

MINI-SENTINEL METHODS DEVELOPMENT

SEQUENTIAL TESTING WORKING GROUP REPORT

Prepared by: Andrea J. Cook, PhD,^{1,2} Ram C. Tiwari, PhD,³ Robert D. Wellman, MS,¹ Susan R. Heckbert, MD, PhD,^{4,5} Lingling Li, PhD,⁶ Patrick J. Heagerty, PhD,² Tracey L. Marsh, MS,⁵ and Jennifer C. Nelson, PhD^{1,2}

Author Affiliations: 1. Biostatistics Unit, Group Health Research Institute, Seattle, WA 2. Department of Biostatistics, University of Washington, Seattle, WA 3. Office of Biostatistics, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD 4. Departments of Epidemiology and Pharmacy, University of Washington, Seattle, WA 5. Group Health Research Institute, Seattle, WA 6. Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA

May 12, 2011

Mini-Sentinel is a pilot project sponsored by the [U.S. Food and Drug Administration \(FDA\)](#) to inform and facilitate development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products. Mini-Sentinel is one piece of the [Sentinel Initiative](#), a multi-faceted effort by the FDA to develop a national electronic system that will complement existing methods of safety surveillance. Mini-Sentinel Collaborators include Data and Academic Partners that provide access to health care data and ongoing scientific, technical, methodological, and organizational expertise. The Mini-Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223200910006I.

Mini-Sentinel Methods Development

Sequential Testing Working Group Report

I. OVERVIEW OF STATISTICAL APPROACHES FOR GROUP SEQUENTIAL MONITORING OF POSTMARKET SAFETY SURVEILLANCE DATA: CURRENT STATE OF THE ART FOR USE IN MINI-SENTINEL.....	3
A. INTRODUCTION.....	3
B. METHODS.....	4
1. <i>Data Specifications and Notations</i>	4
2. <i>Sequential Testing Framework</i>	4
3. <i>Group Sequential Statistical Methods</i>	5
C. HYPOTHETICAL DATA EXAMPLE.....	11
D. SUMMARY OF SEQUENTIAL METHODS BEFORE SIMULATION EVALUATION.....	12
E. TABLES AND FIGURES FOR SECTION I	12
II. PROJECT SUMMARY AND SIMULATION RESULTS	15
A. OUTLINE OF SIMULATION DESIGN.....	16
1. <i>Case 1: Single-Time Exposure</i>	16
2. <i>Case 2: Chronic Exposure</i>	18
B. RESULTS OF SIMULATION	19
1. <i>Case 1: Single-Time Exposure</i>	19
2. <i>Case 2: Chronic Exposure</i>	21
C. DISCUSSION	23
D. TABLES AND FIGURES FOR SIMULATION STUDY SECTION II.....	25
III. REFERENCES	59
IV. APPENDIX: STATISTICAL SOFTWARE CODE FOR THE CSSP AND GSEE.....	61
A. SAS® PROGRAMMING CODE.....	61
1. <i>CSSP</i>	61
2. <i>GS EE</i>	68
B. R PROGRAMMING CODE	83
1. <i>CSSP</i>	83
2. <i>GS EE</i>	88

I. OVERVIEW OF STATISTICAL APPROACHES FOR GROUP SEQUENTIAL MONITORING OF POSTMARKET SAFETY SURVEILLANCE DATA: CURRENT STATE OF THE ART FOR USE IN MINI-SENTINEL

A. INTRODUCTION

Monitoring the safety of marketed medical products is a pressing public health need. Therapeutic and prevention products, such as vaccines, drugs, and devices, go through rigorous clinical trials evaluating efficacy and safety before being approved, but these trials are generally not of sufficient size to systematically detect rare adverse events and do not always include participants similar to the population that receives the products after their marketing. Therefore, the Centers for Disease Control (CDC) and the Food and Drug Administration (FDA) have begun to utilize large multi-site healthcare databases to conduct postmarket surveillance studies for medical product safety. The FDA's Sentinel Initiative is an example of a program designed to improve the evaluation of safety across a large array of FDA-regulated medical products.¹ Mini-Sentinel is a pilot project intended to inform and facilitate development of a fully operational Sentinel System.

This paper describes statistical methods for the evaluation of recently approved products using a prospective cohort observational study design with existing electronic healthcare data for prespecified safety outcomes. The goal of this study design is to detect potential safety concerns quickly by sequentially monitoring effect estimates multiple times throughout a study. The research aim is to determine whether, for a prespecified set of safety outcomes, an excess rate of observed events exists in recipients of the medical product of interest (MPI) compared with a single comparison group. The comparison group is important and can be chosen in several ways. In this report, we consider a concurrent control group defined to be comparable to those taking the MPI after controlling for confounders. For example, when evaluating a new diabetes drug for safety, an appropriate comparison group could be those taking an alternative diabetes drug. However, we would need to control for site and perhaps patient characteristics, since physicians from the various sites contributing data may exhibit different prescribing habits and patient characteristics may be associated with choice of diabetes drug.

This type of safety evaluation has been coined "signal refinement" because potential adverse events are predefined based upon the suggestion of a potential risk observed either during pre-approval, in the FDA's adverse event reporting system (AERS),² or due to known biologic reasons uncovered during the study of similar medical products. This signal refinement stage can be thought of as a preliminary step prior to conducting a more extensive phase IV observational study or confirmatory randomized clinical trial because existing healthcare databases, which are typically constructed for non-research purposes, tend to have issues such as incomplete data, data errors, and lack of information on potential confounders. Several examples of signal refinement studies have been published but this is still a relatively new area.³⁻⁷

Statistical methods used to address hypotheses within postmarket study designs must be able to detect both rare and common adverse events, control for confounding, and maintain the overall type I error across multiple tests. This report describes the current state of statistical methods developed to conduct sequential analysis of prospective cohort data for medical product safety. We present four sequential methods that use different approaches to handle confounding, hold the overall type I error, and have different statistical properties such as time to signal detection and power. Controlling for confounding is a major concern for observational safety surveillance and distinguishes it from the randomized clinical

trial setting in which most sequential monitoring methods have been developed. Further, the outcomes of interest are often rare and therefore inferential properties that hold in randomized trials, such as large sample asymptotics, may not hold in this setting and need to be assessed. We focus on methods already applied to observational safety surveillance studies but, for comparison, also introduce one general method used in randomized clinical trials that is applicable to safety surveillance. We discuss potential limitations of these methods and conclude with discussion of the need for future work tailored to the setting that we describe.

B. METHODS

The electronic data generally captured for signal refinement by systems like Mini-Sentinel are primarily administrative and collected by health plans during the course of routine health care practice. Mini-Sentinel uses a distributed data system, in which individual level data, defined using a common data model, remain at the local site. Distributed programs aggregate event and sample size counts at each site, stratified by exposure group and by confounders. These results are compiled across sites for analysis. Although in some cases analyses may be based on individual level data stripped of direct identifiers, more often de-identified information is combined for central analysis to further protect patient privacy, and thus the focus of this discussion remains on aggregate data.

1. Data Specifications and Notations

We assume that accruing data will be analyzed at specified times ($t=1, \dots, T$). We also assume that each individual i at analysis t ($t=1, \dots, T$), is either exposed to the MPI, $D_i=1$, or not exposed, $D_i=0$, and either has the outcome of interest occurring before the end of analysis t , $Y_i(t)=1$, or does not $Y_i(t)=0$. The exposure time, $E_i(t)$, denotes the cumulative exposure time prior to analysis t . It could be a single-time exposure window (e.g., vaccine: $E_i(t)=1$ for all individuals) or a chronic exposure (time on either MPI or comparator), for which assumptions of the exposure time and outcome relationship must be made (constant risk or change in risk due to exposure duration). For this report we censor participant's exposure time at the date of disenrollment, occurrence of outcome, or discontinuation of use of the initial assigned treatment. Further, participants are censored if they switch exposure groups and begin taking the other medical product (e.g., an exposed individual starts taking the comparator product). These assumptions are consistent with incident user cohort studies that are currently being used in postmarket surveillance.⁸

Further, we assume that a set of baseline confounders, Z_i , is associated with individual i and can be comprised of variables such as age, sex, site, and health conditions. When using aggregate data these confounders are often categorized to form a set of categorical confounders, Z_i^c . For example, a continuous confounder, such as age, can be categorized into 5- or 10-year age groups. Under this data setup, confounding can be addressed by regression, stratification, or matching.

2. Sequential Testing Framework

In a signal refinement study, the overall hypothesis of interest is whether those using the MPI ($D_i=1$) have a higher event rate compared to the unexposed group ($D_i=0$) after accounting for confounding and exposure time. Numerous test statistics (based on the relative risk or hazard ratio, for example) can be derived to evaluate this hypothesis, thus creating different statistical methods. The chosen hypothesis is tested at each analysis t . If the test statistic at analysis t exceeds a predefined critical boundary, $c(t)$, it

signals a significantly elevated rate of events at analysis t ; otherwise, the study continues to the next analysis time until the predefined end of the study. At each analysis more new information accumulates, which may include new participants exposed and unexposed to the MPI since the last analysis, as well as more follow-up or exposure time for participants already included in the previous analysis. Different approaches to incorporating updated data induce different assumptions that need to be accounted for in the calculation of the critical boundary. The critical boundary can be chosen in numerous ways, but it must hold the overall type I error rate across all analyses, taking into account both multiple testing and a skewed testing distribution that conditions on whether or not earlier test statistics exceeded the specified critical value at previous analysis times. A general review of sequential monitoring boundaries has been presented by Emerson et al.⁹ and is beyond the scope of this paper, but we will present approaches specific to the observational surveillance setting and one general method used in randomized clinical trials that is applicable to this area.

3. Group Sequential Statistical Methods

a. Lan-Demets Group Sequential approach using error spending (GS LD)

The first method we consider is a general group sequential method used mainly in randomized clinical trials developed by Lan and Demets¹⁰ using an error spending approach. An error spending approach uses the concept of cumulative alpha or type I error, $\alpha(t)$, defined as the cumulative amount of type I error spent at analysis t and all previous analyses, $1, \dots, t-1$. We assume that $0 < \alpha(1) \leq \dots \leq \alpha(T) = \alpha$, where α is the overall type I error to be spent across the study period. The function $\alpha(t)$ can be any increasing monotonic function that preserves family-wise error, but there are several common approaches including the Pocock¹¹ boundary function $\alpha(t) = \log(1 + (\exp(1) - 1)t/T)\alpha$, the O'Brien-Fleming¹² boundary function $\alpha(t) = 2 \left(1 - \Phi \left(\frac{Z_{1-\alpha/2}}{\sqrt{t/T}} \right) \right)$, and the general power boundary function $\alpha(t) = (t/T)^p \alpha$ for $p > 0$. The most commonly used boundary function for safety evaluation studies has been a flat, Pocock-like boundary on a standardized test statistic scale. This boundary spends α evenly across analyses, given the test statistic is asymptotically normally distributed. Therefore it spends more α at earlier analyses relative to later analyses given the amount of statistical information, or sample size, observed up to time t compared to an O'Brien Fleming boundary, which is commonly used in efficacy studies. This flat boundary has been discussed as Pocock-like, but a Pocock boundary when testing more frequently (quarterly or more often) is not completely flat. For further discussion of boundary shapes and statistical trade-offs between them in practice for postmarket surveillance, see Nelson et al.¹³

Given the error spending boundary function, Lan and Demets developed an asymptotic conditional sequential monitoring boundary for any asymptotically normal test statistic based on independent increments of data.¹⁰ This boundary can be computed and used to compare to almost any standardized test statistic, including one that controls for confounding. For example, when interest is in an adjusted relative risk, $\hat{R}(t)$, or log relative risk, it can be estimated using Poisson regression and a standardized test statistic can be calculated, $Z_{val}(t) = \hat{R}(t) / \sqrt{\text{Var}(\hat{R}(t))}$. The value of $Z_{val}(t)$ can then be compared to the asymptotic conditional monitoring boundary developed by Lan and Demets,¹⁰ resulting in a decision to stop if $Z_{val}(t)$ exceeds the monitoring boundary or to continue collecting additional data. This is an appealing approach because the boundary is very simple to calculate and relies on a well-defined asymptotic distribution. However, in practice, with rare events and frequent testing (small amount of new information between analyses), the asymptotic properties of the boundary fail to hold.

This is similar to the scenario in which an exact test may be preferred to an asymptotically normal test when the sample size is small. The following methods have sought to address the shortcomings of this approach to allow for more precise statistical performance in a wider variety of settings.

b. Group Sequential Likelihood Ratio Test (GS LRT)

The group sequential likelihood ratio test (LRT) approach is a method that has been used in the Vaccine Safety Data Link project to monitor vaccine safety for a single-time vaccine exposure.^{3, 6, 7, 14} The approach uses exposure matching with a fixed matching ratio (1:M) to control for confounding and then computes a LRT statistic. The most commonly used method is the Binomial maxSPRT,¹⁴ which assumes continuous monitoring (i.e., after each matched set of exposed and unexposed come into the dataset, the test statistic is compared to the monitoring boundary).

Specifically, for the maxSPRT method, one creates matched exposure strata, s ($s=1, \dots, S$), such that each exposed individual, with $D_{s1}=1$, is matched to one or more unexposed individuals ($D_{s2}=0, \dots, D_{s(M+1)}=0$) who have the same categorical confounders, Z_i^c . Then the log LRT statistic at each analysis, t , is the following:

$$LLR^1(t) = \log \left(\frac{\left(\frac{Y_{D=1}(t)}{Y(t)} \right)^{Y_{D=1}(t)} \left(\frac{Y_{D=0}(t)}{Y(t)} \right)^{Y_{D=0}(t)}}{\left(\frac{1}{M+1} \right)^{Y_{D=1}(t)} \left(\frac{M}{M+1} \right)^{Y_{D=0}(t)}} \right),$$

where $Y_{D=1}(t) = \sum_{s=1}^{S(t)} \sum_{j=1}^{M+1} Y_{sj} D_{sj}$ and $Y_{D=0}(t) = \sum_{s=1}^{S(t)} \sum_{j=1}^{M+1} Y_{sj} (1 - D_{sj})$ are the number of events observed

among those exposed and unexposed to the MPI up to time t , respectively, and

$Y(t) = Y_{D=1}(t) + Y_{D=0}(t)$ is the total number of events up to time t . Note that $S(t)$ is the number of strata up to time t , which is also the number of exposed participants since we are assuming a fixed matching ratio of 1:M. This particular LRT, which conditions on the total number of events, $Y(t)$, is designed for the rare event case in which only one event is expected to be observed per exposure stratum. One can think of this LRT as comparing the observed proportion of exposed (and unexposed) events out of the total number of events to the expected proportion under the null, which is just $1/(M+1)$ for the exposed participants and $M/(M+1)$ for the unexposed participants.

However, when events are not extremely rare, or when the probability within a stratum of more than one event occurring is not small, the assumptions of this LRT are violated and a more general two-sample binomial likelihood ratio test statistic should be used:

$$LLR^2(t) = \log \left(\frac{\left(\frac{Y_{D=1}(t)}{N_{D=1}(t)} \right)^{Y_{D=1}(t)} \left(1 - \frac{Y_{D=1}(t)}{N_{D=1}(t)} \right)^{N_{D=1}(t) - Y_{D=1}(t)} \left(\frac{Y_{D=0}(t)}{N_{D=0}(t)} \right)^{Y_{D=0}(t)} \left(1 - \frac{Y_{D=0}(t)}{N_{D=0}(t)} \right)^{N_{D=0}(t) - Y_{D=0}(t)}}{\left(\frac{Y(t)}{N(t)} \right)^{Y(t)} \left(1 - \frac{Y(t)}{N(t)} \right)^{N(t) - Y(t)}} \right),$$

where $N_{D=1}(t) = \sum_{s=1}^{S(t)} \sum_{j=1}^{M+1} D_{sj} = S(t)$ and $N_{D=0}(t) = \sum_{s=1}^{S(t)} \sum_{j=1}^{M+1} (1 - D_{sj}) = M \times S(t)$ are the number of people exposed and unexposed to the medical product up to time t , respectively, and $N(t) = N_{D=1}(t) + N_{D=0}(t)$ is the total sample size up to time t . Note that this general LRT incorporates the total sample size, unlike the binomial maxSPRT LRT that is conditional on total number of events. For rare events the performances of each LRT are similar. Further evaluation needs to be conducted to establish the scenarios in which each LRT has better statistical properties.

For the binomial maxSPRT, a Pocock-like boundary has been proposed, $c(t)=a$, which is a flat boundary on the log LRT statistic. One common way to solve for the constant, a , uses an iterative simulation approach similar to the following:

Step 1:

Simulate data assuming H_0 and the observed event rate while controlling for confounding (i.e., using a permutation approach: fix Y_{s1}, \dots, Y_{sM} ($s = 1, \dots, S$) and permute D_{s1}, \dots, D_{sM} to create D_{s1*}, \dots, D_{sM*} so that you hold the exposure strata relationships and thus control for confounding).

Step 2:

Calculate $LLR(t)$ on the simulated dataset.

Step 3:

If $LLR(t) \geq a$ then $Signal_k=1$ and stop loop, otherwise continue to next $t+1$

Step 4:

If $t=T$ then $Signal_k=0$

This process is repeated a large number of times, $Nsim$, and the estimated α -level for the boundary is calculated as $\hat{\alpha} = \sum_{k=1}^{Nsim} Signal_k / Nsim$. One solves for a by repeating the simulation and changing a until $\hat{\alpha} = \alpha$.

This approach is a special case of the general unifying boundary approach developed by Kittleson et al.¹⁵ To allow for the more general approach, define $c(t)=a \times u(t)$, where $u(t)$ is a function dependent upon the proportion of statistical information (e.g., sample size) up to time t and is of the form $u(t)=(N(T)/N(t))^{1-2\Delta}$, where $\Delta>0$ is a fixed parameter depending upon design (e.g., $u(t)=1$ is Pocock and $u(t)=(N(T)/N(t))^{0.5}$ is O'Brien and Fleming). The same approach is used to solve iteratively for a , but the boundary $c(t)$ will now be shaped differently depending upon $u(t)$. We have named this more flexible version of the binomial maxSPRT as the Group Sequential Likelihood Ratio Test (GS LRT). This additional flexibility allows the method to be applied more generally, for example with the Mini-Sentinel pilot, where data are not available as often (potentially quarterly). Further, the shape of boundary can be changed to reflect the desired trade-offs appropriate to the specific safety question of interest. Since the original binomial maxSPRT used a unifying boundary type approach, we have presented it as such here but, as has been shown by others,¹⁶ the error spending approach and unifying approach are complementary and therefore we could have chosen an error spending approach.

A potential limitation of the GS LRT method is the fixed matching ratio. In practice, if there is a need to implement a strict matching criterion due to the need for strong confounding control, then it can be difficult to find M unexposed matches for each exposed participant, especially in the scenario of frequent monitoring. Frequent monitoring typically implies that an exposed participant should be matched to M unexposed participants within the current analysis timeframe. This can lead to loss of matched strata including strata with events. Often the matching criterion is then loosened, leading to less confounding control but a larger matched cohort.

c. *Conditional Sequential Sampling Procedure (CSSP)*

The conditional sequential sampling procedure (CSSP)¹⁷ was specifically developed to handle chronically used exposures, such as drugs that are taken over a period of time. However, the approach is also able to accommodate a single-time exposure such as a vaccine. This method handles confounding using stratification and assumes that the data are aggregated.

Specifically, using categorical confounders, Z_i^c , one stratifies the entire study population (unlike GS LRT, which uses a matched sample). Then at each analysis, t , within each confounder stratum, Z_k^S ($k=1, \dots, K$), one calculates the exposure time, $E_{D,k}(t)$, and number of events, $Y_{D,k}(t)$, among all participants in stratum k on medical product D ($D=0$ (unexposed) or $D=1$ (exposed)) since the previous analysis $t-1$, where $E_{D,k}(t) = \sum_{i=1}^N (E_i(t) - E_i(t-1))I(Z_i^c = Z_k^S \text{ and } D_i = D)$ and

$Y_{D,k}(t) = \sum_{i=1}^N (Y_i(t) - Y_i(t-1))I(Z_i^c = Z_k^S \text{ and } D_i = D)$. Under H_0 (no relationship between exposure to the MPI and the outcome conditional on strata) the conditional distribution of

$Y_{D=1,k}(t) | Y_{D=1,k}(t) + Y_{D=0,k}(t)$ is *Binomial* $\left(Y_{D=1,k}(t) + Y_{D=0,k}(t), \frac{E_{D=1,k}(t)}{E_{D=1,k}(t) + E_{D=0,k}(t)} \right)$, which is based

on the proportion of exposure time observed for those exposed compared with the total exposure time including exposed and unexposed. Using this stratum-specific conditional distribution, one can simulate the distribution of $Y_{D=1,k}(t)$, the number of outcomes among those on the MPI within each stratum under H_0 given $Y_{D=1,k}(t) + Y_{D=0,k}(t)$.

The test statistic of interest is then the total number of adverse events observed among those exposed up to time t across all strata, $Y_{D=1}(t) = \sum_{k=1}^K Y_{D=1,k}(t)$. The CSSP approach uses an error spending approach in combination with the conditional stratum-specific distributions to create the sequential monitoring boundary. Specifically, it uses the following iterative simulation approach:

Step 1:

Create a single realization of the following dataset of observed exposed counts under H_0 for analysis t , $t=1, \dots, T$ as follows:

For all confounder strata k , simulate $\tilde{Y}_{D=1,k}(t) \sim \text{Binomial}\left(Y_k(t), \frac{E_{D=1,k}(t)}{E_{D=1,k}(t) + E_{D=0,k}(t)}\right)$ if $\tilde{Y}_k(t) > 0$
 else set $\tilde{Y}_{D=1,k}(t) = 0$.

Calculate $\tilde{Y}_{D=1}(t) = \sum_{j=1}^t \sum_{k=1}^K \tilde{Y}_{D=1,k}(j)$ (total number of simulated exposed events at analysis t).

Step 2:

Repeat Step 1 for a large number of realizations, $Nsim$, to create a distribution of total number of exposed events at each analysis, $\tilde{Y}_{D=1}^1(t), \dots, \tilde{Y}_{D=1}^{Nsim}(t)$.

Step 3:

Order $\tilde{Y}_{D=1}^1(t), \dots, \tilde{Y}_{D=1}^{Nsim}(t)$ from smallest to largest and if $\tilde{Y}_{D=1}^j(t) > \tilde{Y}_{D=1}^{(Nsim*(1-\alpha(1)))}(t)$ then signal at analysis t , else continue.

Step 4:

Set the simulated event counts that would have signaled at this analysis,

$\tilde{Y}_{D=1}^{(Nsim*(1-\alpha(t-1)+1))}(t), \dots, \tilde{Y}_{D=1}^{(Nsim)}(t)$, to an extreme value, such as 1,000, so that these realizations will be indicated as having passed the boundary. This allows for a cumulative error spending calculation that incorporates stopping. Otherwise, keep $\tilde{Y}_{D=1}^j(t)$ from step 1 and repeat from 1 at next analysis, $t+1$.

Using this simulation approach explicitly incorporates the sequential monitoring stopping rules. Any form of the cumulative error spending function, $\alpha(t)$, can be assumed as discussed in the section on the Lan-Demets Group Sequential approach using error spending.

This CSSP approach is especially good when studying rare events, but it has limitations when there are too many strata and/or short intervals between analyses. The reason this approach breaks down is because the only informative strata are those that meet the following two criteria: 1) at least one observed event but not all participants observe an event and 2) both an exposed and unexposed participant. Further, each analysis is treated as having separate strata since information from one analysis to the next is being treated as independent. Therefore, the true number of independent strata is $K \times T$ across all analyses. So as both K and T increase, very few strata will be informative. As a result, the test statistic is less stable, which can both influence power and potentially inflate or deflate the type I error. Further, this approach assumes a constant relationship between exposure duration and the probability of an event, which may not be valid. Overall, it has nice properties for the rare event case and will be applicable to postmarket surveillance in settings where testing is not performed highly frequently or when too many confounder strata are required.

d. Group Sequential Estimating Equation Approach (GS EE)

The final approach we will present is one that controls for confounding through regression, unweighted or weighted. It can be applied to either the single exposure time or chronic exposure time settings. It has the flexibility to incorporate different exposure duration relationships but we will focus on a constant relationship (i.e., given exposure duration one assumes a constant rate of disease based just on exposure time). The approach uses a generalized estimating equation (GEE) framework and a score test statistic. Specifically, assume that the mean regression model under the null hypothesis, H_0 , of no relationship between the MPI and the event is $g(E(Y_i(t))) = \beta_0 + \beta_Z Z_i + f_\theta(E_i(t))$, where $g(\cdot)$ is the mean link function, for example, the logit for a logistic model or the logarithm for a Poisson model. The exposure link function, $f_\theta(\cdot)$, would typically be ignored for a single-time exposure or specified as the logarithmic function if using a Poisson model. However, to allow for flexibility this has been kept general.

Given the mean model, the generalized score statistic,¹⁸ $Sc(t)$, can be calculated, with the additional specification for the family from which the data have arisen, for example, a binomial family for logistic regression and a Poisson family for a log regression model. However, a nice property of GEE when using the generalized score statistic is that it only assumes that the mean model is correctly specified.¹⁹

To calculate the sequential monitoring boundary, it has been proposed to use the following permutation data distribution:

Step 1:

At each analysis t , simulate data by fixing $(Y_{N(t-1)+1}, Z_{N(t-1)+1}), \dots, (Y_{N(t)}, Z_{N(t)})$ and permuting $D_{N(t-1)+1}, \dots, D_{N(t)}$ to create $D_{N(t-1)+1}^*, \dots, D_{N(t)}^*$ and calculate $\tilde{Sc}_j(t)$.

Step 2:

Repeat Step 1 for a large number of realizations, $Nsim$, to create a distribution of Score statistics, under H_0 , at each analysis t , $\tilde{Sc}_1(t), \dots, \tilde{Sc}_{Nsim}(t)$.

The boundary can be defined following the unifying boundary formulation as outlined for the GS LRT method or an error spending approach as outlined for GS LD method, except using this permuted dataset and score test statistic. Note that we are not directly estimating the effect of D_i since a score statistic is calculated under H_0 . This allows for the test statistic to have better statistical properties, such as power, when the interest is in comparing alternative hypotheses that are closer to the null (e.g., better power relative to other methods for detecting RR=1.5 versus RR=3.0).²⁰

The potential advantages of this approach compared to the other three approaches are that it may provide more flexible confounder control compared to GS LRT or CSSP and it does not rely as heavily on the asymptotic assumptions needed for the Lan-Demets error spending approach. However, a limitation to this approach, and any regression approach, is that it requires the first analysis to have enough events and observations to estimate the parameters of the mean regression model. This can be difficult for the extremely rare event case, for which the GS LRT or CSSP approach may be preferable. As outlined by Nelson et al.,¹³ it may be advantageous in safety surveillance to delay the first test of the data until an

adequate amount of information has accrued, in which case this method may be applicable in most commonly encountered situations. Further, it requires more computational time than the well-defined asymptotically normal Lan-Demets error spending approach, so under the non-rare event case, the latter approach may be preferable for simplicity. Overall, all four approaches are applicable to the postmarket surveillance setting and a brief summary of assumptions, limitations, and advantages is outlined in **Table 1**.

C. HYPOTHETICAL DATA EXAMPLE

In this section we present a hypothetical sequential monitoring application for which the question of interest is: Does the new drug A (the MPI) have a higher rate of myocardial infarction (MI) compared to drug B. The data are from five sites and the confounders are age, sex, and body mass index (BMI). For de-identification, age is categorized into 5 year categories and BMI into four categories: low ($BMI < 18.5 \text{ kg/m}^2$), normal ($18.5 \text{ kg/m}^2 \leq BMI < 25 \text{ kg/m}^2$), overweight ($25 \text{ kg/m}^2 \leq BMI < 30 \text{ kg/m}^2$), and obese ($30 \text{ kg/m}^2 \leq BMI$).

The study is designed to sequentially monitor up to a total sample size of 10,000 participants assuming a flat, Pocock-style boundary, with the first analysis following accrual of 2,500 participants and then analyses approximately every 417 participants (19 analyses) (**Figure 1**). This scenario is akin to a two-year study with constant accrual of 10,000 participants in which the first analysis occurs after 180 days and each subsequent analysis occurs monthly thereafter. For simplicity, the uptake of each drug is equal and the expected percent with the event, MI, after two years is 5% overall. **Table 2** shows an example of such a dataset.

We now apply three of the four methods discussed previously. We will not apply the *GS LRT* method since it is not applicable outside a single-time exposure setting. For the *GS LD* and *GS EE* methods, one uses the stratum-specific cumulative event data, $Y_{cum,s} = \sum_{j=1}^t Y_{j,s}$, and exposure time data, $Exp_{cum,s} = \sum_{j=1}^t Exp_{j,s}$, at each analysis (**Table 2**: Columns 8 and 10) and fits a Poisson regression model adjusting for age, sex, and BMI categories with $\log(Exp_{cum,s})$ as an offset term. The *GS LD* method then calculates the standardized Wald statistic based on the adjusted relative risk and compares this to the normal approximation boundary developed by Lan-Demets. The *GS EE* method calculates the generalized score test statistic and compares this to the permutation-derived critical boundary. The *CSSP* approach uses the total number of events for those on drug A, $Y_{Cum,D=A}(t) = \sum_{s=1}^{S(t)} Y_{cum,s}(t) I(D_i = A)$, as the test statistic, where $S(t)$ is the total number of strata at analysis t , then calculates an analysis-specific p-value (i.e., the probability of observing this test statistic, or one more extreme, based on the simulated distribution under the null) and compares this p-value to a Pocock error spending boundary. **Figure 1** shows the different boundary shapes for the three methods.

Given these boundaries, **Tables 3a** and **3b** provide an example of the type of monitoring summary one would create for a sequential monitoring study. For this fictitious data example, the actual relative risk was two and all three methods signaled at the second analysis, but results often vary in other data settings. In this case, all methods performed equally well and there was an indication of an elevated rate of MI for those on drug A compared to drug B even after controlling for confounding.

D. SUMMARY OF SEQUENTIAL METHODS BEFORE SIMULATION EVALUATION

We have presented four different group sequential monitoring approaches that are applicable to active postmarket surveillance of administrative observational data. The theoretical underpinnings of each method have been described and illustrated using a hypothetical application. A formal evaluation of these four approaches still needs to be conducted to assess important statistical properties, such as delineation of scenarios in which a given method is appropriate (i.e., it holds the overall type I error and controls for confounding) or outperforms other methods. Performance is often quantified as having higher probability of signaling when a signal exists (power) or how quickly a method detects a signal (time to detection), which are clearly important quantities in safety surveillance. This work was begun in the first year Mini-Sentinel signal detection methods development project that will be presented in Section II.

Other methodological issues still need to be addressed. Open questions include developing better approaches to handle distributed data sources with more nuanced confounding control, extending methods to the survival context for rare adverse events, and controlling for provider or facility effects. Therefore, the statistical methods presented represent a first step toward a general methodology appropriate for the signal refinement surveillance setting.

E. TABLES AND FIGURES FOR SECTION I

Table 1. Overview of the four statistical methods for sequential monitoring including potential advantages and limitations

	Exposure Setting	Confounding Control	Test Statistic	Sequential Boundary Formulation	Potential Advantages	Potential Limitations
GSLD	Single time or chronic exposure	All: Matching, Stratification, Regression	Any standardized test statistic	Error Spending boundary derived using a normal approximation	Easy to apply, flexible confounding control	In very rare event setting, or frequent testing, the normal approximation assumptions may not hold
GSLRT	Single-time exposure	Matching with fixed matching ratio	LRT	Unifying boundary derived using permutation; potential to extend to error spending boundary	Matching provides an appealing interpretation	Information loss due to restricted sample; potential loss of exposed if matching criteria too strict or insufficient confounding control if criteria too loose
CSSP	Single time or chronic exposure	Stratification	Number of events for those on MPI	Error spending boundary derived by conditioning on number of events within strata	Works well for rare adverse events	May not hold type I error when strata are small or if testing is frequent
GSEE	Single time or chronic exposure	Regression	Score Statistic	Unifying boundary or error spending boundary derived using permutation	Flexible confounding control with few assumptions	Requires sufficient outcome data at first analysis to estimate the initial regression parameters

Table 2. Structure of the aggregated data available for analysis in data systems like the Mini-Sentinel Distributed Database.

Analysis	Stratum	Site	AgeCat	Sex	BMIcat	Drug	$Y_{cum,s}^1$	$Y_{t,s}^2$	$Exp_{cum,s}^3$	$Exp_{t,s}^4$
1	1	1	40to45	Male	Normal	A	2	2	250	250
1	1	1	40to45	Male	Normal	B	2	2	280	285
.
1	200	5	70to75	Female	Obese	A	11	11	720	720
1	200	5	70to75	Female	Obese	B	16	16	750	750
2	1	1	40to45	Male	Normal	A	2	0	320	70
2	1	1	40to45	Male	Normal	B	4	2	330	50
.
2	200	5	70to75	Female	Obese	A	13	2	780	60
2	200	5	70to75	Female	Obese	B	17	1	800	50
.
.

¹ $Y_{cum,s}$ is the total cumulative events observed at and before analysis t within each stratum s .
² $Y_{t,s}$ is the events observed only at analysis t within each stratum s .
³ $Exp_{cum,s}$ is the total cumulative exposure time observed at and before analysis t within each stratum s .
⁴ $Exp_{t,s}$ is the exposure time observed only at analysis t within each stratum s .

Table 3a. Examples of monitoring data for the GS LD and GS EE methods when comparing observed test statistics to a standardized test statistic sequential boundary based on outcomes with prevalence 0.05 over the two year study and confounding when the actual adjusted relative risk is 2.

Analysis	Time (mths)	Y_{cum}^1	$Y_{cum,D=A}^2$	Exp_{cum}^3 (person-days)	$Exp_{cum,D=A}^4$ (person-days)	RR_{AtoB}^5	TestStat ⁶	Test Statistic Boundary ⁷	Signal
GS LD									
1	6	73	53	193,373	96,634	1.76	2.09	2.10	No
2	7	97	75	252,559	125,366	2.38	3.48	2.31	Yes
3	8	116	88	314,716	155,774	2.12			
4	9	143	107	379,954	187,629	2.09			
...									
19	24	514	379	1,454,836	703,747	2.06			
GS EE									
1	6	73	53	193,373	96,634	1.76	2.16	2.28	No
2	7	97	75	252,559	125,366	2.38	3.65	2.28	Yes
3	8	116	88	314,716	155,774	2.12			
4	9	143	107	379,954	187,629	2.09			
...									
19	24	514	379	1,454,836	703,747	2.06			

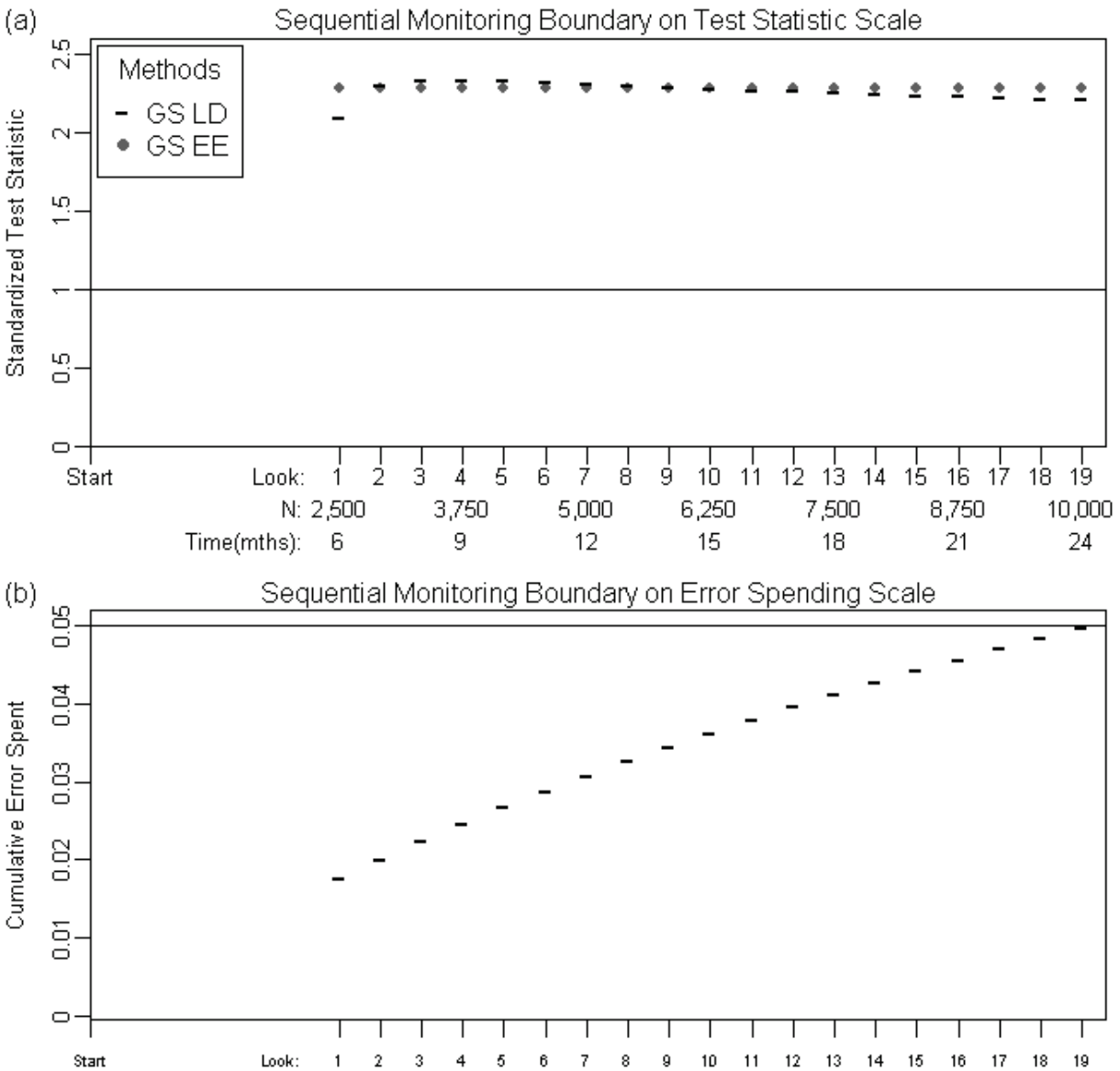
¹ Y_{cum} is the total cumulative events observed at and before analysis t .
² $Y_{cum,D=A}$ is the total cumulative events observed at or before analysis t for those on drug A.
³ Exp_{cum} is the total cumulative exposure time observed at and before analysis t .
⁴ $Exp_{cum,D=A}$ is the total cumulative exposure time observed at and before analysis t for those on drug A.
⁵ RR_{AtoB} is the adjusted relative risk (RR) comparing drug A to drug B at each analysis adjusting for site, age, sex, and BMI category.
⁶ TestStat is the observed test statistic calculated at each analysis and is the Wald based test for GS LD and score based test for GS EE.
⁷ Test Statistic Boundary is the critical boundary in which the Test statistic is compared to indicate if a given analysis has signaled.

Table 3b. Example of monitoring data for the CSSP method when comparing the estimated probability of observing number of observed outcomes in Drug Group A compared to an error spending sequential monitoring boundary based on outcomes with prevalence 0.05 over the two year study and confounding.

Analysis	Time (mths)	Y_{cum}^1	$Y_{cum,D=A}^2$	Exp_{cum}^3 (person-days)	$Exp_{cum,D=A}^4$ (person-days)	Analysis P-value ⁵	Error Spending Boundary ⁶	Signal
CSSP								
1	6	73	53	193,373	96,634	0.020	0.017	No
2	7	97	75	252,559	125,366	0.012	0.020	Yes
3	8	116	88	314,716	155,774			
4	9	143	107	379,954	187,629			
...								
19	24	514	379	1,454,836	703,747			

¹ Y_{cum} is the total cumulative events observed at and before analysis t .
² $Y_{cum,D=A}$ is the total cumulative events observed at or before analysis t for those on drug A.
³ Exp_{cum} is the total cumulative exposure time observed at and before analysis t .
⁴ $Exp_{cum,D=A}$ is the total cumulative exposure time observed at and before analysis t for those on drug A.
⁵ Analysis P-value is the cumulative probability of observing $Y_{cum,D=A}$ Or something more extreme at or before analysis t .
⁶ Error spending boundary is the amount of cumulative alpha one specifies to spend at a given analysis. Given the error spending boundary one computes the current p-value at each analysis and if that current p-value is less than the error spending boundary then the given analysis has signaled.

Figure 1. Sequential Monitoring boundaries for a flat, Pocock-style boundary with a sample size of 10,000 participants with the first analysis after the first 2,500 participants and then approximately every 417 participants (19 analyses) using a) *GS LD* and *GS EE* boundaries based on a standardized test statistic and b) *CSSP* boundary based on the error spending approach.



II. PROJECT SUMMARY AND SIMULATION RESULTS

The goal of this signal detection methods development project was to evaluate and develop statistical approaches for postmarket medical product safety surveillance that specifically focus on confounding control. The first step of the project was to outline preexisting methods applicable to the Mini-Sentinel pilot. We conducted a literature review of established approaches and, based on the assumptions each

made, determined their applicability to Mini-Sentinel. Since the sequential monitoring literature is extensive, we focused on methods for a concurrent control setting that would be flexible enough to extend to a group sequential monitoring approach. We also included a new approach, *GS EE*, that has not been previously published but is being developed by Cook et al.²¹ for the postmarket surveillance context and was improved based on results from this project. The conclusions from the initial literature review are outlined in Section I with four general statistical approaches summarized.

The next step of the project was to conduct a simulation study to evaluate the performance of three of the four methods outlined in Section I: *GS LD*, *CSSP*, and *GS EE*. We did not include the *GS LRT* method because it was not applicable to the chronic exposure setting. Further, evaluation of an exposure-matched study design, such as *GS LRT*, compared to full cohort study design using stratification or regression to control for confounding, requires a different simulation framework to examine the specific trade-offs. This work was outside the scope of this project; however, for completeness we presented the *GS LRT* approach in Section I because it is still applicable to postmarket surveillance for single-time exposure outcomes and is a standard approach used in the vaccine safety datalink (VSD). Design of the simulation and summary of the simulation results are presented in the rest of this section. The FDA has been given code from this project that includes SAS and R code for the *CSSP* method and the *GS EE* method developed from this project. This code is also available at: <http://faculty.washington.edu/acook/software.html>.

A. OUTLINE OF SIMULATION DESIGN

Our simulation study uses a data structure that is similar to that outlined in Section I.B.1. Specifically, we assume that accruing data will be analyzed at specified times ($t=1, \dots, T$) over a two-year study period. We also assume that each individual i at analysis t is either exposed to the MPI, $D_i=1$, or not exposed, $D_i=0$, and either has the outcome of interest occurring before the end of analysis time t , $Y_i(t)=1$, or does not, $Y_i(t)=0$. The exposure time, $E_i(t)$, denotes the cumulative exposure time prior to analysis t . It could be a single-time exposure window (e.g., vaccine: $E_i(t)=1$ for all individuals) or a chronic exposure (time on either MPI or comparator). Further, we assume that a set of baseline confounders, Z_i , is associated with individual i , selected from the following variables: sex, age, and body mass index (BMI). For each set of simulations we will vary the number of baseline confounders, the strength of confounding, the prevalence of outcome, the strength of the relationship between the outcome and MPI (effect size), and the frequency of testing. We will fix the boundary shape to be a Pocock-style boundary, which uses either a Pocock error spending function or a flat boundary on the standardized test statistic of interest when using a unifying boundary framework.¹⁵ The following two subsections outline the simulation set-up for the two general cases: 1) single-time exposure (e.g., vaccines) and 2) chronic exposure (e.g., drugs).

1. Case 1: Single-Time Exposure

Case 1 is the single-time exposure case, in which we assume that individuals are eligible to receive the MPI at a single time and are then followed up for a fixed interval of time in which an event of interest is determined to have or not have occurred. For example, children at 1 year of age receive a single vaccine injection of either the measles, mumps, and rubella (MMR) vaccine or the combination measles, mumps, rubella, and varicella (MMRV) vaccine. We follow each child for 45 days after vaccination to determine whether or not they had a medically treated fever during those 45 days. This case is also relevant for some medical devices, in which patients receiving a medical device or comparison medical

device are followed for 1 month after receipt of the medical device to evaluate whether they had a certain medical complication of interest post-surgery.

For this type of scenario, the outcome of interest, $Y_i(t)$, is a binary outcome that is 1 if the outcome occurred during the fixed follow-up period or 0 otherwise. Further, we can ignore exposure time, $E_i(t)$, because all study participants have the same exposure time. Below is the specific step-wise simulation design for creating a dataset of N study participants for $(i=1, \dots, N)$:

1. Start time, S_i , is the time at which individual i is enrolled in the study and this is uniformly distributed throughout the two-year (720-day) study; $S_i \sim \text{Discrete Uniform}(1, 719)$.
2. Confounder distribution of Z_i is defined in four different configurations:
 - a. Case 1a: Simple binary confounder (sex) with prevalence of 50% making equal distributions of males and females; $Z_i = Z_{1i} \sim \text{Bernoulli}(0.50)$.
 - b. Case 1b: Simple continuous confounder (age) with age distribution uniform between 35 and 65; $Z_i^* \sim \text{Unif}(35, 65)$ but use a 10-year age standardized variable $Z_i = Z_{2i} = (Z_i^* - 50)/10$ so that strength of association between age and outcome or MPI is in terms of relative risks per 10-year age increment.
 - c. Case 1c: Both simple binary and continuous confounders; $\mathbf{Z}_i = (Z_{1i}, Z_{2i})$.
 - d. Case 1d: Simple binary, simple continuous, and a categorical confounder (BMI categories) with BMI prevalence categories (normal ($< 25\text{kg}/\text{m}^2$), overweight ($25\text{--}30\text{kg}/\text{m}^2$), and obese ($\geq 30\text{kg}/\text{m}^2$)) based on the 2009 Behavioral Risk Factor Surveillance System (BRFSS) data for US adults; $\mathbf{Z}_i = (Z_{1i}, Z_{2i}, Z_{3i})$ with $Z_{3i} \sim \text{multinomial}(0.32, 0.37, 0.31)$ for $Z_{3i} \in (0, 1, 2)$.
3. Exposure distribution conditional on confounder distribution:

$D_i | \mathbf{Z}_i \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,z1}Z_{1i} + \beta_{D,z2}Z_{2i} + \beta_{D,z3}Z_{3i}))$ implying $\exp(\beta_{D,0}) = P(D_i | \mathbf{Z}_i = \mathbf{0})$ and $\exp(\beta_{D,z}) = P(D_i | Z_i = c) / P(D_i | Z_i = c-1)$, holding the other Z s fixed. This makes the interpretation of each $\exp(\beta_{D,z})$ for a given Z to be the relative risk per one unit increment in a given Z . For a given confounder case we included only the relevant Z s. For each simulation we further solved for $\beta_{x,0}$ so that the overall probability of D_i was fixed at 0.50 across all simulation configurations.
4. Outcome distribution conditional on exposure and confounder distributions:

$Y_i | D_i, \mathbf{Z}_i \sim \text{Bernoulli}(\exp(\beta_{y,0} + \beta_{y,D}D_i + \beta_{y,z1}Z_{1i} + \beta_{y,z2}Z_{2i} + \beta_{y,z3}Z_{3i}))$ implying $\exp(\beta_{y,0}) = P(Y_i | D_i = 0, \mathbf{Z}_i = \mathbf{0})$, $\exp(\beta_{y,D}) = P(Y_i | D_i = 1, \mathbf{Z}_i) / P(Y_i | D_i = 0, \mathbf{Z}_i)$, and $\exp(\beta_{y,z}) = P(Y_i | D_i, Z_i = c) / P(Y_i | D_i, Z_i = c-1)$, holding the other Z s fixed. Therefore the interpretation of $\exp(\beta_{y,D})$ is the relative risk due to being exposed to the MPI of interest controlling for confounders Z and $\exp(\beta_{y,z})$ is the relative risk of one unit increase of Z controlling for the MPI of interest and holding the other Z s constant. For this part of the simulation we hold the relative risk between each Z and Y constant to simplify the number of parameters changing. Further, to have the overall prevalence fixed across a given simulation, we solve for $\beta_{y,0}$ such that $P(Y(T))$ at the end of the study is fixed.

This simulation set-up allows us to vary the number of confounders, the strength of confounding by changing $\exp(\beta_{y,z})$ or $\exp(\beta_{D,z})$, the effect size or relative risk of Y on D by changing $\exp(\beta_{y,D})$, and the

prevalence of the outcome. Section II.B.1 shows the results of the simulation varying all these parameters.

2. Case 2: Chronic Exposure

Case 2 is the chronic exposure case in which we assume that individuals begin a certain treatment and then continue to receive that exposure until they either experience the event of interest or are censored. Censoring could be administrative due to still being in the study population without an event at a given analysis time or nonadministrative due to disenrollment from the healthcare plan or discontinuation of use of the initial assigned treatment. Further, participants are censored if they switch exposure groups and begin taking the other medical product (e.g., an exposed individual starts taking the comparator medical product). These assumptions are consistent with incident user cohort studies that are currently being used in postmarket surveillance.⁸ For example, if the drug of interest is a diabetes drug and the comparison is a different diabetes drug with myocardial infarction (MI) the outcome of interest, then the outcome data would consist of time to either MI or censoring, allowing participants to have differing follow-up time. Below is the specific step-wise simulation design for creating a dataset of N study participants for $(i=1, \dots, N)$:

1. Start time, S_i , is the time at which individual i is enrolled in the study and this is uniformly distributed throughout the two-year (720-day) study; $S_i \sim \text{Discrete Uniform}(1, 719)$.
2. Confounder distribution of Z_i is as outlined in Section II.A.1 with cases 2a, 2b, 2c, and 2d.
3. Exposure distribution conditional on confounder distribution is as outlined in Section II.A.1.
4. Censoring distribution due to stopping MPI or leaving the study: $C_i \sim \text{Weibull}(2, 200)$, which implies a skewed distribution with heavier censoring earlier, with a mean of 177 days but a long tail allowing individuals to stay on the MPI for a long period of time. Note that we assumed censoring is independent of confounders and exposure.
5. Outcome distribution conditional on exposure and confounder distributions:
 $T_i | D_i, \mathbf{Z}_i \sim \text{exponential}(\exp(\beta_{y,0} + \beta_{y,D} D_i + \beta_{y,z1} Z_{1i} + \beta_{y,z2} Z_{2i} + \beta_{y,z3} Z_{3i}) / 720)$ implying $\exp(\beta_{y,D}) = P(T_i | D_i=1, \mathbf{Z}_i) / P(T_i | D_i=0, \mathbf{Z}_i)$, and $\exp(\beta_{y,z}) = P(T_i | D_i, Z_i=c) / P(T_i | D_i, Z_i=c-1)$, holding the other Z s fixed. Therefore the interpretation of $\exp(\beta_{y,D})$ is the hazard ratio, which in the special case of this exponential setting is also the relative risk for exposure to the MPI of interest while controlling for confounders Z . The factor $\exp(\beta_{y,z})$ is the relative risk of one unit increase of Z controlling for the MPI of interest and holding the other Z s constant.
6. Outcome and exposure duration at a given analysis time t ($t=1, \dots, T$) is $Y_i(t) = I(T_i < C_i \text{ and } T_i < t - S_i)$ and $E_i(t) = \min(t - S_i, C_i, T_i)$. Further, to have the overall prevalence fixed across a given simulation, we solve for $\beta_{y,0}$ such that $P(Y(T))$ at the end of the study is fixed.

This simulation set-up allows us to vary the number of confounders, the strength of confounding by changing $\exp(\beta_{y,z})$ or $\exp(\beta_{D,z})$, the effect size or relative risk of Y on D by changing $\exp(\beta_{y,D})$, and the prevalence of the outcome. Section II.B.2 shows the results of the simulation varying all these parameters.

B. RESULTS OF SIMULATION

In this section we summarize the results of the simulation study. At the beginning of each subsection we provide the main findings and then follow with details of how we came to those conclusions. This is followed with a discussion explaining potential issues with each method leading to the results and a general section of next steps for methods development in this area. Note that all results are based on 500 data replications per a given simulation scenario.

1. Case 1: Single-Time Exposure

a. Changing Probability of Outcome and Confounding Strength

Situation: Comparing methods across probability of the outcome from 0.01 and 0.05 while assuming a Pocock-style boundary, with the first analysis after 180 days and subsequent analyses every 90 days until end of study at two years (720 days). Sample size at end of study is 10,000 participants.

Main Conclusions:

- As probability of outcome increases, power increases.
- As strength of confounding increases, power decreases.
- The three approaches have comparable performance in terms of total power, but the *GS EE* approach has a shorter time to detection compared to the *GS LD* and *CSSP* approaches. The *CSSP* approach has a shorter time to detection compared to the *GS LD* approach in the extremely rare event setting (probability of outcome is 0.01).
- The *GS EE* method has some issues with inflation of the type I error in the presence of very strong confounding (i.e., confounders with $RR=5$). This situation is not likely to occur often, but if strong confounders are expected, the other approaches may be recommended. It is possible that the number of simulations required to estimate the *GS EE* for this setting needs to be explored and could improve performance with respect to type I error.
- As outcomes become more common (probability of outcome is 0.05 vs. 0.01), the three approaches are very comparable when testing frequency is not high and first analysis is a quarter of the way into the total study time.

Tables 1a, 1b, 1c, and 1d assess the single-time exposure case with varying probability of outcome (0.01 and 0.05) and strength of confounding for the four confounder relationship structures, respectively. As shown in Table 1a, for a simple binary confounder the three methods have similar power. However, in cases with a true elevated risk (i.e., $RR(Y|D)>1$), the time to detection is shorter for the *GS EE* method compared to the other approaches. Specifically, for case 1a with a relative risk of 2.0 between Y and D given Z_1 , a relative risk of 5.0 between Y and Z_1 given D, and overall probability of outcome of 0.01, the time to detection of a signal was on average 345.7 days for the *GS EE* approach compared to 363.6 days for *CSSP* and 367.3 days for *GS LD* (Table 1a). However, overall power was the same among the three methods: 89.6%, 90.8%, and 90.8%, respectively. This relationship holds throughout all of Tables 1a-1d, but differences in time to detection are much smaller when the probability of outcome is greater than 0.01.

b. Changing Frequency of Testing, Probability of Outcome, and Confounding Strength

Situation: Compare methods for a two-year (720-day) study using bimonthly (twice per month), monthly, and quarterly testing frequencies assuming a Pocock-style boundary and with the first analysis occurring after 90 days. Sample size at end of study is 10,000 participants for probability of outcome 0.05 (rare setting) and 1,500 participants for probability of outcome of 0.10 (less rare setting).

Main Conclusions:

- As frequency of testing decreases, power generally increases, but not substantially.
- In the rare outcome setting (probability of outcome is 0.05), the three approaches have comparable performance in terms of total power but the *GS EE* approach typically has shorter time to detection compared to the *GS LD* and *CSSP* approaches regardless of frequency of testing. This is due to a relatively higher power for the *GS EE* approach in the first few analyses of the study.
- In the less rare outcome setting (probability of outcome is 0.10), the *GS EE* approach tends to have higher power compared to *GS LD* and *CSSP* approaches, especially in the more frequent testing setting. Further, the *GS EE* approach typically has shorter time to detection compared to the *GS LD* and *CSSP* approaches regardless of frequency of testing.
- The *GS LD* approach has a shorter time to detection compared to the *CSSP* approach with more frequent testing.
- The *CSSP* approach has a shorter time to detection compared to the *GS LD* approach with less frequent testing and fewer confounders, but this pattern reverses when more confounders (i.e., strata) are included.

Tables 2a.1, 2b.1, 2c.1, and 2d.1 assess the single-time exposure case with varying frequency of testing (bimonthly (twice per month), monthly, and quarterly) and strength of confounding for the four confounder relationship structures, respectively, while holding the overall probability of the outcome at 0.05. Tables 2a.2, 2b.2, 2c.2, and 2d.2 show similar results except hold the overall probability of the outcome at 0.10. As shown in Tables 2a.1-2d.1, the overall power tends to be similar across approaches and frequency of testing for the rare outcome situation of 0.05, but as previously noted the time to detection is shorter for the *GS EE* approach relative to the *CSSP* and *GS LD* approaches. The reason for the power and timeliness differences is shown in Figure 1a: the *GS EE* method has higher power in the first 50%-90% of the analysis times compared to the other methods but the other approaches catch up by the end of the study. This occurs in particular for the less frequent testing situation (quarterly and monthly).

Tables 2a.2-2d.2 show a different result. For the situation with a more prevalent outcome (0.10), the power tends to be higher for the *GS EE* approach relative to the other approaches, with sometimes greater than 10% differences, especially with more frequent testing. This results in situations where time to signal detection is up to 50 days shorter, even in the bimonthly testing frequency. Figure 1b showcases these differences for the confounding case 1b situation with a relative risk of 2.0 between Y and D given Z_2 , a relative risk of 2.0 between Y and Z_2 given D , and overall probability of outcome of 0.10 while varying frequency of testing. In the bimonthly testing situation, the *GS EE* approach starts with

higher power at the first analyses and keeps a higher power across all analyses compared to the *CSSP* approach, but then the *GS LD* approach catches up. As frequency goes down, all approaches have comparable power at the end of the study, but the *GS EE* approach has more power earlier on.

2. Case 2: Chronic Exposure

a. Changing Probability of Outcome and Confounding Strength

Situation: Comparing methods across probability of the outcome from 0.01 and 0.05 while assuming a Pocock-style boundary, with the first analysis after 180 days and each subsequent analysis every 90 days until end of study at two years (720 days). Sample size at end of study is 10,000 participants.

Main Conclusions:

- As probability of outcome increases, power increases.
- As strength of confounding increases, power decreases.
- The three approaches have comparable performance in terms of total power, but the *GS EE* approach typically has shorter time to detection compared to the *GS LD* and *CSSP* approaches. The *CSSP* approach has a shorter time to detection compared to the *GS LD* approach, especially in the extremely rare event setting (probability of outcome is 0.01) for the quarterly analyses setting.
- The *GS EE* has issues with inflation of the type I error in the extremely rare event (probability of outcome is 0.01) and not very large sample size ($N=10,000$). This is likely due to difficulties estimating regression parameters in this setting. In general, in any situation where regression would not be advised, the *GS EE* is not recommended. In this situation, when frequency of testing is quarterly the *CSSP* approach may be recommended because the *GS LD* has a slower time to rejection. Further evaluation assessing extremely rare events with larger sample size should be explored since the *GS EE* might be applicable even in the extremely rare event setting (probability of outcome is 0.01).

Tables 3a, 3b, 3c, and 3d assess the chronic exposure case with varying probability of outcome (0.01 and 0.05) and strength of confounding for the four confounder relationship structures, respectively. As shown in Table 3a, with a simple binary confounder the three methods have similar power. However, the time to detection of an actual signal (i.e., $RR(Y|D) > 1$) is shorter for the *GS EE* method than for the other approaches. Specifically, for case 2a with a relative risk of 2.0 between Y and D given Z_1 , a relative risk of 5.0 between Y and Z_1 given D , and overall probability of outcome of 0.01, the time to detection of a signal was on average 381.7 days for the *GS EE* approach compared to 410.0 days for *CSSP* and 416.5 days for *GS LD* (Table 3a). However, overall power was the same between the three methods, at 90.4%, 90.0%, and 89.6%, respectively. This relationship holds throughout all four confounder relationship structures, Tables 3a-3d, but differences in time to detection are much smaller when the overall prevalence changes from the extremely rare situation of 0.01 to the moderately rare one of 0.05. Further, in the extremely rare event setting with sample size of 10,000, as more confounders with a strong confounding relationship are introduced, the *GS EE* approach does not hold the overall type I error. When the relative risk between Y and D equals 1.0 (i.e., no effect), the power and the type 1 error

are equivalent and should equal 0.05 by design; however, values greater than 0.05 were observed in these settings (Tables 3c and 3d).

b. Changing Prevalence of Outcome, Frequency of Testing, and Confounding Strength

Situation: Compare methods for a two-year (720-day) study using bimonthly (twice per month), monthly, and quarterly testing frequencies, assuming a Pocock-style boundary and the first analysis occurring after 90 days. Sample size at end of study is 10,000 participants for probability of outcome 0.05 and 1,500 participants for probability of outcome of 0.10.

Main Conclusions:

- As frequency of testing decreases, power generally increases, but not substantially.
- In the moderately rare outcome setting (probability of outcome is 0.05), the three approaches have overall comparable performance in terms of total power. There is some indication that the *CSSP* has higher power in the more frequent testing setting, but the *GS EE* approach typically has shorter time to detection compared to the *GS LD* and *CSSP* approaches regardless of frequency of testing.
- In the less rare outcome setting (probability of outcome is 0.10), the *GS EE* approach tends to have higher power compared to *GS LD* and *CSSP* approaches, especially in the more frequent testing setting. Further, the *GS EE* approach typically has shorter time to detection compared to the *GS LD* and *CSSP* approaches regardless of frequency of testing.
- The *CSSP* approach tends to have a shorter time to detection compared to the *GS LD* approach, especially in the rarer event setting.
- The *CSSP* approach has issues with holding the type I error in the less rare event setting (probability of outcome is 0.10) and the very frequent testing situation and therefore should not be used in this setting.

Tables 4a.1, 4b.1, 4c.1, and 4d.1 assess the chronic exposure case with varying frequency of testing (bimonthly, monthly, and quarterly) and strength of confounding for the four confounder relationship structures, respectively, while holding the overall probability of the outcome at 0.05. Tables 4a.2, 4b.2, 4c.2, and 4d.2 show similar results with the exception that the overall probability of the outcome is fixed at 0.10. As shown in Tables 4a.1-4d.1, the overall power tends to be similar across approaches, although there is some indication that the *CSSP* has slightly more power in the less frequent testing setting with a moderately rare outcome prevalence of 0.05. But, as previously noted, the time to detection is shorter for the *GS EE* approach relative to the *CSSP* and *GS LD* approaches in the more frequent testing setting. The power and timeliness differences are explained by Figure 2a where the *GS EE* method is seen to have higher power in the first 50%-90% of the analysis times compared to the other methods, with the other approaches catching up by the end of the study. This behavior is particularly pronounced in the less frequent testing situation. These relationships hold as more confounders are introduced.

Tables 4a.2-4d.2 display the results with a less rare outcome (probability of outcome is 0.10). The power is relatively equal across the approaches. Now, when testing less frequently and with more confounders, the time to detection is sometimes shorter for the *CSSP* approach relative to the other approaches.

However, in the more frequent testing situation, the *CSSP* approach is not appropriate due to inflation of type I error. Further, in the more frequent testing situation, the *GS EE* approach tends to have a shorter time to detection than the *GS LD* approach. Figures 2b and 2c showcase these differences for case 2b and case 2d confounding settings with a relative risk of 2.0 between Y and D given confounders, a relative risk of 2.0 between Y and each confounder given D, and overall probability of outcome of 0.10 while varying frequency of testing. In the simple continuous confounder situation (case 2b), Figure 2b shows that in the bimonthly testing situation the three approaches have comparable time to detection. The *GS EE* approach is slightly faster, but as testing frequency decreases, the *CSSP* and *GS EE* approaches have more power earlier on relative to the *GS LD* approach. This leads to a quicker time to rejection for these approaches. Figure 2c shows a similar result except that in the quarterly setting, with many confounders, the *CSSP* approach sometimes has higher power and faster time to detection relative to the other approaches.

C. DISCUSSION

The key finding of this simulation evaluation is that the three group sequential monitoring statistical approaches evaluated have performance properties that differ across study settings. In particular, timeliness of signal detection is especially sensitive to setting and method choice. Therefore, depending on the setting in which one asks the question of interest, one is led to adopting a different statistical approach. We have created general categories of types of settings including: outcome prevalence (very rare, rare, and less rare), confounders (number and strength), frequency of testing (bimonthly, monthly, and quarterly), and amount of time before first analysis (90 days (1/8 into study) versus 180 days (1/4 into study)). The general conclusion was that, in most situations, the *GS EE* approach is more appropriate due to having a faster time to detection compared to the other approaches except in the extremely rare event setting (0.01 with a sample size of 10,000) when issues occurred with holding the overall type I error. In this setting, and in the context of our simulation results, the *CSSP* typically should be recommended.

The reasons for the differences in performance of the statistical approaches are due to the type of test statistic used and the formulation of the boundary for a given method. Specifically, as outlined in Section I, the *CSSP* approach uses the number of events associated with the MPI as the test statistic and compares this to a simulated distribution of the expected number of events assuming no elevated risk is present conditional on confounder and analysis time strata. The *GS EE* method uses a generalized score test statistic based on a regression analysis and assumes that the event rate is equivalent for the exposed and unexposed groups (null hypothesis). While the *GS LD* method can be based on any standardized test statistic, for the simulation study it was based on a regression model using a Wald statistic because this is the most commonly used test statistic in practice. It has already been shown that the generalized score statistic and a Wald statistic should perform similarly, given a large enough sample size, therefore it is unlikely that the test statistic is the reason for the differences between the *GS EE* and *GS LD* approaches. However, in the case of the *CSSP* method, the difference in test statistic is important because it was derived under the rare event setting, where the Binomial and Poisson distributions are approximately equivalent. As the probability of outcome increases, then the *CSSP* approach would not be recommended.

The primary source of discrepancy between times to signal detection across methods lies with the boundary formulation. The *CSSP* and *GS LD* methods both use an error spending approach that relies on the amount of cumulative $\alpha(t)$ to be spent at a given analysis time t . In the standard settings in which

the error spending functions were developed, such as clinical trials, it is common to think of a Pocock error spending curve to spend α resulting in a relatively flat boundary on the standardized test statistic scale. This has nice intuitive properties such as a signal occurring at any given analysis when a similar amount of information informs the signal. However, when analyses are not evenly spaced, such as when the first analysis is delayed until a larger proportion of information has accrued or in the more frequent testing case, this does not hold. As shown in Figure 3, the standardized test statistic boundary derived from a *GS LD* boundary formulation is not flat for the situation with the first analysis at 90 days and subsequent monthly or quarterly analyses. As for the unifying boundary approach, which formulates the boundary on the standardized test statistic scale, the boundary is flat by definition. This resulted in a faster time to detection across almost all situations. A different error spending function could have been created that attempts to approximate a flat boundary but this would need to be done for each setting specifically. Therefore, even though the error spending approach is thought to be intuitive and easy to apply, it does not always result in the intuitive interpretation of the boundary that one is accustomed to. Therefore caution is recommended when using such approaches. It is always important to understand the boundaries on the test statistic scale because this is the amount of statistical information accrued up to a certain point in the study. Since the *CSSP* approach is not based on a standardized test statistic scale, it would be more difficult to derive a boundary based on the error spending function to hold such a specific property. Further evaluation of boundary shapes needs to be conducted in the postmarket surveillance setting to assess trade-offs and advantages across approaches and to not be limited to the infeasible setting of proportional amounts of statistical information across analysis times.

This is the first of many simulation studies that are needed to evaluate different scenarios for settings in postmarket surveillance. In particular, for the chronic exposure setting, this simulation evaluation was limited to the very strong assumption of an exponential distribution for time to event, in which Poisson regression models can be applied. However, in more complicated survival settings such as those that include only proportional hazards assumptions, the *CSSP* and *GS EE* approaches do not apply. The *GS LD* approach can be used with a Cox proportional hazards model but, as has been shown in this evaluation, perhaps better approaches should be developed for the rare adverse event setting with confounding that would improve the time to signal detection.

D. TABLES AND FIGURES FOR SIMULATION STUDY SECTION II

Table 1a. Power, days to rejection, and days to study end across varying strength of confounding and probability of outcome for the simple case of one binary confounder with a binary outcome.

	Effect Size RR(Y D)	Confounder Y and Z		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End
		RR(Y Z ₁)							
GS Wald - Regression and Normal Approx	1	2	0.046	363.9	703.6	0.054	373.3	701.3	
	1	5	0.042	407.1	706.9	0.050	378.0	702.9	
	1.5	2	0.540	416.7	556.2	0.988	285.9	291.1	
	1.5	5	0.482	453.4	591.5	0.994	281.0	283.7	
	2	2	0.898	371.2	406.8	1.000	187.9	187.9	
	2	5	0.908	367.3	399.8	1.000	187.2	187.2	
GS CSSP - Stratification and Error Spending	1	2	0.042	402.9	706.7	0.026	366.9	710.8	
	1	5	0.040	378.0	706.3	0.058	406.5	701.8	
	1.5	2	0.498	425.8	573.5	0.978	281.4	291.1	
	1.5	5	0.530	420.8	561.4	0.988	296.6	301.7	
	2	2	0.896	358.6	396.2	1.000	190.8	190.8	
	2	5	0.908	363.6	396.4	1.000	187.7	187.7	
GS EE Score - Regression and Unifying	1	2	0.062	362.9	697.9	0.042	432.9	707.9	
	1	5	0.060	363.0	698.6	0.046	305.2	700.9	
	1.5	2	0.570	402.6	539.1	0.986	275.7	281.9	
	1.5	5	0.524	385.1	544.5	0.996	284.8	286.6	
	2	2	0.906	355.4	389.7	1.000	188.6	188.6	
	2	5	0.896	345.7	384.7	1.000	190.4	190.4	

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses quarterly (8 analyses), first analysis at second quarter, and 10,000 sample size at end of two years.

Case 1a: $Z_1 \sim \text{Bernoulli}(0.5)$; $D | Z_1 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z} Z_1))$; $Y | D, Z_1 \sim \text{Bernoulli}(\exp(\beta_{Y,0} + \beta_{Y,D} D + \beta_{Y,Z} Z_1))$

$RR(Y|D) = \exp(\beta_{Y,D})$ and $RR(Y|Z_1) = \exp(\beta_{Y,Z})$.

Table 1b. Power, days to rejection, and days to study end across varying strength of confounding and probability of outcome for the simple case of one continuous confounder with a binary outcome.

	Effect Size RR(Y D)	Confounder		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End
		Y and Z	RR(Y Z ₂)						
GS Wald - Regression and Normal Approx	1	2	0.032	371.3	708.8	0.058	450.0	704.3	
	1	5	0.034	428.8	710.1	0.056	353.6	699.5	
	1.5	2	0.416	451.3	608.2	0.978	310.1	319.1	
	1.5	5	0.400	451.8	612.7	0.962	310.0	325.6	
	2	2	0.822	387.6	446.8	1.000	196.4	196.4	
	2	5	0.742	424.3	500.6	1.000	199.8	199.8	
GS CSSP - Stratification and Error Spending	1	2	0.054	426.7	704.2	0.050	399.6	704.0	
	1	5	0.058	403.5	701.6	0.058	412.8	702.2	
	1.5	2	0.506	432.6	574.6	0.982	286.3	294.1	
	1.5	5	0.534	425.4	562.7	0.986	307.2	313.0	
	2	2	0.868	378.0	423.2	1.000	194.8	194.8	
	2	5	0.834	376.0	433.1	1.000	199.3	199.3	
GS EE Score - Regression and Unifying	1	2	0.076	348.2	691.7	0.066	335.5	694.6	
	1	5	0.062	362.9	697.9	0.054	406.7	703.1	
	1.5	2	0.472	405.4	571.5	0.974	278.9	290.3	
	1.5	5	0.488	388.4	558.2	0.968	304.2	317.5	
	2	2	0.852	354.1	408.2	1.000	196.7	196.7	
	2	5	0.828	355.7	418.3	1.000	200.7	200.7	

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses quarterly (8 analyses), first analysis at second quarter, and 10,000 sample size at end of two years.

Case 1b: $Z_2 \sim \text{Unif}(-1.5, 1.5)$; $D | Z_2 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_2} Z_2))$; $Y | D, Z_2 \sim \text{Bernoulli}(\exp(\beta_{Y,0} + \beta_{Y,D} D + \beta_{Y,Z_2} Z_2))$

$RR(Y|D) = \exp(\beta_{Y,D})$ and $RR(Y|Z_2) = \exp(\beta_{Y,Z_2})$.

Table 1c. Power, days to rejection, and days to study end across varying strength of confounding and probability of outcome for the case of one binary confounder and one continuous confounder with a binary outcome.

	Effect Size RR(Y D)	Confounders		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	
		Y and Z ₁ RR(Y Z ₁)	Y and Z ₂ RR(Y Z ₂)							
					P(Y)=0.01			P(Y)=0.05		
GS Wald - Regression and Normal Approx	1	2	2	0.042	471.4	709.6	0.044	347.7	703.6	
	1	2	5	0.018	510.0	716.2	0.054	363.3	700.7	
	1	5	2	0.026	443.1	712.8	0.058	462.4	705.1	
	1	5	5	0.024	450.0	713.5	0.054	373.3	701.3	
	1.5	2	2	0.468	416.5	578.0	0.972	306.3	317.9	
	1.5	2	5	0.356	447.5	623.0	0.950	328.4	347.9	
	1.5	5	2	0.450	422.8	586.3	0.956	316.9	334.6	
	1.5	5	5	0.346	468.7	633.1	0.946	353.9	373.7	
	2	2	2	0.810	398.4	459.5	1.000	194.0	194.0	
	2	5	2	0.726	427.9	508.0	1.000	202.1	202.1	
	2	2	5	0.802	399.1	462.6	1.000	199.6	199.6	
2	5	5	0.698	458.3	537.3	1.000	213.3	213.3		
GS CSSP - Stratification and Error Spending	1	2	2	0.036	460.0	710.6	0.062	333.9	696.1	
	1	2	5	0.062	421.0	701.5	0.082	377.6	691.9	
	1	5	2	0.048	382.5	703.8	0.060	432.0	702.7	
	1	5	5	0.062	357.1	697.5	0.066	392.7	698.4	
	1.5	2	2	0.512	439.5	576.4	0.978	308.1	317.2	
	1.5	2	5	0.466	449.6	594.0	0.982	315.1	322.4	
	1.5	5	2	0.494	435.8	579.6	0.968	319.1	331.9	
	1.5	5	5	0.456	448.4	596.2	0.950	336.9	356.0	
	2	2	2	0.852	398.7	446.2	1.000	195.5	195.5	
	2	5	2	0.834	388.5	443.5	1.000	205.2	205.2	
	2	2	5	0.800	397.4	461.9	1.000	203.9	203.9	
2	5	5	0.808	395.2	457.6	1.000	210.2	210.2		
GS EE Score - Regression and Unifying	1	2	2	0.052	380.8	702.4	0.052	401.5	703.4	
	1	2	5	0.054	316.7	698.2	0.086	326.5	686.2	
	1	5	2	0.060	333.0	696.8	0.056	318.2	697.5	
	1	5	5	0.074	369.7	694.1	0.050	414.0	704.7	
	1.5	2	2	0.490	389.4	558.0	0.966	300.4	314.6	
	1.5	2	5	0.504	379.6	548.5	0.954	310.6	329.4	
	1.5	5	2	0.512	408.2	560.3	0.978	299.1	308.3	
	1.5	5	5	0.458	391.4	569.5	0.946	325.6	346.9	
	2	2	2	0.848	363.2	417.4	1.000	195.1	195.1	
	2	5	2	0.866	360.2	408.4	1.000	196.7	196.7	
	2	2	5	0.778	348.2	430.7	1.000	199.3	199.3	
2	5	5	0.826	338.2	404.6	1.000	209.9	209.9		

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses quarterly (8 analyses), first analysis at second quarter and N=10,000

Case 1c: $Z_1 \sim \text{Bernoulli}(0.5)$; $Z_2 \sim \text{Unif}(-1.5, 1.5)$; $D | Z_1, Z_2 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_1} + \beta_{D,Z_2}))$; $Y | D, Z_1, Z_2 \sim \text{Bernoulli}(\exp(\beta_{Y,0} + \beta_{Y,D} + \beta_{Y,Z_1} + \beta_{Y,Z_2}))$;
 $RR(Y|D) = \exp(\beta_{Y,D})$, $RR(Y|Z_1) = \exp(\beta_{Y,Z_1})$, and $RR(Y|Z_2) = \exp(\beta_{Y,Z_2})$.

Table 1d. Power, days to rejection, and days to study end across varying strength of confounding and probability of outcome for the case of one binary confounder, one continuous confounder, and one categorical confounder with a binary outcome.

	Effect Size RR(Y D)	Confounders			POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End
		Y and Z ₁ RR(Y Z ₁)	Y and Z ₂ RR(Y Z ₂)	Y and Z ₃ RR(Y Z ₃)						
					P(Y)=0.01			P(Y)=0.05		
GS Wald - Regression and Normal Approx	1	2	2	2	0.044	368.2	704.5	0.050	378.0	702.9
	1	2	2	5	0.042	385.7	706.0	0.050	334.8	700.7
	1	2	5	2	0.044	450.0	708.1	0.048	408.8	705.1
	1	2	5	5	0.042	368.6	705.2	0.040	346.5	705.1
	1.5	2	2	2	0.488	467.7	596.9	0.982	290.3	298.1
	1.5	2	2	5	0.416	446.1	606.1	0.988	297.1	302.2
	1.5	2	5	2	0.420	435.4	600.5	0.980	313.2	321.3
	1.5	2	5	5	0.398	444.6	610.4	0.950	307.1	327.8
	2	2	2	2	0.888	377.0	415.4	1.000	189.0	189.0
	2	2	2	5	0.882	383.1	422.8	1.000	193.5	193.5
	2	2	5	2	0.786	394.8	464.4	1.000	194.4	194.4
	2	2	5	5	0.818	402.0	459.9	1.000	201.1	201.1
GS CSSP - Stratification and Error Spending	1	2	2	2	0.058	375.5	700.0	0.048	408.8	705.1
	1	2	2	5	0.044	388.6	705.4	0.034	370.6	708.1
	1	2	5	2	0.056	443.6	704.5	0.046	403.0	705.4
	1	2	5	5	0.044	405.0	706.1	0.068	402.4	698.4
	1.5	2	2	2	0.480	434.3	582.8	0.990	306.7	310.9
	1.5	2	2	5	0.458	455.9	599.0	0.980	323.5	331.4
	1.5	2	5	2	0.468	427.3	583.0	0.974	307.5	318.2
	1.5	2	5	5	0.460	441.0	591.7	0.950	313.4	333.7
	2	2	2	2	0.902	376.4	410.0	1.000	192.2	192.2
	2	2	2	5	0.882	379.2	419.4	1.000	198.9	198.9
	2	2	5	2	0.864	369.4	417.1	1.000	196.6	196.6
	2	2	5	5	0.826	397.0	453.2	1.000	209.2	209.2
GS EE Score - Regression and Unifying	1	2	2	2	0.054	386.7	702.0	0.066	338.2	694.8
	1	2	2	5	0.088	335.5	686.2	0.050	284.4	698.2
	1	2	5	2	0.064	309.4	693.7	0.048	322.5	700.9
	1	2	5	5	0.096	348.8	684.4	0.050	345.6	701.3
	1.5	2	2	2	0.546	398.2	544.3	0.972	281.9	294.1
	1.5	2	2	5	0.536	410.0	553.9	0.976	296.6	306.7
	1.5	2	5	2	0.542	402.2	547.7	0.980	296.8	305.3
	1.5	2	5	5	0.534	388.3	542.9	0.938	304.9	330.7
	2	2	2	2	0.922	358.0	386.3	1.000	190.4	190.4
	2	2	2	5	0.892	338.8	380.0	1.000	192.8	192.8
	2	2	5	2	0.862	336.6	389.5	1.000	191.3	191.3
	2	2	5	5	0.882	360.8	403.2	1.000	197.5	197.5

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses quarterly (8 analyses), first analysis at second quarter, and N=10,000

Case 1d: $Z_1 \sim \text{Bernoulli}(0.5)$; $Z_2 \sim \text{Unif}(-1.5, 1.5)$; $Z_3 \sim \text{Multi}(0.32, 0.37, .31)$; $D|Z_1, Z_2, Z_3 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_1} + \beta_{D,Z_2} + \beta_{D,Z_3}))$;

$Y|D, Z_1, Z_2 \sim \text{Bernoulli}(\exp(\beta_{Y,0} + \beta_{Y,D} + \beta_{Y,Z_1} + \beta_{Y,Z_2} + \beta_{Y,Z_3}))$;

$RR(Y|D) = \exp(\beta_{Y,D})$, $RR(Y|Z_1) = \exp(\beta_{Y,Z_1})$, $RR(Y|Z_2) = \exp(\beta_{Y,Z_2})$, and $RR(Y|Z_3) = \exp(\beta_{Y,Z_3})$.

Table 2a.1. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the simple case of one binary confounder and overall prevalence of 0.05 with a binary outcome.

Freq. of Testing	Effect Size	Confounder Y and Z ₁	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	
	RR(Y D)	RR(Y Z ₁)										
				BIMONTHLY			MONTHLY			QUARTERLY		
GS Wald - Regression and Normal Approx	1	2	0.066	344.1	695.2	0.062	303.9	694.2	0.060	390.0	700.2	
	1	5	0.048	356.3	702.5	0.048	355.0	702.5	0.048	371.3	703.3	
	1.2	2	0.496	375.7	549.2	0.496	382.3	552.5	0.504	406.8	562.1	
	1.2	5	0.466	394.1	568.1	0.478	401.4	567.7	0.466	419.9	580.1	
	1.5	2	0.982	251.7	260.1	0.984	258.5	265.9	0.982	276.2	284.2	
	1.5	5	0.990	245.0	249.8	0.992	248.9	252.7	0.990	265.3	269.8	
GS CSSP - Stratification and Error Spending	1	2	0.036	362.5	707.1	0.046	335.2	702.3	0.040	337.5	704.7	
	1	5	0.040	266.3	701.9	0.034	319.4	706.4	0.046	332.6	702.2	
	1.2	2	0.490	412.2	569.2	0.480	416.6	574.4	0.468	413.5	576.5	
	1.2	5	0.482	398.2	564.9	0.444	402.2	578.9	0.468	416.2	577.8	
	1.5	2	0.980	243.3	252.9	0.996	254.6	256.4	0.990	281.6	286.0	
	1.5	5	0.972	285.5	297.6	0.972	256.2	269.2	0.992	291.4	294.8	
GS EE Score - Regression and Unifying	1	2	0.056	274.3	695.0	0.040	298.5	703.1	0.062	316.5	695.0	
	1	5	0.048	280.0	698.9	0.044	267.3	700.1	0.072	327.5	691.7	
	1.2	2	0.482	349.7	541.5	0.480	362.6	548.5	0.472	387.5	563.0	
	1.2	5	0.482	369.7	551.2	0.508	370.4	542.4	0.460	394.4	570.2	
	1.5	2	0.974	240.8	253.2	0.982	235.2	243.9	0.984	252.4	259.9	
	1.5	5	0.972	237.7	251.2	0.982	230.6	239.4	0.990	250.0	254.7	

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 10,000 sample size at end of two years.

Case 1a: $Z_1 \sim \text{Bernoulli}(0.5)$; $D | Z_1 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_1}))$; $Y | D, Z_1 \sim \text{Bernoulli}(\exp(\beta_{Y,0} + \beta_{Y,D}D + \beta_{Y,Z}Z_1))$

$RR(Y|D) = \exp(\beta_{Y,D})$ and $RR(Y|Z_1) = \exp(\beta_{Y,Z})$.

Table 2a.2. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the simple case of one binary confounder and overall prevalence of 0.10 with a binary outcome.

Freq. of Testing	Effect Size RR(Y D)	Confounder		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End
		Y and Z ₁	RR(Y Z ₁)									
			BIMONTHLY			MONTHLY			QUARTERLY			
GS Wald - Regression and Normal Approx	1	2	0.046	352.2	703.1	0.050	348.0	701.4	0.056	388.9	701.5	
	1	5	0.038	322.1	704.9	0.038	334.7	705.4	0.042	360.0	704.9	
	1.5	2	0.616	387.9	515.4	0.618	391.6	517.0	0.636	414.9	526.0	
	1.5	5	0.564	387.2	532.3	0.578	393.7	531.4	0.582	413.5	541.6	
	2	2	0.964	291.3	306.8	0.958	293.8	311.7	0.972	310.9	322.4	
	2	5	0.940	306.9	331.7	0.940	308.0	332.7	0.950	328.9	348.5	
GS CSSP - Stratification and Error Spending	1	2	0.048	353.7	702.4	0.046	358.7	703.4	0.022	556.4	716.4	
	1	5	0.034	248.8	704.0	0.040	444.0	709.0	0.022	409.1	713.2	
	1.5	2	0.518	365.4	536.3	0.586	394.5	529.3	0.610	413.1	532.8	
	1.5	5	0.526	408.4	556.1	0.504	418.3	568.0	0.582	441.0	557.6	
	2	2	0.922	298.0	330.9	0.958	306.6	323.9	0.958	326.7	343.3	
	2	5	0.912	325.2	360.0	0.944	317.1	339.7	0.944	344.9	365.9	
GS EE Score - Regression and Unifying	1	2	0.048	318.1	700.7	0.046	348.3	702.9	0.060	375.0	699.3	
	1	5	0.062	299.5	693.9	0.056	383.6	701.2	0.062	319.4	695.2	
	1.5	2	0.658	364.2	485.9	0.612	371.3	506.6	0.644	380.1	501.1	
	1.5	5	0.612	359.6	499.4	0.600	372.5	511.5	0.618	359.7	497.3	
	2	2	0.960	251.9	270.6	0.956	267.7	287.6	0.978	290.8	300.2	
	2	5	0.958	268.7	287.6	0.946	285.6	309.1	0.974	314.5	325.1	

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 1,500 sample size at end of two years.

Case 1a: $Z_1 \sim \text{Bernoulli}(0.5)$; $D | Z_1 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_1}))$; $Y | D, Z_1 \sim \text{Bernoulli}(\exp(\beta_{Y,0} + \beta_{Y,D} + \beta_{Y,Z_1}))$

$RR(Y|D) = \exp(\beta_{Y,D})$ and $RR(Y|Z_1) = \exp(\beta_{Y,Z_1})$.

Table 2b.1. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the simple case of one continuous confounder and overall prevalence of 0.05 with a binary outcome.

Freq. of Testing	Effect Size RR(Y D)	Confounder Y and Z ₂		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End
		RR(Y Z ₂)										
			BIMONTHLY			MONTHLY			QUARTERLY			
GS Wald - Regression and Normal Approx	1	2	0.046	273.3	699.5	0.046	373.0	704.0	0.054	320.0	698.4	
	1	5	0.046	331.3	702.1	0.040	321.0	704.0	0.044	360.0	704.2	
	1.2	2	0.434	395.7	579.2	0.462	385.2	565.3	0.454	414.3	581.2	
	1.2	5	0.398	411.9	597.4	0.378	378.4	590.9	0.406	426.9	601.0	
	1.5	2	0.964	272.7	288.8	0.984	275.7	282.8	0.960	298.1	315.0	
	1.5	5	0.958	304.5	321.9	0.948	294.7	316.9	0.956	320.1	337.7	
GS CSSP - Stratification and Error Spending	1	2	0.068	341.9	694.3	0.044	289.1	701.0	0.082	410.5	694.6	
	1	5	0.070	319.7	692.0	0.088	398.2	691.7	0.100	406.8	688.7	
	1.2	2	0.490	374.4	550.7	0.490	404.4	565.4	0.528	430.2	567.0	
	1.2	5	0.450	395.0	573.8	0.520	383.3	544.9	0.512	425.7	569.3	
	1.5	2	0.980	271.3	280.3	0.984	270.4	277.6	0.980	288.0	296.6	
	1.5	5	0.976	271.3	282.1	0.982	280.5	288.4	0.984	297.4	304.2	
GS EE Score - Regression and Unifying	1	2	0.064	280.8	691.9	0.032	258.8	705.2	0.048	206.3	695.3	
	1	5	0.080	220.9	680.1	0.054	250.0	694.6	0.054	286.7	696.6	
	1.2	2	0.458	351.7	551.3	0.438	342.3	554.6	0.482	366.0	549.4	
	1.2	5	0.438	332.9	550.4	0.414	352.0	567.7	0.384	391.9	594.0	
	1.5	2	0.960	233.1	252.5	0.970	252.6	266.6	0.968	280.8	294.8	
	1.5	5	0.958	257.9	277.3	0.952	273.8	295.3	0.944	287.7	311.9	

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 10,000 sample size at end of two years.

Case 1b: $Z_2 \sim \text{Unif}(-1.5, 1.5)$; $D|Z_2 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_2}))$; $Y|D, Z_2 \sim \text{Bernoulli}(\exp(\beta_{Y,0} + \beta_{Y,D}D + \beta_{Y,Z}Z_2))$

$RR(Y|D) = \exp(\beta_{Y,D})$ and $RR(Y|Z_2) = \exp(\beta_{Y,Z})$.

Table 2b.2. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the simple case of one continuous confounder and overall prevalence of 0.10 with a binary outcome.

Freq. of Testing	Effect Size RR(Y D)	Confounder		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End
		Y and Z ₂ RR(Y Z ₂)										
			BIMONTHLY			MONTHLY			QUARTERLY			
GS Wald - Regression and Normal Approx	1	2	0.036	425.8	709.4	0.044	429.5	707.2	0.044	405.0	706.1	
	1	5	0.044	356.6	704.0	0.042	378.6	705.7	0.046	395.2	705.1	
	1.5	2	0.550	407.6	548.2	0.538	415.9	556.4	0.566	436.3	559.4	
	1.5	5	0.464	417.7	579.7	0.474	424.6	580.0	0.462	438.7	590.0	
	2	2	0.930	332.8	359.9	0.918	322.9	355.5	0.936	351.5	375.1	
	2	5	0.896	350.5	389.0	0.888	357.3	397.9	0.912	372.2	402.8	
GS CSSP - Stratification and Error Spending	1	2	0.038	402.6	707.9	0.048	366.2	703.0	0.056	353.6	699.5	
	1	5	0.034	403.2	709.2	0.042	342.9	704.2	0.042	347.1	704.3	
	1.5	2	0.506	374.6	545.2	0.506	400.4	558.3	0.530	439.8	571.5	
	1.5	5	0.432	381.8	573.9	0.452	398.5	574.7	0.496	453.3	587.7	
	2	2	0.858	333.8	388.7	0.904	342.3	378.5	0.940	356.9	378.7	
	2	5	0.818	364.4	429.1	0.860	372.1	420.8	0.874	368.6	412.9	
GS EE Score - Regression and Unifying	1	2	0.060	342.0	697.3	0.072	250.0	686.2	0.074	352.7	692.8	
	1	5	0.066	335.5	694.6	0.078	316.9	688.6	0.070	313.7	691.6	
	1.5	2	0.580	386.7	526.7	0.544	360.3	524.3	0.568	394.2	535.0	
	1.5	5	0.534	408.2	553.5	0.514	355.6	532.7	0.540	400.0	547.2	
	2	2	0.944	326.1	348.1	0.918	294.6	329.5	0.934	318.0	344.5	
	2	5	0.904	333.1	370.3	0.876	305.4	356.8	0.920	347.5	377.3	

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 1,500 sample size at end of two years.

Case 1b: $Z_2 \sim \text{Unif}(-1.5, 1.5)$; $D|Z_2 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_2}))$; $Y|D, Z_2 \sim \text{Bernoulli}(\exp(\beta_{Y,0} + \beta_{Y,D}D + \beta_{Y,Z}Z_2))$

$RR(Y|D) = \exp(\beta_{Y,D})$ and $RR(Y|Z_2) = \exp(\beta_{Y,Z})$.

Table 2c.1. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the case of one binary and one continuous confounder with overall prevalence of 0.05 and a binary outcome.

Freq. of Testing	Effect Size RR(Y D)	Confounders		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	
		Y and Z ₁ RR(Y Z ₁)	Y and Z ₂ RR(Y Z ₂)										
					BIMONTHLY			MONTHLY			QUARTERLY		
GS Wald - Regression and Normal Approx	1	2	2	0.040	403.5	707.3	0.048	287.5	699.2	0.036	430.0	709.6	
	1	2	5	0.044	351.8	703.8	0.046	350.9	703.0	0.046	383.5	704.5	
	1.2	2	2	0.446	385.8	571.0	0.424	376.4	574.3	0.438	410.5	584.5	
	1.2	2	5	0.394	403.2	595.2	0.396	407.0	596.0	0.422	428.2	596.9	
	1.5	2	2	0.980	271.5	280.4	0.962	268.2	285.4	0.980	294.8	303.3	
	1.5	2	5	0.928	306.1	335.9	0.950	301.9	322.8	0.940	331.5	354.8	
GS CSSP - Stratification and Error Spending	1	2	2	0.054	346.7	699.8	0.046	356.1	703.3	0.046	320.9	701.6	
	1	2	5	0.050	280.8	698.0	0.050	357.6	701.9	0.062	435.5	702.4	
	1.2	2	2	0.466	394.9	568.5	0.474	399.4	568.0	0.452	419.3	584.1	
	1.2	2	5	0.444	401.1	578.4	0.522	400.3	553.1	0.530	440.8	572.0	
	1.5	2	2	0.950	271.9	294.3	0.966	277.3	292.3	0.970	295.1	307.8	
	1.5	2	5	0.970	306.2	318.6	0.974	298.1	309.1	0.976	289.5	299.9	
GS EE Score - Regression and Unifying	1	2	2	0.066	235.5	688.0	0.070	337.7	693.2	0.050	374.4	702.7	
	1	2	5	0.056	283.9	695.6	0.066	307.3	692.8	0.054	286.7	696.6	
	1.2	2	2	0.432	321.3	547.7	0.472	340.8	541.0	0.444	392.0	574.4	
	1.2	2	5	0.390	334.4	569.6	0.430	353.6	562.4	0.434	409.8	585.4	
	1.5	2	2	0.970	259.9	273.7	0.960	257.3	275.8	0.972	277.4	289.8	
	1.5	2	5	0.956	246.9	267.7	0.950	273.9	296.2	0.954	290.6	310.3	

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 10,000 sample size at end of two years.

Case 1c: $Z_1 \sim \text{Bernoulli}(0.5)$; $Z_2 \sim \text{Unif}(-1.5, 1.5)$; $D|Z_1, Z_2 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_1} + \beta_{D,Z_2}))$; $Y|D, Z_1, Z_2 \sim \text{Bernoulli}(\exp(\beta_{Y,0} + \beta_{Y,D} + \beta_{Y,Z_1} + \beta_{Y,Z_2}))$;
 $RR(Y|D) = \exp(\beta_{Y,D})$, $RR(Y|Z_1) = \exp(\beta_{Y,Z_1})$, and $RR(Y|Z_2) = \exp(\beta_{Y,Z_2})$.

Table 2c.2. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the case of one binary and one continuous confounder with overall prevalence of 0.10 and a binary outcome.

Effect Size RR(Y D)	Confounders		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	
	Y and Z ₁ RR(Y Z ₁)	Y and Z ₂ RR(Y Z ₂)										
Freq. of Testing	BIMONTHLY			MONTHLY			QUARTERLY					
GS Wald - Regression and Normal Approx	1	2	2	0.052	388.8	702.8	0.044	320.5	702.4	0.052	429.2	704.9
	1	2	5	0.042	315.7	703.0	0.040	417.0	707.9	0.040	315.0	703.8
	1.5	2	2	0.572	384.0	527.8	0.504	394.3	555.8	0.594	408.5	535.0
	1.5	2	5	0.522	412.8	559.7	0.516	420.2	565.3	0.540	429.3	563.0
	2	2	2	0.930	322.5	350.4	0.932	320.5	347.7	0.940	346.8	369.2
	2	2	5	0.880	359.6	402.8	0.854	357.0	410.0	0.888	382.1	419.9
GS CSSP - Stratification and Error Spending	1	2	2	0.028	332.1	709.1	0.030	432.0	711.4	0.042	377.1	705.6
	1	2	5	0.040	410.3	707.6	0.030	418.0	710.9	0.032	410.6	710.1
	1.5	2	2	0.400	387.8	587.1	0.424	409.8	588.5	0.528	426.8	565.2
	1.5	2	5	0.318	412.2	622.1	0.420	387.3	580.3	0.476	447.7	590.4
	2	2	2	0.804	343.2	417.0	0.844	338.2	397.7	0.902	360.6	395.8
	2	2	5	0.698	388.5	488.6	0.826	348.4	413.0	0.886	383.8	422.1
GS EE Score - Regression and Unifying	1	2	2	0.044	253.6	699.5	0.062	300.0	694.0	0.044	310.9	702.0
	1	2	5	0.056	301.1	696.5	0.056	308.6	697.0	0.056	379.3	700.9
	1.5	2	2	0.534	343.2	518.8	0.560	364.0	520.6	0.582	390.0	527.9
	1.5	2	5	0.558	358.4	518.2	0.522	385.6	545.5	0.566	402.6	540.4
	2	2	2	0.942	294.4	319.1	0.932	302.3	330.7	0.938	310.1	335.5
	2	2	5	0.886	308.3	355.3	0.910	289.8	328.6	0.904	346.5	382.3

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 1,500 sample size at end of two years.

Case 1c: $Z_1 \sim \text{Bernoulli}(0.5)$; $Z_2 \sim \text{Unif}(-1.5, 1.5)$; $D | Z_1, Z_2 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_1} + \beta_{D,Z_2}))$; $Y | D, Z_1, Z_2 \sim \text{Bernoulli}(\exp(\beta_{Y,0} + \beta_{Y,D} + \beta_{Y,Z_1} + \beta_{Y,Z_2}))$; $RR(Y|D) = \exp(\beta_{Y,D})$, $RR(Y|Z_1) = \exp(\beta_{Y,Z_1})$, and $RR(Y|Z_2) = \exp(\beta_{Y,Z_2})$.

Table 2d.1. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the case of one binary, one continuous confounder, and one categorical confounder with overall prevalence of 0.05 and a binary outcome.

Effect Size RR(Y D)	Confounders			POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	
	Y and Z ₁ RR(Y Z ₁)	Y and Z ₂ RR(Y Z ₂)	Y and Z ₃ RR(Y Z ₃)										
Freq. of Testing					BIMONTHLY			MONTHLY			QUARTERLY		
GS Wald - Regression and Normal Approx	1	2	2	2	0.036	240.0	702.7	0.038	236.8	701.6	0.058	381.7	700.4
	1	2	2	5	0.050	366.0	702.3	0.052	378.5	702.2	0.062	351.3	697.1
	1	2	5	2	0.060	299.5	694.8	0.054	282.2	696.4	0.042	355.7	704.7
	1	2	5	5	0.058	383.8	700.5	0.056	352.5	699.4	0.038	298.4	704.0
	1.2	2	2	2	0.426	382.7	576.3	0.428	391.3	579.3	0.488	432.3	579.6
	1.2	2	2	5	0.454	382.6	566.8	0.468	393.8	567.4	0.424	425.8	595.3
	1.2	2	5	2	0.470	393.9	566.7	0.472	395.5	566.8	0.470	415.9	577.1
	1.2	2	5	5	0.388	400.2	595.9	0.386	412.4	601.3	0.432	426.7	593.3
	1.5	2	2	2	0.972	260.5	273.4	0.970	261.4	275.2	0.982	275.7	283.7
	1.5	2	2	5	0.970	276.2	289.5	0.968	278.1	292.2	0.974	287.4	298.6
	1.5	2	5	2	0.960	275.0	292.8	0.960	280.0	297.6	0.976	288.8	299.2
1.5	2	5	5	0.954	286.3	306.3	0.952	289.3	310.0	0.962	291.3	307.6	
GS CSSP - Stratification and Error Spending	1	2	2	2	0.034	413.8	709.6	0.032	352.5	708.2	0.050	302.4	699.1
	1	2	2	5	0.018	236.7	711.3	0.038	364.7	706.5	0.046	332.6	702.2
	1	2	5	2	0.046	329.3	702.0	0.042	310.0	702.8	0.036	410.0	708.8
	1	2	5	5	0.026	450.0	713.0	0.042	367.1	705.2	0.030	360.0	709.2
	1.2	2	2	2	0.404	393.6	588.2	0.418	390.1	582.1	0.452	410.2	580.0
	1.2	2	2	5	0.356	416.5	611.9	0.362	400.1	604.2	0.416	446.1	606.1
	1.2	2	5	2	0.398	382.8	585.8	0.416	420.1	595.3	0.444	426.1	589.5
	1.2	2	5	5	0.336	389.3	608.9	0.374	418.2	607.1	0.388	457.9	618.3
	1.5	2	2	2	0.954	295.8	315.3	0.970	288.7	301.7	0.976	290.1	300.4
	1.5	2	2	5	0.942	300.4	324.7	0.944	292.4	316.4	0.972	332.2	343.1
	1.5	2	5	2	0.938	280.6	307.8	0.946	290.7	313.9	0.978	319.0	327.8
1.5	2	5	5	0.926	333.0	361.7	0.936	318.1	343.8	0.958	338.4	354.4	
GS EE Score - Regression and Unifying	1	2	2	2	0.044	297.3	701.4	0.072	278.3	688.2	0.068	317.6	692.6
	1	2	2	5	0.050	283.2	698.2	0.062	301.0	694.0	0.056	334.3	698.4
	1	2	5	2	0.070	330.4	692.7	0.040	225.0	700.2	0.054	310.0	697.9
	1	2	5	5	0.052	222.1	694.1	0.046	240.0	697.9	0.066	308.2	692.8

	1.2	2	2	2	0.444	368.6	564.0	0.456	343.3	548.2	0.488	389.9	558.9
	1.2	2	2	5	0.420	328.9	555.7	0.462	351.7	549.8	0.434	386.1	575.1
	1.2	2	5	2	0.462	341.3	545.0	0.428	368.3	569.5	0.460	355.3	552.2
	1.2	2	5	5	0.384	359.8	581.7	0.422	348.2	563.1	0.408	389.1	585.0
	1.5	2	2	2	0.976	246.0	257.3	0.976	250.3	261.6	0.966	275.4	290.5
	1.5	2	2	5	0.976	248.4	259.7	0.968	260.5	275.2	0.978	266.7	276.7
	1.5	2	5	2	0.960	257.9	276.4	0.978	238.7	249.2	0.970	271.7	285.1
	1.5	2	5	5	0.942	259.6	286.3	0.958	265.4	284.5	0.962	282.2	298.8

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 10,000 sample size at end of two years.

Case 1d: $Z_1 \sim \text{Bernoulli}(0.5)$; $Z_2 \sim \text{Unif}(-1.5, 1.5)$; $Z_3 \sim \text{Multi}(0.32, 0.37, .31)$; $D | Z_1, Z_2, Z_3 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_1}Z_1 + \beta_{D,Z_2}Z_2 + \beta_{D,Z_3}Z_3))$; $Y | D, Z_1, Z_2 \sim \text{Bernoulli}(\exp(\beta_{Y,0} + \beta_{Y,D}D + \beta_{Y,Z_1}Z_1 + \beta_{Y,Z_2}Z_2 + \beta_{Y,Z_3}Z_3))$;

$RR(Y|D) = \exp(\beta_{Y,D})$, $RR(Y|Z_1) = \exp(\beta_{Y,Z_1})$, $RR(Y|Z_2) = \exp(\beta_{Y,Z_2})$, and $RR(Y|Z_3) = \exp(\beta_{Y,Z_3})$.

Table 2d.2. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the case of one binary, one continuous confounder, and one categorical confounder with overall prevalence of 0.10 and a binary outcome.

Effect Size RR(Y D)	Confounders			POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	
	Y and Z ₁ RR(Y Z ₁)	Y and Z ₂ RR(Y Z ₂)	Y and Z ₃ RR(Y Z ₃)										
Freq. of Testing					BIMONTHLY			MONTHLY			QUARTERLY		
GS Wald - Regression and Normal Approx	1	2	2	2	0.038	397.9	707.8	0.038	382.1	707.2	0.062	362.9	697.9
	1	2	2	5	0.064	310.3	693.8	0.068	303.5	691.7	0.042	424.3	707.6
	1	2	5	2	0.048	322.5	700.9	0.060	325.0	696.3	0.038	374.2	706.9
	1	2	5	5	0.040	429.0	708.4	0.040	454.5	709.4	0.042	432.9	707.9
	1.5	2	2	2	0.554	387.1	535.6	0.560	390.5	535.5	0.618	413.6	530.6
	1.5	2	2	5	0.530	412.0	556.7	0.534	415.5	557.4	0.574	423.3	549.7
	1.5	2	5	2	0.556	407.9	546.5	0.556	406.8	545.9	0.554	415.6	551.3
	1.5	2	5	5	0.446	425.5	588.7	0.448	434.1	591.9	0.508	436.9	576.2
	2	2	2	2	0.942	307.4	331.3	0.944	308.6	331.6	0.964	326.4	340.6
	2	2	2	5	0.930	320.1	348.1	0.932	323.8	350.7	0.952	338.6	356.9
2	2	5	2	0.912	318.1	353.5	0.920	323.9	355.6	0.940	337.4	360.4	
2	2	5	5	0.902	333.9	371.7	0.906	336.0	372.1	0.900	341.6	379.4	
GS CSSP - Stratification and Error Spending	1	2	2	2	0.026	378.5	711.1	0.036	398.3	708.4	0.026	422.3	712.3
	1	2	2	5	0.024	326.3	710.6	0.036	268.3	703.7	0.032	382.5	709.2
	1	2	5	2	0.016	313.1	713.5	0.028	304.3	708.4	0.024	412.5	712.6
	1	2	5	5	0.020	397.5	713.6	0.038	330.0	705.2	0.012	435.0	716.6
	1.5	2	2	2	0.210	408.7	654.6	0.320	425.4	625.7	0.448	459.6	603.4
	1.5	2	2	5	0.236	421.8	649.6	0.268	380.8	629.1	0.404	479.0	622.6
	1.5	2	5	2	0.210	369.9	646.5	0.300	411.4	627.4	0.408	430.1	601.7
	1.5	2	5	5	0.166	392.2	665.6	0.204	410.3	656.8	0.312	468.5	641.5
	2	2	2	2	0.556	361.2	520.5	0.762	371.7	454.6	0.890	372.7	410.9
	2	2	2	5	0.480	373.9	553.9	0.622	383.1	510.4	0.818	400.3	458.5
2	2	5	2	0.504	393.0	555.2	0.618	399.9	522.2	0.830	395.4	450.5	
2	2	5	5	0.406	416.6	596.8	0.566	407.7	543.2	0.756	424.3	496.4	
GS EE Score - Regression and Unifying	1	2	2	2	0.042	245.7	700.1	0.060	330.0	696.6	0.050	284.4	698.2
	1	2	2	5	0.056	375.5	700.7	0.074	296.8	688.7	0.066	346.4	695.3
	1	2	5	2	0.058	290.7	695.1	0.062	253.5	691.1	0.048	337.5	701.6
	1	2	5	5	0.068	288.1	690.6	0.050	330.0	700.5	0.036	385.0	707.9
	1.5	2	2	2	0.608	355.7	498.5	0.594	350.7	500.6	0.646	383.1	502.4
	1.5	2	2	5	0.618	357.6	496.1	0.566	363.4	518.2	0.556	386.2	534.4
	1.5	2	5	2	0.542	354.2	521.7	0.522	369.3	536.9	0.570	397.3	536.0
	1.5	2	5	5	0.522	365.3	534.8	0.516	389.7	549.5	0.504	388.2	552.8

	2	2	2	2	0.944	285.2	309.6	0.938	282.1	309.2	0.948	298.9	320.8
	2	2	2	5	0.920	290.1	324.5	0.932	300.1	328.6	0.936	330.6	355.5
	2	2	5	2	0.956	308.2	326.3	0.904	298.1	338.6	0.960	304.9	321.5
	2	2	5	5	0.892	314.4	358.2	0.874	327.3	376.8	0.920	343.6	373.7

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 1,500 sample size at end of two years.

Case 1d: $Z_1 \sim \text{Bernoulli}(0.5)$; $Z_2 \sim \text{Unif}(-1.5, 1.5)$; $Z_3 \sim \text{Multi}(0.32, 0.37, .31)$; $D | Z_1, Z_2, Z_3 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_1}Z_1 + \beta_{D,Z_2}Z_2 + \beta_{D,Z_3}Z_3))$; $Y | D, Z_1, Z_2 \sim \text{Bernoulli}(\exp(\beta_{Y,0} + \beta_{Y,D}D + \beta_{Y,Z_1}Z_1 + \beta_{Y,Z_2}Z_2 + \beta_{Y,Z_3}Z_3))$;

$RR(Y|D) = \exp(\beta_{Y,D})$, $RR(Y|Z_1) = \exp(\beta_{Y,Z_1})$, $RR(Y|Z_2) = \exp(\beta_{Y,Z_2})$, and $RR(Y|Z_3) = \exp(\beta_{Y,Z_3})$.

Table 3a. Power, days to rejection, and days to study end across varying strength of confounding and probability of outcome for the simple case of one binary confounder with a chronic exposure.

	Effect Size RR(Y D)	Confounder Y and Z ₁		Days to Rejection	Days to Study End	Days to Rejection	Days to Study End
		RR(Y Z ₁)	POWER				
				P(Y)=0.01		P(Y)=0.05	
GS Wald - Regression and Normal Approx	1	2	0.054	356.7	700.4	0.050	378.0
	1	5	0.042	424.3	707.6	0.040	288.0
	1.5	2	0.488	475.5	600.7	0.992	327.0
	1.5	5	0.528	501.8	604.8	0.982	319.9
	2	2	0.870	421.7	460.4	1.000	215.6
	2	5	0.896	416.5	448.0	1.000	211.7
GS CSSP - Stratification and Error Spending	1	2	0.064	435.9	701.8	0.050	277.2
	1	5	0.054	340.0	699.5	0.046	328.7
	1.5	2	0.538	465.1	582.8	0.984	336.4
	1.5	5	0.490	472.8	598.9	0.996	335.1
	2	2	0.902	409.3	439.7	1.000	211.7
	2	5	0.900	410.0	441.0	1.000	215.6
GS EE Score - Regression and Unifying	1	2	0.040	373.5	706.1	0.046	317.0
	1	5	0.062	362.9	697.9	0.050	345.6
	1.5	2	0.504	451.4	584.6	0.988	315.7
	1.5	5	0.516	456.3	583.9	0.982	323.3
	2	2	0.894	408.3	441.4	1.000	211.5
	2	5	0.904	381.7	414.2	1.000	212.8

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses quarterly (8 analyses), first analysis at second quarter, and 10,000 sample size at end of two years.

Case 2a: $Z_1 \sim \text{Bernoulli}(0.5)$; $D | Z_1 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_1} Z_1))$; $C \sim \text{Weibull}(2, 200)$; $T | D, Z_1 \sim \text{exponential}(\exp(\beta_{Y,0} + \beta_{Y,D} D + \beta_{Y,Z_1} Z_1) / 720)$; $Y = I(T < C, T < (tl - S))$;
 $D = \min(tl - S, C, T)$

$RR(Y|D) = \exp(\beta_{Y,D})$ and $RR(Y|Z_1) = \exp(\beta_{Y,Z_1})$.

Table 3b. Power, days to rejection, and days to study end across varying strength of confounding and probability of outcome for the simple case of one continuous confounder with a chronic exposure.

	Effect Size RR(Y D)	Confounder Y and Z ₂		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End
		RR(Y Z ₂)							
GS Wald - Regression and Normal Approx	1	2	0.026	436.2	712.6	0.056	340.7	698.8	
	1	5	0.042	480.0	709.9	0.042	432.9	707.9	
	1.5	2	0.424	478.0	617.4	0.964	344.1	357.7	
	1.5	5	0.350	518.9	649.6	0.958	365.3	380.2	
	2	2	0.824	440.4	489.6	1.000	225.2	225.2	
	2	5	0.798	473.0	522.9	1.000	240.5	240.5	
GS CSSP - Stratification and Error Spending	1	2	0.068	407.6	698.8	0.050	367.2	702.4	
	1	5	0.064	424.7	701.1	0.088	421.4	693.7	
	1.5	2	0.520	448.6	578.9	0.980	339.6	347.2	
	1.5	5	0.466	469.7	603.4	0.988	323.2	328.0	
	2	2	0.868	429.7	468.0	1.000	229.7	229.7	
	2	5	0.852	440.9	482.2	1.000	223.9	223.9	
GS EE Score - Regression and Unifying	1	2	0.062	362.9	697.9	0.056	356.8	699.7	
	1	5	0.078	348.5	691.0	0.066	398.2	698.8	
	1.5	2	0.488	440.8	583.7	0.970	340.1	351.5	
	1.5	5	0.462	425.1	583.7	0.970	340.3	351.7	
	2	2	0.830	397.7	452.5	1.000	218.5	218.5	
	2	5	0.858	398.2	443.9	1.000	228.4	228.4	

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses quarterly (8 analyses), first analysis at second quarter, and 10,000 sample size at end of two years.

Case 2b: $Z_2 \sim \text{unif}(-1.5, 1.5)$; $D | Z_2 \sim (\exp(\beta_{D,0} + \beta_{D,Z_2}))$; $C \sim \text{Weibull}(2, 200)$; $T | D, Z_2 \sim \text{exponential}(\exp(\beta_{Y,0} + \beta_{Y,D} + \beta_{Y,Z_2 Z_2}) / 720)$; $Y = I(T < C, T < (tI - S))$;

$D = \min(tI - S, C, T)$

$RR(Y|D) = \exp(\beta_{Y,D})$ and $RR(Y|Z_2) = \exp(\beta_{Y,Z_2})$.

Table 3c. Power, days to rejection, and days to study end across varying strength of confounding and probability of outcome for the case of one binary confounder and one continuous confounder with a chronic exposure.

	Effect Size RR(Y D)	Confounders		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	
		Y and Z ₁	Y and Z ₂							
		RR(Y Z ₁)	RR(Y Z ₂)							
					P(Y)=0.01			P(Y)=0.05		
GS Wald - Regression and Normal Approx	1	2	2	0.042	462.9	709.2	0.054	406.7	703.1	
	1	2	5	0.034	439.4	710.5	0.056	379.3	700.9	
	1	5	2	0.036	500.0	712.1	0.050	414.0	704.7	
	1	5	5	0.028	520.7	714.4	0.062	365.8	698.0	
	1.5	2	2	0.386	484.0	628.9	0.974	351.3	360.9	
	1.5	2	5	0.384	504.4	637.2	0.956	361.5	377.3	
	1.5	5	2	0.444	495.0	620.1	0.954	351.1	368.1	
	1.5	5	5	0.300	525.6	661.7	0.944	379.8	398.9	
	2	2	2	0.802	455.6	508.0	1.000	230.0	230.0	
	2	5	2	0.758	478.5	536.9	1.000	237.1	237.1	
GS CSSP - Stratification and Error Spending	2	2	5	0.806	458.5	509.2	1.000	232.0	232.0	
	2	5	5	0.668	497.2	571.1	1.000	243.9	243.9	
	1	2	2	0.056	421.1	703.3	0.062	426.8	701.8	
	1	2	5	0.058	443.8	704.0	0.076	364.7	693.0	
	1	5	2	0.052	460.4	706.5	0.056	369.6	700.4	
	1	5	5	0.076	395.5	695.3	0.088	370.2	689.2	
	1.5	2	2	0.482	458.6	594.0	0.984	346.8	352.8	
	1.5	2	5	0.436	475.2	613.3	0.980	340.3	347.9	
	1.5	5	2	0.428	461.8	609.5	0.964	336.8	350.6	
	1.5	5	5	0.370	489.4	634.7	0.978	345.3	353.5	
GS EE Score - Regression and Unifying	2	2	2	0.826	424.5	475.9	1.000	234.5	234.5	
	2	5	2	0.812	425.8	481.1	1.000	230.4	230.4	
	2	2	5	0.834	437.7	484.6	1.000	229.7	229.7	
	2	5	5	0.778	444.9	506.0	1.000	237.6	237.6	
	1	2	2	0.060	354.0	698.0	0.056	363.2	700.0	
	1	2	5	0.072	317.5	691.0	0.062	319.4	695.2	
	1	5	2	0.078	383.1	693.7	0.048	360.0	702.7	
	1	5	5	0.106	361.7	682.0	0.070	352.3	694.3	
	1.5	2	2	0.492	431.0	577.8	0.982	339.5	346.3	
	1.5	2	5	0.518	437.5	573.7	0.958	345.2	360.9	
GS EE Score - Regression and Unifying	1.5	5	2	0.534	436.5	568.6	0.958	331.1	347.4	
	1.5	5	5	0.420	429.0	597.8	0.948	343.7	363.2	
	2	2	2	0.842	388.2	440.6	1.000	223.0	223.0	
	2	5	2	0.834	406.0	458.1	1.000	219.2	219.2	
	2	2	5	0.830	396.7	451.6	1.000	230.8	230.8	
	2	5	5	0.814	396.7	456.8	1.000	234.5	234.5	

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses quarterly (8 analyses), first analysis at second quarter and N=10,000

Case 2c: Z₁~Bernoulli(0.5); Z₂~Unif(-1.5,1.5); D|Z₁,Z₂~Bernoulli(exp(β_{D,0}+β_{D,Z1}+β_{D,Z2})); C~Weibull(2,200);

T|D,Z₁,Z₂~exponential(exp(β_{T,0}+β_{T,D}+β_{T,Z1}+β_{T,Z2})/720); Y=(T<C,T<(tl-S)); D=min(tl-S,C,T)

RR(Y|D)=exp(β_{Y,D}), RR(Y|Z₁)=exp(β_{Y,Z1}), and RR(Y|Z₂)=exp(β_{Y,Z2}).

Table 3d. Power, days to rejection, and days to study end across varying strength of confounding and probability of outcome for the case of one binary confounder, one continuous confounder, and one categorical confounder with a chronic exposure.

	Effect Size	Confounders			POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End
		Y and Z ₁	Y and Z ₂	Y and Z ₃						
		RR(Y Z ₁)	RR(Y Z ₂)	RR(Y Z ₃)						
					P(Y)=0.01			P(Y)=0.05		
GS Wald - Regression and Normal Approx	1	2	2	2	0.040	477.0	710.3	0.070	383.1	696.4
	1	2	2	5	0.020	432.0	714.2	0.072	342.5	692.8
	1	2	5	2	0.022	400.9	713.0	0.060	354.0	698.0
	1	2	5	5	0.032	416.3	710.3	0.036	415.0	709.0
	1.5	2	2	2	0.462	486.6	612.2	0.990	334.9	338.8
	1.5	2	2	5	0.444	486.1	616.1	0.988	343.4	347.9
	1.5	2	5	2	0.464	497.3	616.7	0.976	339.0	348.1
	1.5	2	5	5	0.400	509.9	635.9	0.986	353.4	358.6
	2	2	2	2	0.842	449.6	492.3	1.000	215.6	215.6
	2	2	2	5	0.838	455.2	498.1	1.000	223.7	223.7
GS CSSP - Stratification and Error Spending	2	2	5	2	0.832	454.1	498.8	1.000	224.3	224.3
	2	2	5	5	0.758	479.9	538.0	1.000	224.5	224.5
	1	2	2	2	0.060	408.0	701.3	0.070	393.4	697.1
	1	2	2	5	0.032	438.8	711.0	0.056	347.1	699.1
	1	2	5	2	0.038	416.8	708.5	0.078	371.5	692.8
	1	2	5	5	0.074	384.3	695.2	0.072	427.5	698.9
	1.5	2	2	2	0.478	456.0	593.8	0.982	335.1	342.0
	1.5	2	2	5	0.502	468.6	593.8	0.988	338.9	343.4
	1.5	2	5	2	0.544	482.4	590.8	0.996	332.7	334.3
	1.5	2	5	5	0.440	463.1	607.0	0.994	329.8	332.1
GS EE Score - Regression and Unifying	2	2	2	2	0.892	411.1	444.4	1.000	218.2	218.2
	2	2	2	5	0.888	408.2	443.2	1.000	219.2	219.2
	2	2	5	2	0.872	425.4	463.1	1.000	211.0	211.0
	2	2	5	5	0.864	421.3	461.9	1.000	217.1	217.1
	1	2	2	2	0.074	345.4	692.3	0.058	381.7	700.4
	1	2	2	5	0.076	333.9	690.7	0.056	356.8	699.7
	1	2	5	2	0.096	361.9	685.6	0.048	348.8	702.2
	1	2	5	5	0.092	344.3	685.4	0.052	328.8	699.7
	1.5	2	2	2	0.494	436.9	580.1	0.990	331.8	335.7
	1.5	2	2	5	0.506	416.2	566.3	0.984	331.6	337.9
GS EE Score - Regression and Unifying	1.5	2	5	2	0.524	423.2	564.5	0.982	334.2	341.1
	1.5	2	5	5	0.468	425.0	581.9	0.974	318.0	328.5
	2	2	2	2	0.860	385.7	432.5	1.000	214.2	214.2
	2	2	2	5	0.896	397.8	431.3	1.000	218.9	218.9
	2	2	5	2	0.848	398.0	446.9	1.000	218.7	218.7
	2	2	5	5	0.798	406.9	470.2	1.000	219.2	219.2

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses quarterly (8 analyses), first analysis at second quarter, and N=10,000

Case 2d: Z₁~Bernoulli(0.5); Z₂~Unif(-1.5,1.5); Z₃~Multi(0.32,0.37,.31); D|Z₁,Z₂,Z₃~Bernoulli Bernoulli(exp(β_{D,0}+β_{D,Z1}+β_{D,Z2}+β_{D,Z3}));

C~Weibull(2,200); T|D,Z₁,Z₂,Z₃~exponential(exp(β_{V,0}+β_{V,D}+β_{V,Z1}+β_{V,Z2}+β_{V,Z3})/720); Y=I(T<C,T<(tl-S)); D=min(tl-S,C,T)RR(Y|D)=exp(β_{V,D}),

RR(Y|Z₁)=exp(β_{V,Z1}), RR(Y|Z₂)=exp(β_{V,Z2}) and RR(Y|Z₃)=exp(β_{V,Z3}).

Table 4a.1. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the simple case of one binary confounder and overall prevalence of 0.05 with a chronic exposure.

Effect Size	Confounder		POWE R	Days to Rejection	Days to Study End	POWE R	Days to Rejection	Days to Study End	POWE R	Days to Rejection	Days to Study End
	Y and Z ₁	RR(Y Z ₁)									
RR(Y D)	RR(Y Z ₁)		BIMONTHLY			MONTHLY			QUARTERLY		
GS Wald - Regression and Normal Approx	1	2	0.056	327.3	698.0	0.056	322.5	697.7	0.058	356.9	698.9
	1	5	0.042	399.3	706.5	0.042	370.0	705.3	0.040	378.0	706.3
	1.2	2	0.502	422.8	570.8	0.508	436.1	575.8	0.518	468.1	589.5
	1.2	5	0.500	450.1	585.0	0.480	456.9	593.7	0.482	470.9	599.9
	1.5	2	0.992	308.3	311.6	0.992	308.0	311.3	0.988	331.8	336.4
	1.5	5	0.990	310.4	314.5	0.996	302.7	304.3	0.996	327.1	328.7
GS C SSP - Stratification and Error Spending	1	2	0.046	343.7	702.7	0.044	340.9	703.3	0.056	388.9	701.5
	1	5	0.060	353.5	698.0	0.052	375.0	702.1	0.062	418.1	701.3
	1.2	2	0.472	418.5	577.7	0.524	434.3	570.3	0.478	430.4	581.6
	1.2	5	0.420	407.9	588.9	0.486	410.5	569.6	0.506	445.7	581.2
	1.5	2	0.978	310.8	319.8	0.990	297.8	302.0	0.992	321.4	324.5
	1.5	5	0.978	307.4	316.5	0.982	302.4	310.0	0.986	339.6	344.9
GS EE Score - Regression and Unifying	1	2	0.042	295.0	702.2	0.050	290.4	698.5	0.076	333.9	690.7
	1	5	0.070	261.0	687.9	0.058	281.4	694.6	0.044	270.0	700.2
	1.2	2	0.452	397.8	574.4	0.478	394.3	564.3	0.550	434.9	563.2
	1.2	5	0.508	396.9	555.8	0.454	397.3	573.5	0.514	444.7	578.5
	1.5	2	0.978	280.5	290.2	0.992	282.5	286.0	0.986	314.2	319.9
	1.5	5	0.984	277.3	284.4	0.992	285.9	289.4	0.990	311.6	315.7

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 10,000 sample size at end of two years.

Case 2a: $Z_1 \sim \text{Bernoulli}(0.5)$; $D | Z_1 \sim \text{Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z_1} Z_1))$; $C \sim \text{Weibull}(2, 200)$; $T | D, Z_1 \sim \text{exponential}(\exp(\beta_{Y,0} + \beta_{Y,D} D + \beta_{Y,Z_1} Z_1) / 720)$; $Y = I(T < C, T < (tI - S))$;

$D = \min(tI - S, C, T)$

$RR(Y|D) = \exp(\beta_{Y,D})$ and $RR(Y|Z_1) = \exp(\beta_{Y,Z_1})$.

Table 4a.2. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the simple case of one binary confounder and overall prevalence of 0.10 with a chronic exposure.

Effect Size	Confounder Y and Z ₁		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End
	RR(Y D)	RR(Y Z ₁)									
Freq. of Testing	BIMONTHLY			MONTHLY			QUARTERLY				
GS Wald - Regression and Normal Approx	1	2	0.054	397.2	702.6	0.054	402.2	702.8	0.052	446.5	705.8
	1	5	0.042	470.0	709.5	0.042	460.0	709.1	0.040	504.0	711.4
	1.5	2	0.658	446.5	540.0	0.666	445.4	537.1	0.678	459.6	543.4
	1.5	5	0.642	443.0	542.2	0.654	447.7	541.9	0.666	469.5	553.1
	2	2	0.978	341.0	349.3	0.978	343.4	351.7	0.978	358.9	366.8
	2	5	0.980	359.4	366.6	0.980	364.4	371.5	0.978	379.1	386.6
GS CSSP - Stratification and Error Spending	1	2	0.066	319.1	693.5	0.048	388.7	704.1	0.052	401.5	703.4
	1	5	0.058	393.6	701.1	0.050	291.6	698.6	0.036	410.0	708.8
	1.5	2	0.620	416.8	532.0	0.614	428.0	540.7	0.662	444.0	537.3
	1.5	5	0.606	412.3	533.6	0.652	422.4	526.0	0.680	463.0	545.2
	2	2	0.956	327.9	345.1	0.972	327.7	338.6	0.968	350.9	362.7
	2	5	0.968	338.1	350.3	0.968	341.2	353.3	0.986	373.5	378.4
GS EE Score - Regression and Unifying	1	2	0.056	308.6	697.0	0.050	375.6	702.8	0.044	347.7	703.6
	1	5	0.066	315.9	693.3	0.046	354.8	703.2	0.048	393.8	704.3
	1.5	2	0.590	381.8	520.5	0.620	400.2	521.7	0.658	426.7	527.0
	1.5	5	0.630	390.0	512.1	0.596	425.4	544.4	0.622	425.4	536.8
	2	2	0.970	314.9	327.1	0.978	328.7	337.3	0.984	322.1	328.5
	2	5	0.972	305.7	317.3	0.970	317.3	329.3	0.972	327.8	338.8

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 1,500 sample size at end of two years.

Case 2a: Z₁~Bernoulli(0.5); D|Z₁~Bernoulli(exp(β_{D,0}+β_{D,Z1})); C~Weibull(2,200); T|D,Z₁~exponential(exp(β_{v,0}+β_{v,D}D+β_{v,Z1}Z₁)/720); Y=(T<C,T<(tI-S));

D=min(tI-S,C,T)

RR(Y|D)=exp(β_{v,D}) and RR(Y|Z₁)=exp(β_{v,Z1}).

Table 4b.1. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the simple case of one continuous confounder and overall prevalence of 0.05 with a chronic exposure.

Effect Size	Confounder		POWE R	Days to Rejection	Days to Study End	POWE R	Days to Rejection	Days to Study End	POWE R	Days to Rejection	Days to Study End
	Y and Z ₂	RR(Y Z ₂)									
RR(Y D)	RR(Y Z ₂)		BIMONTHLY			MONTHLY			QUARTERLY		
GS Wald - Regression and Normal Approx	1	2	0.028	367.5	710.1	0.028	330.0	709.1	0.030	414.0	710.8
	1	5	0.046	423.9	706.4	0.046	431.7	706.7	0.040	427.5	708.3
	1.2	2	0.466	451.0	594.7	0.510	444.9	579.7	0.508	469.1	592.6
	1.2	5	0.432	461.2	608.2	0.432	448.1	602.5	0.454	477.4	609.8
	1.5	2	0.970	327.1	338.9	0.946	334.6	355.4	0.954	358.1	374.8
	1.5	5	0.978	344.3	352.6	0.962	341.4	355.8	0.964	361.3	374.2
GS CSSP - Stratification and Error Spending	1	2	0.064	244.7	689.6	0.050	418.8	704.9	0.074	391.6	695.7
	1	5	0.096	360.3	685.5	0.110	386.2	683.3	0.076	414.5	696.8
	1.2	2	0.468	421.0	580.1	0.480	416.9	574.5	0.498	456.1	588.6
	1.2	5	0.530	418.5	560.2	0.548	415.1	552.9	0.598	436.8	550.6
	1.5	2	0.976	316.9	326.6	0.978	324.3	333.0	0.988	335.8	340.4
	1.5	5	0.986	315.9	321.5	0.990	319.6	323.6	0.986	327.7	333.2
GS EE Score - Regression and Unifying	1	2	0.074	276.1	687.2	0.046	296.1	700.5	0.056	331.1	698.2
	1	5	0.092	247.5	676.5	0.076	266.8	685.6	0.084	330.0	687.2
	1.2	2	0.436	378.4	571.1	0.462	399.6	572.0	0.472	451.9	593.5
	1.2	5	0.418	350.2	565.4	0.396	367.0	580.2	0.486	383.3	556.4
	1.5	2	0.970	300.6	313.2	0.982	299.9	307.4	0.972	331.7	342.5
	1.5	5	0.958	298.3	316.0	0.948	314.4	335.5	0.972	340.6	351.2

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 10,000 sample size at end of two years.

Case 2b: Z₂~unif(-1.5,1.5); D|Z₂~(exp(β_{D,0}+β_{D,Z2})); C~Weibull(2,200); T|D,Z₂~exponential(exp(β_{Y,0}+β_{Y,D}+β_{Y,Z2})/720); Y=I(T<C,T<(tI-S));

D=min(tI-S,C,T)

RR(Y|D)=exp(β_{Y,D}) and RR(Y|Z₂)=exp(β_{Y,Z2}).

Table 4b.2. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the simple case of one continuous confounder and overall prevalence of 0.10 with a chronic exposure.

Effect Size	Confounder		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End
	RR(Y D)	RR(Y Z ₂)									
Freq. of Testing		BIMONTHLY			MONTHLY			QUARTERLY			
GS Wald - Regression and Normal Approx	1	2	0.050	393.6	703.7	0.050	392.4	703.6	0.058	412.8	702.2
	1	5	0.022	433.6	713.7	0.028	469.3	713.0	0.026	491.5	714.1
	1.5	2	0.594	442.1	554.9	0.606	444.6	553.1	0.602	465.2	566.6
	1.5	5	0.526	454.2	580.2	0.530	452.0	578.0	0.552	473.5	583.9
	2	2	0.940	375.4	396.0	0.946	378.2	396.7	0.948	396.1	412.9
	2	5	0.914	392.8	420.9	0.922	394.3	419.7	0.932	414.8	435.6
GS CSSP - Stratification and Error Spending	1	2	0.070	286.7	689.7	0.046	374.3	704.1	0.056	376.1	700.7
	1	5	0.072	381.3	695.6	0.074	317.0	690.2	0.068	399.7	698.2
	1.5	2	0.594	413.8	538.1	0.586	442.3	557.3	0.638	439.0	540.7
	1.5	5	0.622	409.2	526.7	0.598	424.8	543.5	0.658	457.9	547.6
	2	2	0.960	343.5	358.5	0.946	342.6	362.9	0.958	359.8	374.9
GS EE Score - Regression and Unifying	2	5	0.944	346.1	367.1	0.958	354.2	369.6	0.948	376.3	394.2
	1	2	0.054	308.3	697.8	0.070	327.4	692.5	0.060	699.3	375.0
	1	5	0.066	291.4	691.7	0.054	271.1	695.8	0.076	689.0	312.6
	1.5	2	0.568	397.5	536.8	0.584	411.4	539.8	0.612	527.2	405.0
	1.5	5	0.572	375.1	522.7	0.572	373.1	521.6	0.568	544.0	410.1
	2	2	0.964	318.3	332.7	0.962	339.5	353.9	0.956	366.8	350.6
2	5	0.912	330.6	364.9	0.928	344.9	371.9	0.940	377.5	355.6	

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 1,500 sample size at end of two years.

Case 2b: Z₂~unif(-1.5,1.5); D|Z₂~(exp(β_{D,0}+β_{D,Z2})); C~Weibull(2,200); T|D,Z₂~exponential(exp(β_{Y,0}+β_{Y,D}+β_{Y,Z2})/720); Y=l(T<C,T<(tl-S)); D=min(tl-S,C,T)

RR(Y|D)=exp(β_{Y,D}) and RR(Y|Z₂)=exp(β_{Y,Z2}).

Table 4c.1. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the case of one binary and one continuous confounder with overall prevalence of 0.05 and a chronic exposure.

Freq. of Testing	Effect Size RR(Y D)	Confounders		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End
		Y and Z ₁ RR(Y Z ₁)	Y and Z ₂ RR(Y Z ₂)									
				BIMONTHLY			MONTHLY			QUARTERLY		
GS Wald - Regression and Normal Approx	1	2	2	0.038	388.4	707.4	0.052	345.0	700.5	0.050	446.4	706.3
	1	2	5	0.030	354.0	709.0	0.052	363.5	701.5	0.054	386.7	702.0
	1.2	2	2	0.474	438.7	586.7	0.450	430.8	589.9	0.472	482.0	607.7
	1.2	2	5	0.386	455.8	618.0	0.394	439.9	609.7	0.438	453.3	603.2
	1.5	2	2	0.978	320.6	329.4	0.970	330.1	341.8	0.972	361.7	371.7
	1.5	5	2	0.964	336.3	350.1	0.942	357.0	378.1	0.956	373.4	388.6
GS CSSP - Stratification and Error Spending	1	2	2	0.054	352.2	700.1	0.066	375.5	697.3	0.054	393.3	702.4
	1	2	5	0.078	389.2	694.2	0.100	342.0	682.2	0.088	411.1	692.8
	1.2	2	2	0.450	423.8	586.7	0.498	438.6	579.8	0.518	447.2	578.7
	1.2	2	5	0.510	413.6	563.8	0.568	441.9	562.0	0.568	455.4	569.7
	1.5	2	2	0.968	340.1	352.2	0.958	325.5	342.1	0.984	342.3	348.3
	1.5	2	5	0.980	324.2	332.1	0.970	322.6	334.6	0.978	347.1	355.3
GS EE Score - Regression and Unifying	1	2	2	0.056	312.3	697.2	0.058	262.8	693.5	0.056	337.5	698.6
	1	2	5	0.076	245.1	683.9	0.066	250.0	689.0	0.058	291.7	695.2
	1.2	2	2	0.436	386.6	574.7	0.438	389.7	575.3	0.494	452.6	587.9
	1.2	2	5	0.442	365.0	563.1	0.442	391.4	574.7	0.414	439.6	603.9
	1.5	2	2	0.948	307.2	328.7	0.976	307.0	316.9	0.968	330.2	342.7
	1.5	2	5	0.950	315.6	335.8	0.958	322.5	339.2	0.940	343.9	366.5

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 10,000 sample size at end of two years.

Case 2c: Z₁~Bernoulli(0.5); Z₂~Unif(-1.5,1.5); D|Z₁,Z₂~Bernoulli(exp(β_{D,0}+β_{D,Z1}+β_{D,Z2})); C~Weibull(2,200); T|D,Z₁,Z₂~exponential(exp(β_{T,0}+β_{T,D}+β_{T,Z1}+β_{T,Z2})/720); Y=I(T<C,T<(tl-S)); D=min(tl-S,C,T)

RR(Y|D)=exp(β_{Y,D}), RR(Y|Z₁)=exp(β_{Y,Z1}), and RR(Y|Z₂)=exp(β_{Y,Z2}).

Table 4c.2. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the case of one binary and one continuous confounder with overall prevalence of 0.10 and a chronic exposure.

Freq. of Testing	Effect Size RR(Y D)	Confounders		POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	
		Y and Z ₁ RR(Y Z ₁)	Y and Z ₂ RR(Y Z ₂)										
					BIMONTHLY			MONTHLY			QUARTERLY		
GS Wald - Regression and Normal Approx	1	2	2	0.042	411.4	707.0	0.034	374.1	708.2	0.046	442.2	707.2	
	1	2	5	0.034	412.9	709.6	0.042	425.7	707.6	0.042	467.1	709.4	
	1.5	2	2	0.568	438.3	560.0	0.596	461.8	566.1	0.580	467.7	573.7	
	1.5	2	5	0.504	453.2	585.5	0.534	448.1	574.8	0.508	472.3	594.2	
	2	2	2	0.940	382.8	403.0	0.954	376.4	392.2	0.960	382.9	396.4	
	2	2	5	0.906	401.7	431.6	0.926	394.7	418.8	0.944	398.7	416.7	
GS CSSP - Stratification and Error Spending	1	2	2	0.082	352.0	689.8	0.066	390.9	698.3	0.062	444.2	702.9	
	1	2	5	0.096	354.4	684.9	0.066	370.9	697.0	0.058	372.4	699.8	
	1.5	2	2	0.550	430.1	560.6	0.582	415.6	542.8	0.604	433.3	546.8	
	1.5	2	5	0.588	422.6	545.1	0.598	415.1	537.7	0.654	441.2	537.7	
	2	2	2	0.936	348.9	372.7	0.954	351.4	368.4	0.964	376.6	389.0	
	2	2	5	0.926	346.8	374.5	0.932	346.9	372.3	0.934	360.4	384.1	
GS EE Score - Regression and Unifying	1	2	2	0.060	293.0	694.4	0.064	276.6	691.6	0.036	707.9	385.0	
	1	2	5	0.064	331.9	695.2	0.072	257.5	686.7	0.066	698.2	390.0	
	1.5	2	2	0.582	384.1	524.5	0.570	391.9	533.0	0.618	547.9	441.6	
	1.5	2	5	0.520	401.3	554.3	0.586	398.1	531.4	0.608	550.8	441.7	
	2	2	2	0.918	331.5	363.3	0.962	338.7	353.2	0.944	389.3	369.7	
	2	2	5	0.918	345.9	376.6	0.924	359.7	387.1	0.924	410.4	384.9	

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 1,500 sample size at end of two years.

Case 2c: Z₁~Bernoulli(0.5); Z₂~Unif(-1.5,1.5); D|Z₁,Z₂~Bernoulli(exp(β_{D,0}+β_{D,Z1}+β_{D,Z2})); C~Weibull(2,200); T|D,Z₁,Z₂~exponential(exp(β_{T,0}+β_{T,D}+β_{T,Z1}+β_{T,Z2})/720); Y=I(T<C,T<(tl-S)); D=min(tl-S,C,T)

RR(Y|D)=exp(β_{Y,D}), RR(Y|Z₁)=exp(β_{Y,Z1}), and RR(Y|Z₂)=exp(β_{Y,Z2}).

Table 4d.1. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the case of one binary, one continuous confounder, and one categorical confounder with overall prevalence of 0.05 and a chronic exposure.

Effect Size	Confounders			POWER	Days to Rejection n	Days to Study End	POWER	Days to Rejection n	Days to Study End	POWER	Days to Rejection n	Days to Study End	
	Y and Z ₁	Y and Z ₂	Y and Z ₃										
RR(Y D)	RR(Y Z ₁)	RR(Y Z ₂)	RR(Y Z ₃)	BIMONTHLY			MONTHLY			QUARTERLY			
GS Wald - Regression and Normal Approx	1	2	2	2	0.050	385.8	703.3	0.056	367.5	700.3	0.074	690.1	316.2
	1	2	2	5	0.048	355.0	702.5	0.050	340.8	701.0	0.090	680.4	280.0
	1	2	5	2	0.034	356.5	707.6	0.040	318.0	703.9	0.060	697.1	339.0
	1	2	5	5	0.046	365.9	703.7	0.048	343.7	701.9	0.054	697.0	293.3
	1.2	2	2	2	0.506	424.8	570.6	0.476	420.1	577.3	0.470	580.7	423.6
	1.2	2	2	5	0.452	431.2	589.5	0.494	430.1	576.8	0.466	582.5	424.9
	1.2	2	5	2	0.440	437.9	595.9	0.482	428.6	579.5	0.492	569.9	414.9
	1.2	2	5	5	0.466	422.1	581.2	0.484	421.0	575.3	0.476	585.4	437.1
	1.5	2	2	2	0.982	322.3	329.5	0.980	324.0	331.9	0.978	339.7	331.1
	1.5	2	2	5	0.974	332.9	343.0	0.972	332.5	343.4	0.978	342.7	334.2
	1.5	2	5	2	0.984	318.4	324.9	0.988	318.8	323.6	0.988	324.7	319.9
1.5	2	5	5	0.970	337.3	348.8	0.968	337.4	349.7	0.978	332.8	324.1	
GS CSSP - Stratification and Error Spending	1	2	2	2	0.060	353.0	698.0	0.038	334.7	705.4	0.044	384.5	705.2
	1	2	2	5	0.064	367.0	697.4	0.030	306.0	707.6	0.080	407.3	695.0
	1	2	5	2	0.070	389.6	696.9	0.070	348.9	694.0	0.076	395.5	695.3
	1	2	5	5	0.060	362.5	698.6	0.066	319.1	693.5	0.060	366.0	698.8
	1.2	2	2	2	0.472	431.2	583.7	0.510	417.2	565.6	0.486	452.6	590.0
	1.2	2	2	5	0.456	404.2	576.0	0.468	420.5	579.8	0.512	472.5	593.3
	1.2	2	5	2	0.482	418.8	574.8	0.596	417.5	539.7	0.582	459.6	568.4
	1.2	2	5	5	0.494	420.6	572.1	0.546	426.3	559.6	0.516	435.3	573.1
	1.5	2	2	2	0.980	316.8	324.9	0.978	306.6	315.7	0.988	336.7	341.3
	1.5	2	2	5	0.986	311.5	317.2	0.968	312.1	325.1	0.986	347.2	352.4
	1.5	2	5	2	0.992	312.1	315.3	0.988	314.0	318.8	0.986	324.6	330.1
1.5	2	5	5	0.976	306.0	315.9	0.992	312.8	316.0	0.986	331.3	336.8	
GS EE Score - Regression and Unifying	1	2	2	2	0.050	192.6	693.6	0.048	210.0	695.5	0.074	316.2	690.1
	1	2	2	5	0.068	270.4	689.4	0.068	296.5	691.2	0.090	280.0	680.4
	1	2	5	2	0.070	237.4	686.2	0.064	247.5	689.8	0.060	339.0	697.1
	1	2	5	5	0.070	292.7	690.1	0.042	245.7	700.1	0.054	293.3	697.0
1.2	2	2	2	0.418	380.7	578.2	0.418	382.7	579.0	0.470	423.6	580.7	

	1.2	2	2	5	0.434	366.5	566.6	0.434	386.4	575.2	0.466	424.9	582.5
	1.2	2	5	2	0.430	354.6	562.9	0.458	400.2	573.5	0.492	414.9	569.9
	1.2	2	5	5	0.436	359.7	562.9	0.466	398.1	570.0	0.476	437.1	585.4
	1.5	2	2	2	0.984	295.0	301.8	0.984	295.1	301.9	0.978	331.1	339.7
	1.5	2	2	5	0.974	287.4	298.7	0.982	293.8	301.5	0.978	334.2	342.7
	1.5	2	5	2	0.980	293.8	302.3	0.966	300.4	314.6	0.988	319.9	324.7
	1.5	2	5	5	0.960	299.2	316.0	0.974	298.4	309.4	0.978	324.1	332.8

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 10,000 sample size at end of two years.

Case 2d: $Z1 \sim \text{Bernoulli}(0.5)$; $Z2 \sim \text{Unif}(-1.5, 1.5)$; $Z3 \sim \text{Multi}(0.32, 0.37, .31)$; $D | Z1, Z2, Z3 \sim \text{Bernoulli Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z1}Z1 + \beta_{D,Z2}Z2 + \beta_{D,Z3}Z3))$; $C \sim \text{Weibull}(2, 200)$; $T | D, Z1, Z2, Z3 \sim$

$\text{exponential}(\exp(\beta_{Y,0} + \beta_{Y,D}D + \beta_{Y,Z1}Z1 + \beta_{Y,Z2}Z2 + \beta_{Y,Z3}Z3) / 720)$; $Y = I(T < C, T < (tI - S))$; $D = \min(tI - S, C, T)$

$RR(Y | D) = \exp(\beta_{Y,D})$, $RR(Y | Z1) = \exp(\beta_{Y,Z1})$, $RR(Y | Z2) = \exp(\beta_{Y,Z2})$ and $RR(Y | Z3) = \exp(\beta_{Y,Z3})$.

Table 4d.2. Power, days to rejection, and days to study end across varying frequency of testing and strength of confounding for the case of one binary, one continuous confounder, and one categorical confounder with overall prevalence of 0.10 and a chronic exposure.

Freq. of Testing	Confounders				POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End	POWER	Days to Rejection	Days to Study End
	Effect Size	Y and Z ₁	Y and Z ₂	Y and Z ₃									
	RR(Y D)	RR(Y Z ₁)	RR(Y Z ₂)	RR(Y Z ₃)									
					BIMONTHLY			MONTHLY			QUARTERLY		
GS Wald - Regression and Normal Approx	1	2	2	2	0.050	382.8	703.1	0.046	434.3	706.9	0.040	436.5	708.7
	1	2	2	5	0.022	403.6	713.0	0.024	415.0	712.7	0.040	373.5	706.1
	1	2	5	2	0.032	331.9	707.6	0.036	411.7	708.9	0.044	421.4	706.9
	1	2	5	5	0.038	418.4	708.5	0.020	375.0	713.1	0.040	454.5	709.4
	1.5	2	2	2	0.624	444.4	548.0	0.634	443.0	544.4	0.606	464.3	565.0
	1.5	2	2	5	0.604	448.5	556.0	0.638	435.0	538.2	0.580	470.2	575.1
	1.5	2	5	2	0.578	449.1	563.4	0.608	462.4	563.4	0.600	466.2	567.7
	1.5	2	5	5	0.578	426.1	550.1	0.560	437.1	561.6	0.622	488.8	576.2
	2	2	2	2	0.968	362.3	373.7	0.960	367.0	381.1	0.960	386.3	399.6
	2	2	2	5	0.962	373.2	386.3	0.948	367.3	385.6	0.948	390.4	407.5
2	2	5	2	0.938	372.9	394.4	0.950	355.8	374.0	0.960	384.8	398.2	
2	2	5	5	0.936	365.7	388.4	0.958	376.3	390.7	0.946	393.9	411.5	
GS CSSP - Stratification and Error Spending	1	2	2	2	0.082	330.7	688.1	0.068	306.2	691.9	0.060	378.0	699.5
	1	2	2	5	0.076	350.9	692.0	0.082	262.0	682.4	0.044	380.5	705.1
	1	2	5	2	0.106	275.1	672.8	0.086	313.3	685.0	0.072	355.0	693.7
	1	2	5	5	0.062	303.4	694.2	0.076	263.7	685.3	0.084	428.6	695.5
	1.5	2	2	2	0.612	386.6	516.0	0.612	408.5	529.4	0.676	433.5	526.3
	1.5	2	2	5	0.574	387.4	529.1	0.632	400.0	517.7	0.644	429.6	533.0
	1.5	2	5	2	0.598	402.4	530.1	0.654	388.5	503.2	0.656	430.8	530.3
	1.5	2	5	5	0.572	409.6	542.4	0.570	417.6	547.6	0.642	443.0	542.2
	2	2	2	2	0.938	338.0	361.7	0.962	322.7	337.8	0.956	356.4	372.4
	2	2	2	5	0.940	334.7	357.8	0.952	343.3	361.4	0.956	363.4	379.1
2	2	5	2	0.922	320.7	351.9	0.956	342.1	358.7	0.966	358.9	371.2	
2	2	5	5	0.934	353.3	377.5	0.928	331.3	359.3	0.962	355.5	369.4	
GS EE Score - Regression and Unifying	1	2	2	2	0.060	288.0	694.1	0.060	316.0	695.8	0.052	349.6	700.7
	1	2	2	5	0.062	284.0	693.0	0.062	287.4	693.2	0.070	388.3	696.8
	1	2	5	2	0.062	271.0	692.2	0.056	265.7	694.6	0.044	306.8	701.8
	1	2	5	5	0.064	285.0	692.2	0.048	220.0	696.0	0.048	397.5	704.5
	1.5	2	2	2	0.636	382.7	505.5	0.634	392.3	512.2	0.642	429.3	533.3

	1.5	2	2	5	0.596	403.6	531.5	0.582	415.7	542.9	0.622	427.1	537.8
	1.5	2	5	2	0.586	385.2	523.8	0.606	388.2	518.9	0.586	425.4	547.4
	1.5	2	5	5	0.572	403.7	539.1	0.536	407.6	552.5	0.570	451.9	567.2
	2	2	2	2	0.966	330.1	343.4	0.946	327.6	348.8	0.978	345.8	354.1
	2	2	2	5	0.946	339.6	360.2	0.954	317.3	335.8	0.968	354.4	366.1
	2	2	5	2	0.932	318.6	345.9	0.958	326.4	343.0	0.966	355.7	368.1
	2	2	5	5	0.918	319.5	352.4	0.936	344.9	368.9	0.954	369.4	385.6

* Bold indicates outside +/- 1.5% of the expected type I error of 0.05

Two-Year Study with analyses bimonthly, monthly and quarterly, first analysis after 90 days, and 1,500 sample size at end of two years.

Case 2d: $Z1 \sim \text{Bernoulli}(0.5)$; $Z2 \sim \text{Unif}(-1.5, 1.5)$; $Z3 \sim \text{Multi}(0.32, 0.37, .31)$; $D | Z1, Z2, Z3 \sim \text{Bernoulli Bernoulli}(\exp(\beta_{D,0} + \beta_{D,Z1}Z1 + \beta_{D,Z2}Z2 + \beta_{D,Z3}Z3))$; $C \sim \text{Weibull}(2, 200)$; $T | D, Z1, Z2, Z3 \sim$

$\text{exponential}(\exp(\beta_{Y,0} + \beta_{Y,D}D + \beta_{Y,Z1}Z1 + \beta_{Y,Z2}Z2 + \beta_{Y,Z3}Z3)/720)$; $Y = I(T < C, T < (tl - S))$; $D = \min(tl - S, C, T)$

$RR(Y | D) = \exp(\beta_{Y,D})$, $RR(Y | Z1) = \exp(\beta_{Y,Z1})$, $RR(Y | Z2) = \exp(\beta_{Y,Z2})$ and $RR(Y | Z3) = \exp(\beta_{Y,Z3})$.

Figure 1a. Power curve by number of days on the study for the situation of single-time exposure and simple one binary confounder (case 1a) for *GS LD*, *CSSP*, and *GS EE* methods varying the frequency of testing; a) bimonthly, b) monthly, and c) quarterly. Assumes the effect size is, $RR(Y|D)=1.5$, confounding relationship is $RR(Y|Z_1)=2.0$ and overall prevalence is 0.05.

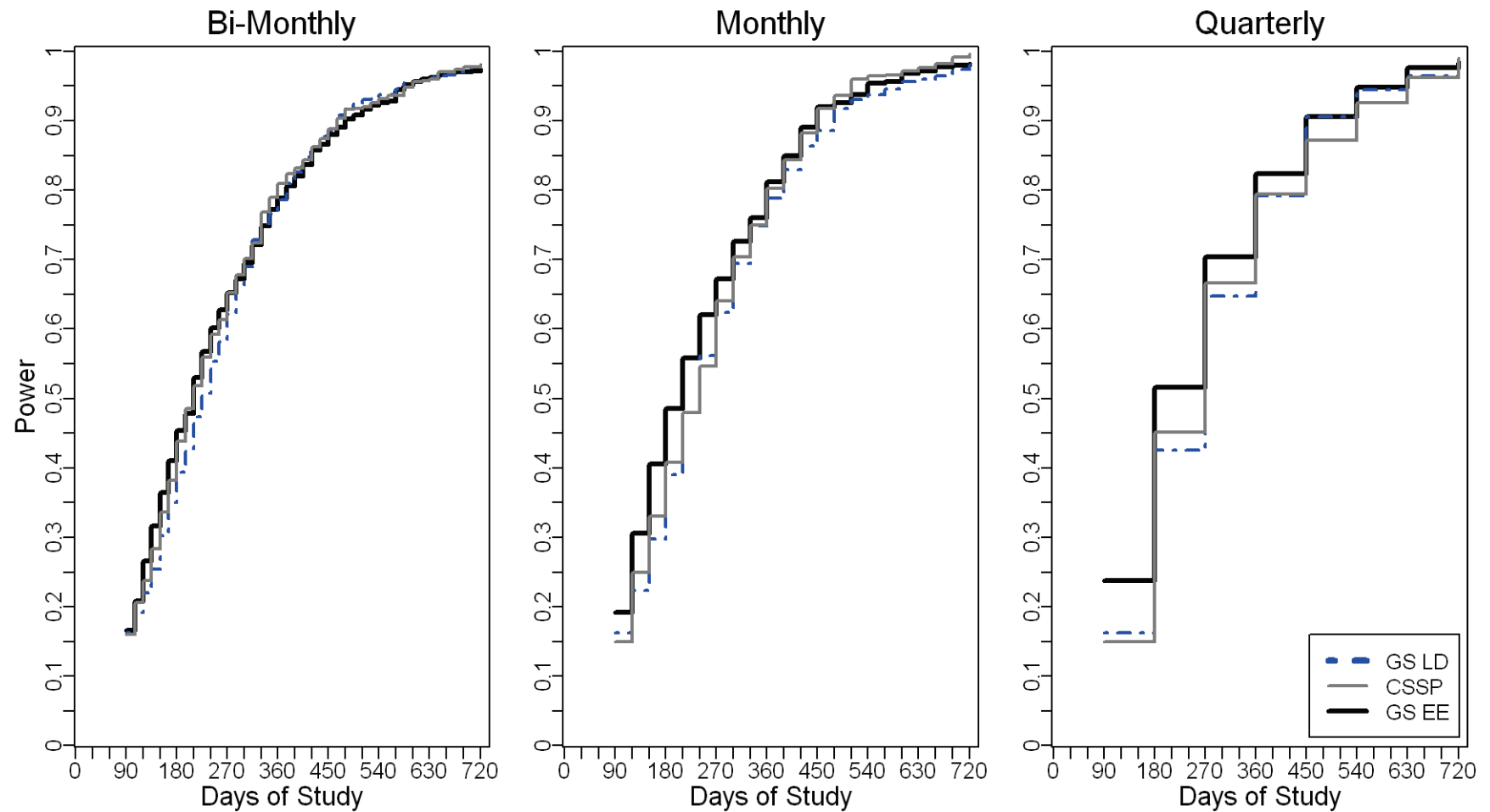


Figure 1b. Power curve by number of days on the study for the situation of single-time exposure and simple one continuous confounder (case 1b) for *GS LD*, *CSSP*, and *GS EE* methods varying the frequency of testing; a) bimonthly, b) monthly, and c) quarterly. Assumes the effect size is, $RR(Y|D)=2.0$, confounding relationship is $RR(Y|Z_2)=2.0$ and overall prevalence is 0.10.

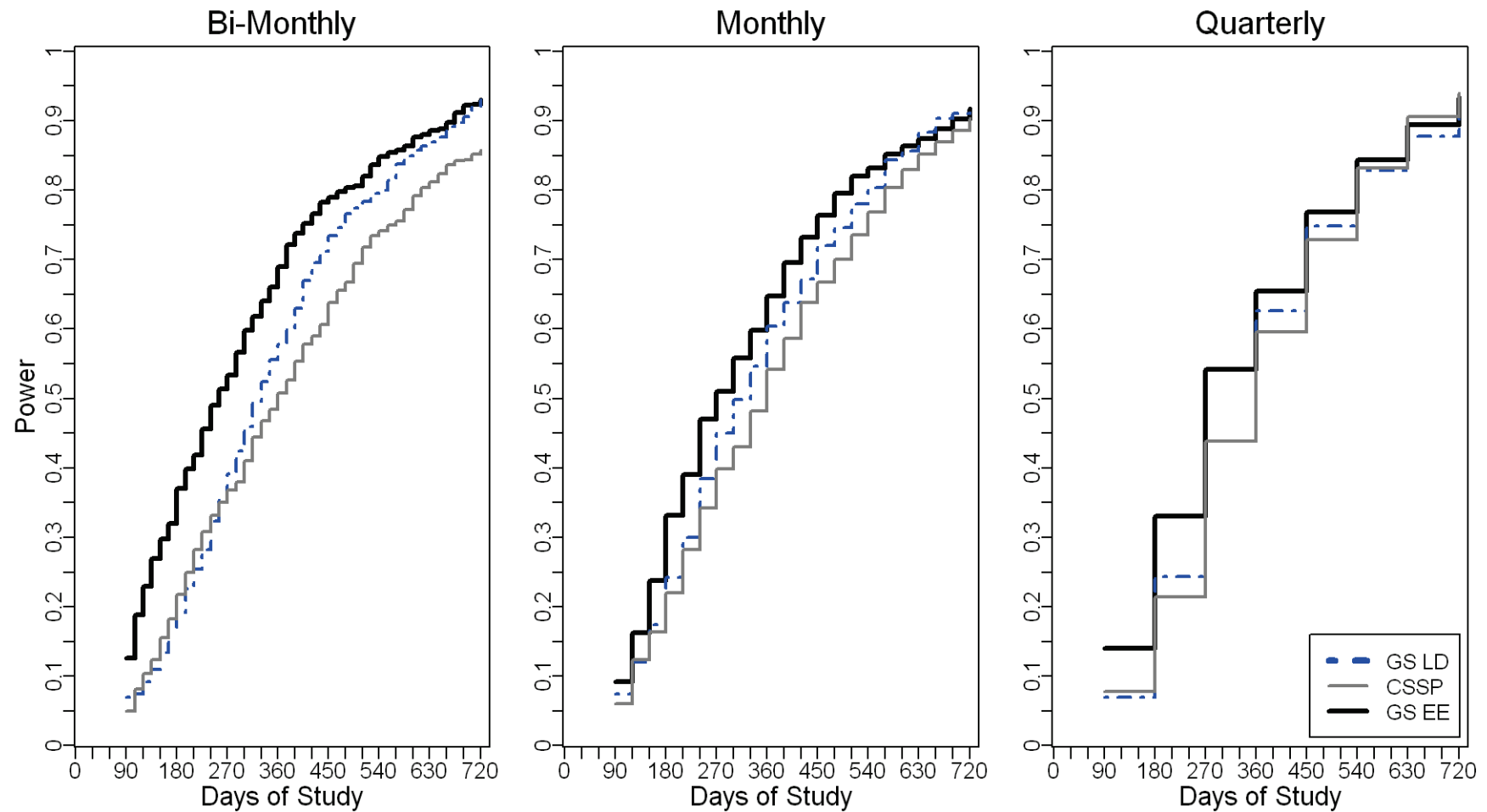


Figure 2a. Power curve by number of days on the study for the situation of chronic exposure time and simple one binary confounder (case 2a) for *GS LD*, *CSSP*, and *GS EE* methods varying the frequency of testing; a) bimonthly, b) monthly, and c) quarterly. Assumes the effect size is, $RR(Y|D)=1.5$, confounding relationship is $RR(Y|Z_1)=2.0$ and overall prevalence is 0.05.

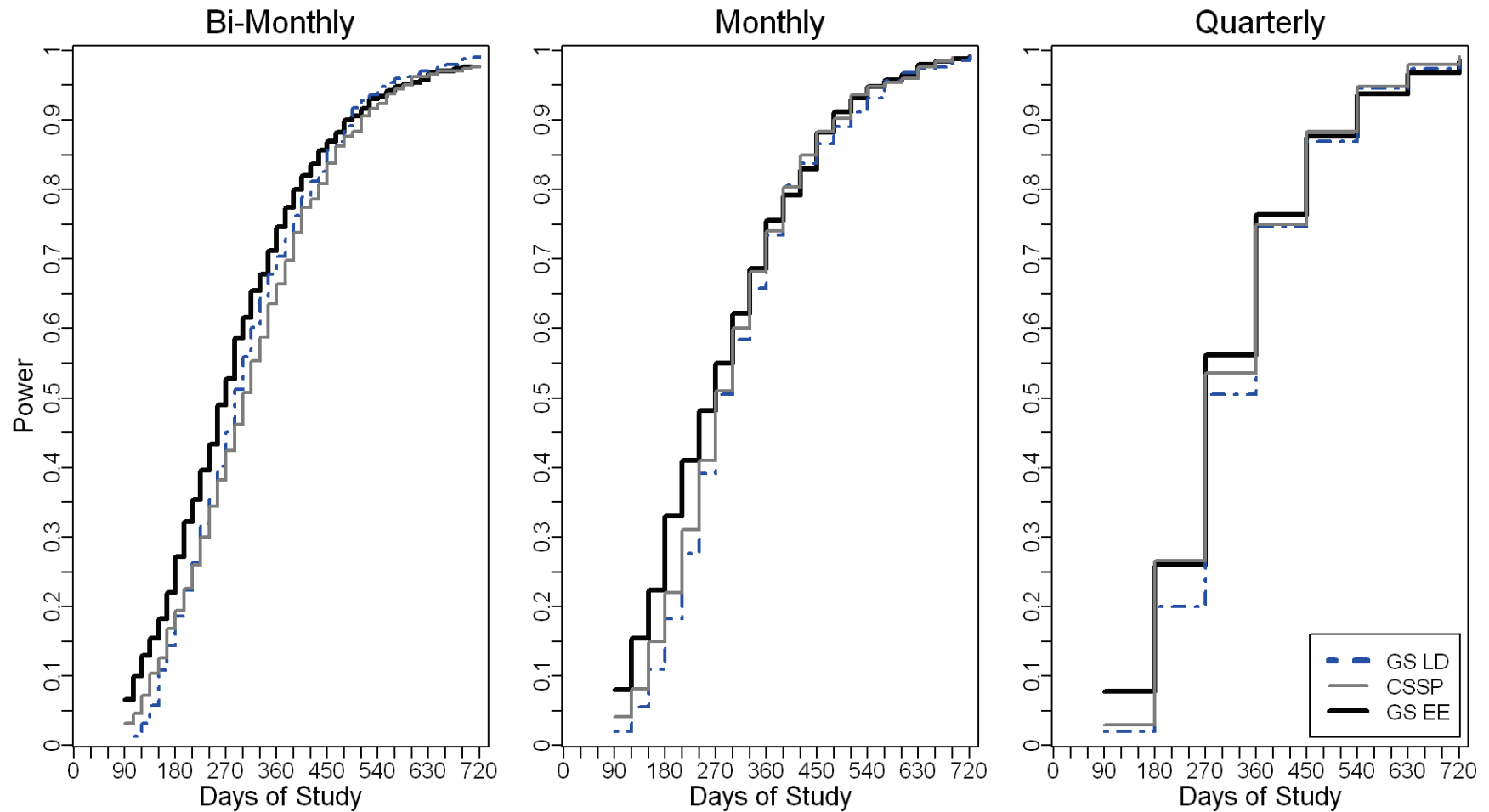


Figure 2b. Power curve by number of days on the study for the situation of chronic exposure time and simple continuous confounder (case 2b) for *GS LD*, *CSSP*, and *GS EE* methods varying the frequency of testing; a) bimonthly, b) monthly, and c) quarterly. Assumes the effect size is, $RR(Y|D)=2.0$, confounding relationship is $RR(Y|Z_2)=2.0$ and overall prevalence is 0.10.

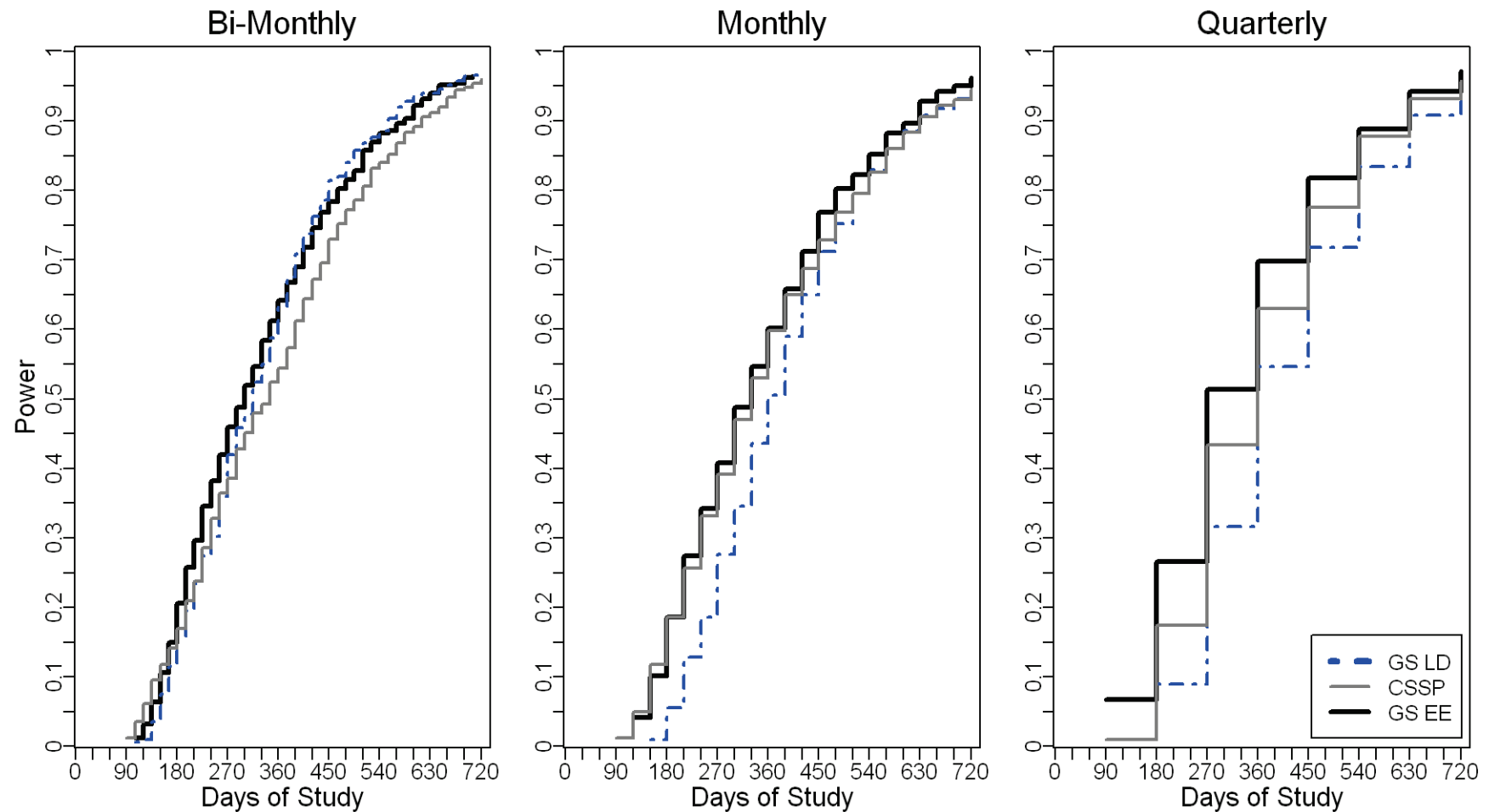


Figure 2c. Power curve by number of days on the study for the situation of chronic exposure time and three confounders (case 2d) for *GS LD*, *CSSP*, and *GS EE* methods varying the frequency of testing; a) bimonthly, b) monthly, and c) quarterly. Assumes the effect size is, $RR(Y|D)=2.0$, confounding relationship is $RR(Y|Z_1)=2.0$, $RR(Y|Z_2)=5.0$, and $RR(Y|Z_3)=5.0$ with overall prevalence is 0.10.

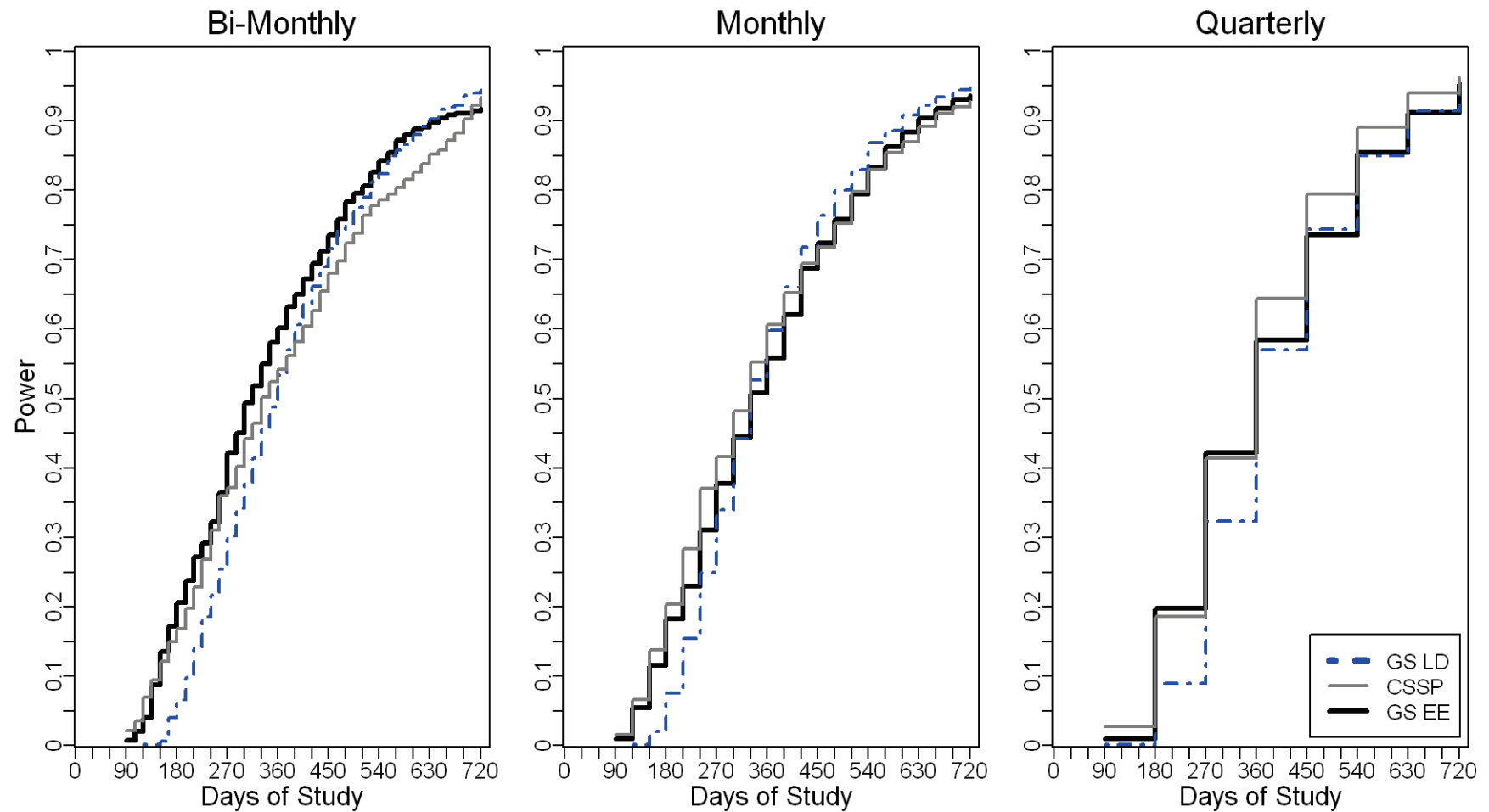
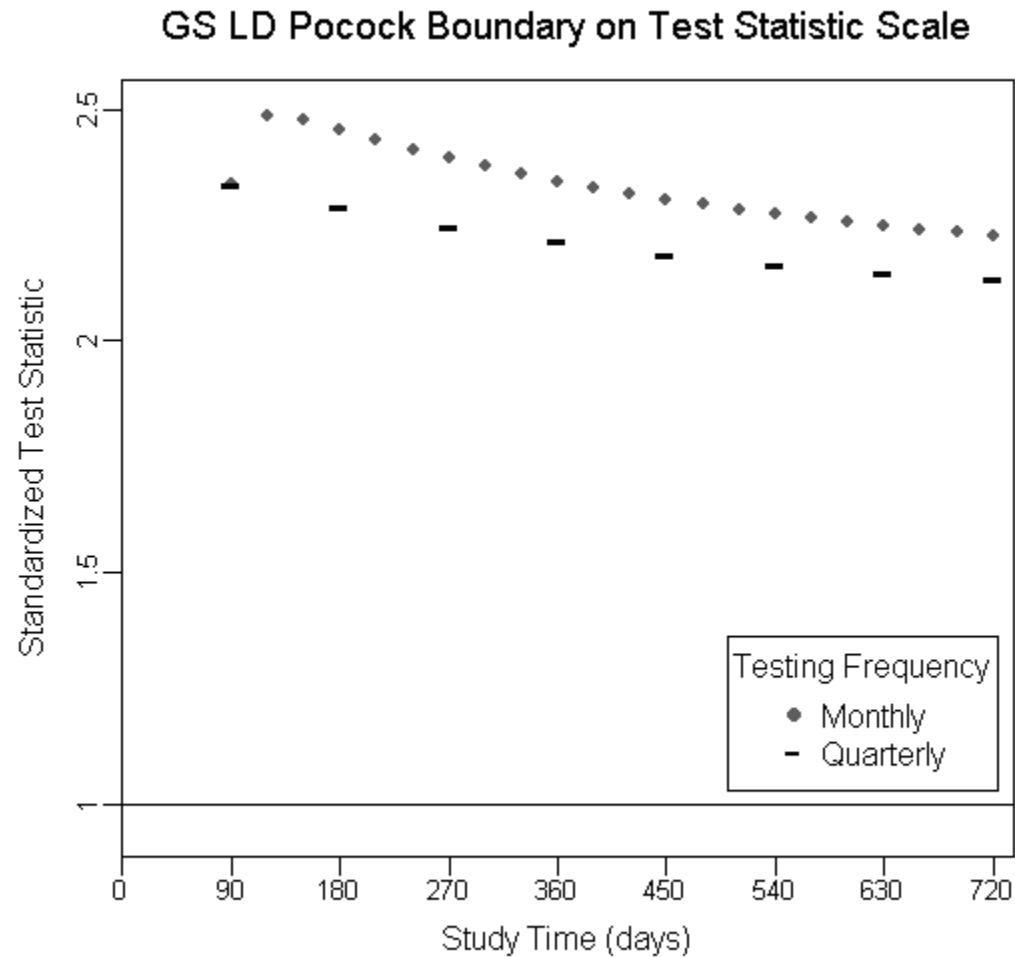


Figure 3. Sequential Monitoring boundaries on the standardized test statistic scale for a flat, Poccock-style, boundary with a sample size of 10,000 participants with the first analysis after the first 90 days and then subsequent analyses either monthly or quarterly.



III. REFERENCES

1. Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System - A National Resource for Evidence Development. *N Engl J Med*. 2011.
2. Adverse Event Reporting System (AERS). FDA CDER, 2011. Available at: <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm>. Accessed 1/27/2011.
3. Belongia EA, Irving SA, Shui IM, et al. Real-time surveillance to assess risk of intussusception and other adverse events after pentavalent, bovine-derived rotavirus vaccine. *Pediatr Infect Dis J*. 2010; 29(1): 1-5
4. Brown JS, Kulldorff M, Chan KA, et al. Early detection of adverse drug events within population-based health networks: application of sequential testing methods. *Pharmacoepidemiol Drug Saf*. 2007; 16: 1275-84.
5. Brown JS, Kulldorff M, Petronis KR, et al. Early adverse drug event signal detection within population-based health networks using sequential methods: key methodologic considerations. *Pharmacoepidemiol Drug Saf*. 2009; 18: 226-34.
6. Klein NP, Fireman B, Yih WK, et al. Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. *Pediatrics*. 2010; 126: e1-8.
7. Lieu TA, Kulldorff M, Davis RL, et al. Real-time vaccine safety surveillance for the early detection of adverse events. *Med Care*. 2007; 45: S89-95.
8. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf*. 2010; 19: 858-68.
9. Emerson SS, Kittelson JM, Gillen DL. Frequentist evaluation of group sequential clinical trial designs. *Stat Med*. 2007; 26: 5047-80.
10. Lan KKG, Demets DL. Discrete Sequential Boundaries for Clinical-Trials. *Biometrika*. 1983; 70: 659-63.
11. Pocock SJ. Interim Analyses for randomized clinical-trials - The Group Sequential Approach. *Biometrics*. 1982; 38: 153-62.
12. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979; 35: 549-56.
13. Nelson JC, Cook AJ, Yu O, et al. Challenges in the design and analysis of sequentially-monitored post-licensure safety surveillance studies using observational health care utilization data. *Pharmacoepidemiol Drug Saf*. 2011. [In submission]

14. Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. A Maximized Sequential Probability Ratio Test for Drug and Vaccine Safety Surveillance. *Sequential Analysis: Design Methods and Applications*. 2011; 30: 58-78.
15. Kittelson JM, Emerson SS. A unifying family of group sequential test designs. *Biometrics*. 1999; 55: 874-82.
16. Burington BE, Emerson SS. Flexible implementations of group sequential stopping rules using constrained boundaries. *Biometrics*. 2003; 59: 770-7.
17. Li LL. A conditional sequential sampling procedure for drug safety surveillance. *Stat Med*. 2009; 28: 3124-38.
18. Rotnitzky A, Jewell NP. Hypothesis-testing of regression parameters in semiparametric generalized linear-models for cluster correlated data. *Biometrika*. 1990; 77: 485-97.
19. Zeger SL, Liang KY, Albert PS. Models for longitudinal data - a generalized estimating equation approach. *Biometrics*. 1988; 44: 1049-60.
20. Lehmann EL, Romano JP. *Testing Statistical Hypotheses (3rd ed)*. New York, NY: Springer Science+Business Media, LLC, 2005.
21. Cook AJ, Wellman RJ, Tiwari RC, Nelson JC. Group Sequential Methods for observational data incorporating confounding through estimating equations with application in Postmarketing Vaccine/Drug Surveillance. 2011. [In preparation]

IV. APPENDIX: STATISTICAL SOFTWARE CODE FOR THE CSSP AND GSEE

Prepared by: Andrea J. Cook, PhD^{1,2}, Robert D. Wellman, MS¹, Tracey L. Marsh, MS³

Author Affiliations: 1. Biostatistics Unit, Group Health Research Institute, Seattle, WA 2. Department of Biostatistics, University of Washington, Seattle, WA 3. Group Health Research Institute, Seattle, WA

Current versions available at: <http://faculty.washington.edu/acook/software.html>

A. SAS® PROGRAMMING CODE

1. CSSP

```

/*
* DRAFT
* Title CSSP Method - Analysis Macro
* Programmer: Tracey Marsh
*
* Description
*     Performs a sequential test on an analytic data set - returns result (signal or not) and summary report.
*     Interim files are stored and required for subsequent looks.
*
* Usage
*     %CSSP_test(look_summary,this_look,info_prop, alpha, bnd_type="Poc", interimLib=work, detail=N);
*
* Arguments
*     look_summary = sas data set - analytic data summary of data new since previous look
*     contains variables as described:
*     look, numeric - positive integers of the form 1-n: defines which look the data is from
*     strata, numeric - positive integers of the form 1-n: defines analysis level subpopulations
*     exp_doi, numeric - positive: total drug of interest exposure since previous look
*     exp_cmp, numeric - positive, >= exp_doi: total exposure (includes drug of interest) since previous look
*     evt_doi, numeric - non-negative: number of adverse events with drug of interest exposure
*     evt_cmp, numeric - non-negative, >= evt_doi: number of adverse events (w or w/o exposure to drug of
*     (interest)
*     this_look = an integer describing the look number - used for labelling and identifying segment of "new"
*     data within look_summary table
*     info_prop = proportion of information cumulatively inclusive of data new to this look.
*         Should be in (0,1], increasing with each look, equal to 1 at final look.
*         Used in determination of boundary for given look.
*     alpha = total amount of type I error cumulative over all looks (i.e. when info_prop = 1).
*     bnd_type = type of error spending function.
*         Supported types: "Poc", "OBF", "Uni" ("Uni" assumed as default).
*     interimLib = SAS library for storing interim data sets and accessing interim data sets from previous looks.
*     detail = Y to produce a more detailed summary report (includes strata level information).
*
* Details
*     The distribution of adverse events for the drug of interest exposure group is simulated under the null
*     hypothesis (no risk difference between exposure to drug of interest and exposure to the comparator) as a
*     binomial event of trials equal to the total number of adverse events observed, with probability equal to

```

```

* the proportional exposure to the drug of interest. The simulation is performed for each strata
* independently, and the total events aggregated. Events are also simulated at each look independently,
* and a cumulative total maintained. (Hence the need to store and access interim tables.) A cumulative p-
* value is inferred from the rank of the number of adverse events observed with exposure to the drug of
* interest within the simulated total events, conditional on having not previously signalled. If the p-value for
* observed events is less than the cumulative error spending function, a signal is indicated.

* Output
* Summary
*
* References
* Li, L. A conditional sequential sampling procedure for drug safety surveillance. Statistics in Medicine 2009;
* 28:3124-3138.
*
* Example
* %CSSP_test(look_summary=sasout.example, this_look=4, info_prop=4/8, alpha=0.05, bnd_type="Poc",
* interimLib=sasout, detail=Y);
*
* see accompanying files "CSSP Method - Example.sas" and "example_data.csv"
*/

```

```

%macro CSSP_test(look_summary=, this_look=, info_prop=, alpha=0.05, bnd_type="Poc", interimLib=work,
detail=N);

```

```

%let nboot = 10000;

```

```

***DEFINE BOUNDARY***;

```

```

data _null_;
    select (UPCASE(&bnd_type.));
        when ("OBF") alpha_spend = 2*( 1- cdf('NORMAL', Quantile('NORMAL',1-
&alpha./2)/(&info_prop.**(0.5))));
        when ("POC") alpha_spend = &alpha.*log(1 + (exp(1)-1)*(&info_prop.));
        when ("UNI") alpha_spend = &alpha.*(&info_prop.);
        otherwise alpha_spend = &alpha.*(&info_prop.);
    end;
    boundary_rank = &nboot.*(1-alpha_spend);
    call symput("cutrank", boundary_rank);
run;

```

```

%put look = &this_look. information proportion = &info_prop. cutrank = &cutrank.;

```

```

***CREATE WIDE DATA SET WITH STRATA STATS***;

```

```

*determine number of strata in dataset;
proc sort data = &look_summary. (keep=strata look where = (look=&this_look.))
    out = strata_unq nodupkey;
    by strata;
run;

```

```

%let dsnid = %sysfunc(open(strata_unq));

```

```

%let nstrata = %sysfunc(attrn(&dsnid.,NOBS));
%let rc = %sysfunc(close(&dsnid.));
%put nstrata &nstrata.;

data strata_stats (keep = look strata e_sum p e_d);
    set &look_summary. (where = (look=&this_look.));
    length esum p1 8.;

    e_sum = evt_doi + evt_cmp;
    p = exp_doi/(exp_doi + exp_cmp);
    e_d = evt_doi;
run;

proc transpose data=strata_stats out=strata_stats_t;
    by look;
run;

data tmp1; set strata_stats_t; if _NAME_='e_sum'; rename col1-col&nstrata. = e_sum1-e_sum&nstrata.; run;
data tmp2; set strata_stats_t; if _NAME_='p'; rename col1-col&nstrata. = p1-p&nstrata.; run;
data tmp3; set strata_stats_t; if _NAME_='e_d'; rename col1-col&nstrata. = e_d1-e_d&nstrata.; run;

data strata_stats (drop = _NAME_);
    merge tmp1 tmp2 tmp3;
    by look;
run;

***SIMULATE DISTRIBUTION OF NEW EVENTS UNDER NULL***;
data dist (keep = look iter events_new) ;
    set strata_stats;
    look = &this_look.;
    array a_esum {&nstrata.} e_sum1-e_sum&nstrata.;
    array a_p {&nstrata.} p1-p&nstrata.;
    array a_ed {&nstrata.} e_d1-e_d&nstrata.;

    *cumulative counts - summed over strata;
    events_new = 0;

    **iter1 holds data counts;
    iter = 1;
    do strata = 1 to &nstrata.;
        if a_esum{strata} >. & a_p{strata} >. & a_ed{strata} >. then
            events_new = events_new + a_ed{strata};
    end;
    output;
    **iter2 on holds simulation of expected counts**;
    do iter = 2 to &nboot.;
        events_new = 0;
        do strata = 1 to &nstrata.;
            if a_esum{strata} >. & a_p{strata} >. & a_ed{strata} >. then do;
                events = a_esum{strata};
                prob = a_p{strata};
            end;
        end;
    end;

```

```

        if prob = 1 then events_new = events_new + events;
        else if prob > 0 & events >0 then events_new = events_new + ranbin(0, events,
prob);
                                end;
                                end;
                                end;
                                output;
                                end;
run;

***CUMULATIVE SIMULATED DISTRIBUTION***;
%if &this_look. = 1 %then %do;
    data dist_lim; set dist; rename events_new = evt_doi_cum; run;
%end;
%else %do;
    data dist_lim (keep = iter look evt_doi_cum);
        merge &interimLib..distsim (in = a) dist;
        by iter;
        if a;
            evt_doi_cum = evt_doi_cum + events_new; *evt_doi_cum is the cumulative number of events
up until time t;
        run;
%end;

***APPLY STOPPING ALGORITHM***;
proc rank data = dist_lim out = dist_w_rank ties = low;
    var evt_doi_cum;
    ranks eventrank;
run;

data &interimLib..distsim (keep = iter look evt_doi_cum) rej(keep = iter look signal) condp (keep = iter signal look
condp eventrank cutrank evt_doi_cum);
    set dist_w_rank;
    retain signal 0;
    cutrank = &cutrank;

    attrib _all_ label = ";

    *first iteration within a sim is the observed data counts;
    if iter=1 then do;
        if eventrank > cutrank then do;
            signal = 1;
            put "SIGNAL - SIGNAL - SIGNAL";
            output rej;
        end;
        else signal = 0;
        condp = 1-(eventrank-1)/&nboot.;
            *the estimated probability of observing "more or equal extreme" outcomes;
            *this is conditional because only events less than boundary stay in set - see next if
clause;
        output condp;
    end;
    *once it signals, nothing will be output - this prevents continued running;

```



```

        if signal = 0 & eventrank <= cutrank then do;
            output &interimLib..distsim;
        end;
run;

***ALPHA ACTUALLY SPENT***;
%let dsnid = %sysfunc(open(&interimLib..distsim));
%let dist_obs = %sysfunc(attrn(&dsnid,NOBS));
%let rc = %sysfunc(close(&dsnid.));

data condp;
    set condp;
    if (signal) then escp = .;
    else escp = 1-&dist_obs./&nboot.;
run;

***STORE INTERIM TABLES***;
%if &this_look.=1 %then %do;
    data &interimLib..condp;
        retain look evt_doi_cum eventrank cutrank condp escp signal;
        set condp (keep = look evt_doi_cum eventrank cutrank condp escp signal); run;
    data &interimLib..analysis_stats; set strata_stats; run;
    data &interimLib..analysis_raw; set &look_summary. (where = (look=&this_look.)); run;
%end;
%else %do;
    proc append
        base = &interimLib..condp
        data = condp (keep = look evt_doi_cum eventrank cutrank condp escp signal);
    quit;
    proc append
        base = &interimLib..analysis_stats
        data = strata_stats;
    quit;
    proc append
        base = &interimLib..analysis_raw
        data = &look_summary. (where = (look=&this_look.));
    quit;
%end;
proc sort data = &interimLib..condp
    out = &interimLib..condp;
    by descending look;
run;

***PRINT SUMMARY REPORT***;
ods noresults;
ods listing close;
data _null_;
    set sashelp.vslib;
    if libname="%upcase(&interimLib.)" then do;
        call symput("outPath",trim(path));
    end;
run;

```

```
ods rtf file = "&outPath.\CSSP Sequential Testing - Look &this_look..rtf";
footnote "CSSP Method - Sequential Testing.sas (&sysdate, &stime)";

data _NULL_;
    Length boundary $3.;
    boundary = &bnd_type.;
    call symput("bound",boundary);
run;

title "Method: CSSP - Look: &this_look. - Information Proportion: &info_prop - Boundary: &bound. - Alpha:
&alpha.";

*high level testing overview;
data condp_sum (keep = look evt_doi_cum p_value bndry_value signal);
    set &interimLib..condp (rename = (condp=p_value));
    bndry_value = 1-cutrank/&nboot.;
run;
data condp_sum;
    retain look evt_doi_cum p_value bndry_value signal;
    set condp_sum;
run;
proc print data=condp_sum;
    title2 "Testing outcomes for all looks through current look (&this_look.)";
run;

*data counts documentation - used in analysis;
proc summary data= &interimLib..analysis_raw nway missing;
    class look stdt eddt;
    var subj: exp: evt;;
    format subj: exp: evt: comma12.0;
    output out=sum1( drop=_type__freq_) sum=;
run;
proc sort data = sum1
    out = sum1d;
    by DESCENDING look;
run;
proc print data=sum1d;
    title2 "Analysis for all looks through current look (&this_look.) based on aggregate counts:";
run;

*data counts documentation - provided this look;
proc summary data= &look_summary. nway missing;
    class look stdt eddt;
    var subj: exp: evt;;
    format subj: exp: evt: comma12.0;
    output out=sum2( drop=_type__freq_) sum=;
run;
proc sort data = sum2
    out = sum2d;
    by DESCENDING look;
run;
proc print data=sum2d;
```

```

        title2 "Aggregate counts provided at current look (&this_look.);";
run;

data diff (drop=subj_c subj_d evt_c evt_d exp_c exp_d);
    merge sum1 (IN=ana) sum2 (IN=new rename = (subj_cmp=subj_c subj_doi=subj_d exp_cmp=exp_c
exp_doi=exp_d evt_cmp=evt_c evt_doi=evt_d));
    by look;
    subj_cmp=-1*(subj_cmp-subj_c);
    subj_doi=-1*(subj_doi-subj_d);
    exp_cmp=-1*(exp_cmp-exp_c);
    exp_doi=-1*(exp_doi-exp_d);
    evt_cmp=-1*(evt_cmp-evt_c);
    evt_doi=-1*(evt_doi-evt_d);
run;

proc print data=diff;
    title2 "Discrepancies between analysis base and aggregate counts provided at current look
(&this_look.);";
run;
*optional summaries by strata;
%IF &detail.=Y %THEN %DO;
proc sort data = &interimLib..analysis_stats
    out = &interimLib..analysis_stats;
    by descending look;
run;
proc print data=&interimLib..analysis_stats;
    title2 "Analysis for current look (&this_look.) based on following data statistics:";
    title3 "";
run;

proc sort data = &interimLib..analysis_raw
    out = &interimLib..analysis_raw;
    by descending look strata;
run;
proc print data=&interimLib..analysis_raw;
    title2 "Analysis for all looks through current look (&this_look.) based on following counts:";
    title3 "";
run;
%END;
ods rtf close;
ods listing;
ods results;
%mend;

```

2. GS EE

```

/*
* Title GSEE Method - Analysis Macro
* Version
* Date
* Author
* Maintainer
* Depends
* License
*
* Description
*       Performs a sequential test on an analytic data set - returns result (signal or not) and summary
*       report.
*
* Usage
*       %GSEE_test(look_raw=dataset, conf_cat=Z1, conf_cont=Z2 Z3, this_look=3, look_plan=myplan,
*       alpha=0.05, delta=0.5, reuse_bndry_vals=Y, family="Poisson", interimLib=sasout, detail=N)
* Arguments
*       look_raw = sas data set - analytic dataset, individual level
*       contains variables as described:
*
*               X - indicator of exposure to drug/product of interest
*               Z1....Zn - covariate combinations
*               S - integer corresponding to first day of exposure time (inclusive)
*                   on or after study start date (which corresponds to S=1)
*               obs_t - total days of exposure time (cumulative through this_look)
*               Y - outcome of interest flag, up to current look time (expect event
occurs on last day of exposure, i.e. further exposure censored at
event)
*       conf_cat - list of variables comprising the categorical confounders (e.g. Z1 Z2 Z3)
*                   standard sas variable names with no commas, quotes or other punctuation in
list, leave blank if none
*
*                   NOTE: in this version categorical confounders must be binomial, i.e.
pass only dummy variables to macro as categorical confounders
*       conf_cont - list of variables comprising the continuous confounders (e.g. Z4 Z5)
*                   standard sas variable names with no commas, quotes or other punctuation in
list, leave blank if none
*
*       this_look = an integer describing the look number - used for boundary calculation, data staging,
and labelling reports
*
*       alpha = total amount of type I error cumulative over all looks (i.e. when info_prop = 1).
*       delta = shape parameter for boundary from unifying family of boundaries
*               (delta = 0.5 is Pocock) and (delta = 0 is Fleming)
*       look_plan = data set containing details of planned looks
*               contains variables as described:
*               look - integer variable describing look number {1,...total_looks}
*               look_sample - number of observed subjects with non-zero exposure
*                   prior to given look day
*                   - look_sample on final look should be total study sample size
*               look_day - most recent study day within data for corresponding look
*               bndry - boundary value for corresponding looks, used in prior analyses

```

```

*           this variable will be ignored if reuse_bndry_vals set to "N"
* reuse_bndry_vals = Y or N - indicates whether use boundary values from prior looks or to
* recalculat anew (based on current data sample)
*           if Y then bndry variable of prior looks table must be populated
*
* family = regression model family
*           Supported families: Binomial (logit link), Poisson (log link)
* interimLib = SAS library for storing interim data sets and output report
*           (NOTE: report cannot be saved to work library)
* detail = Y to produce a look_plan table with boundary used (for reuse in future looks)
*           sas table written to interimLib
*
* Details
*
* Output
*           Summary report with signal, test statistic, exposure time and event counts, and boundary details.
*
* References
*           Cook, A. [need reference]
*
* Example
*           %GSEE_test(look_raw=sasout.example, conf_cat = Z3 Z4, conf_cont = Z1 Z2, this_look=4,
* look_plan=my_plan, alpha=0.05, delta=0.5, reuse_bndry_vals=N, family="Poisson", interimLib=sasout,
* detail=N);
* see accompanying files "GSEE Method - Example.sas" and "example_data.csv"
*/

%macro GSEE_test(look_raw=, conf_cat=, conf_cont=, this_look=, look_plan=, alpha=, delta=, reuse_bndry_vals=,
family=, interimLib=, detail=);

    %let dist_nsim = 1000;

    *creates look_plan_tmp passed to other macros - to cleanup;
    %load_plan(&look_plan.);

    %stage_data(&look_raw.,look_sum, counts, &this_look., look_plan_tmp, &last_look., &family.);

    *** simulate distribution under the null without copious output ***;
    %put OUTPUT:OFF;
    ods listing close;
    ods select none;
    ods noresults;
    options noNotes noMprint;

    %sim_dist(look_sum, &conf_cat., &conf_cont., &dist_nsim., stats_dist, &last_look., &family.);
    %gen_bdry(stats_dist, &dist_nsim., bd, &delta., &alpha., look_plan_tmp, &reuse_bndry_vals.,
&this_look., &last_look.);

    ods listing;
    ods select all;
    ods results;
    options Notes Mprint source source2;

```

```

%put OUTPUT:ON;

*tests data at current look and re-tests prior looks;
%test_all_looks(look_sum, &conf_cat., &conf_cont., signal, bd, &this_look., &alpha., &dist_nsim.,
&family.);
%generate_report(look_sum, signal, counts, &interimLib., &this_look., &last_look., look_plan_tmp, bd,
&alpha., &sample_size., &delta., &dist_nsim., &detail.);
%mend;
/*
* Support Macros Called by GSEE_Test
*/
%macro load_plan(look_plan);

*get overview info from look plan;
proc summary data= &look_plan. nway missing;
var look look_sample look_day;
output out=sum1( drop=_type_ _freq_) max=;
run;
%global last_look sample_size last_day;
data _null_;
set sum1;
call symput("last_look", look);
call symput("sample_size", look_sample);
call symput("last_day", look_day);

run;
data look_plan_tmp ;
set &look_plan. (keep = look_sample look look_day bndry);
length planned_info 8.;

planned_info = look_sample/&sample_size.;

run;

*information proportions;
data fmt_tmp;
set look_plan_tmp (rename = (look=start planned_info=label)) end=eof;

end = start;
fmtname = 'look2info';
type='N';

output;
if eof then do;
hlo = "0";
label = "0";
output;
end;

run;
proc format cntlin=fmt_tmp;
run;

*prior boundary levels;
data fmt_tmp;

```

```

        set look_plan_tmp (rename = (look=start bndry=label)) end=eof;

        end = start;
        fmtname = 'look2bdry';
        type='N';

        output;
        if eof then do;
            hlo = "O";
            label = "0";
            output;
        end;
run;
proc format cntlin=fmt_tmp;
run;

*prior look times;
data fmt_tmp;
    set look_plan_tmp (rename = (look=start look_day=label)) end=eof;

    end = start;
    fmtname = 'look2day';
    type='N';

    output;
    if eof then do;
        hlo = "O";
        label = "0";
        output;
    end;
run;
proc format cntlin=fmt_tmp;
run;

%mend;
%macro stage_data(inData, outData, data_counts, curr_look, look_plan, tot_looks, family);

proc sort data = &inData.
    out = data_tmp;
    by s;
run;

%IF (&curr_look. < &tot_looks.) %THEN %DO;
    *fill out to planned sample size via sampling;
    data _null_;
        set &look_plan. (where = (look = &curr_look.));
        call symput ("true_smpl",look_sample);
    run;
    %local i j needed_smpl;

```

```

%DO i = &curr_look. + 1 %TO &tot_looks.;

    data more;
        set &look_plan. (where = (&i. - 1 <= look <= &i.));
        retain last_sample;

        if _N_=1 then last_sample = 0;
        needed = look_sample - last_sample;

        if (&i. = look) then call symput ("needed_smpl",needed);
        last_sample = look_sample;
    run;

    %IF (&needed_smpl. > 0) %THEN %DO;

        ods noresults;
        ods listing close;
        proc surveysselect data = data_tmp (obs = &>true_smpl.) method=urs sampsize
= &needed_smpl. rep=1 out=new;
            run;
            ods listing;
            ods results;

            data data_tmp_new;
                set new;

                do j = 1 to NumberHits;
                    S = ceil((put(&i.,look2day.)-put(&i.
1,look2day.))*ranuni(0))+put(&i.-1,look2day.);
                    output;
                end;
            run;

            proc append
                base = data_tmp
                data = data_tmp_new (drop = replicate numberhits j);
            quit;
        %END;

    %END;

%END;

data data_tmp (drop = i look_day);
    set data_tmp;

    length e1-e&tot_looks. y1-y&tot_looks. lnk_e1-lnk_e&tot_looks. 8.;
    array exps(&tot_looks.) e1-e&tot_looks.;
    array hois(&tot_looks.) y1-y&tot_looks.;
    array lnk_exps(&tot_looks.) lnk_e1-lnk_e&tot_looks.;

    DO i = 1 TO &tot_looks.;

```



```

*summarize data for all looks up through current;
look_day = input(compress(put(i,look2day.)),best.);
exps(i) = max(0,min(look_day, s + obs_t -1) - s + 1);
hois(i) = y*(s + obs_t - 1 <= look_day);

*link function on exposure;
select(upcase(&family.));
  when('BINOMIAL') lnk_exps(i)=exps(i);
  when('POISSON') do;
    if (exps(i)>0) then lnk_exps(i) = log(exps(i));
    else lnk_exps(i) = .;
  end;
  otherwise do;
    put "ERROR: unknown link type, assuming identity";
    lnk_exps(i)=exps(i);
  end;
end;
END;
run;

data &outData.;
  set data_tmp;

  *need individual fake classes for genmod;
  id = _N_;

  *need a fake class variable for multtest and score statistic calculations;
  dummy=1;

  *use sample number so can reuse scoreStat distribution macro to generate just one statistic;
  _sample_ = 1;
run;

*for report;
proc summary data=&outData. nway missing;
  class X;
  var e1-e&curr_look. y1-y&curr_look.;
  format _numeric_ comma12.0;
  output out=sum1 (drop=_type__freq_) sum=;
run;
proc transpose data=sum1 (keep = X e:) out=exposures (rename=(_0=exp_cmp_1=exp_doi ));
  id X;
  var e;; *sum;
run;
data exposures;
  set exposures;
  look = input(substr(_NAME_,2,1),8.);
run;
proc transpose data=sum1 (keep = X y:) out=events (rename=(_0=evt_cmp_1=evt_doi ));
  id X;
  var y;; *sum;
run;

```

```

data events;
    set events;
    look = input(substr(_NAME_,2,1),8.);
run;

data &data_counts. (drop = _NAME_);
    merge exposures events;
    by look;

    retain end_day;
    if look = 1 then start_day = 1;
    else start_day = end_day + 1;
    end_day = input(compress(put(look,look2day.)),best.);
run;
*cleanup tmp;
/*
proc datasets library=work nolist nodetails;
    delete data_tmp data_tmp_new new fmt_tmp exposures events sum1;
quit;
*/
%mend;

%macro get_null_coefs(dataIn, cat_vars, cont_vars, family, coefsOut);

ods listing close;
ods select none;
ods noresults;

*family based genmod calls;
%IF (%upcase(&family.) = "BINOMIAL") %THEN %DO;
    proc GENMOD data=&dataIn. DESCENDING;
        %IF %TRIM(&cat_vars.) ne "" %THEN %DO;
            class &cat_vars. /param=ref ref=first;
            %END;
        MODEL Y = &cat_vars. &cont_vars. /dist=binomial link=logit;
        ods output parameterestimates=coefs_tmp;

    run;
%END;
%ELSE %IF (%upcase(&family.) = "POISSON") %THEN %DO;
    proc GENMOD data=&dataIn. ;
        %IF %TRIM(&cat_vars.) ne "" %THEN %DO;
            class &cat_vars. /param=ref ref=first;
            %END;
        MODEL Y = &cat_vars. &cont_vars. /dist=poisson link=log offset=lnk_E;
        ods output parameterestimates=coefs_tmp;

    run;
%END;
%ELSE %DO;
    %put "ERROR: unknown family";
%END;

ods listing;

```

```

ods select all;
ods results;

data &coeffsOut. (keep = parameter estimate rename = (estimate=betas));
    set coeffs_tmp (where = (upcase(parameter) not in ('SCALE')) );
run;

*cleanup tmp;
proc datasets library=work nolist nodetails;
    delete coeffs_tmp;
quit;

%mend;

%macro get_stat_distribution(dataIn, cat_vars, cont_vars, permsIn, coeffsIn, num_dist, outSet, family);

    proc iml;

        use &coeffsIn.;
        read all var {betas} into beta_hat;

        all_stats={-5};

        use &dataIn.;
        read all var {dummy &cat_vars. &cont_vars.} into design_null;
        read all var {lnk_e} into lnk_exp;
        read all var {e} into e;
        read all var {y} into y;

        use &permsIn.;
        do i=1 to &num_dist.;

            read all var {X} into X_perm where(_sample_=i);

            %IF %UPCASE(&family.)="BINOMIAL" %THEN %DO;
                mui = exp(design_null*beta_hat)/(1+exp(design_null*beta_hat));
            %END;
            %ELSE %IF %UPCASE(&family.)="POISSON" %THEN %DO;
                mui = exp(design_null*beta_hat)#e;
            %END;
            %ELSE %DO;
                %put "ERROR: unknown family";
            %END;

            ei=(Y-mui);

            %IF %UPCASE(&family.)="BINOMIAL" %THEN %DO;
                sigma2i = mui#(1-mui);
            %END;
            %ELSE %IF %UPCASE(&family.)="POISSON" %THEN %DO;
                sigma2i = mui;
            %END;
        end;
    quit;
%macroend;

```

```

%ELSE %DO;
    %put "ERROR: unknown family";
%END;

design_full = X_perm || design_null;

Ui=design_full#ei;
U=j(1,nrow(Y),1)*Ui;
W=inv(t(design_full)*(design_full#sigma2i));
V=W*(t(Ui)*Ui)*W;
ScT=U*W[1:nrow(W),1]*W[1,1:nrow(W)]*t(U)/V[1,1];

if U[1,1]<0 then ScT = 0;
if ScT = . then ScT = 0;

all_stats=all_stats//ScT;
end;

create &outSet. from all_stats[colname='ScoreStat'];
append from all_stats;
quit;

data &outSet.;
set &outset.;
if _N_ = 1 then delete;
run;

%mend;

*gen stats under the null, for all looks;
*for each sim save max of adjusted stats;
%macro sim_dist(dataIn, vars_cat, vars_cont, dist_nsim, dataOut, tot_looks, family);

data stats_tmp;
    Length look scoreStat _sample_ 8.;
    scoreStat = 0;
    look = 0;
    _sample_ = 0;
run;

%local l;
%DO l = 1 %TO &tot_looks.;
    *permute the exposure indicator variable;
    %IF &l. = 1 %THEN %DO;
        proc multtest data=&dataIn. (where = (e&l. > 0)) perm n=&dist_nsim. noprint
outsamp=all_perms (keep = X _sample_);
            class dummy;
            test mean(X);
        run;
        data all_perms;
            set all_perms;
            look_enter = &l.;
    %END;
%END;

```

```

run;
%END;
%ELSE %DO;
proc multtest data=&dataIn. (where = (e&l. > 0 and e%eval(&l.-1) <=0 )) perm
n=&dist_nsim. noprint outsamp=perms (keep = X _sample_);
class dummy;
test mean(X);
run;
data perms;
set perms;
look_enter = &l.;
run;
proc append
base = all_perms
data = perms;
quit;
%END;
%END;

%DO l = 1 %TO &tot_looks.;

%put boundary look = &l. starting at %sysfunc(datetime()),datetime20.2);

%get_null_coeffs(&dataIn. (where = (e>0) keep = &vars_cat. &vars_cont. e&l. lnk_e&l. y&l.
dummy id rename = (lnk_e&l. = lnk_e e&l. = e y&l. = y ) drop = X), &vars_cat., &vars_cont., &family., betas_null);

*distribution of score statistics under the null;
data look_perms;
set all_perms (where = (look_enter le &l.));
run;
%get_stat_distribution(&dataIn. (where = (e>0) keep = &vars_cat. &vars_cont. e&l. lnk_e&l. y&l.
dummy rename = (lnk_e&l. = lnk_e e&l. = e y&l. = y ) drop = X), &vars_cat., &vars_cont., look_perms , betas_null,
&dist_nsim.,stats_look,&family.);

data stats_look;
set stats_look ;
look = &l.;
_sample_ = _N_;
run;
proc append
base = stats_tmp
data = stats_look ;
quit;
%END; *end look iterations;

*save the statistics and arrange in one column per look;
proc sort data = stats_tmp (where = (_sample_ ne 0))
out = stats_tmp;
by _sample_;
run;

proc transpose data= stats_tmp out= &dataOut. (drop=_name__sample_);

```

```

    by _sample_;
    id look; *across;
    var scoreStat; *sum;
run;

*cleanup tmp;
proc datasets library=work nolist nodetails;
    delete lookperm stats_look stats_tmp;
quit;
%mend;

*from normed stats calculate qtail for alpha level;
%macro gen_bdry(distIn, dist_nsim, bdryOut, delta, alpha, look_plan, reuse_bdry_vals, cur_look, tot_looks);

    %local bdry_calc_first dist_noprevsig last_look;
    %let bdry_calc_first = 1;
    %let dist_noprevsig = &distIn.;
    %IF &reuse_bdry_vals. = Y and &cur_look. ne 1 %THEN %DO;

        %let bdry_calc_first = &cur_look.;
        %let last_look = %eval(&cur_look - 1);

        data dist_noprevsig;
            set &distIn.;

            Length bdry1-bdry&last_look. 8.;
            retain bdry;;
            array bdry(*) bdry1-bdry&last_look.;
            if (_N_ = 1) then do;
                do i = 1 to &last_look.;
                    bdry(i) = (input(compress(put(i,look2bdry.)),best.));
                end;
            end;

            array stats(*) _1- &last_look.;
            Length prev_sig 4.;
            prev_sig = 0;
            do i = 1 to &last_look.;
                if (stats(i) > bdry(i)) then prev_sig = 1;
            end;

            if (prev_sig = 0);
run;

        %let dist_noprevsig = dist_noprevsig;
    %END;

    data dist_max (keep = mx_stat);
        set &dist_noprevsig.;

        *max of array requires at least two elements;

```

```

Length dummystat 8.;

array stats(*) _&bndry_calc_first.-_&tot_looks. dummystat;

do i = 1 to dim(stats)-1;
    *transform statistics to constant boundary scale;
    stats(i) = (input(compress(put(i,look2info.)),best.))**(1-2*&delta.)*stats(i);
end;

dummystat=stats(1);
*on this scale, max over all looks will indicate if ever signalled;
mx_stat = max(of stats(*));
run;

proc sort data = dist_max
    out = dist_max
    ;
    by mx_stat;
run;

%let dsnid = %sysfunc(open(dist_max));
%let dist_left = %sysfunc(attrn(&dsnid.,NOBS));
%let rc = %sysfunc(close(&dsnid.));

data qtail;
    set dist_max;

    retain j m alpha_left p_low val_low;
    if _N_ = 1 then do;
        alpha_left = (&dist_left. - &dist_nsim.*(1-&alpha.))/(&dist_left.);

        m = (1-alpha_left)/4 + 3/8;
        j = floor(&dist_left.*(1-alpha_left) + m);

    end;

    if (_N_ = j) then do;
        p_low = (_N_ - 3/8)/(&dist_left. + 1/4);
        val_low = mx_stat;
    end;
    if (_N_ = j + 1) then do;
        p_high = (_N_ - 3/8)/(&dist_left. + 1/4);
        val_high = mx_stat;

        val_p = val_low + ((1-alpha_left - p_low)/(p_high-p_low))*(val_high - val_low);
        output;
    end;
run;

data &bdryOut.;
    set qtail;
    length bd1-bd&tot_looks. 8.;

```

```

array bdry(*) bd;;

do i = 1 to &tot_looks.;
  if ("&reuse_bdry_vals." = "Y" and i < &cur_look.) then do;
    bdry(i) = (input(compress(put(i,look2bdry.)),best.));
  end;
  else do;
    *transform statistics back to original scale;
    bdry(i) = (input(compress(put(i,look2info.)),best.))**(2*&delta. - 1)*val_p;
  end;
end;
drop i val_p;

run;

proc print data=&bdryOut.;
run;

*cleanup tmp;
proc datasets library=work nolist nodetails;
  delete qtail outranks dist_max fmt_tmp;
quit;
%mend;
%macro test_all_looks(indata, vars_cat, vars_cont, outsigns, bd, tot_looks, alpha, dist_nsim, family);
  data signals;
    look = 0;
    scoreStat = 0;
    boundary = 0;
    signal = 0;
  run;

  %local l currstats currbd;
  %DO l = 1 %TO &tot_looks.;
    data look_data;
      set &indata. (where = (e > 0) keep = &vars_cat. &vars_cont. X e&l. lnk_e&l. y&l.
dummy_sample_rename = (lnk_e&l. = lnk_e e&l. = e y&l. = y ));
    run;

    *if plan ahead - save the betas for these looks when generate boundary;
    %get_null_coefs(look_data, &vars_cat., &vars_cont., &family., betas);
    %get_stat_distribution(look_data, &vars_cat., &vars_cont., look_data, betas,
1,stat_test,&family.);

    data _null_;
      set stat_test;
      call symput("currstat",scoreStat);
    run;
    data _null_;
      set &bd.;
      call symput("currbd",bd&l.);
    run;

    data look_signal;

```



```

        look = &l.;
        scoreStat = &currstat;
        boundary = &currbd.;
        signal = 0;
        if (&currstat. > &currbd.) then do;
            signal = 1;
        end;
run;

proc append
    base = signals
    data = look_signal;
quit;
%END; *end stepping over looks;

proc sort data = signals (where = (look > 0))
    out = &outsigs.;
    by DESCENDING look ;
run;

proc print data=&outsigs.;
run;

proc datasets library=work nolist nodetails;
    delete signals look_signal look_data;
quit;
%mend;

%macro generate_report(look_summary, signals, data_counts, interimLib, this_look, last_look, look_plan, bd,
alpha, max_sample, bnd_type, nboot, detail);
    ods noresults;
    ods listing close;

    %local outPath info_prop;
    data _null_;
        set sashelp.vslib;
        if libname="%upcase(&interimLib.)" then do;
            call symput("outPath",trim(path));
        end;
run;
ods rtf file = "&outPath.\GSEE Sequential Testing - Look &this_look..rtf";
footnote "GSEE Method - Sequential Testing.sas (&sysdate, &systemtime)";

data tmp;
    set &look_plan. (where = (look = &this_look.));
    curr_info = look_sample/&max_sample.;
    call symput("info_prop",curr_info);
run;

    title "Method: GSEE - Look: &this_look. - Information Proportion: &info_prop. - Boundary: &bnd_type. -
Alpha: &alpha.";
    *high level testing overview;

```

```

proc sort data = &signals. (where = (look >0))
  out = &signals.;
  by DESCENDING look;
run;
proc print data=&signals.;
  title2 "Testing outcomes for all looks through current look (&this_look.);";
run;

*data counts documentation - used in analysis;
proc sort data = &data_counts.
  out = &data_counts.;
  by DESCENDING look;
run;
proc print data=&data_counts.;
  title2 "Analysis for all looks through current look (&this_look.) based on aggregate counts:";
run;

*boundary documentation - used in analysis;
proc print data=&bd. (keep = bd:);
  title2 "Analysis for current look (&this_look.) based on complete boundary:";
run;

*optional summaries;
%IF &detail.=Y %THEN %DO;

  data bd_long (keep = look bndry);
    set &bd.;

    array boundary (*) bd:;
    do i = 1 to &last_look.;
      bndry = boundary(i);
      look = i;
      output;
    end;
  run;
  data &interimLib..look_plan_asof_&this_look.;
    merge &look_plan. (IN = a rename =(bndry = bndry_prev)) bd_long (IN = b);
    by look;
  run;
%END;

*cleanup tmp;
proc datasets library=work nolist nodetails;
  delete &signals. &data_counts. &bd. &look_summary.;
quit;
ods rtf close;
ods listing;
ods results;
%mend;

```

B. R PROGRAMMING CODE

1. CSSP

```
#####
## DRAFT OF CODE: Current version available at
## http://faculty.washington.edu/acook/software.html
#####

#####
## DRAFT OF CODE FOR THE CONDITIONAL SEQUENTIAL SAMPLING PROCEDURE
## (CSSP) FOR GROUP SEQUENTIAL MONITORING OF POSTMARKETING SURVEILLANCE
## DATA.
## Programming: Robert D. Wellman
## CSSP method developed by LingLing Li (see references)
##
## Current version available at
## http://faculty.washington.edu/acook/software.html
#####

#####
## DESCRIPTION
## cssp is used to analyze postmarketing surveillance data sequentially
## using stratification for confounding control.
#####

#####
## USAGE
##
## cssp(data,path=NULL,this.look,tstat.sim=10000,alpha=0.05,total.size,
##   current.size,boundary.type=1,rho=1)
#####

#####
## ARGUMENTS
##
## data
## a data frame or list (or object coercible by as.data.frame to a data
## frame) containing variables needed to use method (see details below
## for data specification).
##
## path
## File path indicator folder where interim files are stored. If NULL
## then uses working directory (getwd()).
##
## this.look
## a integer giving the current look in terms of a sequence of looks
## starting from 1. Defaults to 1.
##
## tstat.sim
## an integer giving the number of simulations used to obtain the
```

```

## distribution of the test statistic under the null in each strata.
##
## alpha
## overall alpha level for the study, e.g., 0.05 is commonly used and
## equates to a type I error of 5%.
##
## total.size
## total sample size to be accumulated by at the end of the study
## (last analysis).
##
## current.size
## sample size at the time of this.look.
##
## boundary.type
## an integer giving the desired boundary shape: 1=Pocock,
## 2=O'Brien-Fleming, 3=Power function. If the power function is used,
## an additional parameter, rho, must be specified. Defaults to 1.
##
## rho
## a numeric constant > 0 indicating the desired exponent for
## power function type boundary.
#####

#####
## DETAILS
##
## Data must be formatted as given in the included CSSPexample.Rdata file.
## The data must already be summarized by stratum and include the
## following variables named exactly as specified here:
##
## look integer giving look number in a sequence beginning at 1 with
## no maximum
##
## stratainteger indexing strata valued from 1 to number of strata
##
## subj_cmp integer giving the number of subjects in the comparison
## group within each look and each stratum
##
## subj_doi integer giving the number of subjects in the group on the
## drug of interest within each look and each stratum
##
## exp_cmp numeric giving the aggregated exposure time for the comparison
## group within each look and each stratum
##
## exp_doi numeric giving the aggregated exposure time for the group on
## the drug of interest (DOI) with each look and each stratum
##
## evt_cmp integer giving the aggregated number of events in the comparison
## group within each look and each stratum
##
## evt_doi integer giving the aggregated number of events in the group on
## the drug of interest within each look and each stratum

```

```
#####
#####
## VALUE
##
## cssp returns an object of class "CSSP".
##
## A summary of the analysis by look is printed. The summary function will
## extract the same summary from the CSSP object. After each analysis the
## resulting CSSP object is stored in a user-specified directory. If no
## directory is passed into the cssp function then the current working
## directory is used. The file is saved as "t_sim.Rdata". After the first
## look, cssp will load the previously saved file from the directory
## specified by the user. The augmented object will then be saved over
## the previous version but will contain all information through the
## current look. Directories and file names can be manipulated in
## the code if they need to be modified for a particular situation.
#####

#####
## EXAMPLES
## source("H:/CSSPfunctions.R")
## options(digits=10)
## ex1<-read.csv("H:/example1.csv", header=T)
## look1<-
cssp(ex1,this.look=1,tstat.sim=10000,alpha=0.05,total.size=10000,current.size=1250,boundary.type=1,rho=1)
## look2<-
cssp(ex1,this.look=2,tstat.sim=10000,alpha=0.05,total.size=10000,current.size=2500,boundary.type=2,rho=1)
## summary(look2)
##
##   Events P-value Look Alpha Cum. Alpha Signal Num. Sims
## Look 1   1 0.5539 0.0063 0.0063 0 10000
## Look 2  10 0.2331 0.0063 0.0126 0 10000
#####

#####
## REFERENCES
##
## Li, L. A conditional sequential sampling procedure for drug safety
## surveillance. Statistics in Medicine 2009; 28:3124-3138.
#####

#####
## FUNCTIONS (Load before using cssp)

binom.gen <- function(data, H, R) rbinom(1,data[H],data[R])

comp.r<-function(x,c1,c2) x[,c1]/(x[,c1]+x[,c2])

sim.T<-function(data, H,R,S) sum(rbinom(max(data[,S]),data[,H],data[,R]))
```

```

cssp.current<-function(data,nsim,H,R,S){
  res.sim<-replicate(nsim,sim.T(data, H,R,S))
  return(res.sim)
}

#Error Spending Function (esf)
esf<-function(type=1,t,T,alpha,rho=NULL){
  if(type==1) {alpha*log(1 + (exp(1)-1)*(t/T))}
  else if(type==2) {2*(1-pnorm(abs(qnorm(alpha/2))/(sqrt(t/T))))}
  else if(type==3) {alpha*(t/T)^(rho)}
}

## Class constructor function

CSSP<-function(input.list){
  CSSP<-input.list
  class(CSSP)<-"CSSP"
  return(CSSP)
}

summary.CSSP<-function(x){
  print(x[['Summary']])
}

cssp<-
function(data,path=NULL,this.look,tstat.sim=10000,alpha=0.05,total.size,current.size,boundary.type=1,rho=1){

  if(is.null(path)){path<-getwd()}
  dset<-as.data.frame(data)
  l<-this.look
  if(l==1){
    dset<-subset(dset,look==1)
    alpha.1<-esf(boundary.type,current.size,total.size,alpha,rho)
    cut.rank<-ceiling(tstat.sim*(1-alpha.1))
    dset$h<-apply(dset[,c("evt_doi","evt_cmp")], 1, sum)
    dset$r<-comp.r(dset,"exp_doi","exp_cmp")
    res<-cssp.current(subset(dset,look==l),tstat.sim,H="h",R="r",S="strata")
    t.sim.raw<-list('Look 1'=res)
    save(t.sim.raw,file=paste(path,"/tsimraw.Rdata",sep=""))
    res.sort<-rank(res,ties.method='first')
    a.spent.t<-sum(res.sort>=cut.rank)/tstat.sim
    obs.t<-sum(dset$evt_doi)
    order.obs.t<-sum(res>=obs.t)
    p.val.t<-order.obs.t/tstat.sim
    signal<-p.val.t<a.spent.t
    res1<-ifelse(res.sort>cut.rank, 9999, res)
    output<-list('Summary'=matrix(c(obs.t,p.val.t,a.spent.t,a.spent.t,signal,tstat.sim),nrow=1,
      dimnames=list(c('Look 1'),c('Events','P-value','Look Alpha','Cum. Alpha','Signal',
        'Num. Sims'))),'Look1'=res1)
    save(output,file=paste(path,"/t_sim.Rdata",sep=""))
    #print(output[['Summary']])
  }
}

```

```

class(output)<-"CSSP"
output
}
else if(l>1){
  load(file=paste(path,"/t_sim.Rdata",sep=""))
  load(file=paste(path,"/tsimraw.Rdata",sep=""))
  if(l > dim(output$Summary)[1] + 1){stop('Looks out of sequence.')}
  dset.this<-subset(data,look==l)
  dset.all<-data
  alpha.l<-esf(boundary.type,current.size,total.size,alpha,rho)
  cut.rank<-ceiling(tstat.sim*(1-alpha.l))
  dset.this$h<-apply(dset.this[,c("evt_doi", "evt_cmp")], 1, sum)
  dset.this$r<-comp.r(dset.this, "exp_doi", "exp_cmp")
  res<-cssp.current(subset(dset.this,look==l),tstat.sim,H="h",R="r",S="strata")
  t.sim.raw[[paste('Look',l)]]<-res
  save(t.sim.raw,file=paste(path,"/tsimraw.Rdata",sep=""))
  res.current<-output[[l]] + res
  res.sort<-rank(res.current, ties.method='first')
  a.spent.t<-sum(res.sort>=cut.rank)/tstat.sim
  prev.a.spent.t<-a.spent.t-output[['Summary']][l-1,'Cum. Alpha']
  obs.t<-sum(subset(dset.all,look<=l)$evt_doi)
  order.obs.t<-sum(res.current>=obs.t)
  p.val.t<-order.obs.t/tstat.sim
  signal<-p.val.t<a.spent.t
  res.current<-ifelse(res.sort>cut.rank, 9999, res.current)
  output[['Summary']]<-rbind(output[['Summary']],
  c('Events'=obs.t,'P-value'=p.val.t,'Look Alpha'=prev.a.spent.t,
  'Cum. Alpha'=a.spent.t,'Signal'=signal,'Num. Sims'=tstat.sim))
  output[[paste("Look",l,sep=")]]<-res.current
  rownames(output[['Summary']][l])<-c(paste('Look',l))
  save(output,file=paste(path,"/t_sim.Rdata",sep=""))
  #print(output[['Summary']])
  class(output)<-"CSSP"
  output
}
}
#####

```

2. GS EE

```
#####
## DRAFT OF CODE: Current version available at
## http://faculty.washington.edu/acook/software.html
#####

#####
## CODE FOR THE GS EE COVARIATE ADJUSTED METHOD FOR
## GROUP SEQUENTIALLY MONITORING POSTMARKETING SURVEILLANCE DATA
## By: Andrea J Cook
#####

##### Function to run the GS EE Approach #####
##input: ATime: a vector of analysis times observed and expected
##          to be observed
##          CurTime: which look current analysis is at
##          PrevBounds: give vector of previous boundaries if in
##                      middle of study
##          Nend: specify total sample size at end of study
##                (leave blank if at end of study)
##          propinfo: proportion of statistical information up to each
##                  observed and expected analysis times (0,1]
##                  (example: proportion of total sample size at each look)
##          Y: a vector of observed outcomes
##          X: a vector of indicators of being exposed or unexposed
##          Z: a vector of the confounders
##          S: a vector of the start times
##          E: a vector of the exposure durations observed up to
##            CurTime for each individual
##            (only needed for chronic exposure setting)
##          family: specify "binomial" for single exposure binary outcome
##                  or "poisson" for chronic exposure
##          alpha: total type I error desired to spend across all looks
##          delta: shape parameter for the boundary
##                (delta=.5 is Pocock) (delta=0 is Fleming)
##          nsim: Number of simulations to calculate the boundary
##              (use at least 1000, but higher the better)
##output: maxstatY: a list that returns alpha, delta, signal (T or F),
##          signal time (time of signal or end of study), and
##          test a dataset (ATimet, LLRt, boundt, sigt)
##          at each time point t
#####
seqEEcovAdjUnif<-function(ATime,CurTime=max(ATime),PrevBounds=NULL,
Nend=length(Y),propinfo,Y,X,Z,S,E=rep(1,length(X)),family="binomial",
delta=0.5,alpha=0.05,nsim=10000)
{
  Y<-as.matrix(Y[order(S)])
  X<-X[order(S)]
  Z<-as.matrix(Z)
  Z<-Z[order(S),]
```



```

Z<-as.matrix(Z)
E<-E[order(S)]
S<-S[order(S)]
Zint<-cbind(1,Z)

# Create a vector with amount of sample size observed up to time t
Ncum<-NULL
for(T in 1:length(ATime[ATime<=CurTime]))
{
    Ncum<-c(Ncum,sum(S<=ATime[T]))
}

## BINOMIAL MODEL USING LOGISTIC REGRESSION LINK ##
if(family=="binomial")
{
# If you are in the middle of the study and need to simulate what
# will happen at future looks of the dist of Y, X, and E #
    if(length(Y)<Nend)
    {
        numA<-length(propinfo)
        NcumAll<-c(Ncum, round(Nend*propinfo[(length(Ncum)+1):numA],0))

        Nt<-Ncum[1]
        for(T in 2:numA)
        {
            Nt<-c(Nt,NcumAll[T]-NcumAll[T-1])
        }

        datcur<-gen.dataFO.bin(Nt=Nt,Yt=Y, Xt=X, Zt=Z, Acur=length(Ncum))
        Y<-datcur$Y
        X<-datcur$X
        Z<-datcur$Z
        Ncum<-NcumAll
    }

    betaZ<-NULL
    for(T in 1:length(ATime))
    {
        betaZ<-cbind(betaZ,glm(Y[1:Ncum[T]]~Z[1:Ncum[T],],family=binomial)$coef)
    }
    #Skip looks that you could not estimate beta
    bfound<-apply(is.na(betaZ),2,sum)
    betaZ<-betaZ[,bfound==0]
    ATime<-ATime[bfound==0]
    propinfo<-propinfo[bfound==0]
    Ncum<-Ncum[bfound==0]

    bd<-
seqEEcovAdjUnif.boundary.bin(PrevBound=PrevBounds,propinfo=propinfo,Y=Y,X=X,Z=Z,S=S,Ncum=Ncum,delta=de
lta,betaZ=betaZ,alpha=alpha,nsim=nsim)
    sig<-0
    sigtime<-max(ATime)

```

```

ScT<-NULL

boundScTt<-NULL
boundEst<-NULL
ATimet<-NULL
sigt<-NULL
T<-1
while(T<=length(ATime) & ATime[T]<=CurTime)
{
  ScTcur<-ScTt.bin(Y=Y[1:Ncum[T]],X=X[1:Ncum[T]], Z=Z[1:Ncum[T]],betaZ=betaZ[,T])
  ScT<-c(ScT,ScTcur)
  boundScTt<-c(boundScTt,bd[T])
  ATimet<-c(ATimet,ATime[T])
  sigt<-c(sigt,as.numeric(ScTcur>boundScTt[T]))
  if(ScTcur>boundScTt[T])
  {
    sig<-1
    sigtime<-ATime[T]
    #T<-length(ATime)+1
  }
  T<-T+1
}
test<-data.frame(cbind(ATimet,ScT,boundScTt,sigt))
return(list(alpha=alpha,delta=delta,signal=(sig==1), sigtime=sigtime,test=test,bd=bd))
}
## POISSON MODEL USING LOG LINK ##
if(family=="poisson")
{
# If you are in the middle of the study and need to simulate what # will happen at future looks of the dist of Y, X,
and E #
  if(length(Y)<Nend)
  {
    numA<-length(propinfo)
    NcumAll<-c(Ncum, round(Nend*propinfo[(length(Ncum)+1):numA],0))

    Nt<-Ncum[1]
    for(T in 2:numA)
    {
      Nt<-c(Nt,NcumAll[T]-NcumAll[T-1])
    }
    datcur<-gen.dataFO.pois(Nt,ATime,Yt=Y,Xt=X,Zt=Z,Et=E,St=S,
      Acur=length(Ncum))
    Y<-datcur$Y
    X<-datcur$X
    Z<-datcur$Z
    E<-datcur$E
    S<-datcur$S
    Ncum<-NcumAll
  }

  Etime<-NULL
  Ytime<-NULL

```

```

for(T in 1:length(ATime))
{
  Ecur<-E
  Ecur[(ATime[T]-S+1)<E]<-(ATime[T]-S+1)/[(ATime[T]-S+1)<E]
  Ecur[S>ATime[T]]<-0
  Ycur<-Y
  Ycur[(ATime[T]-S+1)<E]<-0
  Etime<-cbind(Etime,Ecur)
  Ytime<-cbind(Ytime,Ycur)
}
#Skip looks if there are not at least 5 cases ##
sumY<-apply(Ytime,2,sum)
Etime<-Etime[sumY>=5]
Ytime<-Ytime[sumY>=5]
ATime<-ATime[sumY>=5]
propinfo<-propinfo[sumY>=5]
Ncum<-Ncum[sumY>=5]

betaZ<-NULL

for(T in 1:length(ATime))
{
  betaZcur<- glm(Ytime[1:Ncum[T],T]~Z[1:Ncum[T],]+
offset(log(Etime[1:Ncum[T],T])),family="poisson")$coef
  betaZ<-cbind(betaZ,betaZcur)
}
#Skip looks that you could not estimate beta
bfound<-apply(is.na(betaZ),2,sum)
betaZ<-betaZ[bfound==0]
Etime<-Etime[bfound==0]
Ytime<-Ytime[bfound==0]
ATime<-ATime[bfound==0]
propinfo<-propinfo[bfound==0]
Ncum<-Ncum[bfound==0]

bd<-
seqEEcovAdjUnif.boundary.pois(PrevBound=PrevBounds,propinfo=propinfo,Ytime=Ytime,X=X,Z=Z,S=S,Etime=Etime,
Ncum=Ncum,delta=delta,betaZ=betaZ,alpha=alpha,nsim=nsim)
sig<-0
sigtime<-max(ATime)
ScT<-NULL

boundScTt<-NULL
boundEst<-NULL
ATimet<-NULL
sigt<-NULL
T<-1
while(T<=length(ATime) & ATime[T]<=CurTime)
{
  ScTcur<-ScTt.pois(Y=Ytime[1:Ncum[T],T],X=X[1:Ncum[T]],
Z=Z[1:Ncum[T],],E=Etime[1:Ncum[T],T], betaZ=betaZ[,T])
}

```

```

        ScT<-c(ScT,ScTcur)
        boundScTt<-c(boundScTt,bd[T])
        ATimet<-c(ATimet,ATime[T])
        sigt<-c(sigt,as.numeric(ScTcur>boundScTt[T]))
        if(ScTcur>boundScTt[T])
        {
            sig<-1
            sigtime<-ATime[T]
            #T<-length(ATime)+1
        }
        T<-T+1
    }
    test<-data.frame(cbind(ATimet,ScT,boundScTt,sigt))
    return(list(alpha=alpha,delta=delta,signal=(sig==1),
sigtime=sigtime,test=test))
    }
}

### REST OF CODE JUST FOR INTERNAL FUNCTION CALLS #####

#####Generate Data for future Observations #####
## Observations as Unit
## input: Nt: Vector of number of people you plan to observe/have
##         observed at each look
## Yt: Vector of number of drugs you have observed up to time t
## Xt: Vector of indicator of exposed/unexposed you have
##         observed up to time t
## Zt: Matrix of confounders that you have observed up to time t
## Et: Vector of exposure times observed up to time t
## Acur: Current look that we are on at time t
## output: datcur: a list of outcome data Y, X, Z, E, S
#####
gen.dataFO.bin <- function(Nt, Yt, Xt, Zt, Acur)
{
    Y<-Yt
    X<-Xt
    Z<-as.matrix(Zt)

    ## Assumes that current distribution of Y, X, and Z will stay constant
    ## throughout the rest of the study
    ## sample with replacement from current Y, X and Z
    id<-c(1:length(Xt))
    for(i in (Acur+1):length(Nt))
    {
        lvals<-sample(id,Nt[i],replace=T)
        Y<-c(Y,Yt[lvals])
        X<-c(X,Xt[lvals])
        Z<-rbind(Z,as.matrix(Zt[lvals,]))
    }
    return(list(Y=Y,X=X,Z=Z))
}

```

```

gen.dataFO.pois <- function(Nt, ATime, Yt, Xt, Zt, Et, St, Acur)
{
  Y<-Yt
  X<-Xt
  Z<-as.matrix(Zt)
  E<-Et
  S<-St
  ## Assumes that current distribution of X, Z, and E will stay constant ## throughout the rest of the study
  ## sample with replacement from current X, E, and Z
  id<-c(1:length(Xt))
  for(i in (Acur+1):length(Nt))
  {
    lvals<-sample(id,Nt[i],replace=T)
    Y<-c(Y,Yt[lvals])
    X<-c(X,Xt[lvals])
    Z<-rbind(Z,as.matrix(Zt[lvals,]))
    E<-c(E,Et[lvals])
    S<-c(S,runif(Nt[i],ATime[i-1]+1,ATime[i]))
  }
  return(list(Y=Y,X=X,Z=Z,E=E,S=S))
}

```

```

#####Calculate stat#####
#input: St: a vector of score test statistics each look
#      propinfo: vector of total sample size up to look t
#      delta: shape parameter (0.5=Pocock, 0=O'Brien Fleming)
#output: stat: a vector of stat each look (same dimension as score)
#####
stat <- function(St, propinfo, delta)
{
  stat <- (propinfo)^(1-2*delta)*St
  return(stat)
}

```

```

##### Calculate Score Statistic #####
##input: X: a vector indicating if on exposure of interest
##      Y: a vector or matrix of outcomes for look t
##      Z: a matrix of confounders for X at look t
##      E: a vector of exposure durations at look t
##output: ScT: the value of the Score Statistic at each look statistic
#####
## Case 1: Single Exposure Time
ScTt.bin <- function(Y, X, Z,betaZ=NULL)
{
  Y<-as.matrix(Y)
  Zint<-cbind(1,Z)

  # Makes sure that there are at least 5 cases
  if((sum(X*Y)==0)|(sum((1-X)*Y)==0)|sum(Y)<5)
  {
    return(0)
  }
}

```

```

else
{
  n<-length(Y)
  mui<-as.vector(exp(Zint**betaZ)/(1+exp(Zint**betaZ)))
  ei<-as.vector(Y-mui)
  XZ<-cbind(X,Zint)
  Ui<-XZ*ei
  U<-t(rep(1,length(Y))**Ui)
  W<-ginv(t(XZ)**(XZ*mui*(1-mui)))
  V<-W**(t(Ui)**Ui)**W
  ScT<-U**W[,1]**W[1,]**t(U)/V[1,1]
  ScT[(((U[1,1]<0)|(is.na(t(X)**ei)))]<-0
  return(ScT)
}
}

```

Case 2: Chronic Exposure Time

```
ScTt.pois <- function(Y, X, Z, E,betaZ=NULL)
```

```

{
  Y<-as.matrix(Y)
  Zint<-cbind(1,Z)

  # Makes sure that there are at least 5 cases
  if((sum(X*Y)==0)|(sum((1-X)*Y)==0)|sum(Y)<5)
  {
    return(0)
  }
  else
  {
    n<-length(Y)

    mui<-as.vector(exp(Zint**betaZ)*E)
    ei<-as.vector(Y-mui)

    XZ<-cbind(X,Zint)

    Ui<-XZ*ei
    U<-t(rep(1,length(Y))**Ui)
    W<-ginv(t(XZ)**(XZ*mui))
    V<-W**(t(Ui)**Ui)**W
    ScT<-U**W[,1]**W[1,]**t(U)/V[1,1]
    ScT[(((U[1,1]<0)|(is.na(ScT)))]<-0

    return(ScT)
  }
}

```

Find Boundary on the score statistic given unifying boundary function

##input: Ncum: a vector of the cumulative number of people to be

observed at each look

propinfo: vector of total sample size up to look t

```

##          X: a vector of indicator of exposure data observed at each
##          look up to current look
##          Z: a vector of confounder data observed at each look up to
##          current look
##          S: a vector of the start times observed up to current look
##          Etime: a matrix of exposure times observed at each look
##          up to current look
##  delta: shape of the boundary (delta=0.5 is Pocock; 0 is Fleming)
##output: maxstatY: a vector of the critical values at each look t
#####
do.Cr.bin <- function(Ncum,PrevBounds,propinfo,Y,X,Z,S,delta,betaZ)
{
  T<-length(Ncum)
  N<-Ncum[T]
  Ncum0<-c(1,Ncum)

  ScTY<-NULL
  Xcur<-NULL
  for(i in 1:T)
  {
    Xcur<-c(Xcur,sample(X[Ncum0[i]:Ncum0[i+1]],Ncum0[i+1]-
Ncum0[i]+1,replace=F))
    ScTY<-c(ScTY,ScTt.bin(Y=Y[1:Ncum[i]], X=Xcur[1:Ncum[i]],
Z=Z[1:Ncum[i],],betaZ=betaZ[,i]))
  }

  if(!is.null(PrevBounds))
  {
    CurTime<-length(PrevBounds)
    PrevBoundsComp<-c(PrevBounds,rep(10000,T-CurTime))
    sigPL<-as.numeric(sum(ScTY>PrevBoundsComp)>0)
    maxstatY <- max(stat(ScTY[CurTime+1:T], propinfo, delta),na.rm=T)
    return(cbind(sigPL,maxstatY))
  }
  else
  {
    maxstatY <- max(stat(ScTY, propinfo, delta),na.rm=T)
    return(cbind(sigPL=0,maxstatY))
  }
}

do.Cr.pois <- function(Ncum,PrevBounds,propinfo,Ytime,X,Z,S,Etime,
delta,betaZ)
{
  T<-length(Ncum)
  N<-Ncum[T]
  Ncum0<-c(1,Ncum)

  ScTY<-NULL
  Xcur<-NULL
  for(i in 1:T)
  {

```

```

        Xcur<-c(Xcur,sample(X[Ncum0[i]:Ncum0[i+1]],Ncum0[i+1]-
Ncum0[i]+1,replace=F))
        ScTY<-c(ScTY,ScTt.pois(Y=Ytime[1:Ncum[i],i], X=Xcur[1:Ncum[i]],
Z=Z[1:Ncum[i],], E=Etime[1:Ncum[i],i],betaZ=betaZ[,i]))
    }
    if(!is.null(PrevBounds))
    {
        CurTime<-length(PrevBounds)
        PrevBoundsComp<-c(PrevBounds,rep(10000,T-CurTime))
        sigPL<-as.numeric(sum(ScTY>PrevBoundsComp)>0)
        maxstatY <- max(stat(ScTY[CurTime+1:T], propinfo, delta),na.rm=T)
        return(cbind(sigPL,maxstatY))
    }
    else
    {
        maxstatY <- max(stat(ScTY, propinfo, delta),na.rm=T)
        return(cbind(sigPL=0,maxstatY))
    }
}

##### Function to obtain Critical Boundary Values #####
##      input:  PrevBounds: a vector of previous boundaries if not
##              at first look
##      propInfo: Proportion of statistical information observed at
##                 each time look
##      X: a vector of the current indicator of exposure data ##
##          observed up to time t
##      Z: a vector of the confounders observed up to time t
##      S: a vector of the start times observed up to time t
##      E: a vector of the exposure durations observed up to time t
##      Bo: Estimated baseline from model without X
##      Bz: Estimated confounder relationship without X
##      delta: shape parameter of boundary
##              (delta=.5 is Pocock; 0 is Fleming)
##      nsim: Number of simulations to calculate the boundary
##output: maxstatY: a vector of the critical values at each look t
#####
seqEEcovAdjUnif.boundary.bin<-function(PrevBounds,propinfo,
Y,X,Z,S,Ncum,delta,betaZ,alpha=0.05,nsim=1000)
{
    temp<-replicate(nsim,do.Cr.bin(Ncum=Ncum,PrevBounds=PrevBounds,
propinfo=propinfo,Y=Y,X=X,Z=Z,S=S,delta,delta,betaZ=betaZ))
    # Previous Looks alpha spent #
    alphaP<-mean(temp[,1,])
    print(alphaP)
    qtail<-quantile(temp[,2,temp[,1,]==0],1-(alpha-alphaP),na.rm=T,type=9)
    return(CV=c(PrevBounds,
(propinfo^(2*delta-1)*qtail)[length(PrevBounds):length(propinfo)]))
}

seqEEcovAdjUnif.boundary.pois<-function(PrevBounds,propinfo,
Ytime,X,Z,S,Etime,Ncum,delta,betaZ,alpha=0.05,nsim=1000)

```



```
{
temp<-replicate(nsim,do.Cr.pois(Ncum=Ncum,PrevBounds=PrevBounds,propinfo=propinfo,
Ytime=Ytime,X=X,Z=Z,S=S,Etime=Etime,delta,betaZ=betaZ))
  # Previous Looks alpha spent #
  alphaP<-mean(temp[1,])
  print(alphaP)
  qtail<-quantile(temp[2,temp[1,]==0],1-(alpha-alphaP),na.rm=T,type=9)
  return(CV=c(PrevBounds,(propinfo^(2*delta-1)*qtail)[length(PrevBounds):length(propinfo)]))
}
```