Sequential Generalized Likelihood Ratio Tests for Vaccine Safety Evaluation

Mei-Chiung Shih[1,2,*,†], Tze Leung Lai[2,3], Joseph F. Heyse[4] and Jie Chen[5]

[1] *VA Palo Alto Cooperative Studies Program Coordinating Center, Mountain View, CA 94043*

[2] *Department of Health Research and Policy, Stanford University, Stanford, CA 94305*

[3] *Department of Statistics, Stanford University, Stanford, CA 94305*

[4] *Merck Research Laboratories, West Point, PA 19486*

[5] *Abbott Laboratories, Abbott Park, IL 60064*

SUMMARY

The evaluation of vaccine safety involves pre-clinical animal studies, pre-licensure randomized clinical trials and post-licensure safety studies. Sequential design and analysis are of particular interest because they allow early termination of the trial or quick detection that the vaccine exceeds a prescribed bound on the adverse event rate. After a review of recent developments in this area, we propose a new class of sequential generalized likelihood ratio tests for evaluating adverse event rates in two-armed pre-licensure clinical trials and single-armed post-licensure studies. The proposed approach is illustrated using data from the Rotavirus Efficacy and Safety Trial (REST). Simulation studies of the performance of the proposed approach and other methods are also given.

KEY WORDS: Generalized likelihood ratio tests; rotavirus; sequential probability ratio test; vaccine safety.

* Correspondence to: Mei-Chiung Shih, Department of Health Research and Policy, Stanford University, Stanford, CA 94305, U.S.A.

† E-mail: meichiun@stanford.edu

Contract/grant sponsor: NIH/NCRR CTSA, NSF; contract/grant number: UL1 RR025744, DMS-0805879

1. INTRODUCTION

Despite the significant public health impact seen from the introduction of vaccines, the safety of vaccines continues to receive considerable attention, and has raised a variety of issues. First, the withdrawal of a rotavirus vaccine (a tetravalent rhesus-human reassortant rotavirus vaccine, RRV-TV) in 1999 has raised public concerns on vaccine safety [1] and hence the balance of benefit and risk of a vaccine product. Second, unlike any therapeutic products, vaccines are typically given to healthy people and even to vulnerable populations such as infants and young children. In addition, many vaccines are universally recommended and mandated for schooling and some special programs (e.g., military service), where the tolerance of vaccine risk is low. Hence, ensuring vaccine safety is important in public health activities and policies [2, 3].

Vaccine safety evaluation involves pre-clinical animal studies, clinical trials and post-licensure surveillance. Due to obvious limitations, findings in animal studies can only provide biological clues on potential adverse events that may not extrapolate to humans. Pre-licensure vaccine clinical trials usually involve selected populations who receive the vaccine according to a protocol-defined administration method and who are followed for a limited period after vaccination. Some commonly encountered adverse events, such as fever and injection site reaction, are easily observable and documented in vaccine clinical trials; however, a small number of extremely rare and sometimes potentially life-threatening adverse events may not be seen in such trials in spite of their large sample size. Hence, many regulatory agencies require post-licensure monitoring of potential safety issues after the introduction of a new vaccine or vaccine component. Examples of post-licensure vaccine safety surveillance include the Vaccine

Adverse Event Reporting System and Vaccine Safety Datalink in the United States [2, 4]. Sequential methods are increasingly used in vaccine clinical trials and post-licensure monitoring, whose goal is quick detection of the association of adverse events that might be caused by the vaccine. Recently, Davis *et al.* [5], Lieu *et al.* [6] and Heyse *et al.* [7] have proposed to use sequential methods for testing vaccine safety. Though not offered specifically for vaccine safety evaluation, other methods have also been suggested for sequential safety assessment of biopharmaceutical products in clinical trial or post-licensure settings [8, 9, 10, 11, 12].

In this paper we propose a new class of sequential generalized likelihood ratio (GLR) tests for testing the incidence rates of adverse events in vaccine clinical trials and post-licensure surveillance studies. The paper is organized as follows. Section 2 describes typical design considerations for vaccine safety evaluation. Section 3 reviews sequential GLR tests and other sequential tests that have been applied to test vaccine safety. A key ingredient in our proposed GLR tests for vaccine safety, given in Section 4, is the exponential family representation of the rare event sequence under the commonly assumed model of Poisson arrivals of adverse events. Simulation studies are presented in Section 5 to compare the performance of various sequential testing methods, and Section 6 gives an illustrative example from the Rotavirus Efficacy and Safety Trial. Concluding remarks and discussions on the closely related problem of sequential safety surveillance are presented in Section 7.

2. DESIGN CONSIDERATIONS FOR VACCINE SAFETY EVALUATION

Safety profiles of vaccine candidates evolve throughout evaluations in laboratories, animals, phased human clinical trials as well as post-marketing monitoring [2, 13]. It is crucial to

recognize that vaccines are different from most pharmaceutical products in many ways; understanding these differences is important in designing safety studies of vaccines. First, the safety standard is generally higher for vaccines than for drugs. Unlike therapeutic products, vaccines are usually administered to healthy populations, some of whom may be vulnerable children and infants. Some vaccines are universally recommended and as a result are administered to a large number of people. Hence, "first do no harm" is the widespread acceptable principle in public health and a much lower risk tolerance is expected. Second, given that the duration of observation in pre-licensure clinical trials is often less than 30 days (sometimes 42 days) after vaccination, the rarity of certain serious adverse events often necessitates a large sample size. For example, for the REST study of a rotavirus vaccine [7], with a background incidence rate of intussusceptions of 1 in 2000 person years and 42 days of post-vaccination follow-up after each of 3 dose vaccinations, a sample of 60,000 subjects is required in order to observe approximately 10 intussusceptions. Third, vaccines are biologically derived and variations in biological activities can occur. This is further complicated by variations in biological manufacturing processes such as formulation, fermentation and virus sensitivity to storage condition, which together contribute to the variability of biologic activities. These factors may contribute to the adverse experience profile of the vaccine. In addition, many vaccines are combinations of multiple active biologic agents and it is generally difficult, if not impossible, to attribute an adverse event to a particular agent. Finally, unlike drugs for which substitute therapies may be available, vaccines prevent significant morbidity and mortality and usually do not have many alternative options. Hence the decision to withdraw a vaccine should be made with extra care according to risk and benefit balance.

Safety assessment of a vaccine is an ongoing process throughout the product's life cycle. Statistical aspects of design and analysis play an important role in this process. The study design, such as the choice of endpoints, sample size, and study duration, is driven by the objectives, hypotheses, and pre-specified criteria for success, which may vary depending on whether the vaccine is (a) the first vaccine against a particular disease or a vaccine for which a safety issue has been identified for similar vaccine products, or (b) for vulnerable populations, or (c) to be recommended for universal application [2]. Some typical design considerations include the following:

- A meaningful evaluation of the risk of some rare yet serious adverse events to increase the possibility of regulatory approval; this usually requires a large sample size.
- A continuous safety monitoring system to detect increased risk of targeted adverse events as early as possible.
- Criteria for early trial termination due to unsafe outcomes associated with the vaccine during interim monitoring, which would minimize the risk to study participants.
- Immediate communication with the Data and Safety Monitoring Committee when an adverse event is reported.

An example is given in [7] for a rotavirus vaccine trial. The sequential test procedures described in this paper incorporate these design features, as will be explained in the next two sections.

3. REVIEW OF SEQUENTIAL GENERALIZED LIKELIHOOD RATIO TESTS

The use of likelihood ratios as test statistics in sequential analysis has a long history, dating back to Wald's sequential probability ratio test (SPRT) for testing a simple null hypothesis against a simple alternative hypothesis [14]. Suppose $X_1, X_2, \ldots$ are independent random variables with a common density f and one is interested in testing $H_0 : f = f_0$ versus $H_1 : f = f_1$. Let $R_n = \prod_{i=1}^{n} f_1(X_i)/f_0(X_i)$ denote the likelihood ratio based on $X_1, \ldots, X_n$. The SPRT stops sampling at stage

$$T = \inf \Big\{ n \geq 1 : R_n \geq B \ \text{ or } \ R_n \leq A \Big\}, \tag{1}$$

and accepts H_0 (or H_1) if $R_T \leq A$ (or $R_T \geq B$), where A and B are chosen to satisfy the type I and type II error probability constraints $\alpha = \mathrm{pr}_0\{R_T \geq B\}$ and $\tilde{\alpha} = \mathrm{pr}_1\{R_T \leq A\}$. As shown by Wald and Wolfowitz [15], the SPRT minimizes the expected sample size at both H_0 and H_1 among all tests whose error probabilities satisfy $\mathrm{pr}_0\{\text{accept } H_1\} \leq \alpha$ and $\mathrm{pr}_1\{\text{accept } H_0\} \leq \tilde{\alpha}$. The thresholds A and B can be approximated by using Wald's approximations to the error probabilities: $A \approx \log(\frac{\tilde{\alpha}}{1-\alpha})$, $B \approx \log(\frac{1-\tilde{\alpha}}{\alpha})$. Dvoretzky, Kiefer and Wolfowitz [16] extended the SPRT to continuous-time processes with independent increments.

To apply the SPRT to vaccine safety testing, Lieu *et al.* [6] assume that the number N_t of adverse events within d days following vaccination given to m subjects in a clinical trial during the period $[0, t]$ follows a Poisson process with known mean μ_t for the population at risk. For subjects who have received the vaccine, they assume that the mean number of adverse events is still μ_t under H_0 but increases to $\rho\mu_t$ under H_1 with known $\rho > 1$. The stopping rule of the continuous-time SPRT in this case is of the form $T = \inf\{t > 0 : R_t \geq B \text{ or } R_t \leq A\}$, where

the likelihood ratio R_t is the ratio of the density functions of N_t under H_i $(i = 0, 1)$:

$$R_t = \frac{e^{-\rho\mu_t}(\rho\mu_t)^{N_t}/N_t!}{e^{-\mu_t}\mu_t^{N_t}/N_t!} = \rho^{N_t}e^{-(\rho-1)\mu_t}. \tag{2}$$

Because in practice it is often difficult to come up with an appropriate choice of ρ for the alternative hypothesis, Lieu *et al.* [6] maximize (2) over $\rho \geq 1$, yielding

$$\hat{R}_t = \sup_{\rho \geq 1} \rho^{N_t}e^{-(\rho-1)\mu_t} = \exp\{-(\hat{\rho}_t - 1)\mu_t + N_t \log \hat{\rho}_t\}, \tag{3}$$

where $\hat{\rho}_t = \max(1, N_t/\mu_t)$ is the constrained maximum likelihood estimate (MLE) of ρ (≥ 1) at time t. They propose to use the stopping rule

$$\hat{T} = \inf\{t > 0 : \hat{R}_t \geq B\} \tag{4}$$

and to reject H_0 if $\hat{R}_{\hat{T}} \geq B$. They call the test a MaxSPRT and propose to use a truncated version of the test, for which they use Monte Carlo simulations to determine its type I error probability and the power at various alternatives.

For discrete-time observations $X_1, X_2, \ldots$ from an exponential family of densities $f_\theta(x) = \exp\{\theta x - \psi(\theta)\}$ with respect to some measure on the real line, more efficient extensions of the SPRT to composite hypotheses than the MaxSPRT have been introduced in the sequential analysis literature. Lai [17] has given a survey of these tests, which are called "sequential generalized likelihood ratio (GLR) tests". First consider the composite null hypothesis $H_0 : \theta \leq \theta_0$ versus the composite alternative hypothesis $H_1 : \theta \geq \theta_1$, with $\theta_0 < \theta_1$. The sequential GLR test of H_0 versus H_1 stops sampling at stage

$$\tau = \inf\Big\{n \geq 1 : \prod_{i=1}^{n} \left[f_{\hat{\theta}_n}(X_i)/f_{\theta_0}(X_i)\right] \geq B_n^{(0)} \text{ and } \hat{\theta}_n > \theta_0,$$

$$\text{or } \prod_{i=1}^{n} \left[f_{\hat{\theta}_n}(X_i)/f_{\theta_1}(X_i) \right] \geq B_n^{(1)} \text{ and } \hat{\theta}_n < \theta_1 \Big\}, \tag{5}$$

where $\hat{\theta}_n$ is the MLE at stage n, so that the first test statistic in (5) is the GLR statistic for testing θ_0 versus alternatives larger than θ_0, and the second is that for testing θ_1 versus alternatives smaller than θ_1. The test rejects H_i if the GLR statistic for testing θ_i exceeds the threshold $B_\tau^{(i)}$ $(i = 0, 1)$. The test, with $B_n^{(0)} = B_n^{(1)} = 1/c$, has been shown by Schwarz [18] and Wong [19] to be asymptotically Bayes as the cost c per observation approaches 0, for fixed $\theta_0 < \theta_1$, loss $l(\theta)$ at θ of accepting the incorrect hypothesis, and a prior distribution G on the natural parameter space Θ such that $G(I) > 0$ for every open interval $I \subset \Theta$. This asymptotic Bayes property of sequential GLR tests was extended by Lai [20] to the case of testing $H_0 : \theta \leq \theta_0$ versus $H_1' : \theta > \theta_0$, without an indifference zone. The stopping rule still has the form (5) but with $\theta_1 = \theta_0$ and $B_n^{(0)} = B_n^{(1)} = e^{g(cn)}$, i.e.,

$$\tau_c = \inf \left\{ n \geq 1 : \prod_{i=1}^{n} \left[f_{\hat{\theta}_n}(X_i)/f_{\theta_0}(X_i) \right] \geq e^{g(cn)} \right\}, \tag{6}$$

and the test rejects H_0 if and only if $\hat{\theta}_{\tau_c} > \theta_0$, where the function g is the optimal stopping boundary for a continuous-time sequential testing problem associated with an approximating Brownian motion. Note that unlike MaxSPRT that uses the contrained MLE and a time-invariant threshold in (4), the sequential GLR test for testing $H_0 : \theta \leq \theta_0$ versus $H_1' : \theta > \theta_0$ (without an indifference zone) uses the usual GLR statistic (with unconstrained MLE) and a time-varying threshold $e^{g(cn)}$.

Let $I(\theta, \lambda) = E_\theta\{\log[f_\theta(X_1)/f_\lambda(X_1)]\} = (\theta - \lambda)\psi'(\theta) - [\psi(\theta) - \psi(\lambda)]$ be the Kullback-Leibler information number. In the case $B_n^{(0)} = B_n^{(1)} = 1/c$, the stopping rule (5) is bounded above by n^*, where n^* is the smallest integer such that $n^* I(\theta^*, \theta_0) \geq 1/c$ and $\theta^* \in (\theta_0, \theta_1)$

is the minimizer of $\max\{I(\theta,\theta_0), I(\theta,\theta_1)\}$, which is given by the solution of the equation $I(\theta^*,\theta_0) = I(\theta^*,\theta_1)$. For the problem of testing $H_0 : \theta \le \theta_0$ versus $H_1' : \theta > \theta_0$, the stopping rule (5) with $\theta_1 = \theta_0$ and $B_n^{(0)} = B_n^{(1)} = B$ does not have an upper bound when a time-invariant threshold is used. The time-varying threshold $e^{g(cn)}$ in the approximation (6) to the Bayes sequential test of $H_0 : \theta \le \theta_0$ versus $H_1' : \theta > \theta_0$ has the property that $g(t) \to 0$ as $t \to \infty$ and $g(t) \sim \log t^{-1}$ as $t \to 0$, thereby accounting for the time-varying uncertainties in the estimate $\hat{\theta}_n$ of θ as $t = cn$ varies from 0 to ∞.

In the context of conventional clinical trials, even when a sequential design is used to allow early stopping for safety, there is usually an upper bound M on the sample size due to limits on time and resources and to risk-benefit considerations (e.g., when the vaccine is strongly efficacious or when the disease can be fatal). Lai and Shih [21] have pointed out that this upper bound M on the sample size implicitly assumes an alternative θ_1 in testing the one-sided null hypothesis $H_0 : \theta \le \theta_0$ at significance level α. Specifically, the fixed sample test that rejects H_0 if $\sum_{i=1}^{M} X_i \ge c_\alpha$ has maximal power at any alternative $\theta > \theta_0$, and in particular at the alternative $\theta_1 = \theta(M)$ "implied" by M, in the sense that M can be derived from the assumption that this fixed sample size test has some prescribed power $1-\tilde{\alpha}$ at $\theta(M)$. Whereas Lai and Shih [21] have introduced a group sequential modification of (5) and compared it with other group sequential designs in the literature, the design considerations for vaccine safety evaluation trials are different from those for clinical trials to test treatment efficacy. As pointed out in Section 2, vaccine clinical trials involve continuous safety monitoring and fast reporting of adverse events. Therefore a fully sequential design such as (5) is more appropriate than the commonly used group sequential designs for testing of treatment efficacy. In fact, Lieu *et*

al. [6] even assume continuous-time rather than discrete-time data in their stopping rule (4). In the next section we show how the continuous-time information can be represented so that the stopping rule (5) suffices for efficient designs. Note that unlike (5), the SPRT and the MaxSPRT do not have bounded stopping rules. Recognizing this, Lieu *et al.* [6] consider a variant of (4) that stops the trial at time $\tilde{T} = \min(\hat{T}, t^*)$ and rejects H_0 if $\hat{R}_{\tilde{T}} \geq B$, accepting H_0 otherwise.

The sequential GLR test with stopping rule (5) can be readily extended to treat more complex composite hypotheses involving multivariate parameters, as shown in Section 3.4 of [21]. Specifically, consider a multiparameter exponential family of densities $f_\theta(x) = \exp(\theta^T x - \psi(\theta))$ with respect to some measure on $\mathbb{R}^d$, in which θ and x are $d \times 1$ vectors belonging to $\mathbb{R}^d$. To test the null hypothesis $H_0 : u(\theta) \leq u_0$ against the alternative hypothesis $H_1 : u(\theta) \geq u_1$ with $u_0 < u_1$, where u is a smooth real-valued function such that $I(\theta, \lambda)$ is increasing in $u(\lambda)$ for every fixed θ, the sequential GLR test of H_0 versus H_1 stops sampling at stage

$$\tau = \inf\{n \geq 1 : \Lambda_{n,0} \geq b_n^{(0)} \text{ and } u(\hat{\theta}_n) > u_0, \text{ or } \Lambda_{n,1} \geq b_n^{(1)} \text{ and } u(\hat{\theta}_n) < u_1\},$$

where $\hat{\theta}_n$ is the MLE of θ at stage n and $\Lambda_{n,j}$ are the GLR statistics at stage n:

$$\Lambda_{n,j} = n\{\hat{\theta}_n^T \bar{X}_n - \psi(\hat{\theta}_n)\} - \sup_{\theta : u(\theta) = u_j} n\{\theta^T \bar{X}_n - \psi(\theta)\}, \ j = 0, 1.$$

Because the signed-root likelihood ratio statistic $l_{n,j} = \{\text{sign}(u(\hat{\theta}_n) - u_j)\}(2n\Lambda_{n,j})^{1/2}$ behaves like a normal random walk under $u(\theta) = u_j$, $j = 0, 1$, we can approximate $l_{n,j}$ by a sum of independent standard normal random variables under $u(\theta) = u_j$ and thereby determine $b_n^{(0)}, b_n^{(1)}$. In single-armed post-licensure studies in which covariates such as age and vaccination rates may substantially affect adverse event rates, we can use this kind of sequential GLR tests

to incorporate covariate adjustments.

4. EXPONENTIAL FAMILY MODELS OF ADVERSE EVENTS AND ASSOCIATED SEQUENTIAL GLR TESTS

4.1. Adverse events in single-armed post-licensure studies

The single-armed study described by Davis *et al.* [5] is a retrospective study that uses data submitted by the HMOs to the Vaccine Safety Datalink from 1995 through 2000. The data are first segmented into weekly cohorts of vaccinated children. The weekly data are partitioned into a baseline period, which is defined as a period before the introduction of the new vaccine considered in the study, and the surveillance period beginning with the introduction of the vaccine. Each week's dataset is used to count the number of children receiving the vaccine for that week and the number diagnosed with adverse events within 30 days after the vaccination. Thus X_i in this case in Binomial(n_i, p), where n_i is the number of children vaccinated in week i and X_i counts how many of them experience adverse events within the 30-day window. The null hypothesis is $H_0 : p = p_0$, where p_0 is determined from the event rate in the baseline period, and the alternative hypothesis is $H_1 : p = p_1$, where p_1 is based on the effect size that the study wants to detect, e.g., $p_1 = 2p_0$. Davis *et al.* [5] propose to use the SPRT to test H_0 versus H_1 based on the independent binomial random variables X_i with density function proportional to $p^{X_i}(1-p)^{n_i - X_i}$, which belongs to an exponential family with natural parameter $\theta = \log(p/(1-p))$.

As described above, Lieu *et al.* [6] use the number N_t of adverse events in a cohort of vaccinated subjects for sequential testing of vaccine safety. Instead of N_t, which they assume

to be a Poisson process, we can work with the inter-arrival times X_i between successive adverse events. These are independent exponential random variables with means ξ_i. First assume that $\mu_t = \lambda t$ and therefore all the ξ_i are equal to $\xi = 1/\lambda$. The X_i belong to the exponential family $f_\lambda(x) = \lambda e^{-\lambda x}$ with natural parameter $\theta = -\lambda$. The SPRT for testing $H_0 : \lambda \leq \lambda_0$ versus $H_1 : \lambda \geq \lambda_1$ is

$$T = \inf\{n \geq 1 : R_n \geq b \text{ or } R_n \leq a\}, \tag{7}$$

where $R_n = n \log(\lambda_1/\lambda_0) - (\lambda_1 - \lambda_0)S_n$, $S_n = \sum_{i=1}^n X_i$. Since the MLE is $\hat{\lambda}_n = n/S_n$, the stopping rule of the MaxSPRT is

$$\hat{T} = \inf\left\{n \geq 1 : \lambda_0 S_n - n - n \log(\lambda_0 S_n/n) \geq b\right\} \tag{8}$$

for $b > 0$. The stopping rule (5), with $B_n^{(0)} = e^{b_0}$ and $B_n^{(1)} = e^{b_1}$, can be written as

$$\tau = \inf\left\{n \geq 1 : \max_{j=0,1}\left[\lambda_j S_n - n - n \log(\lambda_j S_n/n) - b_j\right] \geq 0\right\}. \tag{9}$$

The more general case in which $X_i \sim \text{Exp}(\lambda_i)$ have rates λ_i varying with i, as in [6], can be converted back to the i.i.d. case by considering $X_i' = X_i/\lambda_i \sim \text{Exp}(1)$.

4.2. Adverse events in pre-licensure randomized clinical trials

Consider a clinical trial in which patients are randomized to receiving vaccine or placebo. Assume that the arrivals of adverse events follow a Poisson process, with rate λ_V for vaccine (V) and λ_C for placebo (C) recipients. This assumption will be relaxed later by allowing the rates to vary with time. When an event occurs, it is associated with either V or C and

$$\text{pr(V | event occurs at time } t \text{ after previous one)} = \frac{\lambda_V e^{-\lambda_V t} \cdot e^{-\lambda_C t}}{(\lambda_V + \lambda_C)e^{-(\lambda_V+\lambda_C)t}} = \frac{\lambda_V}{\lambda_V + \lambda_C}. \tag{10}$$

Suppose adverse events occur at times $T_1 < T_2 < \cdots$, and the event indicator at T_i is $\delta_i = 1$ for V, or 0 for C. Let $\tau_i = T_i - T_{i-1}$. Since the Poisson interarrival times are i.i.d. exponential, it follows from (10) that the likelihood function of (λ_V, λ_C) based on the observations $(T_i, \delta_i), 1 \leq i \leq n$, is

$$\prod_{i=1}^{n} \left[\left(\frac{\lambda_V}{\lambda_V + \lambda_C} \right)^{\delta_i} \left(\frac{\lambda_C}{\lambda_V + \lambda_C} \right)^{1-\delta_i} (\lambda_V + \lambda_C)\, e^{-(\lambda_V + \lambda_C)\tau_i} \right]. \tag{11}$$

The goal of a pre-licensure randomized clinical trial is to show that the vaccine product is effective and safe. The safety objective can be formulated as testing $H_0 : \lambda_V/\lambda_C \leq 1$ versus $H_1 : \lambda_V/\lambda_C \geq \gamma$, where $\gamma > 1$. Let $p = \frac{\lambda_V}{\lambda_V + \lambda_C}$. Then $\lambda_V/\lambda_C \geq \gamma$ if and only if $p \geq \frac{\gamma}{1+\gamma}$. Let $p_0 = 1/2, p_1 = \gamma/(1+\gamma)$. In view of (11), the likelihood ratio statistic for testing H_0 versus H_1 is $\prod_{i=1}^{n} (\frac{p_1}{p_0})^{\delta_i} (\frac{1-p_1}{1-p_0})^{1-\delta_i}$. Hence there is no loss of information in working with the Bernoulli distribution; that is, the actual event times contain no additional information about λ_V/λ_C beyond that provided by the type (V or C) of the events. This argument also applies to $\lambda_{V,i}$ and $\lambda_{C,i}$ that vary with i, since $\prod_{i=1}^{n} (\lambda_{V,i} + \lambda_{C,i}) e^{-(\lambda_{V,i} + \lambda_{C,i})\tau_i}$ is cancelled out in the likelihood ratio statistic, as the δ_i are still independent Bernoulli random variables with means $\pi_i = \lambda_{V,i}/(\lambda_{V,i} + \lambda_{C,i})$.

The SPRT for testing $H_0 : \pi_i \leq p_0$ versus $H_1 : \pi_i \geq p_1$ (for all i) is

$$T = \inf\{n \geq 1 : l_n \geq b \text{ or } l_n \leq a\}, \tag{12}$$

where $l_n = \sum_{i=1}^{n} \{\delta_i \log(\frac{p_1}{p_0}) + (1 - \delta_i) \log(\frac{1-p_1}{1-p_0})\}$ and $a < 0 < b$. The SPRT does not have a bounded stopping rule. The GLR statistic for testing p_j $(j = 0, 1)$ has logarithm

$$l_{n,j} = \sum_{i=1}^{n} \left\{ \delta_i \log(\hat{p}_n / p_j) + (1 - \delta_i) \log\left[(1 - \hat{p}_n)/(1 - p_j) \right] \right\},$$

where $\hat{p}_n = (\sum_{i=1}^{n} \delta_i)/n$. The truncated MaxSPRT has stopping rule $\tilde{T} = \min\{\hat{T}, n^*\}$, where

$$\hat{T} = \inf\Big\{n \geq 1 : l_{n,0} \geq b \text{ and } \hat{p}_n > p_0\Big\}, \tag{13}$$

and rejects H_0 if $l_{\tilde{T},0} \geq b$. The stopping rule (5) of the sequential GLR test is

$$\tau = \inf\{n \geq 1 : l_{n,0} \geq b_0 \text{ and } \hat{p}_n > p_0, \text{ or } l_{n,1} \geq b_1 \text{ and } \hat{p}_n < p_1\}. \tag{14}$$

The stopping rule (14) is bounded above by n^*, where n^* is the smallest integer n such that $nI(p^*) \geq \max(b_0, b_1)$ and $p^* \in (p_0, p_1)$ is the solution of the equation

$$p^* \log(\frac{p^*}{p_0}) + (1-p^*) \log(\frac{1-p^*}{1-p_0}) = p^* \log(\frac{p^*}{p_1}) + (1-p^*) \log(\frac{1-p^*}{1-p_1}),$$

whose common value is denoted by $I(p^*)$. Note that (14) introduces a lower boundary into (13) to allow early stopping for "futility" in the sense that the vaccine is unlikely to be shown unsafe by the prescheduled end of the trial (after observing n^* adverse events).

5. IMPLEMENTATION AND SIMULATION STUDIES

For a given type I error probability α and type II error probability $\tilde{\alpha}$, the thresholds in the stopping rule of the SPRT can be approximated by using Wald's approximations reviewed in the first paragraph of Section 3. The thresholds of the truncated MaxSPRT test and the sequential GLR test can be obtained by solving for the largest positive constants that satisfy the error probability constraints. For example, if $X_i \sim \text{Exp}(\lambda)$, the threshold b of the MaxSPRT truncated at t^* is the solution of $\text{pr}_{\lambda_0}(l_{\tilde{T},0} \geq b) = \alpha$, where $\tilde{T} = \min(\hat{T}, t^*)$ and $\hat{T}$ is given by (8); the thresholds b_0, b_1 of the stopping rule (9) of the sequential GLR test are

the solutions of $\text{pr}_{\lambda_0}(l_{\tau,0} \geq b_0) = \alpha$ and $\text{pr}_{\lambda_1}(l_{\tau,1} \geq b_1) = \tilde{\alpha}$. Because the X_i are independent, these error probabilities can be computed by recursive numerical integration using the Markov property of the random walk l_n or $l_{n,j}$. If X_i is discrete, the integration is replaced with summation. When t^* is large, it is more convenient to use Monte Carlo simulations instead of recursive numerical integration to compute the error probabilities. The Appendix describes an algorithm to implement the GLR test in Section 4.2, for which a software package has been developed using R and is available at the website

http://med.stanford.edu/biostatistics/ClinicalTrialMethodology.html

Example 1. Single-armed post-licensure study.

Consider a single-armed post-licensure study in which the adverse events follow a Poisson process with rate λ_V. Suppose we want to test $H_0 : \lambda_V/\lambda_C = 1$ versus $H_1 : \lambda_V/\lambda_C \geq 3$, with known $\lambda_C = 1$, Type I error probability $\alpha = 0.05$ and Type II error probability $\tilde{\alpha} = 0.1$ at $\lambda_V/\lambda_C = 3$. Table I gives the expected number of adverse events and power for the SPRT, MaxSPRT and the sequential GLR test, whose stopping rules are given by (7)–(9). The SPRT and MaxSPRT are truncated at 1000 events. To determine the thresholds of the stopping rules, the boundary crossing probabilities of the SPRT are obtained by Wald's approximations, and those of MaxSPRT and the sequential GLR test are computed by Monte Carlo using 100,000 simulations. Each result in Table I is based on 50,000 simulations. The SPRT is optimal when the assumed alternative value γ in the likelihood ratio statistic is equal to the actual λ_V/λ_C; recall that $p_1 = \gamma/(1+\gamma)$. The sequential GLR test has comparable expected number of events and power at these values of λ_V/λ_C, except for the case $\lambda_V/\lambda_C = 2$, where the

sequential GLR test has half the expected number of events but also much less power since λ_V/λ_C falls substantially below the lower bound 3 specified by H_1. In contrast, MaxSPRT requires substantially larger expected number of events at $\lambda_V/\lambda_C = 1$ or 2. The maximum number of events is 28 for the sequential GLR test, which is much smaller than 1000 for the truncated SPRT or MaxSPRT.

INSERT TABLE I ABOUT HERE

Example 2. Two-armed pre-licensure randomized clinical trial.

Consider a two-armed pre-licensure randomized study to test $H_0 : \lambda_V/\lambda_C = 1$ versus $H_1 : \lambda_V/\lambda_C \geq 3$, with prescribed Type I error probability $\alpha = 0.05$ and Type II error probability $\tilde{\alpha} = 0.1$ at $\lambda_V/\lambda_C = 3$. Table II gives the expected number of events and power for the SPRT, MaxSPRT and the sequential GLR test, whose stopping rules are given by (12)–(14). The SPRT and MaxSPRT are truncated at 1000 or 100 events (2 cases); 100 is the maximum number of events for the sequential GLR test. To determine the thresholds of the stopping rules, the boundary crossing probabilities of the SPRT are obtained by Wald's approximations, and those of MaxSPRT and the sequential GLR test are computed by using a recursive numerical algorithm described in the Appendix. Table II, whose results are computed by the recursive numerical algorithm, shows the superior performance of the sequential GLR test in two-armed randomized trials, similar to the results in Table I for single-armed post-licensure studies.

INSERT TABLE II ABOUT HERE

6. AN ILLUSTRATIVE EXAMPLE

The Rotavirus Efficacy and Safety Trial (REST) is a blinded, placebo-controlled clinical trial conducted in 11 countries between 2001 and 2004, to assess the efficacy and safety of a pentavalent human-bovine reassortant rotavirus vaccine (RV5). Infants between 6 and 12 weeks of age were randomized at a 1:1 ratio to receive either three doses of RV5 or placebo. All infants were monitored for adverse events during the entire trial duration. The primary safety hypothesis was that RV5 would not increase the risk of intussusception, relative to placebo, within 42 days after any dose. This concern of potential increased risk of intussusception, which is a serious yet uncommon illness with a background incidence rate of 18–56 cases per 100,000 infant years during the first year of life, stems from the withdrawal of a tetravalent rhesus-human reassortant rotavirus vaccine (RRV-TV) in October 1999 when the post-licensure safety surveillance revealed a substantial short-term increase in the risk of intussusception among RRV-TV recipients, primarily in the exposure window 3 - 14 days after the first dose [1, 22]. Details of the REST study design are given in [7] and [23].

Assuming that intussusception occurrences follow a Poisson process, with rate λ_V for vaccine recipients and λ_C for placebo recipients, as in Section 4.2, Heyse *et al.* [7] made use of the fact that conditional on the total number n of intussusception cases from both groups, the number of intussusception cases in the vaccine group is Binomial(n, p), where $p = \lambda_V/(\lambda_V + \lambda_C)$. They therefore applied a repeated significance test that terminates the study after observing a total of n intussusception cases from both groups and declares the vaccine to be unsafe if

$$\text{pr}\{\text{Binomial}(n, p_0) \geq \#_n(V)\} \leq 0.025, \tag{15}$$

where $\#_n(V)$ denotes the number of intussusception cases in the vaccine group among the n intussusception cases. The study is also terminated and declares the vaccine to have a clinically acceptable safety profile if

$$\text{pr}\{\text{Binomial}(n, p_1) \leq \#_n(V)\} \leq 0.025, \tag{16}$$

where $p_1 = 10/11$, corresponding to a 10-fold increase in risk for the vaccine group. Although the nominal significance level of 0.025 in (15) or (16) does not adjust for repeated analysis of the accumulated data, Monte Carlo simulations (involving 10000 random sequences) showed that the probability for the study to stop with a positive conclusion regarding vaccine safety is 0.94 for a vaccine with no increased risk of intussusception, and the probability for the study to declare the vaccine to be unsafe is almost 1 for relative risks of 6 or greater [7]. This conservative approach is appropriate given the nature of the safety evaluation. Section 4.2 provides a methodological innovation that leads to independent Bernoulli random variables without conditioning on the total number of events, thereby making conventional sequential tests directly applicable (to these independent Bernoulli observations).

During the study, all suspected cases of intussusception were promptly reported to, and adjudicated by, an independent, blinded adjudication committee. The study stopped enrollment upon the recommendation of the Data and Safety Monitoring Board (DSMB) when about 70,000 infants had completed their follow-up. At that time, there were 11 confirmed cases of intussusception, 6 in the vaccine group and 5 in the placebo group. Figure 1 summarizes the sequentially accumulated data and the boundaries of (a) the repeated significance test (15)-(16) and (b) the sequential GLR test (14). Here $p_0 = 1/2$ and $p_1 = 10/11$. The lower

boundary of the repeated significance test was crossed and the DSMB recommended to stop the study. If the sequential GLR test (14) had been used instead, the lower boundary would also have been crossed at the same time.

INSERT FIGURE 1 ABOUT HERE

In the REST study, the lower "safe" boundary actually used a group sequential design for the DSMB to conduct interim analysis, starting with a minimum of 60,000 infants and subsequent groups of 10,000 infants. Therefore stopping at the lower boundary involves the total number of intussusception cases of the vaccine and placebo recipients up to the time of each interim analysis. The implementation methods described in Section 5 can be easily modified to handle this situation.

7. DISCUSSION

As noted by Lai [24, p.311], although refinements and modifications of Wald's SPRT for the design of clinical trials had been developed in the 1950s, they received little attention from the biomedical community until the Beta-Blocker Heart Attack Trial (BHAT). The main reason for this lack of interest is that the sample size for a typical trial is too small to allow further reduction while still maintaining reasonable power at the alternatives of interest. BHAT, whose endpoint is time to failure, drew immediate attention to the benefits of sequential methods not because it reduced the number of subjects but because it shortened a four-year study by 8 months in periodic reviews of the patients accrued. The success of BHAT led to the development and increasing use of group sequential designs in phase III clinical trials,

beginning with the influential work of Lan and DeMets [25] that introduced a "type I error spending function" to modify a truncated, fully sequential procedure into a group sequential procedure. The development of vaccine safety tests in the past few years seems to have given fully sequential methods a surge of interest that had been lacking in clinical trials since the 1950s.

The design considerations for vaccine safety evaluation described in Section 2 pave the way for adopting fully sequential tests, beginning with the application of the SPRT by Davis *et al.* [5] described at the end of Section 4.1. Subsequently Lieu *et al.* [6] introduced the MaxSPRT. In Section 3 we have provided an overview of sequential tests of composite hypotheses, showing in particular that the truncated version of MaxSPRT, which has been applied in [6] to post-licensure vaccine safety monitoring, is in fact a sequential GLR test without a lower boundary. As noted in [21] and illustrated in Examples 1 and 2, when a sequential test is truncated, introducing a suitable lower boundary can lead to substantial savings in sample size with little loss of power.

For rare adverse events following vaccination (V) or placebo (C) injection, the effective sample size is the total number of adverse events in the sample of a large number of subjects accrued over a number of years. In Section 4.2 we have shown how this effective sample size can be used to develop an efficient sequential test comparing the event rates of the V and C treatments in a pre-licensure randomized clinical trial. Since the design is information-based, one can adjust, without altering the type I and II error probabilities, the total number of subjects accrued per year and the number of years as the trial progresses, based on the observed adverse event rate of the combined V and C groups as the trial progresses.

The sequential post-licensure studies considered in this paper are either of the type considered by Davis *et al.* [5] or in Phase IV studies, for which the study design typically has a maximum sample size and aims at testing the null hypothesis that the adverse event rate for the vaccinated subjects does not exceed the baseline rate. Another approach to post-licensure vaccine safety monitoring is sequential surveillance, which does not have a maximum sample size and continues until the vaccine is no longer used because of safety or efficacy issues. The analog of the type I error probability of a sequential test is the false alarm rate of a sequential surveillance procedure. One such procedure that has been considered for surveillance in public health is Page's [26] CUSUM (cumulative sum) method; see [27, 28]. Because the CUSUM rule originated from applications to quality control, there are serious limitations in its application to post-licensure vaccine safety surveillance. For example, quality control charts use *average run lengths* (ARL), which are the expected durations to giving an alarm, for their operating characteristics, and the threshold of a CUSUM chart is determined by the ARL to false alarm. The past two decades have witnessed important breakthroughs and major advances in sequential surveillance that has moved far beyond the CUSUM method; see [29, 30, 31] and the references therein. In particular, the ARL to false alarm is replaced by a more flexible false alarm rate in [29], in which more versatile sequential detection and surveillance methods are introduced to replace the CUSUM rule, which like the SPRT, requires complete specification of the baseline and post-change parameters. These methods include the sequential GLR detection rules, which have been extended in [31] to tackle the case where both the baseline and post-change parameters are unknown. Moreover, a theory of sequential surveillance, comparable to the relatively complete theory of sequential detection, is introduced in

[30] for exponential families. While sequential testing can be used to test if the adverse event rate of a large cohort of vaccinated subjects in a post-licensure study differs from the baseline rate, sequential surveillance can be used to detect elevated risks due to environmental, viral or other changes that have affected the approved vaccine's safety and efficacy, or for certain sub-populations with previously undetected risk factors.

ACKNOWLEDGEMENTS

Shih's research was supported by the U.S. Department of Veterans Affairs Cooperative Studies Program. Lai's research was supported in part by the U.S. National Science Foundation under grant number DMS-0805879 and in part by Stanford NIH/NCRR CTSA award number UL1 RR025744. Content is solely the responsibility of the authors and does not represent any official view of the National Center for Research Resources, the National Institutes of Health, the VA Cooperative Studies Program, Merck & Co., Inc., or Abbott Laboratories. The authors thank Balasubramanian Narasimhan of Stanford University for developing the software at the Center for Innovative Study Design website mentioned in Section 5, and Merck Research Laboratories for providing the intussusception data from the REST study.

Appendix: Algorithms for pre-licensure sequential GLR tests

Using the same notation as in §4.2, the sequential GLR test of $H_0 : p \leq \frac{1}{2}(= p_0)$ versus $H_1 : p \geq p_1(= \frac{\gamma}{1+\gamma})$ has stopping rule (14), in which $l_{n,j} = g_{n,j}(S_n)$, where $S_n = \sum_{i=1}^{n} \delta_i$ with $\delta_i \sim \text{Bernoulli}(p)$, and

$$g_{n,j}(s) = s\Big\{ \log\Big(\frac{s/n}{p_j}\Big) - \log\Big(\frac{1-s/n}{1-p_j}\Big)\Big\} + n \log\Big(\frac{1-s/n}{1-p_j}\Big).$$

Find the solution $p^* \in (p_0, p_1)$ of the equation

$$p^* \log(\frac{p^*}{p_0}) + (1-p^*) \log(\frac{1-p^*}{1-p_0}) = p^* \log(\frac{p^*}{p_1}) + (1-p^*) \log(\frac{1-p^*}{1-p_1}),$$

and let I^* denote the common value on either side of the equation. The type I and type II error probabilities can be expressed as $\mathrm{pr}_{p_0}(l_{\tau,0} \geq b_0) = \sum_{1 \leq n \leq n^*} \mathrm{pr}_{p_0}(\tau = n, l_{n,0} \geq b_0, S_n/n > p_0)$ and $\mathrm{pr}_{p_1}(l_{\tau,1} \geq b_1) = \sum_{1 \leq n \leq n^*} \mathrm{pr}_{p_1}(\tau = n, l_{n,1} \geq b_1, S_n/n < p_1)$, respectively, where n^* is the smallest integer n such that $nI^* \geq \max(b_0, b_1)$. We use the following recursive algorithm to compute $\pi_{n,j}(s) = \mathrm{pr}_{p_j}(\tau > n, S_n = s)$ for $j = 0, 1$. Initializing with $\pi_{1,j}(s) = \mathrm{pr}_{p_j}(\delta_1 = s) = p_j^s (1-p_j)^{1-s}$ for $s = 0, 1$, let

$$\pi_{n+1,j}(s) = \begin{cases} p_j \pi_{n,j}(s-1) + (1-p_j)\pi_{n,j}(s) & \text{if } g_{n+1,0}(s) < b_0 \text{ and } g_{n+1,1}(s) < b_1, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\mathrm{pr}_{p_0}(\tau = n+1, l_{\tau,0} \geq b_0, S_\tau/\tau > p_0) = p_0 \pi_{n,0}(s'-1) + (1-p_0)\pi_{n,0}(s')$, where s' is the smallest integer s with $s/(n+1) > p_0$ and $g_{n+1,0}(s) \geq b_0$. Similarly, $\mathrm{pr}_{p_1}(\tau = n+1, l_{\tau,1} \geq b_1, S_\tau/\tau < p_1) = p_1 \pi_{n,1}(s''-1) + (1-p_1)\pi_{n,1}(s'')$, where s'' is the largest integer s with $s/(n+1) < p_1$ and $g_{n+1,1}(s) \geq b_1$.

To solve for smallest b_0 and b_1 such that $\mathrm{pr}_{p_0}(l_{\tau,0} \geq b_0) \leq \alpha$ and $\mathrm{pr}_{p_1}(l_{\tau,1} \geq b_1) \leq \tilde{\alpha}$ for given α and $\tilde{\alpha}$, we can use the following iterative algorithm, initializing with $b_0 = \log(1/\alpha)$, $b_1 = \log(1/\tilde{\alpha})$. Let $\alpha_0 = \alpha$ and $\alpha_1 = \tilde{\alpha}$. If $\mathrm{pr}_{p_j}(l_{\tau,j} \geq b_j^{\mathrm{old}}) > \alpha_j$, increase b_j^{old}, say with step size δ, and use it as b_j^{new}, for $j = 0, 1$. Stop the iterations when $\mathrm{pr}_{p_j}(l_{\tau,j} \geq b_j^{\mathrm{old}}) \leq \alpha_j$. This search for (b_0, b_1) generates a grid of values (b_0^i, b_1^i). We can enlarge or shift the grid into a rectangle that brackets the solution; a rectangle is said to be a bracket if (a) its upper right vertex satisfies $\mathrm{pr}_{p_0}(l_{\tau,0} \geq b_0^i) \leq \alpha$ and $\mathrm{pr}_{p_1}(l_{\tau,1} \geq b_1^i) \leq \tilde{\alpha}$, in which case the vertex is denoted

by $(+,+)$, (b) its lower left vertex satisfies $\mathrm{pr}_{p_0}(l_{\tau,0} \geq b_0^i) > \alpha$ and $\mathrm{pr}_{p_1}(l_{\tau,1} \geq b_1^i) > \tilde{\alpha}$, and (c) none of the other two vertices is $(+,+)$. After finding a bracketing rectangle, the solution can be computed by a fine grid search within the rectangle.

References

[1] Murphy TV, Gargiullo PM, Massoudi MS, Nelson DB, Jumaan AO, Okoro CA, Zanardi LR, Setia S, Fair E, LeBaron CW, Wharton M, Livengood JR; Rotavirus Intussusception Investigation Team. Intussusception among infants given an oral rotavirus Vaccine. *New England Journal of Medicine* 2001; **344**:564–572.

[2] Ellenberg SS, Foulkes MA, Midthun K, Goldenthal KL. Evaluating the safety of new vaccines: Summary of a workshop. *American Journal of Public Health* 2005; **95**:800–807. DOI: 10.2105/AJPH.2004.039438

[3] Wilson CB, Marcuse EK. Vaccine safety–vaccine benefits: Science and the public's perception. *Nature Reviews – Immunology* 2001; **1**:160–165. DOI: 10.1038/35100585

[4] Chen RT, Glasser JW, Rhodes PH, Davis RL, Barlow WE, Thompson RS, Mullooly JP, Black SB, Shinefield HR, Vadheim CM, Marcy SM, Ward JI, Wise RP, Wassilak SG, Hadler SC. Vaccine Safety Datalink project: a new tool for improving vaccine safety monitoring in the United States. The Vaccine Safety Datalink Team. *Pediatrics* 1997; **99**:765–773. DOI: 10.1542/peds.99.6.765

[5] Davis RL, Kolczak M, Lewis E, Nordin J, Goodman M, Shay DK, Platt R, Black S, Shinefield H, Chen RT. Active surveillance of vaccine safety: a system to detect early signs of adverse events. *Epidemiology* 2005; **16**:336–341. DOI: 10.1097/01.ede.0000155506.05636.a4

[6] Lieu TA, Kulldorff M, Davis RL, Lewis EM, Weintraub E, Yih K, Yin R, Brown JS, Platt R; for the Vaccine Safety Datalink Rapid Cycle Analysis Team. Real-time vaccine safety surveillance for the early detection of adverse events. *Medical Care* 2007; **45**:S89–95.

[7] Heyse JF, Kuter B, Dallas MJ, Heaton P. Evaluating the safety of a rotavirus vaccine: the REST of the story. *Clinical Trials* 2008; **5**:131–139. DOI: 10.1177/1740774508090507

[8] Abt K. Poisson sequential sampling modified towards maximal safety in adverse event monitoring. *Biometrical Journal* 1998; **40**:21–41.

[9] Bolland K, Whitehead J. Formal approaches to safety monitoring of clinical trials in life-threatening conditions. *Statistics in Medicine* 2000; **19**:2899–2917.

[10] Goldman AI, Hannan PJ. Optimal continuous sequential boundaries for monitoring toxicity in clinical trials: a restricted search algorithm. *Statistics in Medicine* 2001; **20**:1575–1589. DOI: 10.1002/sim.713

[11] Grigg OA, Farewell VT, Spiegelhalter DJ. Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. *Statistical Methods in Medical Research* 2003; **12**:147–170. DOI: 10.1177/096228020301200205

[12] Hocine MN, Musonda P, Andrews NJ, Farrington CP. Sequential case series analysis for pharmacovigilance. *Journal of Royal Statistical Society* A 2009; **170**:213–236.

[13] Chen RT, Davis RL, Rhodes PH. Special methodological issues in pharmacoepidemiology studies of vaccine safety. In *Pharmacoepidemiology, 4th Edition*. Edited by B.L. Strom. Wiley: New York, 2005; 455–486.

[14] Wald A. Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics* 1945; **16**:117–186.

[15] Wald A, Wolfowitz J. Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistiscs* 1948; **19**:326–339.

[16] Dvoretzky A, Kiefer J, Wolfowitz J. Sequential decision problems for processes with continuous time parameters. *Annals of Mathematical Statistiscs* 1953; **24**:403–415.

[17] Lai TL. Asymptotic optimality of generalized sequential likelihood ratio tests in some classical sequential testing procedures. In *Handbook of Sequential Analysis*. Edited by B.K. Ghosh and P.K. Sen. Dekker: New York, 1991; 121–144.

[18] Schwarz G. Asymptotic shapes of Bayes sequential testing regions. *Annals of Mathematical Statistiscs* 1962; **33**:224–236.

[19] Wong SP. Asymptotically optimal properties of certain sequential tests. *Annals of Mathematical Statistics* 1968; **39**:1244–1263.

[20] Lai TL. Nearly optimal sequential tests of composite hypotheses. *Annals of Statistics* 1988; **16**:856–886.

[21] Lai TL, Shih MC. Power, sample size and adaptation considerations in the design of group sequential clinical trials. *Biometrika* 2004; **91**:507–528.

[22] Murphy TV, Smith PJ, Gargiullo, PM, Schwartz B. The first rotavirus vaccine and intussusception: Epidemiological studies and policy decisions. *Journal of Infectious Diseases* 2003; **187**:1309–1313. DOI: 10.1086/374420

[23] Vesikari T, Matson DO, Dennehy P, Van Damme P, Santosham M, Rodriguez Z, Dallas MJ, Heyse JF, Goveia MG, Black SB, Shinefield HR, Christie CD, Ylitalo S, Itzler RF, Coia ML, Onorato MT, Adeyi BA, Marshall GS, Gothefors L, Campens D, Karvonen A, Watt JP, O'Brien KL, DiNubile MJ, Clark HF, Boslego JW, Offit PA, Heaton PM; Rotavirus Efficacy and Safety Trial (REST) Study Team. Safety and efficacy of a pentavalent human-bovine (WC3) reassortant rotavirus vaccine. *New England Journal of Medicine* 2006; **354**:23–33.

[24] Lai TL. Sequential analysis: some classical problems and new challenges (with discussion). *Statistica Sinica* 2001; **11**:303–408.

[25] Lan KKG, DeMets DL. Discrete sequential boundaries for clinical tirals. *Biometrika* 1983; **70**:659–663.

[26] Page ES. Continuous inspection schemes. *Biometrika* 1954; **41**:100–114.

[27] Sonneson C, Bock D. A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society, Series* A 2003; **166**:5–21.

[28] Frisén M. Optimal sequential surveillance for finance, public health, and other areas (with discussion). *Sequential Analysis* 2009; **28**:310–393.

[29] Lai TL. Sequential changepoint detection in quality control and dynamical systems (with discussion). *Journal of the Royal Statistical Society, Series* B 1995; **57**:613–658.

[30] Lai TL, Liu T, Xing H. A Bayesian approach to sequential surveillance in exponential families. *Communications in Statistics – Theory and Methods* (Special Issue in Honor of S. Zacks) 2009; **38**:2958–2968.

[31] Lai TL. Sequential change-point detection when the pre- and post-change parameters are unknown. *Sequential Analysis* 2010; in press.

Table I. Power and expected number of events for various sequential tests of $H_0 : \lambda_V/\lambda_C = 1$ versus $H_1 : \lambda_V/\lambda_C \geq 3$ in a single-armed post-licensure study.

λ_V/λ_C	GLR[§]	SPRT[†]			MaxSPRT[‡]
		$\gamma = 2.0$	3.0	5.0	
		(a) Expected total number of events			
1.0	3.6	9.7	4.4	2.5	957.9
2.0	7.4	14.7	9.8	4.9	20.4
3.0	6.7	8.7	7.1	5.7	9.6
4.0	5.2	7.1	5.5	4.9	6.7
5.0	4.2	6.4	4.8	4.2	5.4
		(b) Probability of rejecting H_0			
1.0	0.049	0.041	0.038	0.030	0.049
2.0	0.523	0.949	0.648	0.305	1.000
3.0	0.900	1.000	0.966	0.734	1.000
4.0	0.956	1.000	0.997	0.925	1.000
5.0	0.957	1.000	1.000	0.980	1.000

[§] The thresholds $b_0 = 3.435, b_1 = 1.822$ are chosen such that $p_{\lambda_V/\lambda_C=1}(\text{ reject } H_0) \leq 0.05$, $p_{\lambda_V/\lambda_C=3}(\text{ accept } H_0) \leq 0.10$.

[†] Truncated at $n^* = 1000$; γ is the assumed alternative value of λ_V/λ_C in the likelihood ratio statistic.

[‡] Truncated at $n^* = 1000$; the threshold $b = 4.306$ is chosen such that $p_{\lambda_V/\lambda_C=1}(\text{ reject } H_0) \leq 0.05$.

Table II. Power and expected number of events for various sequential tests of $H_0 : \lambda_V/\lambda_C = 1$ versus $H_1 : \lambda_V/\lambda_C \geq 3$ in a two-armed pre-licensure clinical trial.

λ_V/λ_C	GLR[§]	$\text{SPRT}_1^\dagger$			$\text{SPRT}_2^\dagger$			$\text{MaxSPRT}_1^\ddagger$	$\text{MaxSPRT}_2^\ddagger$
		$\gamma = 2.0$	3.0	5.0	$\gamma = 2.0$	3.0	5.0		
				(a) Expected total number of events					
1.0	17.4	37.0	16.2	8.3	35.8	16.2	8.3	957.4	96.5
2.0	29.4	45.2	27.6	14.4	43.4	27.4	14.4	63.8	49.2
3.0	21.8	26.2	20.3	14.2	26.2	20.3	14.2	28.2	24.5
4.0	16.5	20.3	15.9	12.4	20.3	15.9	12.4	19.3	17.1
5.0	13.6	17.6	13.7	11.0	17.6	13.7	11.0	15.4	13.9
				(b) Probability of rejecting H_0					
1.0	0.041	0.044	0.043	0.044	0.042	0.043	0.044	0.050	0.048
2.0	0.642	0.914	0.647	0.398	0.860	0.639	0.398	1.000	0.865
3.0	0.931	0.994	0.926	0.729	0.993	0.925	0.730	1.000	0.998
4.0	0.979	0.999	0.978	0.873	0.999	0.978	0.873	1.000	1.000
5.0	0.991	1.000	0.992	0.932	1.000	0.992	0.932	1.000	1.000

[§] The thresholds $b_0 = 3.466, b_1 = 2.773$ are chosen such that $p_{\lambda_V/\lambda_C=1}(\text{ reject } H_0) \leq 0.05$, $p_{\lambda_V/\lambda_C=3}(\text{ accept } H_0) \leq 0.10$.

[†] Truncated at $n^* = 1000$ (SPRT_1) or $n^* = 100$ (SPRT_2); γ is the assumed alternative value of λ_V/λ_C in the likelihood ratio statistic. The thresholds $b = -2.251,\ a = 2.890$ are obtained by using Wald's approximations to boundary crossing probabilities.

[‡] Truncated at $n^* = 1000$ (MaxSPRT_1, $b = 4.130$) or $n^* = 100$ (MaxSPRT_2, $b = 3.466$). The threshold b is chosen such that $p_{\lambda_V/\lambda_C=1}(\text{ reject } H_0) \leq 0.05$.

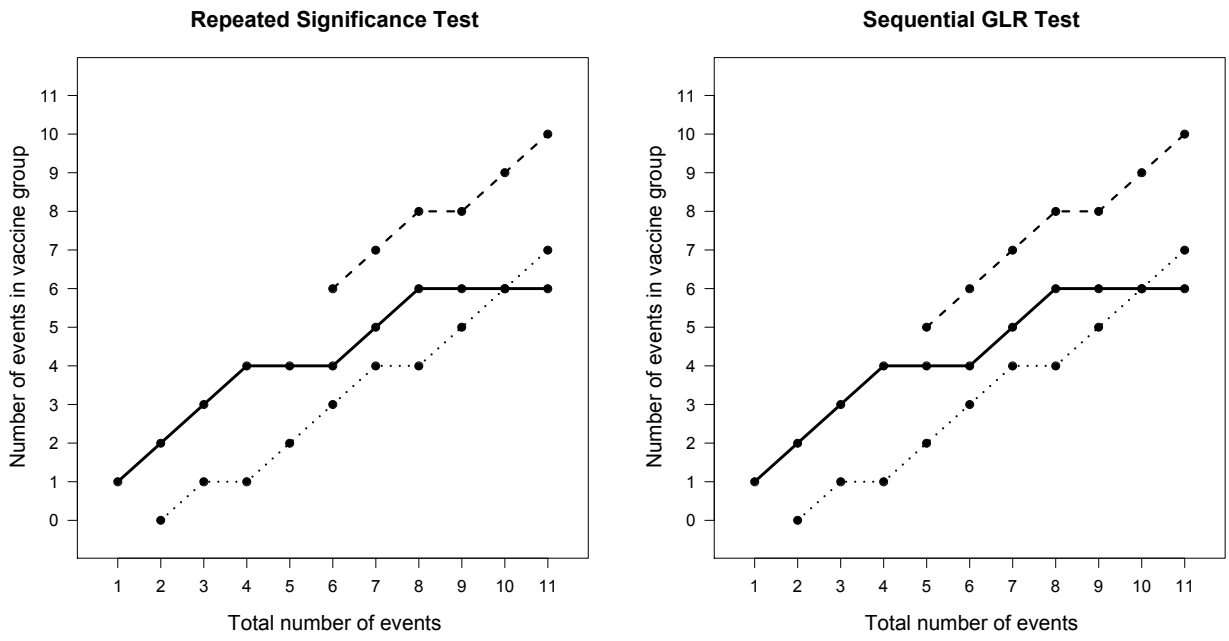

Figure 1. Stopping boundaries of the repeated significance tests (left panel) and sequential GLR test (right panel) for the REST study, where the unsafe boundaries are in dashed lines and the safe boundaries in dotted lines. Also given are the observed data (solid lines).