Research Article

Statistics in Medicine

Received 27 December 2014, Accepted 19 October 2015 Published online 11 November 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6805

Composite sequential Monte Carlo test for post-market vaccine safety surveillance

Ivair R. Silva[*†]

Group sequential hypothesis testing is now widely used to analyze prospective data. If Monte Carlo simulation is used to construct the signaling threshold, the challenge is how to manage the type I error probability for each one of the multiple tests without losing control on the overall significance level. This paper introduces a valid method for a true management of the alpha spending at each one of a sequence of Monte Carlo tests. The method also enables the use of a sequential simulation strategy for each Monte Carlo test, which is useful for saving computational execution time. Thus, the proposed procedure allows for sequential Monte Carlo test in sequential analysis, and this is the reason that it is called 'composite sequential' test. An upper bound for the potential power losses from the proposed method is deduced. The composite sequential design is illustrated through an application for post-market vaccine safety surveillance data. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: alpha spending; valid p-value; power loss

1. Introduction

Sequential statistical analysis is an important tool for detecting serious adverse events associated with a new commercialized drug but not detected during phase 3 of clinical trials [1–5]. In post-marked vaccine safety surveillance, the objective is to detect an increased risk for the occurrence of certain adverse events as early as possible. Sequential statistical analysis is an efficient way to repeatedly analyze the data as these accumulate along the time while keeping the type I error probability under the desired nominal significance level. Under the null hypothesis, if the drug/vaccine is safe, the relative risk associated with the monitored population is not greater than 1. In general, at each moment t_i of time previously scheduled, for $i = 1, \ldots, G$, the cumulative data are summarized through a test statistic. The test statistic at t_i, say U_{t_i}, is then used for drawing a decision about accepting/rejecting the null hypothesis.

For some inferential problems, the shape of the probability distribution of U_{t_i} is unknown. In such cases, conventional Monte Carlo testing can be conducted to assess the theoretical p-value. A well-known property of the conventional Monte Carlo test is that for a single test, the nominal significance level is analytically guaranteed [6]. But, this property does not hold for the overall significance level if multiple conventional Monte Carlo tests are performed for a sequence of nonindependent test statistics. Li and Kulldorff [7] handled this problem by defining a flat Monte Carlo threshold for monitoring the maximum log-likelihood ratio over the G moments of test. Although simple and intuitive, the method of Li and Kulldorff [7] does not favor choosing different type I error probabilities for each individual test.

Because the different chunks of data are usually correlated among each other, the true management of the type I error probability of each single Monte Carlo test is a difficult task. But, it is worth offering a solution for this problem because the type I error management is valorous for the sequential analysis practice. For example, if only 10 people enter in the monitoring system in the very first test, then the statistical power is probably small even for large true relative risks. Thus, if the control of the overall significance level is imperative, it is more prudent to use a small amount of type I error probability for

Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, U.S.A.
Department of Statistics, Federal University of Ouro Preto, Ouro Preto, MG, Brazil.
[*]*Correspondence to: Ivair R. Silva, Department of Statistics, Federal University of Ouro Preto, Campus Morro do Cruzeiro, CEP 35400 000, Ouro Preto, MG, Brasil.*
[†]*E-mail: ivairest@gmail.com*

the first test in order to save a larger probability amount for future tests. Because of the cumulative data, subsequent tests are based on larger sample sizes; therefore, they can produce a better statistical power, for the same amount of type I error probability, than a first test based on a small sample such as of 10 people. But if, instead of statistical power, expected time to signal is the most important performance measure in the analysis, then the analyst may prefer to spend a larger amount of type I error probability in the initial tests than in later tests.

The point is how to define the threshold of a sequential analysis when Monte Carlo simulation is used to calculate the p-value in each of the several tests. In this paper, a valid Monte Carlo p-value is introduced in order to perform sequential analysis through Monte Carlo simulation. The method enables an arbitrary choice for the amount of type I error probability to be used in each look at the data. Additionally, the procedure guarantees a true control on the power loss followed by the Monte Carlo approach. For the cases where the calculation of the test statistic is computationally intensive, a sequential Monte Carlo simulation design can be adopted in order to save execution time in each individual test. Thus, the method is called 'composite sequential' because it favors to apply a sequential Monte Carlo design at each test of a sequential analysis. All results are valid for the general case of any test statistic.

The content of this material is organized in the following way: Section 2.1 presents an overview of the sequential analysis theory immersed in the post-market vaccine and drug safety surveillance context. Section 2.2 offers a description for both conventional and sequential Monte Carlo tests. A new method for true management of the type I error probability in each Monte Carlo test is described in Section 3.1. The performance of the new method, in terms of statistical power, is demonstrated in Section 3.2. Section 4 introduces the composite sequential procedure. An example of application is offered in Section 5 for vaccine safety surveillance data of adverse events after Pediarix vaccination (GlaxoSmithKline, Philadelphia, PA, USA).

2. Overview of sequential analysis and Monte Carlo testing

This section presents some definitions and notations involving the sequential analysis field and the Monte Carlo test approach. The sequential analysis subject is described in the context of post-market vaccine safety surveillance. These contents are necessary in the construction of the methods introduced in Sections 3 and 4.

2.1. Sequential analysis for drug/vaccine safety surveillance

Sequential analysis can be performed by a continuous or a group sequential fashion. Let U_t be a non-negative stochastic process describing the number of adverse events appearing during the time interval $(0, t]$.

Definition 1 (Group sequential analysis)
For a set of constants $A_1, \ldots, A_G$ and a sequence $\{t_i\}_{i=1}^{G}$ of times, a group sequential analysis design is any procedure that rejects the null hypothesis if $U_{t_i} \geqslant A_i$ for some $i \in [1, \ldots, G]$.

Definition 2 (Continuous sequential analysis)
For a continuous real-valued function $B(t)$, a continuous sequential analysis design is any procedure that rejects the null hypothesis if $U_t \geqslant B(t)$ for some $0 < t < L$, where L is an arbitrary constant representing the maximum length of surveillance.

The thresholds A_i and $B(t)$ are established according to logistic/financial characteristics related to each specific application, and it may respond to some probabilistic requirements, such as a desired significance level, statistical power, and expected time to detect an elevated risk for the occurrence of an adverse event.

There is a number of group sequential methods currently available in the literature, but the two prominent methods, especially for clinical trials, are the Pocock test [8] and the O'Brien and Fleming test [9]. Although quite important for the sequential analysis field, these two methods are not described here in detail because, usually, they can be performed through analytical calculations, and then Monte Carlo is not needed.

For continuous sequential analysis, an important method is the maximized sequential probability ratio test (MaxSPRT). MaxSPRT was developed for the prospective rapid cycle vaccine safety surveillance and implemented by the Centers for Disease Control and Prevention-sponsored Vaccine Safety Datalink [10], and it has been in use for monitoring increased risks of adverse events in post-market safety surveillance [1, 2, 4, 5]. In the vaccine safety surveillance context, C_t is the random variable that counts the number

of adverse events in a known risk window from 1 to W days after a vaccination that was administered in a period $[0, t]$. Commonly, under the null hypothesis, C_t is supposed to have a Poisson distribution with mean μ_t, where μ_t is a known function of the population at risk, adjusted for age, gender, and any other covariates of interest. Under the alternative hypothesis, C_t is Poisson with mean $R\mu_t$, where R is the unknown increased relative risk due to the vaccine. The MaxSPRT statistic is given by $U_t = (\mu_t - c_t) + c_t \log(c_t/\mu_t)$, when $c_t \geqslant \mu_t$, and $U_t = 0$, otherwise.

Maximized sequential probability ratio test is feasible only if μ_t is known. In practice, μ_t is usually estimated through observational information from historical data sets. But, if μ_t is unknown, an alternative is to use the conditional MaxSPRT (cMaxSPRT) introduced by Li and Kulldorff [7]. Historical data have a determinant assignment in the cMaxSPRT. Consider the existence of two samples, one from a period previous to the surveillance beginning, called historical data, and a second sample, which by its turn is collected while the surveillance proceeds. Denote the total person time in the historical data by V, and use P_k to denote the cumulative person time until the arrival of the kth adverse event from the surveillance population. Given fixed observed values for c and k, assume that V follows a Gamma distribution with shape c and scale $1/\lambda_V$, and P_k Gamma-distributed with shape k and scale $1/\lambda_P$. Then, the cMaxSPRT test statistic, here denoted by U_k, is expressed as follows:

$$U_k^* = I\left(\frac{k}{c} > \frac{P_k}{V}\right)\left[c\log\frac{c(1+P_k/V)}{c+k} + k\log\frac{k(1+P_k/V)}{(P_k/V)(c+k)}\right]. \quad (1)$$

Let K denote the maximum number of observed events to interrupt the surveillance without rejecting the null. The joint probability distribution of $\tilde{U}_K = \left(U_1^*, U_2^*, \ldots, U_K^*\right)$ can be formally obtained, but its cumulative density function is fairly complicated, and the intricacy increases rapidly with K. Then, Li and Kulldorff [7] suggested the use of Monte Carlo simulation to find a flat critical value, $B(t) = CV$, for the test statistic $U = \max\left\{\tilde{U}_K\right\}$. Observe that the only random variable in U_k^* is the ratio P_k/V. Because P_k and V are Gamma distributed, the distribution of the ratio $\lambda_0 P_k/\lambda_0 V$ does not depend on the unknown parameter λ_0. Consequently, the cumulative distribution of U depends on the observed values c and k only. Thus, samples from U, for fixed (c, K), can be obtained by generating independent values from two Gamma distributions with shape c and scale 1, and with shape k and scale 1, respectively.

Maximized sequential probability ratio test and cMaxSPRT are based on a flat threshold in the scale of the log-likelihood ratio. But although simple and intuitive, the flat signaling threshold is not appropriate if one wants to ensure a rigorous control on the amount of alpha spending in each one of G group sequential tests.

An alternative way of defining the signaling threshold is based on the concept of 'alpha spending function' [11]. Basically, the alpha spending function dictates the amount of type I error probability that is to be used, or spent, at each one of the tests. Denote the single probability of rejecting the null hypothesis at the jth test by α_j, with $j = 1, 2, \ldots, G$, then $\sum_{j=1}^{G} \alpha_j \leqslant \alpha$, where α is the overall significance level. Jennison and Turnbull [11] offer a rich review of the existing proposals to choose the shape of the alpha spending function. The alpha spending concept favors for great flexibility when planning a surveillance, but this concept can be difficult to use with Monte Carlo p-values. This is the problem addressed in Section 3. Aiming a friendly presentation of the solution, an overview of Monte Carlo testing appears to be necessary.

2.2. *Monte Carlo testing*

Let U denote a real-valued test statistic. When the shape of the probability distribution of U is unknown but samples of U can be obtained by simulation under the null hypothesis, a Monte Carlo design can be used to calculate the p-value. The null hypothesis is rejected if the Monte Carlo p-value is smaller than or equal to α, where $\alpha \in (0, 1)$ is a desired significance level. Monte Carlo testing can be broadly categorized into two different types: the conventional (n-fixed) Monte Carlo test and the sequential (random n) Monte Carlo test. The n-fixed Monte Carlo test was proposed by Dwass [12], introduced by Barnard [13], and extended by Hope [14] and Birnbaum [15]. Let u_0 denote an observed value of the test statistic U for a realized data set. Also, use $U_1, U_2, \ldots, U_{n-1}$ to denote a sample of test statistics generated by Monte Carlo simulation under the null hypothesis, where n is a predefined and arbitrary integer. The conventional Monte Carlo p-value is then calculated in the following way: $P_n = (Y + 1)/n$, where $Y = \sum_{t=1}^{n-1} I_{\{U_t \geqslant u_0\}}(U_t)$. Thus, the null hypothesis, H_0, is to be rejected if $P_n \leqslant \alpha$. A well-known property of this n-fixed Monte Carlo test procedure is that it promotes a valid p-value, that is, $Pr(P_n \leqslant \alpha | H_0) \leqslant \alpha$ [6].

If the process of simulating the test statistic is time consuming, the option is to adopt a sequential Monte Carlo design. Unlike the conventional approach, in the sequential approach, the total number of simulations is no longer a fixed number n, but, instead, it is a random variable. With a sequential procedure, the simulations are interrupted as soon as an evidence about accepting/rejecting H_0 is identified, and then a considerable amount of time can be saved in the simulation process. An optimal design for performing sequential Monte Carlo tests was introduced by Silva and Assunção [16]. Their proposal, denoted here by MC_o, minimizes the expected number of simulations for any alternative hypothesis. MC_o also ensures arbitrary bounds for potential power losses in comparison with the conventional Monte Carlo test. The MC_o procedure plays a fundamental role in the construction of the composite approach introduced in Section 4; therefore, an overview of this optimal procedure is offered here. For this, define $Y_t = \sum_{j=1}^{t} I_{\{U_j \geqslant u_0\}}(U_j)$, $t = 1, \ldots, m$, where m is an arbitrary integer representing an upper bound for the number of simulations to be generated. The MC_o method requires the calculation of the upper and the lower sequential risks, defined respectively by

$$R_u(t) = \sum_{x=Y_t}^{x_0} \sum_{y=0}^{y_0} C_y^{m-1} C_x^t B(x+y+1, m-x-y+1), \text{and} \tag{2}$$

$$R_l(t) = \frac{Y_t}{t+1} - \sum_{x=0}^{Y_t-1} \sum_{y=0}^{y_0} C_y^{m-1} C_x^t B(x+y+1, m-x-y+1), \tag{3}$$

where $C_x^n = n!/[x!(x-x)!]$, $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, $x_0 = \min\{\lfloor \alpha(m+1) \rfloor - 1, t\}$, and $y_0 = \min\{\lfloor \alpha(m+1) \rfloor - 1 - x, m-t\}$. To ensure the same power than the n-fixed Monte Carlo testing, the maximum number of simulations, m, has to be equal to $(n-1)$. The simulation is interrupted as soon as the decision function ψ_t turns out to be equal to 1 or 2. At the tth simulation, that is, after having simulated $U_1, U_2, \ldots, U_t$, the null hypothesis is not rejected if $\psi_t = 1$; it is rejected if $\psi_t = 2$, and the simulation process proceeds while $\psi_t = 0$, where

$$\psi_t = \begin{cases} 0, \text{ if } R_u(t) > \delta_u, R_l(t) > \delta_l \text{ and } t < m, \\ 1, \text{ if } t > 20 \text{ and } R_u(t) \leqslant \delta_u \text{ or if } t = m, \\ 2, \text{ if } R_l(t) \leqslant \delta_l, \end{cases} \tag{4}$$

with $\delta_u = \epsilon/m$ and $\delta_l = 0.001/\lfloor \alpha(m+1) \rfloor$. This test criterion ensures that the power loss, with respect to the n-fixed Monte Carlo test, is not greater than $(\epsilon \times 100)\%$ [16]. Also, a valid p-value can be calculated for this sequential design. If t^* is the number of simulations at the interruption moment, a valid sequential p-value is

$$\hat{P} = \begin{cases} Y_{t^*}/t^*, & \text{if } \psi_{t^*} = 1, \\ (Y_{t^*}+1)/(t^*+1), & \text{if } \psi_{t^*} = 2. \end{cases} \tag{5}$$

The construction of a Monte Carlo design for group sequential analysis does not demand the use of sequential Monte Carlo designs like MC_o. In fact, the management of arbitrary type I error probabilities at each one of the G tests can be handled with the n-fixed Monte Carlo design, and indeed, this is the mechanism used for constructing the new method introduced in the next section. MC_o shall be necessary only for the composite sequential procedure introduced later with Section 4.

3. Conventional Monte Carlo test for group sequential analysis

Let G denote the maximum number of hypothesis tests scheduled for a group sequential analysis. This section introduces a valid design for an arbitrary management of the alpha spending at each of G n-fixed Monte Carlo tests.

3.1. A valid Monte Carlo alpha spending for sequential analysis

Define $\tilde{\mathbf{U}} = (U_1, \ldots, U_G)$, the G-dimensional vector of test statistics for sequential analysis, that is, the jth entry of this vector is a one-dimensional test statistic for the jth test, with $j = 1, \ldots, G$, and

$\tilde{\mathbf{u}} = (u_1, \ldots, u_G)$, an observed value of $\tilde{\mathbf{U}}$. Thus, let $\tilde{\mathbf{U}}_i = (U_{1,i}, \ldots, U_{G,i})$ denote a G-dimensional Monte Carlo copy of $\tilde{\mathbf{U}}$ generated under H_0, with $i = 1, \ldots, m$. Define the following Monte Carlo measures:

$$\begin{aligned} X_1 &= \sum_{t=1}^{m} I_{\{U_{1,t} \geqslant u_1\}}(U_{1,t}), \\ X_2 &= \sum_{t=1}^{m} I_{\{U_{1,t} \geqslant u_1, U_{2,t} \geqslant u_2\}}\left(U_{1,t}, U_{2,t}\right), \\ &\vdots \\ X_G &= \sum_{t=1}^{m} I_{\left\{\bigcap_{j=1}^{G} U_{j,t} \geqslant u_j\right\}}(U_{1,t}, \ldots, U_{G,t}). \end{aligned} \tag{6}$$

Use x_j to denote the observed value of X_j for a given simulated sequence of vectors $\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_m$. The exact conditional p-value, associated with the jth test and denoted by P_j, is given by

$$P_j = Pr\left(U_{j,t} \geqslant U_j \middle| \bigcap_{k=1}^{j-1} U_{k,i} \geqslant u_k \right), j = 2, \ldots, G, \text{ and } P_1 = Pr\left(U_{1,t} \geqslant u_1\right). \tag{7}$$

By construction, X_j follows a binomial distribution with parameters (x_{j-1}, p_j), where p_j is a realized value of P_j and $x_0 = m$. A valid conditional Monte Carlo p-value for the jth test, $\hat{P}_j$, can be calculated as follows:

$$\hat{P}_j = \begin{cases} (X_1 + 1)/(m + 1), & \text{for } j = 1, \\ (X_j + 1)/(x_{j-1} + 1), & \text{for } j = 2, \ldots, G. \end{cases} \tag{8}$$

The null hypothesis is to be rejected at the jth test if $\hat{P}_j \leqslant \theta_j$, where $\theta_j \in (0, 1)$ is a threshold fixed by the user according to the desired alpha spending, α_j, planned for test j. Note that in the particular case of a nonsequential analysis, that is, $G = 1$, $\hat{P}_1$ coincides with the conventional Monte Carlo p-value, and then it is sufficient to set $\theta_1 = \alpha$ for ensuring an actual alpha level test.

Observe that the event $\left\{\hat{P}_j \leqslant \theta_j\right\}$ occurs if, and only if, $X_j \leqslant \left\lfloor \theta_j(x_{j-1} + 1) - 1 \right\rfloor = h_j$; then h_j can be interpreted as a threshold in the scale of X_j, where $\lfloor a \rfloor$ is the greatest integer smaller than or equal to a, $a \in \mathbb{R}$. Then, for $G = 1$, $h_1 = h = \lfloor \alpha(m + 1) - 1 \rfloor$. If $G > 1$, the threshold h_j is to be chosen in a way to comply with the desired alpha spending, α_j.

Given a sequence of observed conditional p-values, $p_1, \ldots, p_G$, the overall probability of rejecting H_0, denoted here by $\pi(p_1, \ldots, p_G)$, can be expressed as follows:

$$\begin{aligned} \pi(p_1, \ldots, p_G) &= Pr\left(X_1 \leqslant h_1\right) + \sum_{j=2}^{G} Pr\left(\bigcap_{k=1}^{j-1} X_k > h_k, X_j \leqslant h_j \right) \\ &= Pr\left(X_1 \leqslant h_1\right) + \sum_{x_1 = h_1 + 1}^{m} Pr\left(X_2 \leqslant h_2 | X_1 = x_1\right) Pr\left(X_1 = x_1\right) + \\ &\quad + \cdots \sum_{x_1 = h_1 + 1}^{m} \sum_{x_2 = h_2 + 1}^{x_1} \cdots \sum_{x_G = h_{G-1} + 1}^{x_{G-1}} Pr\left(X_G \leqslant h_G\right) \prod_{j=1}^{G-1} Pr\left(X_j = x_j\right). \end{aligned} \tag{9}$$

The probability of rejecting H_0 can be rewritten as a sum of G power spending terms in the following way:

$$\pi(p_1, \ldots, p_G) = \sum_{j=1}^{G} \pi_j(p_1, \cdots, p_j). \tag{10}$$

The type I error probability associated with the jth test is calculated by integrating each one of the terms $\pi_j(p_1, \ldots, p_j)$ with respect to the null probability measure $F_{\tilde{P}_j}(\tilde{p}_j | H_0)$ of $\tilde{P}_j$ as follows:

$$\pi_{0,j} = \int_0^1 \int_0^1 \cdots \int_0^1 \pi_j\left(p_1, \ldots, p_j\right) F_{\tilde{P}_j}\left(\tilde{p}_j | H_0\right) d\tilde{p}_j. \tag{11}$$

Therefore, the overall type I error probability is equal to

$$\pi_0 = \sum_{j=1}^{G} \pi_{0,j}. \tag{12}$$

The distribution $F_{\tilde{P}_j}(\tilde{p}_j|H_0)$ is an important term for the calculation in expression (11). For $j = 2, \ldots, G$, define $u_j = F_{0,j}^{-1}\left(1 - p_j | \cap_{k=1}^{j-1} \{U_k \geqslant u_k\}\right)$, and make $u_1 = F_{0,1}^{-1}(1 - p_1)$, where $F_{0,j}$ is the probability distribution of U_j under the null hypothesis. Now, everything is prepared for the assertion and proof of the theorem later.

Theorem 3.1
For any test statistic $U_{j,i}$ directed to a group sequential analysis, let $P_1, \cdots, P_G$ be a sequence of exact conditional p-values that can be analytically expressed according to (7). Then, the joint probability distribution of $\tilde{P}_G = (P_1, \cdots, P_G)$, under the null hypothesis, is

$$F_{\tilde{P}}(\tilde{p}|H_0) = \prod_{j=1}^{G} p_j. \tag{13}$$

The theorem offers the general shape for the joint probability distribution of the theoretical vector of exact p-values under the null hypothesis. Note that there are no assumptions for the marginal distribution of the $U_j's$ or even for the joint distribution of $\tilde{\mathbf{U}}$.

Proof

$$\begin{aligned}
F_{\tilde{P}}(\tilde{p}|H_0) &= Pr\left[P_1 \leqslant p_1, \cdots, P_G \leqslant p_g|H_0\right] \\
&= Pr\left[P_1 \leqslant p_1|H_0\right) Pr\left(P_2 \leqslant p_2|H_0, P_1 \leqslant p_1\right] \cdots \\
&\cdots Pr\left[P_G \leqslant p_G|H_0, \cap_{j=1}^{G-1}\{P_j \leqslant p_j\}\right] \\
&= Pr\left[U_1 \geqslant F_{0,1}^{-1}(1 - p_1)\right] Pr\left[U_2 \geqslant F_{0,2}^{-1}\left(1 - p_2 \left|U_1 \geqslant u_1\right)\right| P_1 \leqslant p_1, H_0\right] \cdots \\
&\cdots Pr\left[U_G \geqslant F_{0,G}^{-1}\left(1 - p_G \left|\cap_{j=1}^{G-1}\{U_j \geqslant u_j\}\right)\right| \cap_{j=1}^{G-1} P_j \leqslant p_j\right] \\
&= \left[1 - F_{0,1}\left(F_{0,1}^{-1}(1 - p_1)\right)\right]\left[1 - F_{0,2}\left(F_{0,2}^{-1}(1 - p_2 \left|U_1 \geqslant u_1\right)\right| P_1 \leqslant p_1\right)\right] \cdots \\
&\cdots \left[1 - F_{0,G}\left(F_{0,G}^{-1}\left(1 - p_G \left|\cap_{j=1}^{G-1}\{U_j \geqslant u_j\}\right)\right)\right| \cap_{j=1}^{G-1} P_j \leqslant p_j\Bigg)\Bigg] \\
&= \prod_{j=1}^{G} p_j.
\end{aligned}$$

□

From expression (11) and using Theorem 3.1, the exact alpha spending at test j can be calculated using the following expression:

$$\pi_{0,j} = \sum_{x_1=h_1+1}^{m} \sum_{x_2=h_2+1}^{x_1} \cdots \sum_{x_{j-1}=h_{j-1}+1}^{x_{j-2}} \frac{(h_j + 1)}{\prod_{k=1}^{j}(x_{k-1} + 1)}, \tag{14}$$

where $x_0 = m$.

The tuning parameter θ_j is established such that $\pi_{0,j} \leqslant \alpha_j$. Recall that $h_j = \lfloor \theta_j (x_{j-1} + 1) - 1 \rfloor$; then $\pi_{0,j}$ is a nondecreasing function with respect to θ_j. Thus, a trivial numerical procedure, like the bisection method, will always give a feasible solution for θ_j given an arbitrary α_j. The idea is to find out the largest value of θ_j for which $\pi_{0,j} \leqslant \alpha_j$. Table I brings θ_j solutions for some common parameterization choices. The values in this table were obtained under a uniform alpha spending, that is, $\alpha_j = \alpha/G$, where α is the overall desired significance level. The solutions were possible through the bisection method and

Table I. Threshold settings to use the Monte Carlo alpha spending for global significance levels of $\alpha = 0.1, 0.05, 0.025, 0.01$, $m = 999$, and $G = 1, \ldots, 5$.

	G	θ_1	θ_2	θ_3	θ_4	θ_5
$\alpha = 0.1$	1	0.100000	na	na	na	na
	2	0.050000	0.055556	na	na	na
	3	0.033333	0.037200	0.041150	na	na
	4	0.025000	0.028241	0.032100	0.039999	na
	5	0.020000	0.022900	0.026600	0.033590	0.042000
$\alpha = 0.05$	1	0.050000	na	na	na	na
	2	0.025000	0.028241	na	na	na
	3	0.016666	0.019340	0.022880	na	na
	4	0.012500	0.014925	0.018182	0.023900	na
	5	0.010000	0.012269	0.015345	0.020500	0.033000
$\alpha = 0.025$	1	0.025000	na	na	na	na
	2	0.012500	0.014925	na	na	na
	3	0.008333	0.010485	0.013330	na	na
	4	0.006250	0.008260	0.010850	0.015150	na
	5	0.005000	0.006912	0.009259	0.013150	0.026300
$\alpha = 0.01$	1	0.010000	na	na	na	na
	2	0.00500	0.006912	na	na	na
	3	0.003333	0.004191	0.0051516	na	na
	4	0.002500	0.003144	0.004319	0.005309	na
	5	0.002000	0.003540	0.003572	0.009000	0.015000

Here, the alpha spending used for each group testing is $\alpha_j = \alpha/G$.

implemented in the R software [17]. This table was built for $m = 999$, and the solutions for larger m values will differ only slightly in a way that does not affect the significance level in each test. Thus, the values shown in this table can be used for larger values of m such as 9999 or 99,999.

The uniform alpha spending is a simple and intuitive choice for cases where the analyst has no practical concerns about the statistical performance of the analysis besides the global level of the sequential test. But, the Monte Carlo procedure introduced in this section is actually advantageous when the analyst indeed desires to establish a personalized, and maybe uncommon, alpha spending function. Many shapes can be designed for the alpha spending function given a maximum number G of tests and an overall α level, and it depends on each problem according to financial, logistical, or ethical issues. For example, suppose that for $G = 5$ and $\alpha = 0.05$, the desired alpha spending is to be managed in the following way: $\alpha_1 = 0.02$, $\alpha_2 = 0.015$, $\alpha_3 = 0.01$, $\alpha_4 = 0.003$, and $\alpha_5 = 0.002$. In this setting, the associated thresholds are $\theta_1 = 0.02$, $\theta_2 = 0.01767$, $\theta_3 = 0.015345$, $\theta_4 = 0.0096$, and $\theta_5 = 0.0085$.

3.2. *Statistical power performance*

The variability introduced by the Monte Carlo approach can lead to reductions in terms of statistical power with respect to the theoretical exact sequential test. This section develops an analytical expression to bound such potential power losses. This is possible by assuming a regular class of probability distributions for the exact conditional p-value.

3.2.1. A class of distributions for the conditional p-value. By definition, the exact conditional p-value, expression (7), is a random variable. In order to study the performance of the sequential Monte Carlo test, Fay and Follmann [18] introduced a class of p-value distributions with shape of the form

$$H_{\alpha,1-\beta}(p) = 1 - \Phi\left\{\Phi^{-1}(1-p) - \Phi^{-1}(1-\alpha) + \Phi^{-1}(\beta)\right\}, \tag{15}$$

where $\Phi(.)$ is the cumulative distribution function of a standard normal distribution and β is the type II error probability. The rationale behind the use of this class is that among the possible real scenarios where (15) holds, one can find out regular distributions families like the cases where the test statistic either follows the standard normal distribution under the null hypothesis and an $N(\mu, 1)$ under the alternative, or a central $\chi^{(2)}_1$ under the null hypothesis and a noncentral $\chi^{(2)}_1$ under the alternative hypothesis.

If the power function of a Monte Carlo test design is nonincreasing with p, the greatest power loss occurs when β is equal to 0.5, that is, the worst case is $H_{\alpha,0.5}(p)$. To see this, let $h(p)$ be a probability density function related to a member inside this class. Indexed by α and β, a particular member of this class is related to the following probability density function:

$$h_{\alpha,1-\beta}(p) = \exp\left\{-\frac{1}{2}\left[\Phi^{-1}(\beta) - \Phi^{-1}(1-\alpha)\right]\left[\Phi^{-1}(\beta) - \Phi^{-1}(1-\alpha) + 2\Phi^{-1}(1-p)\right]\right\}, \tag{16}$$

where Φ^{-1} is the inverse function of the standard normal cumulative distribution function $\Phi(\cdot)$. The first derivative of $h_{\alpha,1-\beta}(p)$ with respect to p is equal to

$$h'_{\alpha,1-\beta}(p) = \frac{\left[\Phi^{-1}(\beta) - \Phi^{-1}(1-\alpha)\right]}{\phi_Z(\Phi^{-1}(1-p))} h_{\alpha,1-\beta}(p), \tag{17}$$

where $\phi_Z(\cdot)$ is the density function of the standard normal distribution. For $1-\beta \geqslant \alpha$, we have $h'_{\alpha,1-\beta}(p) \leqslant 0$ for all $p \in (0,1)$. If, for a fixed p, the Monte Carlo test is decreasing with p, then the worst case probability distribution $H_{\alpha,1-\beta}(p)$ is the one that maximizes $h_{\alpha,1-\beta}(p)$ for $p = \alpha$; that is, the evaluation of the worse is made by finding out the β value that maximizes $h_{\alpha,1-\beta}(\alpha) = \exp\left\{-0.5\left[\left(\Phi^{-1}(\beta)\right)^2 - \left(\Phi^{-1}(1-\alpha)\right)^2\right]\right\}$ with respect to β. Observe that because of the symmetry property of the normal distribution, the minimum value of $\left(\Phi^{-1}(\beta)\right)^2$ occurs for $\beta = 0.5$ because $\Phi^{-1}(0.5) = 0$; then, for any $\alpha \leqslant 1-\beta$, the analytical solution that maximizes $h_{\alpha,1-\beta}(\alpha)$, with respect to β, is $\beta = 0.5$.

3.2.2. Upper bounds for potential power losses. Assume that the conditional probability distribution of P_k, given $P_{k-1}, \ldots, P_1$, represented by $F_{P_{k^-}}$, is shaped according to (15). This assumption is adequate, for example, for the cases where U_j is a Markovian stochastic process, that is, the conditional distribution of U_j, given the pass, only depends on the observed statistic of the immediately predecessor test. Under these terms, it seems to be reasonable to assume that the same family of distribution is adequate for each of the conditional distributions.

The function $\pi_k(p_1, \ldots, p_k)$, $k \leqslant G$, is decreasing with p_k and, for $j < k$, $\pi_k(p_1, \ldots, p_k)$ is increasing with p_j. Recall that if the Monte Carlo power function for the kth test is decreasing with p_k, then the worst distribution scenario of the type (15) occurs for $\beta = 0.5$. If $(1-\beta) \geqslant \alpha$, because $\pi_k(p_1, \ldots, p_k)$ is increasing with p_j ($j < k$) and, from expression (17), $h_{\alpha,1-\beta}(p)$ is nonincreasing with p, the worst case in the kth test for $F_{P_{j^-}}$, with $j < k$, is the uniform(0,1) distribution. But, the analytical manipulation of the Monte Carlo power by using $H_{\alpha,1-\beta}(p)$ is intractable. To circumvent this problem, and following the suggestion from Fay and Follmann [18], a $Beta(a,b)$ distribution can be used to approximate $H_{\alpha,1-\beta}(p)$. Denote this approximating function by $\tilde{H}_{\alpha;0.5}(p)$. This approximation is chosen in a way that the expected value of p coincides with that from $H_{\alpha,1-\beta}(p)$, and $\tilde{H}_{\alpha,1-\beta}(\alpha) = H_{\alpha,1-\beta}(\alpha) = 1-\beta$. Thus, a lower bound, π_k^*, for the power of the kth Monte Carlo test, is given by

$$\pi_k^* = \sum_{x_1=h_1+1}^{m} \sum_{x_2=h_2+1}^{x_1} \cdots \sum_{x_{k-1}=h_{k-1}+1}^{x_{j-2}} \sum_{x_k=0}^{h_k} \frac{C_{x_k}^{x_{k-1}} B(a+x_k, b+x_{k-1}-x_k)}{B(a,b)\prod_{j=1}^{k-1}\left(x_j+1\right)}. \tag{18}$$

For evaluating the power difference between the Monte Carlo and the exact approaches, it is helpful to obtain the expression for the exact power of the kth test, denoted here by $\beta_{A,k}$, and which can be expressed by

$$\beta_{A,k} = \int_{\phi_1}^{1}\int_{\phi_2}^{1}\cdots\int_{\phi_{k-1}}^{1}\int_{0}^{\phi_k} \tilde{H}_{\alpha;0.5}(p)dp = F_H(\phi_k|a,b)\prod_{j=0}^{k-1}(1-\phi_j), \tag{19}$$

where $\phi_0 = 0$, $F_H(\phi_k|a,b)$ is the $Beta(a,b)$ distribution evaluated at ϕ_k, and $\phi_1, \ldots, \phi_k$ are chosen in a way to ensure the desired alpha spending for the kth test. Thus, from (18) and (19), an upper bound for the power loss of the proposed Monte Carlo design is as follows:

$$\max_k \left\{\beta_{A,k} - \pi_k^*\right\}. \tag{20}$$

Let $\beta_{0,k}$ be the type I error probability in the kth test, which is obtained by integrating the multivariate uniform $(0,1)^k$ density, in the rejection region, for the kth test. For $k = 1$ and for an alpha spending of α_1, the value of ϕ_1 is α_1. For $k = 2$,

$$\beta_{0,2} = \int_{\phi_1}^{1} \int_{0}^{\phi_2} dp_2 dp_1 = \phi_2(1-\phi_1) = \phi_2(1-\alpha_1),$$
$$\Rightarrow \text{for an alpha spending of } \alpha_2, \text{ we have } \phi_2 = \alpha_2/(1-\alpha_1).$$

By recursive application of the previous reasoning, the alpha spending at the kth test, for $k = 1, \cdots, G$, is given by

$$\beta_{0,k} = \int_{\phi_1}^{1} \int_{\phi_2}^{1} \cdots \int_{\phi_{k-1}}^{1} \int_{0}^{\phi_k} dp_k \cdots dp_1 = \phi_k \prod_{i=0}^{k-1} (1-\phi_i). \tag{21}$$

Then, if α_k is the alpha spending in the kth test, the value of ϕ_k is explicitly obtained:

$$\phi_k = \frac{\alpha_k}{\left(1 - \sum_{i=0}^{k-1} \alpha_{i-1}\right)}, \tag{22}$$

with $\alpha_0 = 0$.

For the special case of a constant alpha spending, that is, $\alpha_j = \alpha/G = \phi$, and from expression (21), the tuning parameters ϕ_j are given by

$$\phi_j = \phi - (j-1)\phi^2, \text{ for } j = 1, \cdots, G. \tag{23}$$

Each combination of α and G leads to a particular ϕ value, which, by its turn, leads to particular values of a and b to be used in the approximating distribution $\tilde{H}_{\alpha,1-\beta}(p)$. For example, for $\alpha = 0.05$ and $G = 1$, the associated solution is $a = 0.359$ and $b = 2.523$, which gives an upper bound for the power loss equal to 0.0187. For $G = 5$ and $\alpha_j = \alpha/5$, the solution is $a = 0.3$ and $b = 7.6$, which gives a power loss upper bound of 0.018.

If U_j is computationally hard to obtain, a sequential strategy can be adopted in order to save execution time in each single Monte Carlo test. Complicated calculation as, for example, those involved in lengthy inferential procedures usually leads to the situation where the test statistic distribution is unknown and Monte Carlo is required. For example, the multivariate problem of finding spatial clusters of disease outbreak requires the use of computationally intensive testing procedures, like the flexibly shaped scan statistic proposed by Tango and Takahashi [19]. The execution time of the flexibly shaped scan can take as long as 2 weeks depending on the application. Another example is the tree-based scan statistic for database disease surveillance proposed by Kulldorff *et al.* [20], which allows for the study of relation among many databases related to a sort of potential risk factors, and then it can take a long time to run. Computational speed is an important issue even today after the many advances in processing speed of computational resources. The big data analysis is an example of why saving execution time in numerical intensive procedures is still important. The next section is devoted to construct a method for performing sequential Monte Carlo tests in sequential analysis. The method favors to save execution time in computational intensive problems that might appear in practice.

4. On a composite approach: sequential Monte Carlo for sequential analysis

The composite sequential procedure introduced in this section is indicated for applications where computational time is a relevant issue. As already stressed earlier, the term 'composite' was chosen to emphasize that the method is a sequential procedure because of two different reasons: for each test, among a sequence of G tests, a sequential Monte Carlo test is considered. There is a lack in the literature concerning a comprehensive method of performing sequential Monte Carlo tests in the sequential analysis context. This composite scenario has already been treated by Jennison [21], but his proposal was concerned only on saving execution time in a single test. The problem of how to ensure the overall nominal significance level after multiple sequential Monte Carlo tests remains an open question. This section introduces a procedure to address this challenge. For the kth sequential test, define

$$X_{t,k} = \sum_{i=1}^{t} I_{\left\{\bigcap_{k=1}^{k} U_{k,i} \geqslant u_k\right\}}(U_{1,i}, \cdots, U_{k,i}), \tag{24}$$

which is the random variable to be tracked in the kth test after the tth simulation. Now, the concept of sequential risk introduced in Section 2.2 is extended for the sequential analysis context. For $c_{0,t} = \min\left\{\left\lfloor\theta_k(x_{t,k-1}+1)\right\rfloor - 1, t\right\}$ and $y_{0,t} = \min\left\{\left\lfloor\theta_k(x_{t,k-1}+1)\right\rfloor - 1 - x, x_{t,k-1} - t\right\}$, the upper and the lower conditional sequential risks are defined by

$$CR_u(t,k) = \sum_{x=x_{t,k}}^{c_{0,t}} \sum_{y=0}^{y_{0,t}} C_y^{x_{t,k-1}} C_x^t B(x+y+1, x_{t,k-1} - x - y + 1), \text{ and} \tag{25}$$

$$CR_l(t,k) = \frac{x_{t,k}}{t+1} - \sum_{x=0}^{x_{t,k}-1} \sum_{y=0}^{y_{0,t}} C_y^{x_{t,k-1}-1} C_x^t B(x+y+1, x_{t,k-1} - x - y + 1), \tag{26}$$

where $x_{t,k-1}$ is the observed value of $X_{t,k-1}$ for $k = 1, \cdots, G$, with $x_{t,0} = t$. The optimal MC_o procedure described in Section 2.2 can be applied in order to save execution time in the kth test. The power loss from such approach is not greater than 1%. The Monte Carlo simulation for the kth test is interrupted as soon as the decision function, $\psi_{t,k}$, is equal to 1 or 2. The null hypothesis is not rejected if $\psi_{t,k} = 1$, and H_0 is rejected if $\psi_{t,k} = 2$, where

$$\psi_{t,k} = \begin{cases} 0, \text{ if } CR_u(t,k) > \delta_u, CR_l(t,k) > \delta_{l,k} \text{ and } t < m, \\ 1, \text{ if } t > 20 \text{ and } CR_u(t,k) \leqslant \delta_u, \\ 2, \text{ if } CR_l(t,k) \leqslant \delta_{l,k} \text{ or if } t = m, \end{cases} \tag{27}$$

with $\delta_u = 0.01/m$ and $\delta_{l,k} = 0.001/\left\lfloor\theta_k(m+1)\right\rfloor$. The threshold θ_k is chosen in a way to ensure the desired alpha spending, α_k. If t^* is the number of simulations at the interruption moment, a valid conditional sequential p-value, for the kth sequential test, is given by

$$\bar{P}_k = \begin{cases} X_{t^*,k}/t^*, & \text{if } \psi_{t^*,k} = 1, \\ \left(X_{t^*,k}+1\right)/(t^*+1), & \text{if } \psi_{t^*,k} = 2. \end{cases} \tag{28}$$

5. Monitoring adverse events after vaccination

The composite sequential method is illustrated here through an application for real data. This application is meant to mimic a post-market vaccine safety surveillance analysis, and the test statistic adopted is the cMaxSPRT statistic.

5.1. Data description

This illustrative example uses a time series data of health insurance claims from the Centers for Disease Control and Prevention-sponsored Vaccine Safety Datalink project. The series has 82 weekly entries, where each entry counts the number of adverse events associated with neurological symptoms within 28 days after Pediarix vaccination, which is manufactured by GlaxoSmithKline. This vaccine, with a simple injection, protects children from five different diseases: diphtheria, tetanus, whooping cough, hepatitis B, and polio. Figure 1(A) shows the time series formed by these 82 observed counts of adverse events indexed by week.

Recall that the cMaxSPRT requires the existence of a historical data. Once the objective here is not actually to analyze this data set but to use it as a realistic scenario to illustrate the applicability of the composite sequential approach, the data set was divided in two subsamples, each with 41 entries, where the first batch of 41 observations is taking as the historical sample and the second subsample as the surveillance sample. The surveillance sample is represented with the dark down triangles shown in Figure 1(A). Because these both samples came from the same population under the same vaccination, it is natural to expect that a test for $H_0 : \lambda_V = \lambda_P$ against $H_1 : \lambda_V < \lambda_P$ may not reject the null. We use this controlled application as a motivation to verify if the data presented some change with respect to the relative risk

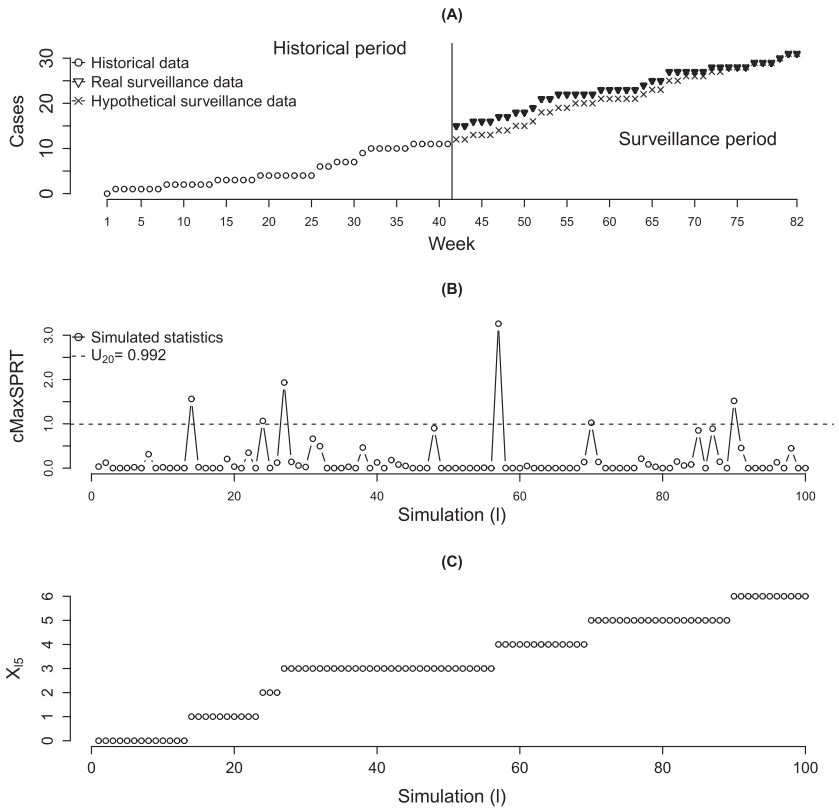

Figure 1. Real and hypothetical scenarios for the observed cMaxSPRT statistics at the fifth sequential test (A). The first hundred simulated values under the null (B) and the $X_{l,5}$ trajectory (C) are also available.

of having an adverse event during the weeks. An abrupt change in the cumulative cases occurred around 42nd week, and then it can be of interest to test if the relative risk after that point is increased.

5.2. *Tuning parameters settings*

The assumption of a Gamma distribution for P_k implies an exponential distribution, with mean λ_P, for each arrival time of the k independent adverse events. Because the distribution of the ratio P_k/V, under H_0, does not depend on λ_P, the tth Monte Carlo value from the null distribution of U_k can be obtained by generating k independent values from an exponential distribution with parameter $1/\lambda_P = 1$. The sum of such generated values represents the tth Monte Carlo cumulative person time ($P_{t,k}$) for the surveillance period. Similarly, the sum of c independent values from an exponential distribution, with parameter $1/\lambda_V = 1$, represents the tth Monte Carlo cumulative person time (V_t) for the historical data. The values $P_{t,k}$ and V_t are then used to calculate the tth Monte Carlo copy ($U_{t,k}$) of the cMaxSPRT statistic.

The analysis considers a scenario with a maximum of five tests ($G = 5$) and an overall significance level of 0.05 ($\alpha = 0.05$). To further explore the flexibility offered by the Monte Carlo method in which concerns the management of the type I error, instead of a simple uniform function, here, the non-naive alpha spending of the end of Section 3.1 is considered: $\alpha_1 = 0.02$, $\alpha_2 = 0.015$, $\alpha_3 = 0.01$, $\alpha_4 = 0.003$, and $\alpha_5 = 0.002$, which leads to the thresholds $\theta_1 = 0.02$, $\theta_2 = 0.01767$, $\theta_3 = 0.015345$, $\theta_4 = 0.0096$, and $\theta_5 = 0.0085$. Each test presents less than 1% of power loss; then the overall power loss is bounded at 5%. For this application, $m = 100,000$ and the threshold parameters are fixed according to Table (I). For the group sequential design, assume that a test is to be performed after each chunk of four adverse events, which means that the maximum length of surveillance is $K = 20$.

5.3. *Data analysis results*

In the historical data, it was observed a total of $c = 11$ adverse events accumulated after $v = 5.3415$ (observed V) person-time amount. Observed values of P_k and U_{0,k_j}, $j = 1, \ldots, 5$, are presented in the

Table II. Part I of this table gives the real and hypothetical scenarios for the observed cMaxSPRT statistics at each of the five sequential tests; part II shows the three first simulated cMaxSPRT values, under H_0, for the sequential analysis over the hypothetical data.

Part I – test statistic information				Part II – simulation photograph				
Test (j)	k_j	P_{k_j}	U_{0,k_j}	$t =$	1	2	3	…
	Real scenario			V_t	11.214	11.031	8.772	…
1	4	0.183	5.297					
2	8	1.507	1.914	$P_{t,4}$	4.597	4.265	5.793	…
3	12	2.638	1.776	$P_{t,8}$	8.483	7.765	7.665	…
4	16	3.766	1.750	$P_{t,12}$	12.622	10.246	13.308	…
5	20	5.942	0.890	$P_{t,16}$	14.337	14.321	14.903	…
	Hypothetical scenario			$P_{l.20}$	18.494	16.690	19.418	…
1	4	1.195	0.319	$U_{t,4}$	0.000	0.000	0.099	…
2	8	1.935	1.065	$U_{t,8}$	0.000	0.002	0.000	…
3	12	3.494	0.748	$U_{t,12}$	0.000	0.074	0.000	…
4	16	4.486	1.005	$U_{t,16}$	0.055	0.042	0.000	…
5	20	5.784	0.992	$U_{t,20}$	0.034	0.122	0.000	…

The historical information used here are $c = 11$ and $v = 5.3415$.

upper block (real scenario) of Table II, part I. The first four adverse events concentrated in a short person-time amount, equal to 0.183. This high frequency is reflected through a large observed value of $U_{0,4} = 5.297$, which is the observed test statistic for the first group sequential test ($j = 1$). This result indicates that there may exist an inflation in the relative risk at some point after the last cases observed in the historical period.

The first group sequential test ($j = 1$) is based on the Monte Carlo simulations $U_{1,4}$, $U_{2,4}$, and so on. For the calculations shown in this example, the simulations were possible by generating random numbers from an exponential distribution through the 'rexp' function of the R software. The simulated values were then combined in order to form the $U_{t,4}$ according to the instructions given in the first paragraph of Section 5.2. The MC_o signaling thresholds, δ_u and $\delta_{l,4}$, were not touched before $t = 99,999$ in this first test, and the observed p-value was equal to $566/100,000 = 0.00566$ ($x_{99999,4} = 565$). Thus, because the observed p-value is smaller than the threshold $\theta_1 = 0.02$, H_0 is to be rejected; that is, for a significance level of 0.02, we should conclude that λ_P is greater than λ_V.

From a practical perspective, the analyst should stop the sequential surveillance and conclude that H_0 is to be rejected. However, in order to further explore this application, fake data showing a different line shape are considered for the surveillance period. The trajectory of this new series of cumulative cases is represented with the symbol '×' in Figure 1(A) in the surveillance period. Note that the jumping after week 41 is not present in this fake series. The summary for the observed statistics from these new data is offered in the lower block (hypothetical scenario) of Table II, part I. To help with the understanding behind each step in this composite sequential analysis, Table II also presents, in part II, the first three Monte Carlo cumulative person-time values for both historical (V_t) and surveillance (P_{t,k_j}) periods, as well as the related U_{t,k_j}. This information is used to obtain $X_{t,j}$. For example, with $j = 1$, the $X_{t,1}$ value is obtained by comparing the line labeled as $U_{t,4}$ with the first value, 0.319, of the Hypothetical scenario of Table II.

For the first test ($k_1 = 4$), $CR_u(t,k)$ turned out smaller than $\delta_u = 0.01/99,999 = 10^{-7}$ just after the 20-s simulated statistic, which returned $x_{22,1} = 6$, leading to a p-value of $\bar{p}_1 = x_{22,1}/t^* = 6/22 \approx 0.273$. Thus, H_0 was not rejected at the first group sequential test. For the second test ($k_2 = 8$), $CR_u(t,k)$ turned out smaller than 10^{-7} after the ninth simulated statistic with $x_{31,2} = 4$. The observed p-value was $\bar{p}_2 = (x_{31,2})/(x_{22,1} + 9) = 4/(15) =\approx 0.267$. Tests number three ($k_1 = 12$), four ($k_1 = 16$), and five ($k_1 = 20$) also crossed the threshold ($\delta_u = 10^{-7}$) before the simulation of number 99,999. Then, H_0 is not rejected with these fake data. Figure 1(B) shows the trajectory for the first hundred simulated values $U_{1,20}, U_{2,20}, \ldots, U_{100,20}$. These values are used in the fifth sequential test ($j = 5$). Figure 1(C) presents the values of $X_{t,5} = \sum_{i=1}^{t} I(U_{t,20} \geqslant 0.992)$ while it evolves with t.

6. Conclusions

Monte Carlo simulation was demonstrated to be a feasible approach for sequential analysis. The true control of the type I error probability associated with a sequence of tests is guaranteed. More importantly, such control is valid for any statistical dependence structure among the distributions of the test statistics calculated in each look at the data. Although inspired by the post-market vaccine safety surveillance problem, the method can be directly applied for sequential analysis of any other sort of problem. The real application shown in this paper is not an example of computationally intensive problem. The evaluation of the real gains with the composite design, in terms of expected signal time, should be explored for time-consuming statistical tools, such as the tree-based scan statistic for database disease surveillance. This is a lack left for future studies.

Acknowledgements

This research was funded by the National Institute of General Medical Sciences, USA, grant #RO1GM108999. Additional support was provided by Fundação de Amparo à Pesquisa do Estado de Minas Gerais, Minas Gerais, Brasil (FAPEMIG).

References

1. Lieu T, Kulldorff M, Davis R, Lewis E, Weintraub E, Yih W, Yin R, Brown J, Platt R. Real-time vaccine safety surveillance for the early detection of adverse events. *Medical Care* 2007; **45**(S):89–95.
2. Klein NP, Fireman B, Yih WK, Lewis E, Kulldorff M, Ray P, Baxter R, Hambidge S, Nordin J, Naleway A, Belongia EA, Lieu T, Baggs J, Weintraub E, Vaccine Safety Datalink. Measles-mumps-rubella-varicella combination vaccine and risk of febrile seizures. *Pediatrics* 2010; **126**:e1–e8.
3. Avery T, Vilk Y, Kulldorff M, Li L, Cheetham T, Dublin. S. Prospective, active surveillance for medication safety in population-based health networks: a pilot study. *Pharmacoepidemiol Drug Saf* 2010; **19**:S304.
4. Belongia EA, Irving SA, Shui IM, Kulldorff M, Lewis E, Yin R, Lieu TA, Weintraub E, Yih WK, Li R, Baggs J, Vaccine Safety Datalink Investigation Group. Real-time surveillance to assess risk of intussusception and other adverse events after pentavalent, bovine-derived rotavirus vaccine. *Pediatric Infectious Disease Journal* 2010; **29**:1–5.
5. Yih W, Nordin J, Kulldorff M, Lewisc E, Lieua T, Shia P, Weintraube E. An assessment of the Safety Datalink of adolescent and adult tetanus-diphtheria-acellular pertussis (TDAP) vaccine, using active surveillance for adverse events in the Vaccine Safety Datalink. *Vaccine* 2009; **27**:4257–4262.
6. Silva I, Assunção R, Costa M. Power of the sequential Monte Carlo test. *Sequential Analysis* 2009; **28**(2):163–174.
7. Li L, Kulldorff M. A conditional maximized sequential probability ratio test for pharmacovigilance. *Statistics in Medicine* 2010; **29**:284–295.
8. Pocock S. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977; **64**:191–199.
9. O'Brien P, Fleming T. A multiple testing procedure for clinical trials. *Biometrics* 1979; **35**:549–556.
10. Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Sequential Analysis* 2011; **30**:58–78.
11. Jennison V, Turnbull B. *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall/CRC: London, 1999. ISBN 0-8493-0316-8.
12. Dwass M. Modified randomization tests for nonparametric hypotheses. *Annals of Mathematical Statistics* 1957; **28**: 181–187.
13. Barnard GA. Contribution to the discussion of professor Bartlett's paper. *Journal of the Royal Statistical Society* 1963; **25B**:294.
14. Hope A. A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society* 1968; **30B**:582–598.
15. Birnbaum ZW. Computers and unconventional test statistics. In *Reliability and Biometry: Statistical Analysis of Lifelength*, 5th ed., Proschan F, Serfling RJ (eds). Soc. Indust. Appl. Math.: Philadelphia, 1974; 441–458.
16. Silva I, Assunção R. Optimal generalized truncated sequential Monte Carlo test. *Journal of Multivariate Analysis* 2013; **121**:33–49.
17. Team RC. R: a language and environment for statistical computing. *R Foundation for Statistical Computing* Vienna, Austria 2014. http://www.R-project.org [Accesses on 18 November 2014].
18. Fay M, Follmann D. Designing Monte Carlo implementations of permutation or bootstrap hypothesis tests. *The American Statistician* 2002; **56**(1):63–70.
19. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics* 2005; **4**:4–11.
20. Kulldorff M, Fang Z, Walsh S. A tree-based scan statistic for database disease surveillance. *Biometrics* 2003; **59**:323–331.
21. Jennison C. Bootstrap tests and confidence intervals for a hazard ratio when the number of observed failures is small, with applications to group sequential survival studies. *Computing Science and Statistics* 1992; **22**:89–97.