g., cumulative event counts (Y_t), MaxSPRT, Pocock, O'Brien-Fleming, and Wang-Tsiatis). Key features include: − Exact calculation of statistical power − Calculation of expected time to signal − Computation of expected sample sizes − Sample size determination for sequential designs − Alpha spending plan design and optimization − One-tailed and two-tailed testing capabilities − Functions for sequential analysis of Poisson data (Analyze.Poisson) − Functions for sequential analysis of binary data (Analyze.Binomial) − Performance evaluations for Poisson thresholds (Performance.Threshold.Poisson) − Performance evaluations for binomial thresholds (Performance.Threshold.Binomial) − Support for multiple weighted outcomes − Construction of fixed-width and fixed-accuracy confidence intervals − Visualization and monitoring of sequential test boundaries and results

All these capabilities are seamlessly available irrespective of the chosen test statistic scale.

In this paper, we offer a detailed explanation of the steps and challenges associated with the implementation of MaxSPRT on SAEFVIC's reporting database. SAEFVIC is the jurisdictional vaccine safety surveillance service of Victoria, Australia. Its AEFI and AESI reporting database (Clothier et al, 2011) does not currently possess dose administration information. Dose administration data is collected from the Australian Immunisation Register. At the time of writing, SAEFVIC only had access to COVID-19 vaccines, in near real-time. Hence, the denominator of the MaxSPRT calculation - the expected number of events - requires estimation or conditioning to deal with its associated uncertainty. Therefore, in our implementation, we use the concept of cMaxSPRT, similar to the approach adopted by the US CDC (Center for Biologics Evaluation and Research, 2021) for surveillance of COVID-19 vaccine safety using insurance claim data. In both cases, the denominator is not known in real-time and at least not during the analysis. Being able to use a statistical method e.g. MaxSPRT in the absence of dose distributed denominator is therefore a practical advantage for spontaneous reporting systems to function as stand-alone signal detection systems.

B Data Processing and Environment Setup Implementation

In this section, we have explained the steps of implementation in a particular surveillance setting (in our case SAEFVIC).

B.1 Preprocessing

SAEFVIC data are prepared prior to processing. The pre-processing involves de-identification, noise handling, and missing data management. We have used the features of the MaxSPRT framework to deal with the pre-preprocessing challenges, especially, the missing data management. We allow analysers to opt-in methods to fill in missing values with some appropriate alternative values or simply discard the rows/records with missing values.

To justify filling in missing values, which is questionable from the point of view of epidemiology, this is logical from a data analysis perspective in certain scenarios. As an example, if we discard the records for some missing values in gender but the gender-specific stratification is not required in the calculation or analysis, then removing that record is not worthwhile. Rather, discarding the value adds bias to the numerator and denominator calculations. However, it is up to the user of the framework to decide whether to use the filling-in method or not.

We first de-identify the data to preserve the privacy of the reporters and vaccinees. In the framework, we allow for the de-identification of the data, which is quite straightforward, by removing all demographics and identifiable information about the vaccine and the reporter. The analyser can also mention if any further information needs to be removed for the de-identification process other than the demographics. In the de-identification process, we assign a unique identifier to each report for record-keeping purposes. We keep the columns as specified in Table 2.

Next, we dealt with noisy data. We treat noisy data in a special manner by initially considering them as missing values. In this paper, noisy data are defined as values or inputs susceptible to error due to human and system error. First, we run a routine data cleaning for logic consistency, missing values, and filtering process, where any value that is beyond the range and domain of the corresponding data attribute is identified as noisy data. We first replaced a noisy value with missing values. Then treated them the same way as missing values.

In the third preprocessing step, we dealt with missing values. A common practice of dealing with missing values in data processing is to delete rows with noisy or missing values. However, in the case of vaccine safety surveillance, every report is important, so it is necessary to manage, instead of removing, missing records. We needed to carefully deal with missing records and use effective ways to fill them in. Of course, fill-in methods sometimes affect some important properties of the data such as mean, median, mode, standard deviation, and variance of the data. Therefore, the choice of the right fill-in method that has the least negative impact on the data is essential. Nevertheless, we leave it to the analyser to use fill-in, discard the record, or use any specific way of dealing with the missing and noisy data. In case, the analyser chooses to fill in missing values, then one of the following could be used: (i) default value/s, (ii) data-type specific alternative value, (iii) the analyser provided alternative values, or (iv) a method to get appropriate alternative values provided by the analyser.

To the best of our knowledge, there is no best or single approach to dealing with missing values. Hence, in our framework, we allowed several methods to fill in missing values. An end user of the framework can opt into a specific method through parameterisation if they want to override the default one in their MaxSPRT analysis. Moreover, our data processing unit is capable of accepting new methods of filling in the missing values. The methods integrated into the current implementation that are used to fill in missing values in our data are actually the available methods in the Python Pandas package for fill-in such as "replace" with mean, median, mode, or some specific value like zero; and "interpolate" which also comes with a handful of options[1].

To explain the preprocessing process on our data, the following steps are taken. As an example, for different columns, we have used different default fill-in methods for illustration purposes. For missing "Age" and "TTO" values, we used linear interpolation as default. For "Sex" attribute, the missing values can be either replaced by randomly chosen value between "M" and "F" or as per the AESI we can decide if bias for a particular "Sex" is to be considered. By bias, we mean, some AESIs are more common or only for one sex (male or female) in comparison to the other. For treating missing values of "Time to vaccine", "Reporting time" and "Submission time", we converted the times to double precision real numbers and then applied linear interpolation. However, this part is quite tricky and crucial, as these values will be used extensively in calculating most of the important terms in MaxSPRT. Also, there is an inherent order of the three values by definition, and the order is *Time to vaccine ≤ Reporting time ≤ Submission time*. Conducting the fill-in should not violate this order constraint. There is another constraint on "Time to vaccine", "Reporting time", and "TTO" that is *Reporting time – Time to vaccine ≤ TTO*. Therefore, the interpolation of "Time to vaccine", "Reporting time", and "TTO" should satisfy this constraint as well.

We transformed the three timestamp values while dealing with missing values. In addition, we ran another iteration of data transformation on our data before we started processing them for analysis. The transformation includes a common and unique format of timestamps for all three timestamps we have used in our data, and TTO values are rounded up to the nearest integer value.

To perform age-based and sex-based MaxSPRT analysis, we use age and sex columns. However, age-based analysis requires binning (or grouping or stratification of ages). We considered stratification as a preprocessing step. Here, we added a column to the data, "Age-Strata" having nine different strata, namely 0–9, 10–19, 20–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80+. This grouping can be done on more granular levels and also in a dynamic way rather than the static grouping as shown here. The alternative ways of age grouping also added variations to the analysis, as we mentioned before. Hence, we kept the scope to feed in new binning techniques as well as new or different age strata in our system.

Finally, we discarded the reports where TTO is beyond the risk window, and this is an important step. Note that this is applicable if it is decided that the reactions that were observed after the specified and well-defined risk window are not considered in surveillance.

As a pre-processing step of the data, we assessed that the values/rates need to be scaled by the length of risk window. This is because these values are computed in terms of person-time on a scale of 100,000 people per annum, and the risk windows are expressed in days. As an example, if the rate is 5.0 for 100,000 people-year and the risk window is 40 days long, scaling it down to $\frac{5 \times 40}{100000 \times 365}$ or $\frac{5 \times 40}{365}$ for 100k people is required to be used in MaxSPRT analysis.

Once the scaling is performed, we call them a per-day comparator to estimate the expected number of events. This

---

[1]Interested readers are recommended to see Pandas manual or https://www.geeksforgeeks.org/working-with-missing-data-in-pandas/.

is referred to as the expected rate of AEFI approximated for the currently vaccinated population which is not known, at least during the time of analysis.

We performed age-based stratification of the whole MaxSPRT analysis. Hence, this parameter offers another layer of computation where the rates vary for different age groups. In the case of no age-based stratification, we assumed a single and common rate for all age groups. This allows for facilitating the use of age groups and avoiding the age groups without any major reworking in the implementation, especially in the other modules of the implementation that are developed to allow age grouping in mind. Note that in cases of age grouping, all the calculations are performed at a more granular level and all the parameters used, including this one, are taken/calculated based on the age ranges of the groups.

Interestingly, in the case of some parameters that are not available for age groups, we used the common and general values for all ages. In our current implementation, we have allowed both, adding age-based and general values for this parameter. However, for brevity and simplicity, this paper uses a single and general value for the parameter.

B.2 Processing of SAEFVIC Data

After the preprocessing of data, we started calculating the metrics that are required to perform MaxSPRT analysis and detect signals. The analysis is required to be performed in a well-defined frame of weeks for the sake of tracking the computation and facilitating weekly analysis. As a next step, we need a set of continuous windows in the analysis, where the data is divided into different windows of weeks based on their occurrence dates. Our implementation does not rely on any explicit length of the analysis or size of the data frame, or count of windows in the data frame. Moreover, our current implementation allows the analysers or epidemiologists to set their preferred start date, end date, length of analysis or count of windows in the frame.

In this task, the main challenge was to take the starting date of the first week. Then the formation of the windows was quite straightforward. The first window was formed by adding 7 days to the start date of the analysis. This gives us another date. Hence, our windows are defined by the start and end date of the week (does not matter what day of the week it is). Then, adding 1 to the last date of the first week gives us the first date of the second week. We make a set of windows each of 7 days or 1 week long by continuously following the steps. A point worth noting is that the start date of analysis is by default taken from the data by finding the earliest vaccination date. This can also be replaced by the earliest report date or report submission date. On the other hand, the frame ends with a week having an end date that is not later than six days from the latest date of reporting found in the data. Hence, our implementation does not rely on any explicit length of the analysis, the size of the data frame, or the count of windows in the data frame. Moreover, our current implementation allows analysts or epidemiologists to set their preferred start date, end date, length of analysis, or count of windows in the frame.

Finally, while computing various matrices like $p(s,t)$, l_{stiwa}, Y, μ, and LLR, the data is further processed and new columns are added to it as required. We described this further processing and amendment of the columns to the data later, in Appendix B.

B.3 Environment Setup and tools

Silva et al. (Silva et al, 2021) and Greene et al (Greene et al, 2011) recommended the use of R programming language as there is a package called "Sequential" in R developed especially for sequential analysis and contains all the required functionalities for sequential analysis. The Sequential R package provides a powerful framework for implementing sequential analysis through defining essential design parameters such as Type I and II error rates, surveillance length, and statistical power. It enables users to define a concave, convex, or linear alpha spending function that dynamically distributes the Type I error among several looks and automatically modifies decision thresholds in response to progress tracking. This adjustment reduces false positives while assisting in the earliest possible signal detection. In order to ensure effective surveillance design, the program additionally computes and optimises Expected Time to Signal (ETS) and Expected Sample Size (ESS). Combining these components improves overall statistical power in continuous or group sequential vaccine safety monitoring and allows for flexible adjustment of significance thresholds. However, Kulldorff Kulldorff et al (2011), Silva Silva et al (2019) and Li Li and Kulldorff (2010) provided recalculated values for specific parameters such as ETS, ESS and CV for different alpha and significant level that if someone wants to use those values they do not need to regenerate the values to save time for analysis.

We have used Python and various packages (e.g., Panda) of Python that facilitate easy data handling and extensibility. For better error handling, extensibility, and code reuse; and to follow software engineering best practice and principles, we adopted the Object-Oriented Programming style. This allows new methods to be injected easily into the implementation. As an example, "R" codes can be easily integrated by embedding into the Python code segment. Hence, the "R" sequential package and other required "R" codes to deal with CV calculation and α spending plan is added as an embedded module in the python implementation of MaxSPRT. We have integrated the "R"-scripting facility and embedded the "Sequential" package in our current implementation to provide maximum facility to the analysers.

Our choice of Object-oriented programming style also helps in implementing the different variations of sequential analysis by adding more than one method for the same measure/metric like $P(s,t)$, θ, and l_{stiwa}. Therefore, in the later stage especially during MaxSPRT analysis different variations of MaxSPRT can be designed by just choosing parameter values and more variations can be added without breaking any working and existing code/module.

C Implementation and code repository

In order to provide repeatable and expandable vaccination safety monitoring, we used the R Sequential package to create our sequential analysis architecture. Numerous techniques are supported by the implementation, such as binomial-type sequential testing, Poisson-based alpha spending approaches, MaxSPRT, and cMaxSPRT. Our method is intended to be schema-flexible and generalised, reducing rigid data dependencies and facilitating interoperability across various datasets and surveillance contexts.

Table X (see Figure) provides a summary of the SAEFVIC database design, which serves as the basis for the underlying data structure used for the sequential analysis. Our implementation only needs a limited number of fields to work

**Table 9** Demonstration of l_{stiwa} calculation and exposed person-time

VID	TTO	Week, t = 1	Week, t = 2	Week, t = 3
V1	5	$l_{sti(w=0)a} = 5$		
V2	4		$l_{sti(w=0)a} = 4$	
V3	12	$l = 1$	$l_{sti(w=1)a} = 7$	$l_{sti(w=2)a} = 4$
V4	22	$l_{sti(w=0)a} = 3$	$l_{sti(w=1)a} = 7$	$l_{sti(w=2)a} = 4$
V5	11	$l_{sti(w=0)a} = 2$	$l_{sti(w=1)a} = 7$	$l_{sti(w=2)a} = 2$
V6	10		$l_{sti(w=0)a} = 4$	$l_{sti(w=1)a} = 6$
V7	6		$l = 1$	$l_{sti(w=1)a} = 5$
V8	10			$l_{sti(w=0)a} = 4$
Weekly Sum l_{stiwa}		11	30	25

(Each week spans the days: Sat, Sun, Mon, Tue, Wed, Thu, Fri.)

Table 10 Demonstration of event counts and person-time calculation for concurrent Data assuming 7 days of risk window.

VID / Measure	TTO	Week, t = 1	Week, t = 2	Week, t = 3
V1	5			
V2	4			
V3	12			
V4	22			
V5	11			
V6	10			
V7	6			
V8	10			
Unvaxed events		4	1	0
Unvax Person-days		44	16	3
Risk events		1	2	1
Risk Person-days		12	28	13
Control events		0	2	4
Control Person-days		0	12	40

(Each week spans the days: Sat, Sun, Mon, Tue, Wed, Thu, Fri.)

Table 11 Demonstration of Different $p(s,t)$ Calculations

	For s = 10				Using Historical Data				Using current data		
					Idea 1	Idea 2	Idea 3		Idea 4		Idea 5
t	Expected Count (ER)	Accumulated Count (AER)	Observed count (OR) [up to S]	Accumulated count (AOR)	p(i) [proportion of data arriving at week i]	p(s, t) [sum of all p(i) from current to last week]	p(s, t) [AOR / AER]	p(s, t) [AER / Total ER]	p(i) [proportion of data arriving at week i]	p(s, t) [sum of all p(i) from current to last week]	Hazel's $p(s,t)$ expectation: OR/ER (1.B)
3	6	6	6	6	6/332 = 0.0181	1	6/6 = 1	6/332 = 0.02	6/265 = 0.02	1.00	6/6 = 1
4	25	31	24	30	25/332 = 0.075	0.98	30/31 =0.97	31/332 = 0.09	24/265 = 0.09	0.98	24/25 = 0.96
5	68	99	58	88	68/332 = 0.205	0.91	88/99 = 0.89	99/332 = 0.30	58/265 = 0.22	0.89	58/68 = 0.85
6	72	171	62	150	72/332 = 0.217	0.70	150/171 =0.88	171/332 = 0.52	62/265 = 0.23	0.67	62/72 = 0.86
7	88	259	78	228	88/332 = 0.265	0.49	228/259 = 0.88	259/332 = 0.78	78/265 = 0.29	0.43	78/88 = 0.89
8	21	280	19	247	21/332 = 0.063	0.22	247/280 = 0.88	280/332 = 0.84	19/265 = 0.07	0.14	19/21 = 0.91
9	25	305	9	256	25/332 = 0.075	0.16	256/305 = 0.84	305/332 = 0.92	9/265 = 0.03	0.07	9/25 = 0.36
10	27	332	9	265	27/332 = 0.081	0.08	265/332 = 0.8	332/332 = 1.00	9/265 = 0.03	0.03	9/27 = 0.33
Total	332		265								

properly, even if the whole schema includes extensive metadata such as vaccine identities, vaccination and reporting timestamps, demographic information, and thorough reaction descriptions. To be more precise, any dataset that contains the following fields is adequate for doing reliable sequential analysis:

(i) 'VacID' (unique report identifier), (ii) either 'Time of Vaccination' or a comparable timestamp (e.g., 'Vaxdate'), (iii) a reaction event date (from which Time to Onset (TTO) can be derived), and (iv) 'Reactions' (the event or condition of interest).

Because of this generalisation, the system can be used in a variety of ways, including when complete clinical or demographic data is not available, as is frequently the situation in post-market vaccine surveillance across jurisdictions. The code is modular, enabling customisation of event definitions, stratification, and filtering to meet various surveillance objectives. It is also flexible with respect to input formats—for instance, gender can be represented as Male/Female/Others, 0/1/2, or M/F/O, and if Date of Birth is provided instead of age, the system adjusts calculations accordingly.

Interested people can get the whole implementation in our public GitHub repository, which includes visual and textual output generation code, parameter configurations, and preprocessing scripts: (https://github.com/OurRepoName)

D Additional Vaiations of Sequential Analysis

Some additional potential variations that can be added are as follows:

1. Based on different ways of calculating the proportion of data complete, as shown in Table 5, there are five variations using the data available to us. More variations can be added based on the nature, characteristics, source, and purpose of the data.
2. Another prime source of variation in MaxSPRT is obtained through the proportion of data complete calculation based on the data source. Our options are using either concurrent data or historical data. We have added both in our current implementation.

3. Based on the calculation of the length of exposure. Currently, we use observed exposed person time. This exposed person-time can be calculated in a different and perhaps more reasonable way that adds more effectiveness in the calculation of the length of exposure and thus the expected number of events.

4. CDC has not used risk window slicing in (Center for Biologics Evaluation and Research, 2021) but Greene et al. used risk window slicing in (Greene et al, 2011). Hence, the risk window slicing can add another variation in the sequential analysis to the conditional MaxSPRT.

5. A new thread to the variations can be weighted average of expected event counts calculated using the formulas used in Definitions 8 (Appendix A, Terminology). This is only possible when baseline data is available. The efficacy of this idea needs extensive attention in terms of theoretical development and experimental analysis.