



Published in final edited form as:

Methodol Comput Appl Probab. 2018 June ; 20(2): 739–750. doi:10.1007/s11009-017-9586-z.

Type I Error Probability Spending for Post-Market Drug and Vaccine Safety Surveillance With Poisson Data

Ivair R. Silva¹

¹Department of Statistics, Federal University of Ouro Preto, Ouro Preto, Minas Gerais, Brazil

Abstract

Statistical sequential hypothesis testing is meant to analyze cumulative data accruing in time. The methods can be divided in two types, group and continuous sequential approaches, and a question that arises is if one approach suppresses the other in some sense. For Poisson stochastic processes, we prove that continuous sequential analysis is uniformly better than group sequential under a comprehensive class of statistical performance measures. Hence, optimal solutions are in the class of continuous designs. This paper also offers a pioneer study that compares classical Type I error spending functions in terms of expected number of events to signal. This was done for a number of tuning parameters scenarios. The results indicate that a log-exp shape for the Type I error spending function is the best choice in most of the evaluated scenarios.

Keywords

Sequential probability ratio test; Expected number of events to signal; Log-exp alpha spending

1 Introduction

In statistical inference, the classical non-sequential approach for hypothesis testing consists on defining the test criterion as a function of an unique batch of data, and hence only one test is performed. Differently, in sequential hypothesis testing multiple tests are performed sequentially, each based on information formed by a combination of new data with data observed in previous tests. Although the historic development of sequential analysis theory can be undoubtedly credited to challenges appearing in clinical trials (Jennison and Turnbull 2000), applications for post-market drug and vaccine safety surveillance of adverse events is now in the fringe of the sequential analysis practice (Davis et al. 2005; Lieu et al. 2007; Kulldorff et al. 2011; Yih et al. 2011; Leite et al. 2016).

In the context of post-market drug and vaccine safety surveillance, a very useful model is the Poisson distribution. Following the construction used by Leite et al. (2016), Silva and Kulldorff (2015), Kulldorff et al. (2011), and Yih et al. (2011), let X_t denote the number of adverse events observed within the $(0, t]$ time interval after a vaccination, where X_t follows a Poisson distribution with average $(R \times \mu_t)$. In this construction, μ_t is a known function of the

ivairrest@gmail.com.

Electronic supplementary material The online version of this article

exposed population, and R is the relative risk reflecting the effect of the vaccine in the risk of occurring an adverse event. Basically, under the null hypothesis, the vaccine is safe, hence the relative risk is smaller than or equal to 1. Otherwise, the expected number of events is greater than μ_b , i.e. $R > 1$. This problem motivates the discussion in the present paper, which is directed to one tailed-tests with hypothesis of the form:

$$H_0: R \leq 1 \text{ against } H_1: R > 1. \quad (1)$$

Usually, sequential methods for testing the hypotheses in Eq. 1 are based on monitoring a real valued test statistic, say $W(X_t)$, $t \in (0, T]$, in comparison to a signaling threshold CV_t . The null hypothesis is rejected for the first t such that $W(X_t) \geq CV_t$, or the null is taken as true if $W(X_t) < CV_t$ for each $t \in (0, T]$. The tuning parameter T , called maximum length of surveillance, is an arbitrary constant defined by the investigator. The time dependent function CV_t is specified in a way that the overall Type I error probability is safely smaller than or equal to a desired significance level, α . This type of threshold-based testing design is in the essence of established methodologies, such as e.g. Wald's SPRT test (Wald 1945), the method of Pocock (1977), the O'Brien & Fleming's statistic (O'Brien and Fleming 1979), the maximized sequential probability ratio test (MaxSPRT) introduced by Kulldorff et al. (2011), and the modified MaxSPRT introduced by Gombay and Li (2015).

An important design issue is the choice between continuous or group sequential fashions. With the continuous design, each $t \in (0, T]$ can lead to the null rejection. Unlike, under a group sequential design, only selected points $t_1 < t_2 < \dots < t_G$ are taken as testing times. Naturally, the choice between continuous versus group designs depends on logistical/ financial aspects of data collection since continuous analysis can only be applied if it is feasible to observe the arrival time of each single new observation. But, in situations where the choice between continuous and group approaches is feasible prior to start the surveillance, a question that emerges is if one should be adopted in detriment of the other. Based on simulation studies, the works of Zhao et al. (2012) and Nelson et al. (2012) seemed to suggest that group sequential designs should be preferred even when data is extracted continuously. But, the work of Silva and Kulldorff (2013) put a light in the subject. They used analytical arguments to show that, if expected number of events to signal and expected length of surveillance are the design criterion for fixed power and alpha level, then continuous designs should always be used in place of group designs if logistically/ financially possible. It is worth noting that similar evaluations for other statistical performance measures, such as the risk function, are still pending in the literature. In order to cover this lack, the present paper shows that continuous sequential analysis is also uniformly better than group sequential in terms of expected loss (risk function), expected number of events to signal, and expected number of events.

Another important design issue is the shape of CV_t . While some classical methods establish flat thresholds in the scale of $W(X_t)$ (Wald 1945; Pocock 1977; Kulldorff et al. 2011), others are defined under increasing or decreasing CV_t functions (O'Brien and Fleming 1979; Gombay and Li 2015). But, all these methods can be rewritten in terms of a more comprehensive approach, the approach based on the Type I error probability spending function. Type I error spending functions, usually denoted by $S(t)$, are meant to dictate the

amount of Type I error probability to be applied for each $t \in (0, T]$. A traditional choice is the so-called power-type function:

$$S(t) = \alpha \times \left(\frac{t}{T}\right)^\rho, \quad \rho > 0, \quad t \in (0, T]. \quad (2)$$

A number of Type I error spending functions are available in the literature. The comparison among many $S(t)$ shapes has historically been made having expected sample size as the statistical performance measure of interest (Lan and DeMets 1983; Kim and Demets 1987). Unfortunately, the comparison of $S(t)$ shapes in terms of expected number of events to signal (instead of expected sample size) has not been explored for the Poisson data yet. This paper brings a comparison study among four established Type I error spending functions. This is done having expected number of events to signal as the statistical performance measure of interest. Our results show that the log-exp function introduced by Lan and DeMets (1983) seems to be the best one.

This material is organized in the following way: the next section extends the results of Silva and Kulldorff (2013) about the optimality of continuous sequential analysis over group sequential analysis. Section 3 presents analytical expressions for statistical power calculation of continuous sequential designs. Section 4 compares four different Type I error probability spending functions in terms of expected number of events to signal for fixed statistical power. Section 5 closes the paper with the last considerations.

2 Continuous Designs are Uniformly Better than Group Designs

Continuous and group sequential designs are now formally defined.

Definition 1 (Group Sequential Design)

For a set of constants $a_1 \ a_2 \ \dots \ a_G$, and a sequence $\{t_i\}_{i=1}^G$ of times, where $t_G = T$, a group sequential analysis design for Poisson data is a procedure that takes H_1 as true if $W(X_{t_i}) \geq a_i$ for some $i \in \{1, \dots, G\}$, and takes H_0 as true otherwise.

Definition 2 (Continuous Sequential Design)

Let CV_t denote a continuous function in the $(0, T]$ interval. A continuous sequential analysis design for Poisson data is any procedure that takes H_1 as true if $W(X_t) \geq CV_t$ for some $t \in (0, T]$.

The notion of “uniformly better design” is another important topic for the goals of this paper. It was introduced by Silva and Kulldorff (2013) having T , significance level, power, expected time to signal, and expected length of surveillance as the target statistical performance measures, and now we extended this definition to arbitrary set of statistical performance measures.

Definition 3 (Uniformly Better Design)

Let D_1 and D_2 denote two sequential analysis designs for hypothesis testing for an unknown P -dimensional parameter θ . Let $\vec{M}(\theta) = (M_1(\theta), \dots, M_K(\theta))$ denote a K -dimensional row vector of statistical performance measures, and for $i = 1, 2$, let $\vec{M}_i(\theta) = (M_{1,i}(\theta), \dots, M_{K,i}(\theta))$ denote the vector of measures associated to D_i . Thus, D_1 is uniformly better than D_2 over $\vec{M}(\theta)$ if $M_{j,1}(\theta) \leq M_{j,2}(\theta)$ for each $j = 1, \dots, K$, and $M_{j,1}(\theta) < M_{j,2}(\theta)$ for some j .

Evidently, the parameter of interest in the present work is the one-dimensional relative risk, i.e. here $\theta = R$. For the target performance measures, define the 8-dimensional vector:

$$\vec{M}(R) = (T, \beta(1), \beta(R), E[\tau|H_0 \text{ rej.}, R], E[X|H_0 \text{ rej.}, R], E[\tau|R], E[X|R], \phi(R)), \quad (3)$$

where $\beta(1)$ is the overall Type I error probability (which equals to the desired α level), $\beta(R)$ is the overall statistical power, $E[\tau|H_0 \text{ rej.}, R]$ is the expected time to signal, $E[X|H_0 \text{ rej.}, R]$ is the expected number of events to signal, $E[\tau|R]$ is the expected time of surveillance, $E[X|R]$ is the expected number of events. The term $\phi(R)$ represents the risk function, also called expected loss, defined by:

$$\phi(R) = l_1(R) \times \beta(R) \times I(R \leq 1) + l_2(R) \times [1 - \beta(R)] \times I(R > 1), \quad (4)$$

where $l_j(R)$, $j = 1, 2$, is an arbitrary real-valued function of R that measures the undesirable effects, losses, followed by the Type I and Type II errors. The loss function is arbitrated by the analyst in such a way that $l_1(R)$ represents the loss followed by the Type I error, and $l_2(R)$ represents the loss followed by the error of Type II. For example, one way to construct a loss function is to consider the financial costs involved in the development and implementation of a recently approved and commercialized vaccine.

As proved by Silva and Kulldorff (2013), continuous sequential analysis is uniformly better than group sequential over the measures $T, \beta(1), \beta(R), E[\tau|H_0 \text{ rej.}, R]$ and $E[\tau|R]$. Now we extend their results for the performance vector in Eq. 3.

Theorem 1 For a Poisson stochastic process X_t indexed by continuous or discrete time, let

$$\vec{M}_1(R) = (T_1, \beta_1(1), \beta_1(R), E_1[\tau|H_0 \text{ rej.}, R], E_1[X|H_0 \text{ rej.}, R], E_1[\tau|R], E_1[X|R], \phi_1(R)) \quad (5)$$

denote the performance measure of an arbitrary group sequential design D_1 that takes H_1 as true for large values of X_t . Thus, there always exists a continuous sequential design, with performance measure

$$\vec{M}_2(R) = (T_2, \beta_2(1), \beta_2(R), E_2[\tau|H_0 \text{ rej.}, R], E_2[X|H_0 \text{ rej.}, R], E_2[\tau|R], E_2[X|R], \phi_2(R)) \quad (6)$$

such that:

- i. $T_2 = T_1$,
- ii. $\beta_2(1) = \beta_1(1)$,
- iii. $\beta_2(R) = \beta_1(R)$,
- iv. $E_2[\tau | H_0 \text{ rej.}, R] < E_1[\tau | H_0 \text{ rej.}, R]$,
- v. $E_2[X | H_0 \text{ rej.}, R] < E_1[X | H_0 \text{ rej.}, R]$,
- vi. $E_2[\tau | R] < E_1[\tau | R]$,
- vii. $E_2[X | R] < E_1[X | R]$, and
- viii. $\varphi_2(R) = \varphi_1(R)$.

Therefore, for each group sequential design, there always exists a continuous sequential design which is uniformly better over the vector of performance measures in Eq. 3.

Proof For a given group sequential design constructed according to Definition 1, consider an associated continuous sequential design having $T = t_G$ and, if $t \in (t_{j-1}, t_j]$, then $CV_t = a_j$.

With this associated continuous design, the proof of item (i) is trivial.

Items (ii) and (iii) follow from the fact that the associated continuous design rejects H_0 if, and only if, the group sequential also does, which holds for each $R \in (0, \infty)$.

To prove items (iv) to (vii), let τ_c denote the time when the associated continuous design is stopped. If the null hypothesis is rejected, then $\tau_c = \inf\{t \in (0, T] : W(X_t) \geq CV_t\}$. It means that, if $\tau_G \in \{t_1, \dots, t_G\}$ is the stopping time of the group sequential design, then $\tau_c \leq \tau_G$. In addition, define $j = \min\{i : W(X_{t_i}) \geq a_i\}$, therefore $t_{j-1} < \tau_c < t_j$ with $t_0 = 0$. Because X_t is a Poisson stochastic process, there exists an $t_i < m < t_{i+1}$ such that $E[X_m | R] - E[X_{t_i} | R] > 0$ and hence $Pr(X_m - X_{t_i} > k | R) > 0$ for any value of k , i.e. the probability that the continuous sequential analysis rejects the null hypothesis at time m is $Pr(X_{t_i} < a_i, X_m \geq a_{i+1} | R) > 0$, which implies that $E[\tau_c | R] < E[\tau_g | R]$. Then expected time to signal and expected number of events to signal with the associated continuous design are smaller than those from the group design.

Item (viii) follows from item (iii). \square

It is important to stress that the proof of the Theorem does not require the presence of all the 8 measures of Eq. 3. Furthermore, follows also from this proof that the continuous design is uniformly better than the group over any subset of $\vec{M}(R)$.

Corollary 1 *If X_t is a Poisson stochastic process, then for each group sequential design there always exists an uniformly better group sequential design over any subset of measures taken from the vector in Eq. 3.*

Under a practical perspective, Theorem 1 can be used to support the practice of adopting a continuous sequential approach in post-market-safety surveillance, where data is collected in

a near continuous matter (Silva and Kulldorff 2013). Under a theoretical perspective, its importance relates to the implication that optimal sequential analysis designs are in the class of continuous designs.

Corollary 2 *An optimal sequential hypothesis testing design for hypotheses of the form in Eq. 1, and having one or more of the performance measures in Eq. 3, is a continuous sequential design.*

Inspired by the corollary above, hereinafter the content of this paper is directed for continuous sequential analysis. To do so, next section uses recursive expressions to derive analytical calculation of power and expected number of events to signal followed by continuous sequential designs. Such expressions shall be useful in the subsequent sections.

3 Recursive Power Function

Assume that $W(X_t)$ is: (i) monotone increasing with μ_t for each fixed $X_t = x > 0$, and (ii) monotone increasing with x for each fixed $X_t = x > 0$ and $\mu_t > 0$. With this, for a given threshold CV_b , let τ_x denote the arrival time of the x -th adverse event for $x = 1, \dots, N$, where:

$$N = \max\{x \in \mathbb{N} : W(x) \leq CV_t\}.$$

Hence, there always exist thresholds $0 < \mu_{(1)} < \mu_{(2)} < \dots < \mu_{(N)}$, with $\mu_{(N)} = T$, such that the probability of rejecting H_0 can be written as:

$$\Pr[\text{rej. } H_0 | R] = \Pr[W(X_t) \geq CV_t \text{ for some } t \in (0, T) | R] \tag{7}$$

$$= \Pr[X_t \geq W^{-1}(CV_t) | R] \tag{8}$$

$$= \Pr[\cup_{x=1}^N \{\tau_x \leq \mu_{(x)}\} | R]. \tag{9}$$

where $\mu_{(x)} = \sup\{t : x = W^{-1}(CV_t)\}$, and $W^{-1}(CV_t)$ is the value y such that $W(y) = CV_t$.

Since $X_t \sim \text{Poisson}(\mu_t)$, write $\mu_t = \lambda t$, where λ is a known constant, and the time index t is a deterministic function of intrinsic characteristics of the analyzed phenomena, like for example the effects of covariates at each analysis combined with the proportion of exposed individuals of the population. Thus, the joint density of the vector (τ_1, \dots, τ_N) is:

$$f_{\tilde{\tau}}(y_1, \dots, y_N | R) = (R\lambda)^N e^{-y_N R\lambda} I(y_N > 0). \tag{10}$$

Consider the transformation $\tilde{T} = \lambda \tilde{\tau} = (T_1, \dots, T_N)$, which has density:

$$f_{\vec{T}}(t_1, \dots, t_N | R) = R^N e^{-t_N R} I(t_N > 0). \quad (11)$$

Hence, the overall probability of rejecting H_0 , denoted by $\beta(R)$, is given by:

$$\begin{aligned} \beta(R) &= \Pr[\text{rej. } H_0 | R] = \Pr\left(\bigcup_{x=1}^N \{T_x \leq \mu_{(x)}\} | R\right) \\ &= \sum_{x=1}^N \Pr(N = x | R), \end{aligned} \quad (12)$$

where N is the cumulative number of observed events at to the signaling time. For $N = 1$:

$$\Pr(N = 1 | R) = \Pr(T_1 \leq \mu_{(1)}) = 1 - e^{-\mu_{(1)} R}.$$

For $N=2$:

$$\begin{aligned} \Pr(N = 2 | R) &= \Pr(T_1 > \mu_{(1)} \cap T_2 > \mu_{(2)}) \\ &= \int_{\mu_{(1)}}^{\mu_{(2)}} \int_{\mu_{(1)}}^{t_2} R^2 e^{-R t_2} dt_1 dt_2 \\ &= \Pr(\mu_{(1)} \leq T_2 \leq \mu_{(2)} | R) - R^{-1} \mu_{(1)} \Pr(\mu_{(1)} \leq T_1 \leq \mu_{(2)} | R). \end{aligned}$$

For $N=3$:

$$\begin{aligned}
\Pr(N=3 \mid R) &= R^{-1} \Pr(T_1 > \mu_{(1)} \cap T_2 > \mu_{(2)} \cap T_3 > \mu_{(3)}) \\
&= \int_{\mu_{(2)}}^{\mu_{(3)}} \int_{\mu_{(2)}}^{t_3} \int_{\mu_{(1)}}^{t_2} R^3 e^{-Rt_3} dt_1 dt_2 dt_3 \\
&= \Pr(\mu_{(2)} \leq T_3 \leq \mu_{(3)} \mid R) - R^{-1} \mu_{(1)} \Pr(\mu_{(2)} \leq T_2 \leq \mu_{(3)} \mid R) \\
&\quad + R^{-2} (\mu_{(1)} \mu_{(2)} - \mu_{(2)}^2 / 2) \Pr(\mu_{(2)} \leq T_1 \leq \mu_{(3)} \mid R).
\end{aligned}$$

For N=4:

$$\begin{aligned}
 \Pr(N = 4 | R) &= R^{-1} \Pr(T_1 > \mu_{(1)} \cap \dots \cap T_4 > \mu_{(4)}) \\
 &= \int_{\mu_{(3)}}^{\mu_{(4)}} \int_{\mu_{(3)}}^{t_3} \int_{\mu_{(2)}}^{t_3} \int_{\mu_{(1)}}^{t_2} R^4 e^{-Rt_1} dt_1 dt_2 dt_3 dt_4 \\
 &= \Pr(\mu_{(3)} \leq T_4 \leq \mu_{(4)} | R) - R^{-1} \mu_{(1)} \Pr(\mu_{(3)} \leq T_3 \leq \mu_{(4)} | R) \\
 &\quad + R^{-2} (\mu_{(1)} \mu_{(2)} - \mu_{(2)}^2 / 2) \Pr(\mu_{(3)} \leq T_3 \leq \mu_{(4)} | R) \\
 &\quad - R^{-3} \left[\frac{\mu_{(3)}^3}{3!} - \frac{\mu_{(1)} \mu_{(3)}^2}{2} + \mu_{(3)} \left(\mu_{(1)} \mu_{(2)} - \frac{\mu_{(2)}^2}{2} \right) \right] \\
 &\quad \times \Pr(\mu_{(3)} \leq T_1 \leq \mu_{(4)} | R).
 \end{aligned}$$

Thus, a recursive expression, with respect to N, can be written to express $\beta(R)$:

$$\begin{aligned}
 \beta(R) &= \sum_{N=1}^N \sum_{i=1}^N (-1)^{i+1} \psi_i \Pr(\mu_{(N-1)} \leq T_{N+1-i} \leq \mu_{(N)} | R) \\
 &= \sum_{N=1}^N \beta_N(R).
 \end{aligned} \tag{13}$$

where $\mu_0 = 0$, $\psi_1 = 1$, and for $i = 2, \dots, N$,

$$\psi_i = \sum_{j=1}^{i-1} (-1)^{j+1} \frac{(r \mu_{(i-1)})^j}{j!} \psi_{i-j}.$$

It merits remark that $\beta_N(1)$ is the punctual Type I error probability spending associated to the N (th) event, which depends only on $\mu_{(N)}, \mu_{(N-1)}, \dots, \mu_{(1)}$. This way, given a target Type I error probability spending $S(t)$, one can ways elicit $\mu_{(N)}$ such that $\beta_N(1)$ equals to $S(\mu_{(N)})$ for each $N = 1, \dots, N$.

The expected number of events to signal can also be calculated in a recursive way using the following:

$$E[X | H_0 \text{ rejected}] = \frac{\sum_{N=1}^N N \left[\sum_{i=1}^N (-1)^{i+1} \psi_i \Pr(\mu_{(N-1)} \leq T_{N+1-i} \leq \mu_{(N)} | R) \right]}{\beta(R)}. \quad (14)$$

4 Comparing Type I Error Spending Functions

The expressions for power and expected number of events to signal in Eqs. 13 and 14, respectively, are used in the present section to perform a comparison study among four Type I spending functions. The first one is the power-type function already shown in the Introduction:

$$S_1(t) = \alpha \times \left(\frac{t}{T} \right)^\rho, \rho > 0. \quad (15)$$

According to Kim and Demets (1987), Jennison and Turnbull (1989), and Jennison and Turnbull (1990), under appropriate choices for ρ the function $S_1(t)$ can mimic Pocock's and O'Brien & Fleming's methods with reasonable precision.

The second Type I error spending considered in this section was firstly suggested by Lan and DeMets (1983) as a good approximation for the O'Brien & Fleming's test, and it is given by:

$$S_2(t) = 2 - 2 \times \Phi(x_\alpha \times \sqrt{T/t}), \quad (16)$$

where x_α is such that $\Phi^{-1}(1 - \alpha/2) = x_\alpha$.

Our third function, also introduced by Lan and DeMets (1983), is the log-exp function, which is meant to approximate the Pocock's test, and it is given by:

$$S_3(t) = \alpha \times \log \left\{ 1 + [\exp(1) - 1] \times \frac{t}{T} \right\}. \quad (17)$$

Finally, the fourth function considered in this comparison study, suggested by Hwang et al. (1990), is also designed to mimic Pocock's test, and its shape is as follows:

$$S_4(t) = \alpha \times \frac{1 - \exp\left\{-\frac{t}{T}\gamma\right\}}{1 - \exp\{-\gamma\}}, \quad \gamma \in (0, 0.5]. \quad (18)$$

All these functions were extensively studied by Lan and DeMets (1983), Kim and Demets (1987), Jennison and Turnbull (1989), Jennison and Turnbull (1990), and Jennison and Turnbull (2000) to explore the balance among statistical power, expected sample size, and maximum length of surveillance. In general, convex shapes lead to small expected sample sizes. Jennison and Turnbull (2000) shows that, for fixed statistical power, the function $S_1(t)$ promotes small expected samples sizes using ρ around 1.5 or 2.

The maximum length of surveillance and the expected sample size are historically used as target measures in sequential analysis. This is due to challenges arising in clinical trials, where it is usually difficult to recruit many patients for the clinical experiment. Differently, in post-market drug and vaccine safety surveillance the maximum sample size can be arbitrarily large when a surveillance system is up and running. By another hand, since a lot of people is exposed to the drug, a delayed signalization of actual risks from the drug can potentially cause real damages to the health of many patients (Silva and Kulldorff 2015). In this scenario, the meaningful performance measure is no longer maximum length of surveillance or expected sample size, but expected number of events to signal for fixed power. As stressed earlier in this paper, there are no studies comparing Type I error spending functions in terms of the trade-off between expected number of events to signal and power.

4.1 Tuning Parameters Settings

A total of 600 scenarios were considered, each corresponding to a specific choice among $\alpha = 0.01, 0.05, R = 1, 1.1, \dots, 3$, and $T = 10, 20, \dots, 100$. For each scenario, power and expected number of events to signal were calculated for each of the four Type I error functions. Note that $S_1(t)$ and $S_4(t)$ define families of functions since different shapes are obtained if different values of ρ and γ are fixed. Because of this, each of the 600 designs described above were evaluated in the following set of meaningful values: $\rho = 0.1, 0.2, \dots, 2$, and $\gamma = 0.1, 0.2, \dots, 0.5$.

All calculations were executed using the **R** language (Team 2015), and can be reproduced with the code available as a Supplementary Material.

4.2 Results

We first investigate the best choice among ρ and γ values for using $S_1(t)$ and $S_4(t)$, respectively. We start with $S_1(t)$. After analyzing the overall numerical evolution of the expected number of events to signal as a function of power, we find out that, in general, the choice $\rho = 0.5$ is always either the best or near to the best choice, depending on the actual value of R . Unfortunately, due to space limitations, we are not able to show figures and tables containing all results from each of the tuning parameters explored in this work. But, some numerical results can be shown in order to illustrate the general conclusions. For example, Fig. 1 brings expected number of events to signal as a function of power for $\rho = 0.1, 0.5, 0.7, 1, 1.5$. The x-axis, given in the log-scale, represents the actual power for each

of the different values of T . For $R = 1.5$, the best curve is associated to $\rho = 0.7$, but the curve for $\rho = 0.5$ is very close to it for each power. For scenario $R = 2$, the choice $\rho = 0.5$ is the best one. These results are uniformly observed for each of the tuning parameters considered.

Concerning the best choice for γ in $S_4(t)$, the global behavior can be illustrated considering the sample of γ values of 0.1, 0.3, and 0.7. Figure 2 illustrates the general behavior that identifies $\gamma = 0.5$ as the best choice in terms of minimizing expected number of events to signal for fixed power.

Now we compare $S_1(t)$, $S_2(t)$, $S_3(t)$ and $S_4(t)$. For this, we use $\rho = 0.5$ and $\gamma = 0.5$. Figure 3 shows the comparison using $R = 1.5, 2$. It is evident that the $S_3(t)$ shape represented by the dotted line, called log-exp function, leads to the smallest expected number of events to signal among these four functions. Therefore, the final conclusion after all these massive numerical investigation is that the log-exp function should be adopted for sequential analysis with Poisson data when expected number of events to signal is the meaningful statistical performance measure.

5 Last Comments

The behavior of the expected number of events to signal in a continuous sequential surveillance is interesting. Because the continuous approach allows to look to the data as many times as wanted, the maximum length of surveillance is increased comparatively to a group sequential testing. This sample size enlargement also increases the expected length of surveillance. This explains why expected number of events to signal and expected sample size are not positively related, which, by its turn, explains the discordance between the findings of the present work and the findings of many authors. While former works indicate usage of a convex Type I error spending function, such as promoted by the O'Brien & Fleming's method, the present work suggests usage of a concave function. The former is concentrated on minimizing expected length of surveillance, and the later is directed to minimize expected number of events to signal. Therefore, if expected sample size or maximum length of surveillance are used as design criterion, then O'Brien & Fleming's test is known to be the good choice. Instead, if expected number of events to signal is the design criterion, then the concave shape promoted by the log-exp function $S_3(t)$ should be used.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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, Brazil (FAPEMIG)

References

Davis R, Kolczak M, Lewis E, Nordin J, Goodman M, Shay D, Platt R, Black S, Shinefield H, Chen R (2005) Active surveillance of vaccine safety: a system to detect early signs of adverse events. *Epidemiology* 16:336–341 [PubMed: 15824549]

- Gombay E, Li F (2015) Sequential analysis: design methods and applications. *Seq Anal* 34:57–76
- Hwang IK, Shih WJ, DeCani JS (1990) Group sequential designs using a family of type I error probability spending functions. *Stat Med* 9:1439–1445 [PubMed: 2281231]
- Jennison C, Turnbull BW (1989) Interim analyses: the repeated confidence interval approach (with discussion). *J R Stat Soc B* 51:305–361
- Jennison C, Turnbull BW (1990) Statistical approaches to interim monitoring of medical trials: a review and commentary. *Stat Sci* 5:299–317
- Jennison C, Turnbull MB (2000) Group sequential methods with applications to clinical trials Chapman and Hall/CRC, London
- Kim K, Demets DL (1987) Design and analysis of group sequential tests based on the type I error spending rate function. *Biometrika* 74(1):149–154
- Kulldorff M, Davis R, KM, Lewis E, Lieu T, Platt R (2011) A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Seq Anal* 30:58–78
- Lan KKG, DeMets DL (1983) Discrete sequential boundaries for clinical trials. *Biometrika* 70(3):659–663
- Leite A, Andrews N, Thomas S (2016) Near real-time vaccine safety surveillance using electronic health records: a systematic review of the application of statistical methods. *Pharmacoepidemiol Drug Saf* 25:225–237 [PubMed: 26817940]
- Lieu T, Kulldorff M, Davis R, Lewis E, Weintraub E, Yih W, Yin R, Brown J, Platt R (2007) Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care* 45(S):89–95
- Nelson J, Cook A, Yu O, Dominguez C, Zhao S, Greene S, Fireman B, Jacobsen S, Weintraub E, Jackson L (2012) Challenges in the design and analysis of sequentially monitored postmarket safety surveillance evaluations using electronic observational health care data. *Pharmacoepidemiol Drug Saf* 21(S1):62–71 [PubMed: 22262594]
- O'Brien P, Fleming T (1979) A multiple testing procedure for clinical trials. *Biometrics* 35:549–556 [PubMed: 497341]
- Pocock S (1977) Group sequential methods in the design and analysis of clinical trials. *Biometrika* 64:191–199
- Silva I, Kulldorff M (2013) Sequential-package R Foundation for Statistical Computing, Vienna
- Silva I, Kulldorff M (2015) Continuous versus group sequential analysis for post-market drug and vaccine safety surveillance. *Biometrics* 71(3):851–858 [PubMed: 26011024]
- Team RC (2015) R: a language and environment for statistical computing R Foundation for Statistical Computing, Vienna
- Wald A (1945) Sequential tests of statistical hypotheses. *Ann Math Stat* 16:117–186
- Yih W, Kulldorff M, Fireman B, Shui I, Lewis E, Klein N, Baggs J, Weintraub E, Belongia E, Naleway A, Gee J, Platt R, Lieu T (2011) Active surveillance for adverse events: the experience of the vaccine safety datalink project. *Pediatrics in press*
- Zhao S, Cook A, Jackson L, Nelson J (2012) Statistical performance of group sequential methods for observational post-licensure medical product safety surveillance: a simulation study. *Statistics and Its Interface* 5:381–390

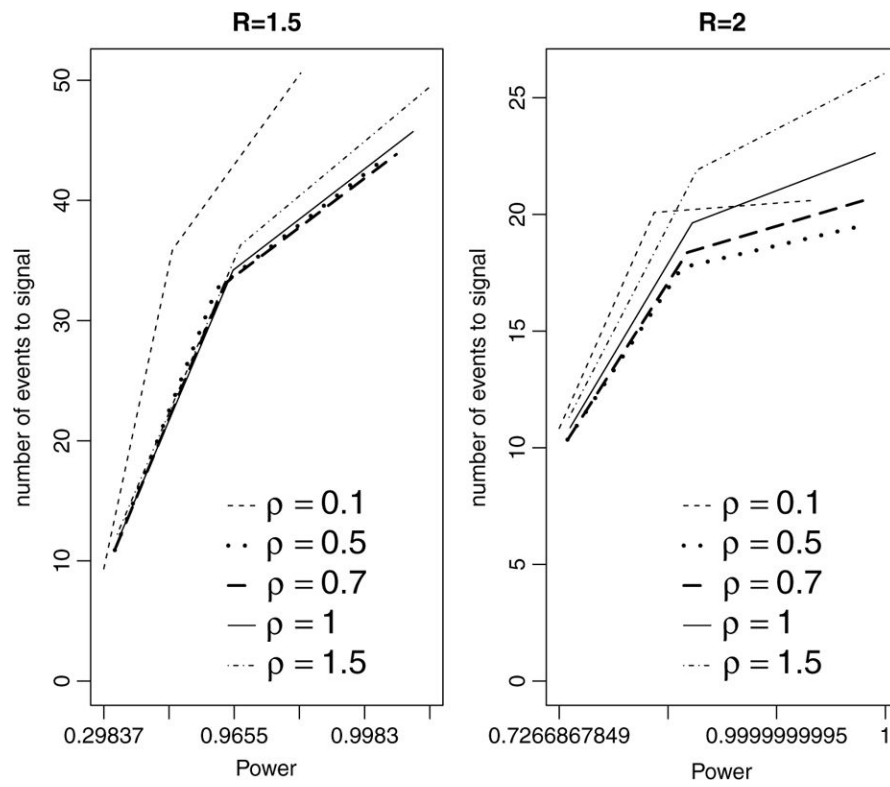


Fig. 1. Expected number of events to signal versus power using the $S_1(t)$ shape based on $\rho = 0.1, 0.5, 0.7, 1, 1.5$. The tuning parameters used were $R = 1.5, 2, \alpha = 0.05$, and selected maximum length of surveillance T . Note the log-scale for power

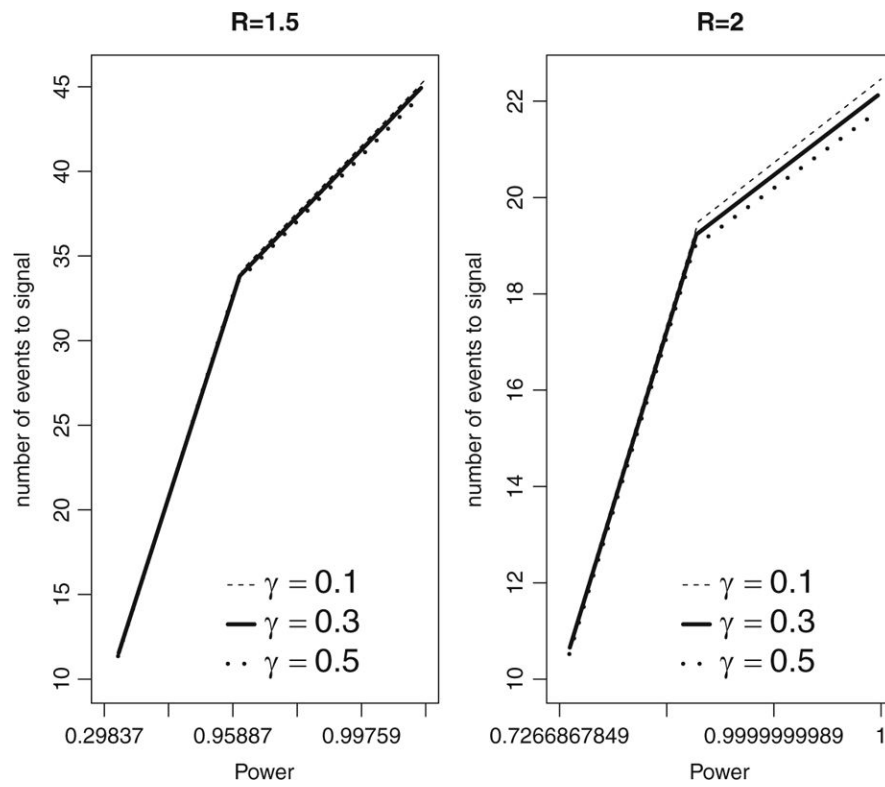


Fig. 2. Expected number of events to signal versus power using the $S_2(t)$ shape based on $\gamma = 0.1, 0.3, 0.5$. The tuning parameters used were $R = 1.5, 2, \alpha = 0.05$, and selected maximum length of surveillance T . Note the log-scale for power

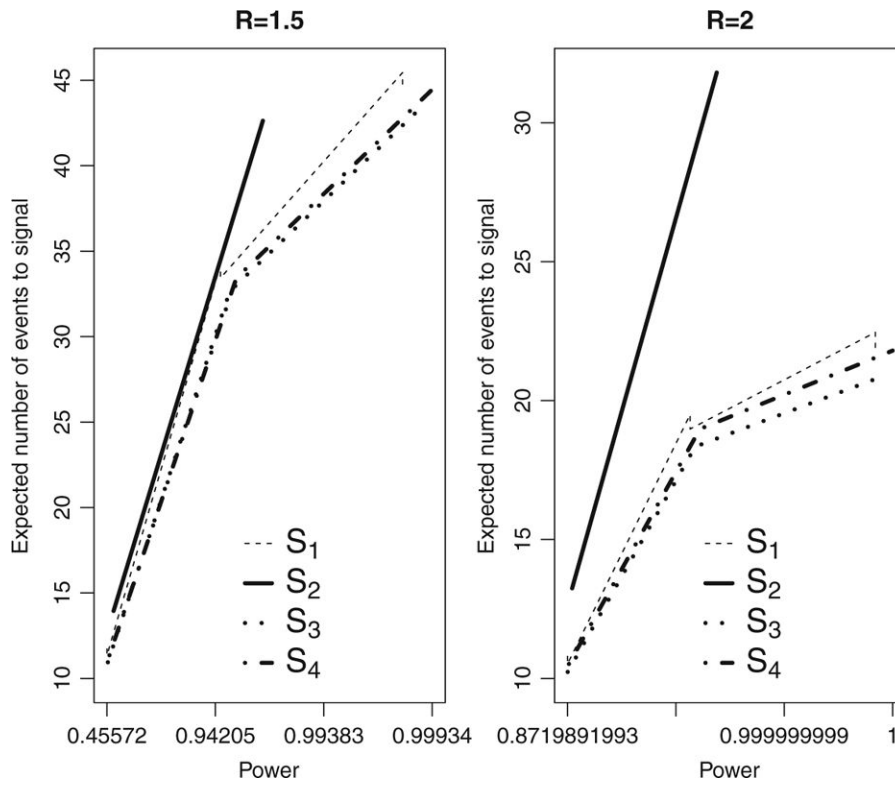


Fig. 3. Expected number of events to signal versus power comparing the shapes $S_1(t)$ with $\rho = 0.5$, $S_2(t)$, $S_3(t)$, and $S_4(t)$ with $\gamma = 0.5$. Each design is based on the tuning parameters $R = 1.5, 2$, $\alpha = 0.05$, and selected maximum length of surveillance T . Note the log-scale for power