Sequential Analysis

ISSN: 0747-4946 (Print) 1532-4176 (Online) Journal homepage: https://www.tandfonline.com/loi/lsqa20

A Maximized Sequential Probability Ratio Test for Drug and Vaccine Safety Surveillance

Martin Kulldorff , Robert L. Davis , Margarette Kolczak† , Edwin Lewis , Tracy Lieu & Richard Platt

To cite this article: Martin Kulldorff , Robert L. Davis , Margarette Kolczak† , Edwin Lewis , Tracy Lieu & Richard Platt (2011) A Maximized Sequential Probability Ratio Test for Drug and Vaccine Safety Surveillance, Sequential Analysis, 30:1, 58-78, DOI: 10.1080/07474946.2011.539924

To link to this article: https://doi.org/10.1080/07474946.2011.539924

Copyright Taylor and Francis Group, LLC

Published online: 19 Jan 2011.

Submit your article to this journal

Article views: 8521

View related articles

Citing articles: 120 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=lsqa20

Sequential Analysis, 30: 58–78, 2011
Copyright © Taylor & Francis Group, LLC
ISSN: 0747-4946 print/1532-4176 online
DOI: 10.1080/07474946.2011.539924

A Maximized Sequential Probability Ratio Test for Drug and Vaccine Safety Surveillance

Martin Kulldorff[1], Robert L. Davis[2], Margarette Kolczak[3,†], Edwin Lewis[4], Tracy Lieu[1], and Richard Platt[1]
[1]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care, Boston, Massachusetts, USA
[2]Center for Health Research Southeast, Kaiser Permanente Georgia, Atlanta, Georgia, USA
[3]Immunization Safety Branch, National Immunization Program, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
[4]Division of Research, Kaiser Permanente Northern California, Oakland, California, USA

Abstract: Because of rare but serious adverse events, pharmaceutical drugs and vaccines are sometimes withdrawn from the market, either by a government agency such as the Food and Drug Administration (FDA) in the United States or by the manufacturing pharmaceutical company. In other cases, a drug may be generally safe but increase the risk for serious adverse events for certain subpopulations such as pregnant women or people with heart problems. Due to limited sample size and selected study populations, rare adverse events are often impossible to detect during phase 3 trials conducted before the drug is approved for general use. It is then important to conduct post-approval drug safety surveillance, using, for example, health insurance claims data. In such surveillance, the goal should be to detect serious adverse events as early as possible without too many false alarms, and it is then natural to use a continuous or near-continuous sequential test procedure that reevaluates the data on a daily or weekly basis.

In this article, we first show that Wald's classical sequential probability ratio test (SPRT) for continuous surveillance is very sensitive to the choice of relative risk required in the specification of the alternative hypothesis, making it difficult to use for drug and vaccine safety surveillance. We instead propose the use of a maximized sequential probability ratio test (MaxSPRT) based on a composite alternative hypothesis, which works well across a range of relative risks. We illustrate the use of this method on vaccine safety surveillance and compare it with the classical SPRT.

[†]Deceased.

Received January 22, 2010, Revised September 24, 2010, October 28, 2010, Accepted October 30, 2010

Recommended by Nitis Mukhopadhyay

Address correspondence to Martin Kulldorff, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA 02215, USA; E-mail: martin_kulldorff@hms.harvard.edu

A table of critical values for the MaxSPRT is provided, covering most parameter choices relevant for vaccine and drug safety surveillance. The critical values are based on exact numerical calculations. We also calculate the statistical power, the expected time until the null hypothesis is rejected, and the average length of surveillance.

Keywords: Drug safety; Pharmacovigilance; Rapid cycle analysis; Sequential analysis; Surveillance.

Subject Classifications: 62L10; 92C50.

1. INTRODUCTION

The early detection of unexpected adverse events is very important in both drug and vaccine safety surveillance. Though common adverse events are often detected during phase 2 and 3 clinical trials, rare but serious adverse events may go undetected due to limited sample size. Other adverse events may go undetected if they only affect a subpopulation that was excluded from the clinical trial. To catch these types of adverse events it is important to conduct post-marketing drug and vaccine safety surveillance (Davis et al., 2005; DuMouchel, 1999; O'Neill and Szarfman, 2001; Szarfman et al., 2002). This can be done by monitoring adverse events among patients receiving these drugs/vaccines as part of their regular medical care, using, for example, observational health insurance claims data. Even when no adverse events are found, it is important to do this type of surveillance to ensure the public that new drugs and vaccines are not only effective but also safe, so that patients do not avoid taking important and life-saving drugs/vaccines due to safety concerns.

In order to detect a problem with an adverse event as early as possible, the ideal is to do near-continuous monitoring of patients as they receive the drug or vaccine under study, generating an adverse event signal if and when the number of adverse events are so great that they are unlikely to be due to chance alone. For such continuous sequential analyzes, Wald (1945, 1947) proposed a sequential probability ratio test (SPRT), where a signal is generated if the likelihood ratio exceeds a certain predetermined value, and the observation ends if the likelihood falls below another predetermined lower bound. The key aspect of this method is that the p-values are adjusted for looking at the data in a continuous fashion, or as often as the investigator wishes (i.e., multiple testing). Note that this is not surveillance in the sense of detecting a change, so cumulative sum (CUSUM) and other sequential quality control methods are not suitable in our context. Rather, we want to do surveillance to monitor for an inherent safety problem that is always present in the drug rather than detecting a suddenly occurring safety problem due to, for example, a manufacturing problem in a new batch of the drug.

Sequential probability ratio tests have been extended and refined in various ways, including Bayesian approaches (Lechner, 1962; Peskir and Shiryaev, 2000), and both the theoretical and practical aspects of the field have been summarized in excellent books by Ghosh et al. (1997), Jennison and Turnbull (2000), Mukhopadhyay and de Silva (2009), and Govindarajulu (2004), among others. Particularly relevant to this article are efforts to deal with composite alternatives in various settings such as binomial proportions (Hoel et al., 1976; Joanes, 1972; Meeker, 1981), normally distributed data (Lachin, 1981; van der Tweel et al.,

1996), Poisson data (Abt, 1998), variance components (Ghosh, 1965), functions of unknown parameters (Bangdiwala, 1982), as well as more general models using asymptotic optimality (Lai, 1988; Schwarz, 1962), stepwise sequential probability ratio tests (Huang, 2004), cost functions (Holm, 1985; Schipper et al., 1997), or various approaches that reduce composite hypotheses to simple hypotheses (Ghosh, 1970; Lai, 2001; Wald, 1947). Lai (2001, section 2) provides an excellent review.

One problem with Wald's classical sequential probability ratio test is that the result is highly dependent on the relative risk used to specify the alternative hypothesis. In this article we illustrate this problem in the context of vaccine safety surveillance, showing that an unfortunate choice of the relative risk for the alternative hypothesis may either delay the detection of an important signal or completely miss it. We propose instead the use of a maximized sequential probability ratio test (MaxSPRT), where the alternative hypothesis is composite rather than simple, with the relative risk defined as being greater than one rather than a specific value. Moreover, because there is no reason to stop the study if the drug has a beneficial effect, we only use one critical value boundary to reject the null hypothesis when an excess risk is found, in combination with an upper boundary on the total length of surveillance. Because we do not use acceptance and rejection boundaries that remain unchanged over time, MaxSPRT is a "generalized sequential probability ratio test," as defined by Weiss (1953, p. 273). Because we are using a likelihood ratio with a composite alternative, MaxSPRT is also a "sequential generalized likelihood ratio test," a term first used by Siegmund and Gregory (1980, p. 1223).

Asymptotic results are available for some sequential generalized likelihood ratio tests (e.g., Lai, 1988; Schwarz, 1962; Woodroofe, 1978), but those results are not applicable for calculating critical values for the MaxSPRT. Nor are asymptotic approximations needed, because it is possible to obtain the critical values to any desired precision using iterative numerical calculations. This is done for two probability models relevant for drug and vaccine safety surveillance. In a Poisson model, the number of adverse events at each time is compared with covariate adjusted expected counts based on, for example, historical data or the scientific literature. In a binomial model, the number of adverse events among exposed individuals or time periods is compared with the number of adverse events among matched controls or matched time periods. We provide tables with critical values for different surveillance parameters, so that the users do not have to do any computation of their own, except for calculating the test statistic itself, which can be done using a pocket calculator or a spreadsheet. We also calculate the statistical power, the expected time until the null hypothesis is rejected, and the average length of surveillance.

The MaxSPRT was developed in response to direct vaccine safety surveillance needs in the Centers for Disease Control and Prevention (CDC)-sponsored Vaccine Safety Datalink (VSD) and, as such, it is already in practical use. In this article, the method is illustrated using historical data on fever and neurological symptoms after Pediarix™ vaccination (GlaxoSmithKline Biologicals, Rixensart, Belgium). Though this was the first test application, several real-time vaccine safety surveillance applications of the MaxSPRT have already been published in the medical literature (Belongia et al., 2010; Klein et al., 2010; Lieu et al., 2007; Yih et al., 2009), citing a working version of this article.

For the rest of the article, we first briefly describe the Pediarix vaccine safety data used to illustrate the methods. We then describe Wald's classical SPRT and

show how seemingly contradictory results are obtained depending on the choice of alternative hypothesis, explaining why that happens. Next we present a maximized SPRT with a composite alternative, for both Poisson- (Section 4) and binomial- (Section 5) type data. We then apply the Poisson-based maximized SPRT to the same Pediarix data, comparing it with the results from the classical SPRT. We end with a discussion.

2. VACCINE DATA

To illustrate the use of both the classical SPRT and the MaxSPRT for vaccine safety surveillance, we have applied them using a historical time series of health insurance claims data from the CDC-sponsored VSD project. With these data, we mimic a prospective weekly surveillance system for evaluating whether there is increased risk of either fever or neurological symptoms within 28 days after Pediarix vaccination. Manufactured by GlaxoSmithKline, Pediarix is a combination vaccine that with a single injection protects children from five different diseases: diphtheria, tetanus, whooping cough, hepatitis B, and polio. The VSD project and the data it uses have been described in detail by Chen et al. (1997) and Davis et al. (2005). Here we only give a brief overview.

Started in 1991, the Vaccine Safety Datalink is a collaborative project between CDC and eight different health plans: Group Health Cooperative of Puget Sound (Seattle, Wash.); Harvard Pilgrim Health Care/Harvard Vanguard Medical Associates (Boston, Mass.); Health Partners (Minneapolis, Minn.), Kaiser Permanente Colorado (Denver, Colo.), Marshfield Clinic (Marshfield, Wis.), Northern California Kaiser Permanente (Oakland, Calif.), Northwest Kaiser Permanente (Portland, Ore.), and Southern California Kaiser Permanente (Torrance, Calif.). Together, these plans cover approximately 650,000 children in the United States under the age of six, 3.5% of the total United States population in that age group. As part of the project, immunizations of these children are automatically tracked. Moreover, information about disease diagnoses made during routine medical care at hospitals, emergency departments, and outpatient clinics is available. The data are recorded for all medical events, so that one can tally the number of adverse events seen within a risk window of a fixed number of days after vaccination.

3. WALD'S SEQUENTIAL PROBABILITY RATIO TEST

3.1. Mathematical Definition

Sequential analysis was first developed by Wald (1945, 1947), who introduced the SPRT for continuous surveillance. The likelihood-based SPRT proposed by Wald is very general in that it can be used for many different probability distributions. In our setting, it is defined as follows.

Let C_t be the random variable representing the number of adverse events within D days following a vaccination (or drug prescription) that was given during the time period $[0, t]$, and let c_t be the corresponding observed number of adverse events. Note that time is defined in terms of the time of the vaccination rather than the time

of the adverse event and that, hence, we actually do not know the value of c_t until time $t + D$.

Under the null hypothesis (H_0), C_t follows a Poisson distribution with mean μ_t, where μ_t is a known function reflecting the population at risk. In our setting, μ_t reflects the number of people who received the drug/vaccine during the time interval $[0, t]$ and a baseline risk for those individuals, adjusting for age and gender. Under the alternative hypothesis (H_A), the mean is instead $RR\mu_t$, where RR is the increased relative risk due to the drug/vaccine. Note that $C_0 = c_0 = \mu_0 = 0$.

With the classical SPRT, tests are performed continuously at every time point $t > 0$ as additional data are collected. The test statistic is the likelihood ratio, which for the Poisson distribution is defined as

$$LR_t = \frac{P(C_t = c_t \mid H_A)}{P(C_t = c_t \mid H_0)} = \frac{e^{-RR\mu_t}(RR\mu_t)^{c_t}/c_t!}{e^{-\mu_t}\mu_t^{c_t}/c_t!} = e^{(1-RR)\mu_t}(RR)^{c_t}$$

or, equivalently, as the test statistic is often defined using the log-likelihood ratio

$$LLR_t = ln(LR_t) = (1 - RR)\mu_t + c_t ln(RR)$$

This test statistic is sequentially monitored for all values of $t > 0$, until either $LLR_t \geq ln[(1 - \beta)/\alpha]$, in which case the null hypothesis is rejected, or until $LLR_t \leq ln[\beta/(1 - \alpha)]$, in which case it is accepted. With this stopping rule, the null hypothesis will be falsely rejected with probability α when it is true (type 1 error), and the alternative hypothesis will be falsely rejected with probability β when it is true (type 2 error), although it should be noted that these are approximate results (Wald, 1945, 1947). Note that $LLR_0 = 0$.

As an example, for $\alpha = 0.05$ and $\beta = 0.20$, the upper and lower rejection levels are 2.77 and -1.56, respectively. We will use these two values for the SPRT throughout the article. The SPRT is designed for continuous monitoring, but in practice it is often evaluated at frequent but discrete time intervals, resulting in a slightly conservative test procedure. In this article we use it for weekly data.

3.2. Pediarix Vaccination Safety Surveillance

The first question we will ask using the historical vaccine data is if there is an increased risk of fever during the 4 weeks following Pediarix vaccination. The top left of Figure 1 shows the result of the classical SPRT. With an alternative hypothesis of $H_A : RR = 2.0$, there is enough evidence to reject the alternative hypothesis after 7 weeks, with the conclusion that there is no evidence that Pediarix increases the risk of fever. With $H_A : RR = 1.2$, we get the opposite result, with a rejection of the null hypothesis after 13 weeks, with the conclusion that Pediarix increases the risk of fever.

Why do we get these seemingly contradictory results? Suppose that the true $RR = 1.2$. If the alternative hypothesis is $H_A : RR = 2$; then there is more evidence for the null hypothesis than for the alternative hypothesis and, hence, the alternative hypothesis will be rejected. If the alternative is $H_A : RR = 1.2$, then there is more evidence for the alternative hypothesis than for the null hypothesis, and the null hypothesis will be rejected. Hence, which hypothesis is rejected depends on the

alternative hypothesis chosen. This makes perfect mathematical sense, but the practical implications are disturbing, because we do not know beforehand what excess relative risk we should look for.

One option would be to take a conservative approach by always choosing a very low relative risk for the alternative hypothesis, so that any true relative risk below that threshold is clinically unimportant and uninteresting to detect. That can also lead to problems, though. In the top right part of Figure 1, the results are shown when using the classical SPRT to evaluate an increased risk of neurological symptoms during the 4 weeks following Pediarix vaccination. With $H_A : RR = 1.2$, there is some evidence of an excess risk, and after 65 weeks there is enough evidence to reject the null hypothesis. If, instead, we use an alternative model with $H_A : RR = 2.0$, then the null hypothesis is rejected after 32 weeks.

What is going on now? Suppose the true $RR = 2$. If the alternative model is $H_A : RR = 1.2$, then it is almost as bad as the null model with $RR = 1$, so the

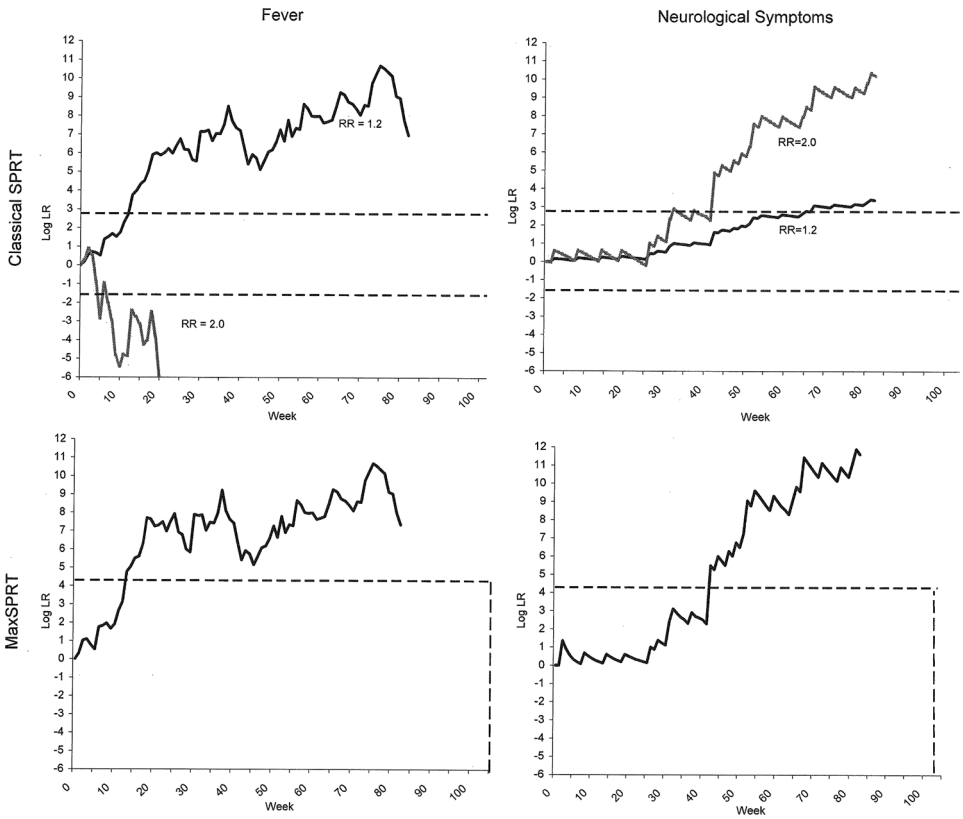

Figure 1. Analyses of the safety of Pediarix^{TM} vaccination with respect to fever (left) and neurological symptoms (right) during the 28 days following vaccination, using the classical SPRT (top) with different relative risks defining the alternative hypothesis ($RR = 1.2$ and $RR = 2.0$) and the MaxSPRT (bottom) with a composite alternative ($RR > 1$). The dashed lines are the critical value bounds. The solid lines are values of the log-likelihood ratio test statistics. The final point estimates for the true relative risk were 1.16 for fever and 2.75 for neurological symptoms.

log-likelihoods are similar and the log-likelihood ratio stays close to zero, and it will take a long time until we reach the upper boundary to reject the null hypothesis. If the alternative model is instead $H_A : RR = 2$, then there is much more evidence for the alternative than for the null hypothesis, resulting in a larger log-likelihood ratio, and the null will be rejected much sooner. Again, this makes perfect mathematical sense though the practical consequences are worrisome, because the time until we detected a serious risk would be longer when using a low conservative relative risk for the alternative hypothesis than if we had used a higher relative risk. Note that if we are only concerned about power, but not in the time it takes to signal, this is not a problem and we can safely use the classical SPRT with a simple alternative chosen as the lowest relative risk of interest Wald (1947, p. 73).

Another way to look at this latter problem is in terms of statistical power and sample size. If we want to detect a true relative risk of 1.2 with 80% power ($\beta = 0.20$), then we need a larger sample size than if we want to detect a true relative risk of 2.0 with the same power. Hence, with $H_A : RR = 1.2$, we would expect to have to wait longer until the null is rejected. One option around this problem would be to modify the SPRT so that the likelihood calculations are based on the lowest relative risk of interest to detect (e.g., $H_A : RR = 1.2$) and the threshold is calculated to guarantee the desired power for a higher relative risk (e.g., 80% power for $H_A : RR = 2$). Though it would be fairly easy to use computer simulations to calculate the correct critical values for any combination of power and pairs of relative risks, we think that a more natural approach is to use an MaxSPRT with a composite alternative hypothesis, as described below.

4. A MAXIMIZED SEQUENTIAL PROBABILITY RATIO TEST: POISSON DATA

For drug and vaccine safety surveillance, we propose the use of an MaxSPRT, with a composite alternative hypothesis $H_A : RR > 1$. The test statistic is the maximum likelihood under the composite alternative hypothesis divided by the likelihood under the simple null hypothesis (Lorden, 1973), and we reject the null hypothesis if the test statistic reaches the critical value before an upper limit on the length of surveillance is reached.

4.1. Log Likelihood Ratio

For the Poisson model, the MaxSPRT likelihood ratio based test statistic is

$$\begin{aligned} LR_t &= \max_{H_A} \frac{P(C_t = c_t | H_A)}{P(C_t = c_t | H_0)} = \max_{RR>1} \frac{e^{-RR\mu_t}(RR\mu_t)^{c_t}/c_t!}{e^{-\mu_t}\mu_t^{c_t}/c_t!} \\ &= \max_{RR>1} e^{(1-RR)\mu_t}(RR)^{c_t} \end{aligned}$$

The maximum likelihood estimate of RR is c_t/μ_t when $c_t \geq \mu_t$, so

$$LR_t = e^{\mu_t - c_t}(c_t/\mu_t)^{c_t}$$

when $c_t \geq \mu_t$ and $LR_t = 1$ otherwise. Equivalently, when defined using the log-likelihood ratio

$$\begin{aligned} LLR_t &= ln(LR_t) = \max_{RR>1}[(1-RR)\mu_t + c_t ln(RR)] \\ &= (\mu_t - c_t) + c_t ln(c_t/\mu_t) \end{aligned}$$

when $c_t \geq \mu_t$ and $LLR_t = 0$ otherwise. Note that the maximum likelihood estimate is unique and that it is also the minimum variance unbiased estimator. Though the same is true for the binomial probability distribution in the next section, the proposed approach may not work for other probability models for which these properties do not hold.

4.2. Critical Values

When defining the critical values for the test statistic there are a few different options. One is to use the classical SPRT approach and reject the null when the *LLR* reaches an upper bound and accept the null when the *LLR* reaches a lower bound. An alternative approach is to use any generalized SPRT (Weiss, 1953), where the bounds are not constant over time. There are pros and cons with different rejection and acceptance boundaries, and the choice will depend on the application. We calculate critical values when there is a constant upper bound to reject the null hypothesis, no lower bound to reject the alternative hypothesis, and the alternative is rejected if and only if the surveillance has reached a predetermined upper limit on the length of surveillance, defined in terms of the expected number of events accrued under the null hypothesis. That is, the upper limit is defined in terms of sample size rather than calendar time. For drug and vaccine safety surveillance we like such boundaries. Because we are doing observational surveillance using data that are collected regardless, there is no harm in continuing the surveillance if the drug/vaccine is safe except for minor data analytic costs. It also puts an upper limit on the length of surveillance, which the classical approach does not have.

In Table 1 we present the upper bounds used for the rejection of the null hypothesis for different upper limits on the maximum length of surveillance. These critical values are based on numerical calculations using the R software language (R Development Core Team, 2009). To do these calculations, first note that the time when the critical value is reached and the null hypothesis is rejected can only happen at the time when an adverse event occurs. For a specified critical value V and upper limit UL on the length of surveillance, it is then possible to calculate α, the probability of rejecting the null, using an iterative approach, as follows. Let s_n be the latest possible time when the null will be rejected based on n adverse events. This means that

$$LLR_{s_n} = (\mu_{s_n} - n) + nln(n/\mu_{s_n}) = V$$

and, hence,

$$\mu_{s_n} = -n * W(-e^{-1-V/n})$$

Table 1. Critical values for the MaxSPRT-based log-likelihood ratios for Poisson data. T is the upper limit on the length of surveillance, expressed in terms of the expected number of events under the null.

T	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
0.1	2.044069	4.119293	6.579669
0.2	2.266893	4.179630	6.754862
0.5	2.637928	4.483740	7.034472
1	2.853937	4.670428	7.172614
1.5	2.964971	4.778944	7.278202
2	3.046977	4.862223	7.341453
2.5	3.110419	4.924475	7.397851
3	3.162106	4.971792	7.445736
4	3.245004	5.040311	7.518319
5	3.297183	5.091907	7.569312
6	3.342729	5.136461	7.608607
8	3.413782	5.206326	7.673013
10	3.467952	5.260513	7.724863
12	3.511749	5.302914	7.767520
15	3.562591	5.351279	7.814719
20	3.628123	5.414770	7.877573
25	3.676320	5.463382	7.924478
30	3.715764	5.502563	7.962688
40	3.774663	5.561620	8.022182
50	3.819903	5.605972	8.067072
60	3.855755	5.642209	8.102340
80	3.910853	5.697631	8.157530
100	3.952321	5.738974	8.199403
120	3.985577	5.772435	8.232827
150	4.025338	5.812121	8.272692
200	4.074828	5.862113	8.322983
250	4.112234	5.899824	8.360938
300	4.142134	5.929897	8.391288
400	4.188031	5.976241	8.438008
500	4.222632	6.011088	8.473183
600	4.250310	6.039013	8.501314
800	4.292829	6.081871	8.544590
1,000	4.324917	6.114225	8.577253

where W is Lambert's W-function, the inverse of the function $f(x) = xe^x$, so that $W(xe^x) = x$. Lambert's W is also called the product log.

At time s_1 there are either zero events, in which case the surveillance continues, or at least one event, in which case the null is rejected, and it is easy to calculate these Poisson-based probabilities. If there were zero events at time s_1, we calculate the probability of zero, one, or two or more events at time s_2, which are also Poisson-based probabilities. If there were zero or one event at time s_2, we calculate the probabilities of $0, 1, 2$, or $3+$ events at time s_3, and so on. Having n or more events at time s_n is always an absorbing state leading to the rejection of the null hypothesis. When $s_n > UL$ we stop, after first calculating the probability of having n or more adverse events at time UL given the probabilities at time s_{n-1}.

With this procedure we can calculate α for any specified critical value, but it is really the reverse that we want, calculating the critical value for any specified α. This is done through interpolation. Suppose that we want the critical value for $\alpha = 0.05$. First calculate $\alpha(V_1)$ and $\alpha(V_2)$ for two reasonable guesses on the critical value. Then calculate

$$V_{i+1} = V_i - (V_i - V_{i-1})\frac{\alpha(V_i) - 0.05}{\alpha(V_i) - \alpha(V_{i-1})}$$

and repeat this iteratively until the desired precision on α has been obtained. If the initial values are estimated based on the critical value from closely related upper limits, the procedure will converge to a precision of 0.00000001 within approximately three to four iterations. For any initial values in the [2, 10] range, it will usually converge in at most seven iterations.

Note that these numerical calculations only have to be done once for each α and *UL* pair. Hence, users of the MaxSPRT do not need to do their own numerical calculations, as long as they use one of the upper limits presented in Table 1.

As expected, the longer we are willing to do the surveillance, the larger the *LLR* must be before we reject the null hypothesis. This is because there is more multiple testing that needs to be adjusted for. Hence, the less willing we are to stop early and accept the null hypothesis, the longer it takes to reach the critical value needed to stop and reject the null hypothesis.

It is important to note that the null hypothesis should be rejected as soon as the *LLR* reaches the critical value, even if it subsequently drops below again. Allowance for this is taken into account when the critical value is determined. In fact, due to the randomness of the data, it is rather typical that the *LLR* falls below the critical value soon after it reaches it for the first time but then climbs above again and stays above.

4.3. Statistical Power

In addition to ensuring the correct α level, it is important to consider the statistical power to reject the null for different true relative risks. This is presented in Table 2, based on exact numerical calculations similar to those used for the critical values. The power is obviously higher when the true relative risk is larger, so the main interest is to compare the power for different upper bounds on the length of surveillance. As expected, the power for the maximized SPRT is higher when T, the expected number of events defining the maximum length of surveillance, is larger. This is natural because the sample size is allowed to grow larger before the surveillance ends. In fact, the power obtained is a natural criterion to use when selecting the upper limit on the length of surveillance. The trade-off is that we must be willing to collect data for a longer time period if the null is not quickly rejected. Hence, the choice of the upper limit on surveillance length is the classical trade-off between sample size and power, although we often do not have to utilize the full sample size that we allow for.

Table 2. Statistical power for the Poisson-based MaxSPRT for different true relative risks. The type 1 error is $\alpha = 0.05$. T is the upper limit on the length of surveillance, expressed in terms of the expected number of events under the null.

	True relative risk					
T	1.2	1.5	2	3	5	10
0.1	0.060	0.075	0.100	0.148	0.242	0.449
0.2	0.062	0.081	0.115	0.187	0.337	0.648
0.5	0.066	0.093	0.147	0.273	0.532	0.899
1	0.070	0.107	0.185	0.379	0.729	0.987
1.5	0.073	0.118	0.221	0.475	0.852	0.9987
2	0.076	0.130	0.255	0.561	0.924	0.9999
2.5	0.078	0.140	0.289	0.637	0.962	0.999990
3	0.081	0.151	0.323	0.703	0.981	0.999999
4	0.086	0.172	0.390	0.809	0.996	1
5	0.089	0.190	0.447	0.876	0.9992	1
6	0.093	0.208	0.500	0.920	0.9998	1
8	0.100	0.244	0.600	0.970	0.9999943	1
10	0.107	0.280	0.685	0.989	0.9999998	1
12	0.114	0.315	0.756	0.996	1	1
15	0.123	0.367	0.836	0.9993	1	1
20	0.138	0.450	0.921	0.99997	1	1
25	0.153	0.526	0.963	0.999999	1	1
30	0.167	0.596	0.984	1	1	1
40	0.196	0.713	0.997	1	1	1
50	0.225	0.803	0.9996	1	1	1
60	0.254	0.868	0.99994	1	1	1
80	0.311	0.944	0.999999	1	1	1
100	0.368	0.978	1	1	1	1
120	0.424	0.992	1	1	1	1
150	0.505	0.998	1	1	1	1
200	0.623	0.99990	1	1	1	1
250	0.722	0.9999952	1	1	1	1
300	0.800	0.9999998	1	1	1	1
400	0.903	1	1	1	1	1
500	0.956	1	1	1	1	1
600	0.981	1	1	1	1	1
800	0.997	1	1	1	1	1
1,000	0.9996	1	1	1	1	1

4.4. Signal Timeliness and Length of Surveillance

In sequential analyses it is not only the α level and statistical power that are important but also the time it takes to reject the null hypothesis when the alternative is true. Conditioned on the null being actually rejected, the top part of Table 3 shows the expected time until rejection for different parameter values. The bottom part of the table shows the expected length of surveillance, until either the null hypothesis is rejected or accepted. In some applications, it is also important to consider the time until the alternative hypothesis is rejected when the null hypothesis is true, but in drug and vaccine safety surveillance that is not a major concern.

Table 3. Average length of surveillance for the Poisson-based MaxSPRT. The top part of the table is the time until a signal is generated rejecting the null hypothesis. The lower part of the table is the time until the end of surveillance, either because of a signal or because of reaching the upper limit on the length of surveillance. The type 1 error is $\alpha = 0.05$. T is the upper limit on the length of surveillance. All times are expressed in terms of the expected number of events under the null hypothesis.

	True relative risk						
T	1	1.2	1.5	2	3	5	10
Average time until H_0 is rejected							
1	0.22	0.25	0.30	0.35	0.39	0.37	0.22
2	0.40	0.51	0.63	0.75	0.79	0.62	0.24
5	0.96	1.38	1.82	2.09	1.78	0.83	0.26
10	1.83	2.99	4.02	4.13	2.45	0.87	0.27
20	3.48	6.70	8.68	6.96	2.67	0.91	0.28
50	8.06	19.76	20.45	8.94	2.82	0.96	0.30
100	15.13	43.78	29.93	9.30	2.92	0.99	0.31
200	28.37	89.73	33.00	9.62	3.01	1.02	0.32
500	64.95	171.79	34.40	10.01	3.12	1.06	0.33
1,000	121.49	196.27	35.37	10.27	3.20	1.08	0.33
Average time until surveillance ends							
1	0.96	0.95	0.93	0.88	0.77	0.54	0.23
2	1.92	1.89	1.82	1.68	1.32	0.72	0.24
5	4.80	4.68	4.40	3.70	2.18	0.83	0.26
10	9.59	9.25	8.33	5.98	2.53	0.87	0.27
20	19.17	18.17	14.91	8.00	2.67	0.91	0.28
50	47.90	43.20	26.28	8.96	2.82	0.96	0.30
100	95.76	79.28	31.46	9.30	2.92	0.99	0.31
200	191.42	131.25	33.02	9.62	3.01	1.02	0.32
500	478.25	186.15	34.40	10.01	3.12	1.06	0.33
1,000	956.07	196.59	35.37	10.27	3.20	1.08	0.33

4.5. Aggregated Data

The MaxSPRT is, just as the classical SPRT, formulated for data that are continuously collected and evaluated. In drug and vaccine surveillance, it is often more practical to collect data on a slightly aggregate basis such as weekly or monthly counts. If the log-likelihood ratio is only evaluated at the end of each week, the MaxSPRT will be slightly conservative in that the probability of rejecting the null when it is true is somewhat less than the nominal α level. It will also result in a slight delay in detecting a true signal. A slightly modified approach, which will maintain the correct α level, is to randomly allocate the adverse event observations within the expected counts accrued during that week by using a uniform distribution. For example, suppose there were two observed adverse events compared to 1.2 expected during the first week of surveillance. Each of the two observed events will then be randomly assigned in the [0, 1.2] interval according to the uniform distribution, independently of each other.

In most practical settings, either approach will work fine without major differences in the results as long as the level of aggregation is modest. The MaxSPRT should not be used when the sequential testing is done in a less frequent manner though, such as once every year, because it would then adjust for more multiple testing than necessary. It is then more appropriate to use group sequential methods (Jennison and Turnbull, 2000).

5. A MAXIMIZED SEQUENTIAL PROBABILITY RATIO TEST: BINOMIAL DATA

Reliable estimates for the expected number of events are not always available before the start of drug and vaccine safety surveillance. An alternative design is then to collect information about potential adverse events from both exposed and unexposed times. For example, in a self-control design, we may compare an exposed time period after vaccination with an unexposed time period before vaccination from the same individual or with an unexposed time period long after vaccination. Alternatively, we may compare individuals exposed to the drug/vaccine with matched unexposed individuals. Unless the unexposed time period is much longer than the exposed, we cannot use the Poisson distribution. We should instead use a binomial probability model when calculating the log-likelihood function and the critical values. The upper limit on the length of surveillance will also be different and will now be defined in terms of the number of adverse events seen. That is, we would continue the surveillance until either there is a signal rejecting the null hypothesis or when we have observed a total of N adverse events in the exposed and unexposed time periods combined. In essence, we have a number of coin tosses (adverse events), which may either turn up as head or tail (exposed or unexposed). Under the null hypothesis, the probability of a head is known to be p, where $p = 0.5$ for a 1:1 matching ratio when the exposed and unexposed time periods are of the same length, $p = 0.25$ for a 1:3 matching ratio, etc.

Other than these differences, the principles behind the MaxSPRT are the same for Poisson- and binomial-type data.

5.1. Log-Likelihood Ratio

Let n be the number of adverse events seen so far during the sequential data collection, and among those n events, let $c_n \leq n$ be the number of adverse events during the exposed time periods. Let z be the length of the matched unexposed time period divided by the length of the exposed time period. Conditional on the number of adverse events n, we can then write the likelihood ratio for the binomial model as:

$$LR_n = \max_{H_A} \frac{P(C_n = c_n | H_A)}{P(C_n = c_n | H_0)} = \max_{RR>1} \frac{[RR/(z+RR)]^{c_n}[z/(z+RR)]^{n-c_n}}{[1/(z+1)]^{c_n}[z/(z+1)]^{n-c_n}}$$

The maximum likelihood estimate of RR is $zc_n/(n-c_n)$. So

$$LR_n = \frac{(c_n/n)^{c_n}[(n-c_n)/n]^{n-c_n}}{[1/(z+1)]^{c_n}[z/(z+1)]^{n-c_n}}$$

when $zc_n/(n - c_n) > 1$ and $LR_n = 1$ otherwise. Equivalently, when defined using the log-likelihood ratio

$$\begin{aligned} LLR_n &= ln(LR_n) \\ &= c_n ln\left(\frac{c_n}{n}\right) + (n - c_n) ln\left(\frac{n - c_n}{n}\right) - c_n ln\left(\frac{1}{z+1}\right) - (n - c_n) ln\left(\frac{z}{z+1}\right) \end{aligned}$$

when $zc_n/(n - c_n) > 1$ and 0 otherwise.

Table 4. Critical values for the log-likelihood ratios from the MaxSPRT for binomial data for 1, 2, and 3 unexposed individuals per exposed, respectively. N is the upper limit on the length of surveillance, defined in terms of the observed number of adverse events. For small N and small α levels it is not always possible to reject the null even if all adverse events are among the exposed. Such combinations of parameter values make the MaxSPRT non applicable (n/a).

	Matching ratio 1:1			Matching ratio 1:2			Matching ratio 1:3		
N	0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
3	n/a	n/a	n/a	2.19723	n/a	n/a	2.77259	n/a	n/a
4	n/a	n/a	n/a	2.19723	n/a	n/a	2.77259	4.15889	n/a
5	2.77259	n/a	n/a	2.29791	4.39445	n/a	2.77259	4.15889	5.54518
6	2.77259	n/a	n/a	2.29791	4.39445	n/a	2.77259	4.15889	5.54518
8	2.77259	4.15889	n/a	2.90393	4.39445	6.59168	2.77259	4.15889	6.93148
10	2.77259	4.15889	6.23833	2.90393	4.39445	6.59168	2.77259	4.15889	6.93148
12	2.77259	4.27363	6.23833	3.19516	4.39445	6.59168	2.77259	4.39445	6.93148
15	2.89118	4.50710	6.72326	3.19516	4.39445	6.59168	2.77259	4.45847	6.93148
20	3.09884	4.85204	6.93148	3.29584	4.39445	6.89371	2.77259	4.51579	6.93148
25	3.13949	4.85204	6.93148	3.29584	4.59581	7.04215	2.87683	4.51579	6.93148
30	3.39139	4.85204	6.93148	3.29584	4.86283	7.04215	3.00043	4.60292	6.93148
40	3.46574	4.85989	7.34969	3.29584	4.93395	7.31447	3.23539	4.79047	6.97759
50	3.46574	4.88272	7.36129	3.29584	5.08104	7.41709	3.31895	4.92501	7.08667
60	3.46574	4.96051	7.44608	3.29584	5.08160	7.63273	3.33085	5.05772	7.18758
80	3.46574	5.11924	7.62462	3.30986	5.17987	7.69029	3.33085	5.23455	7.37322
100	3.46574	5.31315	7.62462	3.43691	5.30927	7.69029	3.45219	5.36765	7.56519
120	3.47863	5.37043	7.62462	3.48031	5.45081	7.69029	3.45660	5.47676	7.67217
150	3.60597	5.44610	7.64717	3.53437	5.49307	7.69029	3.52540	5.51207	7.70334
200	3.68065	5.51372	7.77200	3.59470	5.49307	7.69029	3.64449	5.54518	7.77539
250	3.75290	5.54518	7.87014	3.67774	5.49307	7.76739	3.72423	5.54518	7.83649
300	3.82197	5.54518	7.94491	3.73387	5.49307	7.78945	3.78278	5.54518	7.89501
400	3.89723	5.54518	7.98030	3.78859	5.49307	7.90802	3.86715	5.54518	7.97126
500	3.95630	5.55891	8.05717	3.87392	5.53040	7.99256	3.92779	5.54518	8.05510
600	3.97652	5.62593	8.12388	3.93147	5.55664	8.04060	3.98762	5.54518	8.11146
800	4.06641	5.68788	8.19592	4.02452	5.62280	8.07877	4.07561	5.60324	8.21196
1,000	4.12966	5.76441	8.25052	4.09409	5.67413	8.14459	4.11634	5.64641	8.23855
1,200	4.15888	5.79584	8.30479	4.12633	5.71949	8.20833	4.15889	5.69864	8.23855
1,500	4.15888	5.83888	8.31777	4.14389	5.78536	8.25265	4.15889	5.73464	8.30161
2,000	4.15888	5.91773	8.31777	4.18194	5.83545	8.31836	4.15889	5.74860	8.31777

5.2. Critical Values

The critical values for the MaxSPRT for binomial data are provided in Table 4. Note that the critical values are often identical for different values of the upper limit on the survival length N. This is because of the discrete nature of the data. For example, with $N = 10$ there are only $2^{10} = 1{,}024$ possible outcomes of the surveillance, because each of the 10 adverse events will either be during an exposed or an unexposed time period. This discreteness also means that the actual α level is usually somewhat smaller than the nominal 0.05 but never higher.

The critical values for the binomial model were calculated analytically, using an iterative Markov chain approach, and hence there is no need for computer simulations or approximate asymptotic results. Because of the discrete nature of the data, there are only a finite number of values that the likelihood can take. For each of these likelihood values l, a separate Markov chain is constructed. The state space of the Markov chain is (n, c_n), where $n > 0$ and $0 \leq c_n \leq n$. For the value n, the probability for each state can easily be computed iteratively from the probabilities for the value $n - 1$, with the initial condition that $P[(0, 0)] = 1$. Those states for which $LLR_n(c_n) \geq l$ are absorbing states. By summing the probabilities of the absorbing states we get the α level when the likelihood value l is used as the critical value.

6. EXAMPLE: PEDIARIX VACCINE SAFETY SURVEILLANCE

We applied the Poisson-based MaxSPRT to the same Pediarix data that we analyzed using the classical SPRT in Section 3. As the upper limit on the length of surveillance we choose 800 and 15 expected events for fever and neurological

Table 5. Number of weeks until a signal is seen for the MaxSPRT, with different upper limits on the length of surveillance, and for the classical SPRT, with different relative risks used for the alternative hypothesis. For values in bold, the null hypothesis was rejected, indicating that the vaccine causes fever. For values in italic, the null hypothesis was accepted, indicating that the vaccine does not increase the risk of fever/neurological symptoms. After 82 weeks of surveillance, the observed relative risk was 1.16 for fever and 2.7 for neurological symptoms.

	Fever		Neurological	
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
MaxSPRT				
$T \sim 2$ years	**13**	**17**	**42**	**42**
$T \sim 1$ year	**13**	**17**	**42**	**42**
$T \sim 3$ months	**13**	**17**	**32**	**42**
Classical SPRT				
$RR = 1.05$	**36**	**73**	>82	>82
1.1	**16**	**30**	>82	>82
1.2	**13**	**16**	**65**	>82
1.5	**13**	**13**	**42**	**52**
2.0	*7*	*7*	**32**	**42**
5.0	*1*	*1*	*13*	*13*
10.0	*1*	*1*	*6*	*6*

symptoms, respectively, corresponding to approximately to 2 years of surveillance. The results are shown at the bottom of Figure 1.

For fever, the MaxSPRT rejects the null hypothesis after 13 weeks at the $\alpha = 0.05$ level, due to 97 observed cases when 69.7 were expected under the null, with $RR = 1.39$ and $LLR = 4.78$. For neurological symptoms, the MaxSPRT rejects the null hypothesis after 42 weeks at the $\alpha = 0.05$ level, due to 15 observed cases when 5.5 were expected under the null, with $RR = 2.7$ and $LLR = 5.51$.

In Table 5, we compare the results when using the MaxSPRT with different upper limits on the length of surveillance and the classical SPRT for different relative risks used for the alternative hypothesis. Results are provided for α levels of 0.05 and 0.01. The power for the classical SPRT depends on the true relative risk but is set to be 0.80 for the alternative chosen. The power for the MaxSPRT depends on the upper limit of the length of surveillance as well as on the true relative risk as shown in Table 2. Note that under normal circumstances one would do at most one of these analyses using prespecified parameter values, and we only present the multiple results for methodological comparisons.

7. DISCUSSION

The VSD project uses weekly data to rapidly detect any vaccine safety problems. The expected number of adverse events under the null hypothesis is typically very small each week, and the number of weekly analyses is at least 100. This means that we have near real-time surveillance, and it is then more appropriate to use sequential methods for continuous surveillance rather than the group sequential methods that are commonly used for clinical trials.

In this article we demonstrated an inherent problem when utilizing Wald's classic SPRT for continuous surveillance of vaccine and drug adverse events. We then presented a maximized SPRT that uses a composite rather than a simple alternative hypothesis. The MaxSPRT was explored for two different probability models using the Poisson and binomial distributions respectively, but the general approach can be used for other distributions such as the hypergeometric, suitable for other types of data. The MaxSPRT has been shown to work well for vaccine safety surveillance, with good statistical power and timeliness until signals are generated.

Though the focus of this article is methodological, the clinically relevant findings of our analysis deserve a brief comment. First, as in any disease surveillance setting, it is important to realize that a statistical signal may either be due to a true excess risk or to other issues, including systematic differences in coding or diagnostic practices. A signal is hence a call for a detailed epidemiological study rather than proof of a clinical problem. Mild fever is a known side effect of Pediarix vaccination (Partridge and Yeh, 2003), so it is not surprising that we see a 16% elevated risk in our data. For neurological symptoms, we found that the excess number of cases is at least partly explained by changes made in the medical health records encounter forms affecting two different neurological symptoms.

7.1. Abt's SPRT

Ours is not the first Poisson-based sequential probability ratio test with a composite alternative to be used for drug safety monitoring. Abt (1998) provided an important

first step in that direction, using a different approach. For some values of a and b specified by the user, with $0 < a < 1$ and $0 < b < 1$, define $R(a, b, t)$ as

$$R(a, b, t) = \arg\min_{RR>1} \frac{ln(\frac{1-b}{a}) + \mu_t(RR - 1)}{ln(RR)}$$

This means that, for each time t, $R(a, b, t)$ is the value of the relative risk that minimizes the number of cases needed to reject the null hypothesis of the classical SPRT with $\alpha = a$ and $\beta = b$. The test statistic is then defined as

$$A_t = \frac{P[C_t = c_t \mid H_A : RR = R(a, b, t)]}{P(C_t = c_t \mid H_0)} = e^{[1-R(a,b,t)]\mu_t}[R(a, b, t)]^{c_t}$$

The upper and lower rejection limits are set to be $ln[(1 - b)/a]$ and $ln[b/(1 - a)]$, respectively, as with the classical SPRT. As Abt (1998) pointed out, because of the minimization done when calculating $R(a, b, t)$, a and b no longer represent the approximate type 1 and 2 errors. Rather, for any pair (a, b), the true type 1 and 2 errors are calculated using simulations. For example, with $a = 0.07$ and $b = 0.08$, the type 1 error is $\alpha = 0.1$ and the type 2 error is $\beta = 0.05$ (Abt, 1998).

The difference between Abt's SPRT and the MaxSPRT is that the former finds the relative risk that minimizes the number of cases needed to reject the null hypothesis with the classical SPRT, whereas the latter defines the test statistic by maximizing the likelihood over different relative risk parameter values. This latter approach is the standard way to deal with composite alternative hypotheses, through the creation of a likelihood ratio test statistic (Lehmann, 1986).

7.2. Rejection and Acceptance Regions

In this article we have defined the critical bounds so that the null is rejected when the *LLR* reaches a certain fixed value, and the null is accepted when the prespecified upper limit on the length of surveillance is reached. This is a natural choice for drug and vaccine safety applications but not the only option. The MaxSPRT can be used with any other type of critical bounds as well, including the traditional upper and lower bounds used by the classical SPRT as well as various generalized SPRT rejection regions of triangular or other shapes. Calculating and providing tables for critical values, statistical power, and timeliness for such versions of the MaxSPRT is an important area for further work. In post-marketing safety surveillance, the main issue is the time to signal, and there is not only a trade-off between type 1 error, overall power, and timeliness to signal but, equally or more important, between the timeliness to signal for different true excess risks. Because the objectives are very different, these trade-offs are very different for post-marketing safety surveillance versus pre-marketing clinical trials.

7.3. Critical Values

Though the critical values are based on extensive numerical calculations for the Poisson model and nontrivial analytical calculations for the binomial model, a nice

feature of the MaxSPRT is that the users do not have to do any of these calculations themselves but, rather, can simply use the tables provided in this article in the same old-fashioned way that we used to do for most statistical distribution functions. The only exception is if the user wants to use some other parameter values for the α level, for the upper limit on the length of surveillance, or for the matching ratio. The exact values of these design parameters are not critical for most applications, though, and for drug and vaccine safety surveillance it will almost always be possible to choose suitable parameter values from those included in the tables.

7.4. Weekly Vaccine Safety Surveillance

The examples provided in this article used historical data to mimic a real-time surveillance system. The CDC-sponsored VSD project has been or is currently using MaxSPRT for weekly surveillance of the safety of meningococcal (Lieu et al., 2007), tetanus-diphtheria-pertussis (Yih et al., 2009), rotavirus (Belongia et al., 2010), measles-mumps-rubella-varicella (Klein et al., 2010), human papillomavirus, seasonal influenza, and H1N1 influenza vaccines. For most of the vaccines, no safety problems have been detected. For the combined measles-mumps-rubella-varicella vaccine, the MaxSPRT detected an increases risk of febrile seizures, leading the Advisory Committee on Immunization Practices to revise their recommendations for its use (Klein et al., 2010). The method has also been evaluated for drug safety surveillance using historical data from the HMO Research Network (Brown et al., 2007).

For the Poisson-based MaxSPRT it is necessary to choose a comparison group to calculate the expected counts. Likewise, for the binomial-based MaxSPRT, it is necessary to choose a set of matched unexposed time periods or individuals for each exposed person. As in any observational study, different designs are prone to different types of confounding and bias. One option is to choose a historical comparison group of people having received the old vaccine that the new vaccine under surveillance is meant to replace in a Poisson MaxSPRT analysis. For example, when monitoring the safety of the measles-mumps-rubella-varicella vaccine, historical recipients of the older measles-mumps-rubella were used as the control group. This helps to ensure that the two populations are reasonably similar, but there could be bias due to, for example, temporal changes in disease incidence or disease coding practices. To overcome the latter, one may instead use concurrent matched controls and a binomial MaxSPRT, comparing individuals receiving the vaccine with age- and gender-matched controls who had a well-care visit around the same time. This resolves the issue of temporal trends, but individuals receiving the vaccines may be generally healthier or less healthy than their matched controls, introducing a different type of bias. A third option that has been used is a sequential self-control design, where the number of adverse events in an exposed time window just after vaccination is compared to the number of adverse events in an unexposed time window either before the vaccination or long after vaccination. This removes any bias due to differences between individuals. If a pre-vaccination comparison window is used, though, there is a potential for confounding due to indication or contraindication because a person diagnosed with the adverse event of interest may be more or less prone to receive the vaccination, creating biased results. If a comparison window long after vaccination is used, that reduces the timeliness of the

surveillance system. Moreover, both self-control designs could suffer from bias if there is seasonal variation in both the vaccine administration and the adverse event. The severity of each of these potential sources of confounding depends on both the vaccine and the adverse event under surveillance. Because different designs are prone to different types of bias, it is sometimes worthwhile to use more multiple designs for the same vaccine and adverse event pair. In two medically oriented papers, the pros and cons of these different designs are discussed in more detail when used for vaccine and drug safety surveillance (Brown et al., 2007; Lieu et al., 2007).

ACKNOWLEDGMENTS

This work was supported in part by the Centers for Disease Control and Prevention through the Vaccine Safety Datalink Project and in part by grant HS10391 to the HMO Research Network Center for Education and Research on Therapeutics (CERTs) from the Agency for Health Care Research and Quality. We thank Ruihua Yin for data support, Elizabeth Pfoh for creating the figure, and Paul Gargiullo for valuable comments on an earlier draft.

REFERENCES

Abt, K. (1998). Poisson Sequential Sampling Modified Towards Maximal Safety in Adverse Event Monitoring, *Biometrical Journal* 40: 21–41.

Bangdiwala, S. I. (1982). A Sequential Likelihood Ratio Test for General Hypotheses, *Sequential Analysis* 1: 57–80.

Belongia, E. A., Irving, S. A., Shui, I. M., Kulldorff, M., Lewis, E., Yin, R., Lieu, T. A., Weintraub, E., Yih, W. K., Li, R., Baggs, J., and the Vaccine Safety Datalink Investigation Group (2010). Real-Time Surveillance to Assess Risk of Intussusception and Other Adverse Events after Pentavalent, Bovine-Derived Rotavirus Vaccine, *Pediatric Infectious Disease Journal* 29: 1–5.

Brown, J. S., Kulldorff, M., Chan, K. A., Davis, R. L., Graham, D., Pettus, P. T., Andrade, S. E., Raebel, M. A., Herrinton, L., Roblin, D., Boudreau, D., Smith, D., Gurwitz, J. H., Gunter, M. J., and Platt, R. (2007). Early Detection of Adverse Drug Events within Population-Based Health Networks: Application of Sequential Testing Methods, *Pharmacoepidemiology and Drug Safety* 16: 1275–1284.

Chen, R. T., Glasser, J. W., Rhodes, P. H., Davis, R. L., Barlow, W. E., Thompson, R. S., Mullooly, J. P., Black, S. B., Shinefield, H. R., Vadheim, C. M., Marcy, S. M., Ward, J. I., Wise, R. P., Wassilak, S. G., Hadler, S. C., and the Vaccine Safety Datalink Team (1997). Vaccine Safety Datalink Project: A New Tool for Improving Vaccine Safety Monitoring in the United States, *Pediatrics* 99: 765–773.

Davis, R. L., Kolczak, M., Lewis, E., Nordin, J., Goodman, M., Shay, D. K., Platt, R., Black, S., Shinefield, H., and Chen, R. T. (2005). Active Surveillance of Vaccine Safety: A System to Detect Early Signs of Adverse Events, *Epidemiology* 16: 336–341.

DuMouchel, W. (1999). Bayesian Data Mining in Large Scale Frequency Tables, with an Application to the FDA Spontaneous Reporting System, *American Statistician* 53: 177–190.

Ghosh, B. K. (1965). Sequential Range Tests for Components of Variance, *Journal of American Statistical Association* 60: 826–836.

Ghosh, B. K. (1970). *Sequential Tests for Statistical Hypothese*, Reading, MA: Addison-Wesley.

Ghosh, M., Mukhopadhyay, N., and Sen, P. K. (1997). *Sequential Estimation*, New York: Wiley.

Govindarajulu, Z. (2004). *Sequential Statistics*, Singapore: World Scientific Publishing Company.

Hoel, D. G., Weiss, G. H., and Simon, R. (1976). Sequential Tests for Composite Hypotheses with Two Binomial Populations, *Journal of the Royal Statistical Society* B38: 302–308.

Holm, S. (1985). On the Optimality of Differentiated SPR Tests of Composite Hypotheses, *Metrika* 32: 15–33.

Huang, W. (2004). Stepwise Likelihood Ratio Statistics in Sequential Studies, *Journal of Royal Statistical Society* B66: 401–409.

Jennison, C. and Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*, Boca Raton, FL: Chapman and Hall/CRC.

Joanes, D. N. (1972). Sequential Tests of Composite Hypotheses, *Biometrika* 59: 633–637.

Klein, N. P., Fireman, B., Yih, W. K., Lewis, E., Kulldorff, M., Ray, P., Baxter, R., Hambidge, S., Nordin, J., Naleway, A., Belongia, E. A., Lieu, T., Baggs, J., and Weintraub, E., for the Vaccine Safety Datalink (2010). Measles-Mumps-Rubella-Varicella Combination Vaccine and the Risk of Febrile Seizures, *Pediatrics* 126: e1–e8.

Lachin, J. M. (1981). Sequential Clinical Trials for Normal Variates Using Interval Composite Hypotheses, *Biometrics* 37: 87–101.

Lai, T. L. (1988). Nearly Optimal Sequential Tests of Composite Hypotheses, *Annals of Statistics* 16: 856–886.

Lai, T. L. (2001). Sequential Analysis: Some Classical Problems and New Challenges, *Statistica Sinica* 11: 303–351.

Lechner, J. A. (1962). Optimum Decision Procedures for a Poisson Process Parameter, *Annals of Mathematical Statistics* 33: 1384–1402.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, Second edition. New York: Springer-Verlag.

Lieu, T. A., Kulldorff, M., Davis, R. L., Lewis, E. M., Weintraub, E., Yih, K. W., Yin, R., Brown, J. S., and Platt, R. (2007). Real-Time Vaccine Safety Surveillance for the Early Detection of Adverse Events, *Medical Care* 45: S89–S95.

Lorden, G. (1973). Open-Ended Tests for Koopman-Darmois Families, *Annals of Statistics* 1: 633–643.

Meeker, W. Q. (1981). A Conditional Sequential Test for the Equality of Two Binomial Proportions, *Applied Statistics* 30: 109–115.

Mukhopadhyay, N. and de Silva, B. M. (2009). *Sequential Methods and Their Applications*, Boca Raton, FL: Chapman and Hall/CRC.

O'Neill, R. T. and Szarfman, A. (2001). Some U.S. Food and Drug Administration Perspectives on Data Mining for Pediatric Safety Assessment, *Current Therapeutic Research* 62: 650–663.

Partridge, A. and Yeh, S. H. (2003). Clinical Evaluation of a DTaP-HepB-IPV Combined Vaccine, *American Journal of Managed Care* 9: S13–S22.

Peskir, G. and Shiryaev, A. N. (2000). Sequential Testing Problems for Poisson Processes, *Annals of Statistics* 28: 837–859.

R Development Core Team. (2009). *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing.

Schipper, M., den Hartog, J., and Meelis, E. (1997). Sequential Analysis of Environmental Monitoring Data: Optimal SPRTs, *Environmetrics* 8: 29–41.

Schwarz, G. (1962). Asymptotic Shape of Bayes Sequential Testing Regions, *Annals of Mathematical Statistics* 33: 224–236.

Siegmund, D. and Gregory, P. (1980). A Sequential Clinical Trial for Testing $p1 = p2$, *Annals of Statistics* 8: 1219–1228.

Szarfman, A., Machado, S. G., and O'Neill, R. T. (2002). Use of Screening Algorithms and Computer Systems to Efficiently Signal Higher-Than-Expected Combinations of Drugs and Events in the U.S. FDA's Spontaneous Reports Database, *Drug Safety* 25:381–392.

van der Tweel, I., Kaaks, R., and van Noord, P. A. H. (1996). Comparison of One-Sample Two-Sided Sequential t-Tests for Application in Epidemiological Studies, *Statistics in Medicine* 15: 2781–2795.

Wald, A. (1945). Sequential Tests of Statistical Hypotheses, *Annals of Mathematical Statistics* 16: 117–186.

Wald, A. (1947). *Sequential Analysis*, New York: Wiley.

Weiss, L. (1953). Testing One Simple Hypothesis Against Another, *Annals of Mathematical Statistics* 24: 273–281.

Woodroofe, M. (1978). Large Deviations of Likelihood Ratio Statistics with Applications to Sequential Testing, *Annals of Statistics* 6: 72–84.

Yih, W. K., Nordin, J. D., Kulldorff, M., Lewis, E., Lieu, T., Shi, P., and Weintraub, E. (2009). An Assessment of the Safety of Adolescent and Adult Tetanus-Diphtheria-Acellular Pertussis (Tdap) Vaccine, Using Active Surveillance for Adverse Events in the Vaccine Safety Datalink, *Vaccine* 27: 4257–4262.