# MINI-SENTINEL METHODS

# SAFETY SIGNALING METHODS FOR SURVIVAL OUTCOMES TO CONTROL FOR CONFOUNDING IN THE MSDD

**Prepared by:** Andrea J Cook, PhD,[1,2] Robert D Wellman, MS,[1] Azadeh Shoaibi, PhD MHS,[3] Ram C Tiwari, PhD,[4] Susan R. Heckbert, MD, PhD,[5] Lingling Li, PhD,[6] Rima Izem, PhD,[4] Rongmei Zhang, PhD,[4] 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 Medical Policy Center for Drug Evaluation and Research, FDA, Silver Spring, MD 4. Office of Biostatistics, Center for Drug Evaluation and Research, FDA, Silver Spring, MD 5. Department of Epidemiology, University of Washington, Seattle, WA 6. Department of Population Medicine, Harvard Pilgrim Health Care Center and Harvard Medical School, Boston, MA

**October 9, 2015**

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

## Safety Signaling Methods For Survival Outcomes To Control For Confounding In The MSDD

### Table of Contents

# I. BACKGROUND, OBJECTIVES AND DELIVERABLES OF THE TASK ORDER

In previous Mini-Sentinel workgroups[1,2], statistical methods were developed, evaluated, and applied to sequentially monitor rare event outcomes that occur acutely following medical product exposure with adjustment for confounders. None of these approaches were explicitly designed for sequential testing with survival-type data. For example, the PRISM Year 2 workgroup developed and evaluated a causal method using inverse probability treatment weighting (IPTW) for vaccine safety evaluations in a distributed setting (PRISM Year 2: "Enhancing current sequential analytic techniques to improve causal inference"[1]). This method can be applied for a single point-in-time analysis or in a sequential monitoring framework. However, this approach was developed for a one-time (i.e., vaccine) exposure and for events that occur shortly following exposure (e.g., seizure within 1-42 days). Extending these methods to allow for chronically used exposures (e.g., drugs) and events that may occur further in time from the initiation of drug use (e.g., acute myocardial infarction) is critical and requires survival techniques.

This workgroup was tasked to first review the statistical and epidemiology literature on methods for survival outcomes that incorporate adjustment for confounders using a causal inference approach. The workgroup then assessed the methods applicability to Mini-Sentinel and made recommendations on which strategies were best suited for use within this setting, where rare events, a distributed data environment, and sequential testing introduce new complications. Both design-based (e.g., matching, stratification) and analysis-based (e.g., inverse probability weighting) confounder adjustment approaches were to be considered since each of these approaches has different strengths and weaknesses. To focus the workgroup and to avoid overlap with ongoing work in Mini-Sentinel using design based approaches through matching, we concentrated our review on approaches using Cox's Proportional Hazards (PH) models[3] with direct adjustment for confounders. We focused on methods that would be viable in the distributed data setting (e.g. individual-level data remains at the healthcare site behind firewalls and only de-identified data is shared across sites). Barriers to effective data sharing, such as privacy concerns and proprietary information policies, make pooling of individual-level data across sites rarely used unless deemed critical to the question of interest.

The second aim of the task order was to develop new approaches tailored to the Mini-Sentinel setting. From our literature review we decided to compare methods that aggregate data (Section II.C.1) as a form of de-identification or conduct site-specific Cox's PH regression models and share summary statistics across sites (Section II.C.2). These methods have not been evaluated in the rare event setting of Sentinel and would be classified as new statistical approaches. We further extended the methods to sequential monitoring using different boundary formations (Section II.D).

The third aim of the task order was to evaluate via simulation the most promising existing approaches and new approaches tailored to the Mini-Sentinel setting. To this end we conducted a formal statistical evaluation comparing the new approaches with the gold standard approaches if we did not have the distributed data setting (Section III). We also compared different sequential monitoring boundary approaches. Specifically, we assessed when exact testing methods were needed instead of simpler boundaries based on normal approximation methods, the assumptions for which may not hold in the rare event setting.

The simulation evaluation and methods development was extensive in this task order. It was an iterative process in which we would propose a new method and assess the performance via simulation, and then improve on the new method for problems observed from the simulation study. Due to the time consuming nature of the simulation study and methods development, we were not able to apply the new methods to an existing Mini-Sentinel example (Aim 4). Further, because we were developing new methods we did not actually have an existing Mini-Sentinel dataset that had the information necessary to apply the methods. This task order was not designed or funded to conduct a new data pull. We have instead outlined the type of data and structure that would be required to conduct the methods assessed in this task order. Further, as our final deliverable, we have created R code that can be used on a dataset with the structure we have outlined.

# II. STATISTICAL METHODS

For this task order we will propose and evaluate several group sequential methods assuming Cox's PH regression models. We will first present several methods for a one time analysis tailored to the distributed data setting and rare event setting. We will then discuss two standard approaches to incorporate group sequential monitoring.

## A. STANDARD SURVIVAL METHODS IN THE NON-DISTRIBUTED DATA SETTING

We will present standard survival regression methods that are typically applied to datasets in which individual level data could be shared across sites without concerns about patient privacy or concerns that the data are proprietary. These gold standard methods will then be compared to methods that account for the distributed data setting of Mini-Sentinel and assess if there is any strong evidence of loss of information or bias.

### 1. Standard Cox PH regression

Assume for at the end of the study follow-up period at site *s* (*s* = 1, … , *S*), we observe data from participant *i* (*i*= 1,… , $n_s$) that has either received the exposure of interest, $X_{si} = 1$, or the comparator, $X_{si} = 0$. Furthermore, each person has a set of baseline confounders, $\boldsymbol{Z}_{si}$, $\delta_{si}$ indicating whether they have experienced the outcome before the end of the study follow-up period and 0 otherwise and $T_{si}$ for time to event or censoring.

Consider a Cox's PH regression model for a single site,

$$\lambda(T_{si}, \delta_{si}|X_{si}, \boldsymbol{Z}_{si}) = \lambda_0(T_{si})exp[\beta_X^s X_{si} + \boldsymbol{\beta}_z^s \boldsymbol{Z}_{si}], \qquad (1)$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard function, $\beta_X^s$ is the site-specific log(HR) comparing the exposure of interest to the comparator, and $\boldsymbol{\beta}_z^s$ is a $1\ x\ p$ vector of unknown regression parameters for site *s*.

Now extend this model to the standard centralized analysis setting in which data is collected across sites and analysis is done in a centralized location. The standard approach would now be to add a set of site indicator variables, $\boldsymbol{S}_{si}$, and directly adjust for them in the regression model. Specifically, we would fit the following model:

$$\lambda(T_{si}, \delta_{si}|X_{si}, \boldsymbol{Z}_{si}, \boldsymbol{S}_{si}) = \lambda_0(T_{si}) exp\left[\beta_X\, X_{si} + \boldsymbol{\beta}_z\, \boldsymbol{Z}_{si} + \boldsymbol{\beta}_S\, \boldsymbol{S}_{si}\right]. \quad (2)$$

We would estimate the regression model using standard partial maximum likelihood estimation to derive the fitted estimates $\hat{\beta}_x$ , $\widehat{\boldsymbol{\beta}}_z$ , and $\widehat{\boldsymbol{\beta}}_S$ .

For a given analysis time we would be interested in assessing the following hypothesis: $H_O: \beta_X = 0$ versus $H_A: \beta_X > 0$. To assess this hypothesis we would derive a test statistic. One standard test statistic is the Wald test statistic, $\hat{\beta}_X / \sqrt{\hat{V}(\hat{\beta}_X)}$. However, it is more common to form a score test statistic (a.k.a. Log Rank Statistic) since it is relatively more powerful, while still being straightforward to calculate. The corresponding Log Rank test statistic is:

$$LR(a) = \frac{\sum_{\{s,i:\delta_{si}=1\}} \left[ X_{si} - \frac{\sum_{\{k,l:T_{kl} \geq T_{si}\}} X_{kl} \exp\left(\widehat{\boldsymbol{\beta}}_z^{(0)} \boldsymbol{Z}_{kl} + \widehat{\boldsymbol{\beta}}_S^{(0)} \boldsymbol{S}_{kl}\right)}{\sum_{\{k,l:T_{kl} \geq T_{si}\}} \exp\left(\widehat{\boldsymbol{\beta}}_z^{(0)} \boldsymbol{Z}_{kl} + \widehat{\boldsymbol{\beta}}_S^{(0)} \boldsymbol{S}_{kl}\right)} \right]}{\sqrt{\sum_{\{s,i:\delta_{si}=1\}} \left[ \frac{\sum_{\{k,l:T_{kl} \geq T_{si}\}} X_{kl} \exp\left(\widehat{\boldsymbol{\beta}}_z^{(0)} \boldsymbol{Z}_{kl} + \widehat{\boldsymbol{\beta}}_S^{(0)} \boldsymbol{S}_{kl}\right)}{\sum_{\{k,l:T_{kl} \geq T_{si}\}} \exp\left(\widehat{\boldsymbol{\beta}}_z^{(0)} \boldsymbol{Z}_{kl} + \widehat{\boldsymbol{\beta}}_S^{(0)} \boldsymbol{S}_{kl}\right)} - \left( \frac{\sum_{\{k,l:T_{kl} \geq T_{si}\}} X_{kl} \exp\left(\widehat{\boldsymbol{\beta}}_z^{(0)} \boldsymbol{Z}_{kl} + \widehat{\boldsymbol{\beta}}_S^{(0)} \boldsymbol{S}_{kl}\right)}{\sum_{\{k,l:T_{kl} \geq T_{si}\}} \exp\left(\widehat{\boldsymbol{\beta}}_z^{(0)} \boldsymbol{Z}_{kl} + \widehat{\boldsymbol{\beta}}_S^{(0)} \boldsymbol{S}_{kl}\right)} \right)^2 \right]}}$$

where $\widehat{\boldsymbol{\beta}}_z^{(0)}$ and $\widehat{\boldsymbol{\beta}}_S^{(0)}$ are the fitted parameter estimates of model (2) under $H_O$ that $\beta_X = 0$. Large positive values of *LR*(*a*) signify that the exposure of interest has a higher hazard ratio compared to a comparator.

This standard regression approach can be directly used when individual level data is available. However, it may be necessary when conducting multi-site analyses, or desirable for patient privacy reasons, to limit the amount of information transferred across sites. In Section B we will discuss potential alternatives to limiting data and still conducting appropriate regression analyses that will perform well even in the rare event setting.

### 2. Site-Stratified Cox PH regression

Instead of adjusting for site in the mean model as outlined in model (2) in Section II.A.1, another common method to account for confounding by site is to use a site-stratified Cox PH regression model (see model (3)). The site-stratified cox model makes a proportional hazard assumption in each site but allows for different baseline hazards between sites. In addition, this model allows for between site heterogeneity of the confounding estimates. This approach better accounts for differences across sites compared with adjusting directly for site and therefore, reduces potential bias if there are different relationships between those that receive the exposure versus comparator. The disadvantage of this approach is there may be some loss of power/efficiency relative to adjusting for site in the situation without site heterogeneity. In the simulation study in Section III, we will assess whether this loss of power/efficiency occurs in the rare event setting. The specific form of the Cox PH regression model is,

$$\lambda(T_{si}, \delta_{si}|X_{si}, \boldsymbol{Z}_{si}) = \lambda_{s0}(T_{si}) exp\left[\beta_x^{(Strat)} X_{si} + \boldsymbol{\beta}_z^{(Strat)} \boldsymbol{Z}_{si}\right], \text{ for } s=1,\ldots,S. \quad (3)$$

In Section II.B.2 we will propose a site-stratified approach tailored to the distributed data setting. In the simulation study we will compare the standard site-stratified Cox PH model to the new distributed data setting method.

## B. EXTENSIONS TO THE DISTRIBUTED DATA SETTING

There are several approaches to extend Cox PH regression methods to the distributed data setting. We will first discuss a method for de-identifying individual level data that uses standard Cox PH regression. Then we will discuss how to instead conduct Mantel-Haenszel[4] type test statistics using site-specific regression models that may be more appropriate when site heterogeneity is expected.

### 1. Standard Cox PH regression with de-indentified data

Assume that one can deidentify data by categorizing all confounders and time. Then, de-identified, categorized and data are shared from different sites to fit models described in previous section. Specifically, we will categorize continuous confounders (e.g., age is categorized to age1=35-39, age2=40-44, ...). We can do something similar with time, but the statistical implications may be more complicated. We will propose initially to discretize time into categories such as week. Implications of the categorization of time will be both outcome ties and interval censoring. We account for outcome ties in estimating $\beta_X$ using Effron's method. Due to the rare event setting, we expect very few outcomes, so the implications of ties may be limited. For more common event settings, we would not advise to categorize time since having more ties increases estimate variability. For simplicity, we will initially also ignore implications of interval censoring unless we face problems in reserving statistical properties such as type I error and bias. We expect not taking interval censoring into account may lead to issues with variance estimation (type I error inflation), but fewer issues with bias.

### 2. Mantel-Haenszel type test statistic in distributed data setting

To limit data transmission, an alternative to categorizing all confounders and time is for each site to run a site-specific model and to transmit centrally only summary statistics. Specifically, one could fit the site-specific Cox Regression model (1) at each analysis time *a*. From this model one can summarize the findings from site *s* using either the adjusted log hazard ratio, $\hat{\beta}_X^s$, or site specific log rank test statistic, $\widehat{LR}$. To create a single stratified estimate across sites, apply standard stratified or meta-analysis estimation of the form,

$$\hat{\theta} = \frac{\sum_{s=1}^{S} w_s \hat{\theta}^s}{\sum_{s=1}^{S} w_s} \qquad (4)$$

where $w_s$ is either sample size at site *s* up to analysis *a* or the reciprocal of the variance of $\hat{\theta}^s$ and $\hat{\theta}^s$ is the test statistic of interest (e.g. $\widehat{LR}^s$). The estimated variance of $\hat{\theta}$ is

$$\hat{V}(\hat{\theta}) = \frac{\sum_{s=1}^{S} w_s^2 \hat{V}(\hat{\theta}^s)}{\left[\sum_{s=1}^{S} w_s\right]^2}.$$

Note that when the test statistic is the LR, then the $\hat{V}(\widehat{LR}^s) = 1$. The advantage of using stratified estimation is that it handles effect modification due to site. Sites often have different prescribing and coding patterns and therefore differential site effects are highly likely. The disadvantage is that site-

specific regression models are less statistically efficient and less stable, especially in the rare event setting. We will explore the loss of power in the rare event setting in the simulations (Section III). Note that there are other methods available to combine the estimates across sites, such as random effects. For estimation purposes and computational simplicity to develop exact boundaries as described in Section II.C.2, we used this fixed effect approach to combine information.

## C. EXTENDING TO THE GROUP SEQUENTIAL TESTING SETTING

We will now discuss how to incorporate active surveillance in which we will sequentially monitor over time for association of exposure with elevated risk of a given outcome. We will first present a method that develops boundaries assuming normal approximation theory and then discuss a method more tailored to rare outcomes.

### 1. Group Sequential Cox's PH using Lan Demets Normal Approximation Boundary Approach

The normal approximation boundary approach computes group sequential boundaries on the error spending scale originally developed by Lan and Demets[5] for randomized clinical trials. Error spending uses the concept of cumulative alpha or type I error, $\alpha(a)$, defined as the cumulative amount of type I error spent up to analysis a (a = 1,…, A). We assume that $0 < \alpha(1) \leq \cdots \leq \alpha(A) = \alpha$, where $\alpha$ is the overall type I error specified to be spent across the study period. There are several commonly used models for $\alpha(a)$ including the Pocock boundary function[6] $\alpha(a) = \log\left(1 + (\exp(1) - 1)\, n_a/n_A\right)\alpha$, O'Brien-Fleming boundary function[7] $\alpha(a) = 2\left(1 - \varphi\left(\frac{Z_{1-\alpha}}{\sqrt{n_a/n_A}}\right)\right)$, and the general power boundary function $\alpha(a) = \left(n_a/n_A\right)^p \alpha$ $p > 0$. For safety evaluations a flat, Pocock-like boundary on a standardized test statistic scale has often been used.[8]

Given a specified error spending function, Lan and Demets[5] developed a conditional sequential monitoring boundary, referred to here as GS LD, for any asymptotically normal standardized test statistic, $W_{LD}(a) \sim N(0,1)$ as $n_a \to \infty$, based on independent increments of data. This boundary can be computed and used to compare to any standardized test statistic that is asymptotically normal, including one that controls for confounding. For this task order we will use the Log Rank test statistic and Mantel-Haenzel stratified test statistics as outlined in previous sections.

Error-spending 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 (which produces small amounts of new information between analyses) the asymptotic properties of the boundary may fail to hold. The following section will describe a more recent method to address the shortcomings of this approach for the rare event setting.

### 2. Group Sequential Cox's PH using Exact Boundary Calculations

We will now describe a method that is an extension of the Group Sequential GEE method[9] using a Cox's PH model adjusting for confounders with a Log Rank test statistic. For our boundary formulation we will modify a well-established simulation approach initially proposed by Wang and Tsiatis[10] and extended in the context of an unifying family of boundaries by Kittelson and Emerson.[11] This approach allows for the application of a wide range of commonly used boundary shapes including the Pocock-like boundary[6] and

the O'Brien and Fleming-like boundary.[7] Specifically, the boundary is defined as $b(a) = \omega u(a)$ where $u(a)$ is specified as a function dependent on *a* and comes from the unifying boundary family (specifically for Pocock-like $u(a) = 1$ and O'Brien and Fleming-like $u(a) = \sqrt{n_A/n_a}$ and $\omega$ is solved iteratively by permuting the data under $H_O$ to hold the type I error at $\alpha$.

To form a boundary it is necessary to define a test statistic, the variability of the test statistic over time, the shape of the boundary, the number of analysis times, $\alpha$-level (type I error), and either end of study sample size or overall power. We first assume that the end of study sample size with number of observations per analysis time is known, and we allow power to vary. For observational studies to determine the variability of the test statistic over time, one must also assume the distribution at each analysis time of all variables in the model including outcome, exposure and all confounders. We then discuss how to alter this boundary selection process to incorporate earlier non-pre-specified analysis times, variable number of observations $n_a$ per analysis time, and unknown future distributions of the outcome, exposure, and confounders.

To accommodate rare events, we propose to use a permutation approach for boundary formation which has advantageous non-parametric assumptions. Under the null that $\beta_X(a) = 0$ for all *a*, outcome given confounders, $T_{si}(a), \delta_{si}(a)|\boldsymbol{Z}_{si}$, is independent of exposure $X_{si}$. Note that now we define , $T_{si}(a)$, as the time to event or censoring where censoring occurs at analysis time *a* instead of end of the entire study period. Similarly the event, $\delta_{si}(a)$, must occur before the end of analysis time *a*. Therefore, we can permute observed exposures, $\boldsymbol{X}_s$, within site *s* while fixing the observed set of outcomes and confounders ($\boldsymbol{T}_s, \delta_s, \boldsymbol{Z}_s$). Since we are analyzing data at times *a* = 1, … , *A*, and in practice the variability in the proportion exposed may directly affect the variability of the test statistic, it is important to permute *X* within analysis time *a*. To do this, we assume that the data are ordered by time of entry into study such that for analysis at time *a* the new data observed at analysis time *a* since *a*-1 is indexed by $\{n_{s,a-1} + 1\}$ to $\{n_{s,a}\}$and for the first analysis time has index {1} to $\{n_{s,1}\}$. Given this ordering of the data the permutation approach proceeds as follows:

**Permutation Data Approach (at end of study):**

Step 1: Within each analysis time *a* and site *s* create permuted exposure data $\boldsymbol{X}_s^{(j)}$ $(j = 1, \dots, N_p)$, by permuting observed exposures, $(X_{n_{s,a-1}+1}, \dots, X_{n_{s,a}})$, to form $\boldsymbol{X}_s^{(j)} = \left(X_{n_{s,a-1}+1}^{(j)}, \dots, X_{n_{s,a}}^{(j)}\right)$.

Step 2a (De-identified Cox PH analysis): Calculate the permuted test statistic $\theta^{(j)}(a) = LR^{(j)}(a)$ on the permuted exposure data, $\boldsymbol{X}_s^{(j)}$, and observed outcome and confounder data.

Step 2b (Mantel-Haenszel analysis): Calculate the site-specific test statistic $\theta_s^{(j)}(a)$ (e.g. $\theta_s^{(j)}(a) = LR_s(a)$) on the sites permuted exposure data and observed outcome and confounder data. Then centrally combine the site specific test statistics to form a stratified test statistic $\theta^{(j)}(a) = \frac{\sum_{s=1}^{S} w_s \theta_s^{(j)}(a)}{\sum_{s=1}^{S} w_s}$.

Step 3: If $\theta^{(j)}(a) \geq \omega * u(a)$ then denote $C_j = 1$ (e.g. permuted data *j* has crossed boundary) and stop, otherwise continue to next *a* + 1.

If *a* = *A* then $C_j$ = 0 indicating permuted data set *j* did not signal.

This process is repeated a large number, $N_p$, times ($j$ = 1, … , $N_p$). Then the estimated $\alpha$-level for the boundary is calculated as $\hat{\alpha} = \sum_{j=1}^{N_p} {C_j}/{N_p}$. Repeat the simulation steps 1-2 until $\hat{\alpha} = \alpha$ by decreasing, or increasing, the value of $\omega$ in the boundary calculation $\omega * u(a)$.

This simulation framework requires that we have a complete dataset, ($\boldsymbol{X}_s, \boldsymbol{T}_s, \delta_s, \boldsymbol{Z}_s$), for all observations $i$ = 1, … , $n_A$. However, this is not practical at earlier analysis times $a < A$. To solve this, at times $a < A$, we can instead make assumptions about how the data will look at future analysis times. Specifically, we will assume that future data will look like the current outcome, exposure, and confounder data. To approximate the future distributions of $\boldsymbol{T}, \delta, \boldsymbol{Z}$, and $\boldsymbol{X}$, we can sample the future observations, $n_a$+1 to $n_A$, by sampling with replacement from the observed ($X_{si}, T_{si}, \delta_{si}, \boldsymbol{Z}_s$) ($i$ = 1, … , $n_a$). This will create the complete dataset necessary to perform the permutation approach described previously for all analysis times. Note that this is a conservative assumption since we will assume only new individuals are entering the population and not that for our current population we will observe more exposure information. Therefore we will likely have higher boundaries in earlier analyses then what will be necessary to preserve the overall type I error. We will eventually use this preserved alpha in future analyses as we observe more information and boundaries are updated.

In practice, at each new analysis time, we keep the prior critical values c(1), … , c($a$-1) since these were the signaling thresholds used at previous analysis times and each analysis is defined to be conditional on the prior analyses. Using these values, we will then solve for the current analysis time critical value c(a) using the newly updated observed information. These new data may have a different distribution of outcomes, exposures, and confounders compared to what we had assumed during previous analysis times, and the sample sizes may be different than initially planned. Thus it is important to note our method allows for both different than expected outcome, exposure, and confounder distributions along with sample size at a given analysis time. However, changes in the distribution affect the variability of the estimator and therefore the corresponding signaling threshold c($a$), which is defined to maintain $\alpha(A) = \alpha$. Therefore, at each analysis time we will update the boundary to maintain the original boundary family, but moving it slightly compared to the initially planned threshold in order to keep the overall type I error constant as the variability of the data changes over time.

## III. SIMULATION STUDY

The following simulation study was conducted to evaluate the operating characteristics of the new distributed methods compared to standard Cox PH models. The purpose of this simulation was to compare the gold standard methods in which individual-level data could be pooled across Data Partners to conduct either standard Cox PH models adjusting for site or site stratified Cox PH models, to the new distributed data methods that attempt to either de-identify data through aggregation or use stratified analysis approaches in which only summary data is shared across sites. It is important to understand the operating characteristics of proposed estimators in order to effectively evaluate quantities such as the false positive rate (type I error), study power, and the average time to detection of a true signal.

Section III.A below details the different scenarios evaluated. In the simulation evaluations, we vary the two-year incidence of outcome in the unexposed portion of the population from 1% to 5%, the

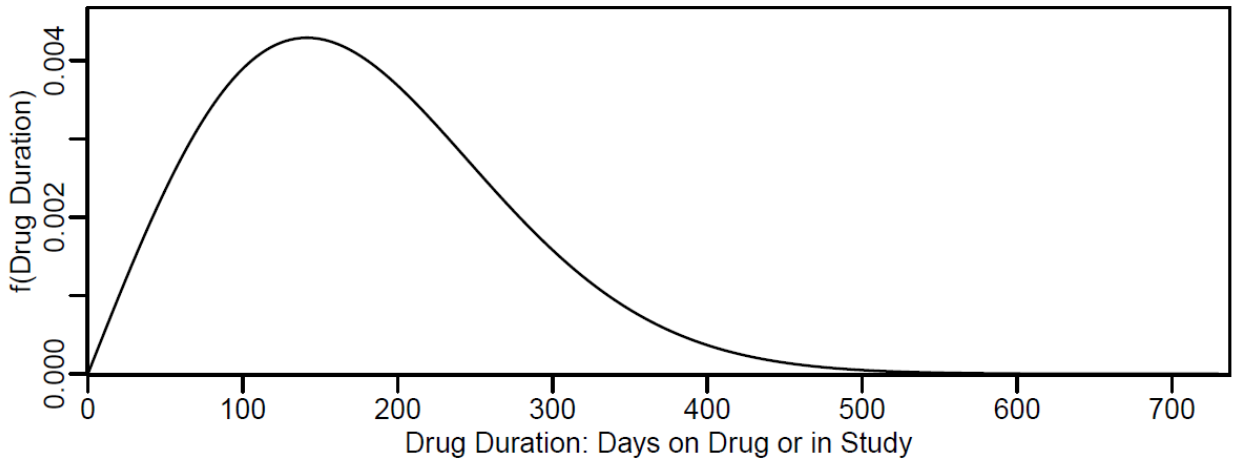
proportion of exposed individuals in the study population from 10% to 50%, and the strength of age and site confounders. The total end of two-year sample size is fixed at 10,000 across simulations. All simulation results are based on 500 replicate datasets and were conducted using R version 3.0.2.

## A. DATA STRUCTURE

Below is the specific step-wise simulation design for creating a dataset of *N* study participants for (*i=1,…,N*);

1. Start date, $D_i$, is the time at which individual *i* enters the study, and this is uniformly distributed throughout the two-year (720 day) study period, $D_i$ ~ Discrete Uniform(1,719);
2. Site distribution, $S_i$ is three equal-sized sites. The variable $S_i$ is generated from the fixed distribution of equal-sized sites, and then the two corresponding binary (dummy) variables, $S_{i1}$ and $S_{i2}$, are generated for use in regressions and calculations using design matrices.
3. Confounder distributions : The simulations performed for this study included a single continuous confounder (age), $Z_i$, which is distributed as Uniform(35,65) and then centered at 50 and scaled so that a one-unit change is equivalent to 10 years. For aggregated methods, $Z_i$, is categorized into 5 binary (dummy) variables for 5-year age increments.
4. Exposure distribution conditional on confounders:
   $log(X_{si}|Z_{si}, S_{1si}, S_{2si}) = \exp(\beta_{x,0} + \beta_{x,z}Z_{si} + \beta_{x,s1}S_{1si} + \beta_{x,s2}S_{2si})$
   where for each coefficient other than $\beta_{x,0}$, exp($\beta$) is the relative risk of exposure associated with a one unit change in that particular variable holding the other variables fixed. We varied the strength of the exposure/confounder association to be a relative risk of 1.3 and 1.5. For each simulation we solved for $\beta_{x,0}$ so that the overall probability of *X* was fixed at either 50% or 10% across all simulation configurations.
5. Overall outcome distribution conditional on exposure and confounders: To define the survival outcome, we needed to specify the censoring and time to event distributions.
   a. First, we specified that the censoring distribution is $C_{si} \sim Weibull(2,200)$ which is independent of covariates. This translates to an average censoring time of 177 days (SD 93 days) and is depicted below in Figure 1. This would depict a medical product that is typically taken for a longer duration.

**Figure 1. Censoring Distribution of Weibull(2, 200).**



b. Second, we specified the time-to-event distribution. For this simulation study, we assumed that:
$E_{si} \sim exp(\beta_{y,0} + \beta_{y,x}X_{si} + \beta_{y,z}Z_{si} + \beta_{y,s1}S_{1si} + \beta_{y,s2}S_{2si})$
where for each coefficient other than $\beta_{y,0}$, $\exp(\beta_{y,.})$, is the relative risk of the outcome associated with a one unit change in that particular variable holding the other variables fixed. We varied the strength of outcome and age relationship to be 1.3 and 1.5 and the relationship between outcome and sites to be (1.2, 1.2) and (1.2, 0.8). We varied the association between outcome and exposure of interest to be 1, 1.5, and 2. For each simulation we solved for $\beta_{y,0}$ so that the overall probability of the outcome at the last analysis *A*, $\delta_{si}(A) = I(E_{si} < C_{si} \cap E_{si} < t_A - D_{si})$, was fixed at either 1% or 5% across all simulation configurations.

6. Sequential Monitoring Outcome Definition: At analysis *a* (*a=1, …, A*) occurring at time $t_a$, we use the subset of data with start date before analysis time ($D_{si} < t_a$). Given the time-to-event, $E_{si}$ , and censoring, $C_{si}$, distributions described previously across the entire study period, we define the outcomes and exposure duration at a given analysis time as:
    a. Individual-level data:
        i. Time to event or censoring at analysis time *a*: $T_{si}(a) = \min(t_a - D_{si}, C_{si}, E_{si})$
        ii. Outcome indicator at analysis time *a*: $\delta_{si}(a) = I(E_{si} < C_{si} \cap E_{si} < t_a - D_{si})$
    b. De-identified time-aggregated data
        i. Time to event or censoring at analysis time *a*: $T^c_{si}(a)$is defined as categorizing $T_{si}(a)$ into 7-day bins.
        ii. Outcome indicator at analysis time *a*: $\delta^c_{si}(a) = I(E_{si} < C_{si} \cap E_{si} < T^c_{si}(a))$
7. Frequency of Testing and Boundary Shape: Through all simulations we assumed a two-year study with a 6 month lag and 6 quarterly looks thereafter. We used a Pocock or Pocock-like boundary throughout evaluations since this is what has been used most commonly in safety surveillance (add refs).

## B. RESULTS

We first present the results for the scenario with the more common outcome incidence of 0.05. In this scenario, we expect the new distributed methods and standard Cox PH models to perform similarly, and we expect not to see issues with using the normal approximation Lan Demets boundary. The second set of simulations assesses the performance of the methods in the less common outcome incidence scenario of 0.01, in which performance was expected to have some issues for the distributed methods and for the normal approximation boundary formation.

### 1. More Common Outcome Frequency of 0.05

Table 1, Table 2, and Table 3 show the results for the more common outcome scenario using Lan Demets boundary formation. In Table 1, we evaluated whether the methods appropriately held the type I error if we simulated data with no association between the exposure and outcome of interest, when varying the strength of confounding and the prevalence of exposure. If the method is performing appropriately, we would expect the type I error to be held at 0.05. As shown in Table 1, there is no strong indication that any of the methods are not holding the type I error in this more common outcome scenario and therefore all methods are valid.

In Table 2 and Table 3 we assess the performance of the methods when there is an association between the exposure and outcome of interest with RR=1.5. Table 2 evaluates power and Table 3 shows time to signal detection or end of study without a signal. As shown in Table 2, as exposure prevalence increases, power increases as expected. Looking across methods, there does not seem to be any indication that power decreases as we aggregate data (confounders, follow-up time, or both) or conduct stratified Mantel-Haenszel type estimates. Further, in Table 3 there is no indication of different time to study end across methods. Therefore, both methods for de-identified data proposed in this write-up perform well for the more common outcome case with few confounders.

## 2. Rare Outcome Frequency of 0.01

Table 4, Table 5, and Table 6 show the results for the less common outcome scenario using Lan Demets boundary formation. Table 4 shows the results for assessing type I error when there is no relationship between outcome and exposure of interest. Tables 5 and 6 show the results assessing study power and time to signal detection for an association between exposure and outcome with RR=2. In the scenario with 0.50 exposure prevalence (the bottom section of Table 4), type I error seems to be slightly lower than 0.05 across all strength of confounding scenarios, but there is no strong differences across methods (based on results across columns). In this high exposure prevalence scenario, there are no strong differences across methods in terms of power (Table 5, bottom half). However, when comparing time to study end (Table 6, bottom half) there is some indication that methods using non-site stratified estimates have slightly faster time to study end. Specifically, for the first row of the prevalence of exposure=0.50 section, the time to study end of the Cox method with individual-level continuous data is 368 days compared to 391 using the site-stratified Cox method with individual-level continuous data. Aggregating, but without a stratified estimate, yields the same result (mean time to study end of 368) as the individual-level continuous estimate, and the stratified estimate on the aggregated data is 392 days to study end similar to the stratified Cox estimate. Mantel-Haenszel site stratified results are consistently in between the Cox site stratified and Cox with site adjustment results. Therefore, for this scenario, the Cox aggregated method for confounder adjustment may be preferable.

Assessing the low exposure prevalence scenario of 0.10, there is an indication of large issues with type I error (Table 4, top half). Type I error is consistently elevated especially for the Mantel-Haenszel type estimator, but is even high in the gold standard case using individual level data with continuous time and confounders. Therefore, this may indicate issues with using a normal approximation boundary formation in the rare outcome and rare exposure setting.

We then conducted a brief simulation study using exact boundaries to see if the type I error was properly held for the rare outcome case in which the normal approximation boundary methods did not perform well. Table 7 displays the type I error results from this evaluation showing that type I error is held much closer to 0.05 level, but in a few circumstances there is still elevated type I error for the new proposed distributed methods relative to standard Cox PH methods with continuous time and confounders. However, as exposure prevalence becomes more common all methods had appropriate type I error when using exact boundaries and were closer to 0.05 compared to the normal approximation methods. Table 8 and Table 9 show power and time to study end using the exact methods and results show some slight power advantages using Aggregated time and confounders relative to Mantel Haenszel type estimate but the advantages are modest.

**Table 1. Type I error for two site confounders and age using Lan Demets Boundary Formation, for outcome frequency=0.05 and no association between exposure and outcome**

| RR (Y\|X, Z, and Site) | | | | RR (X\|Z and Site) | | | TYPE I ERROR | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Continuous T Continuous Z | | Aggregate T Continuous Z | Continuous T Aggregate Z | Aggregate T Aggregate Z | | Distributed |
| X | Z | Site 1 | Site 2 | Z | Site 1 | Site 2 | Cox | Strat. Cox | Cox | Cox | Cox | Strat. Cox | MH Strat. |
| **Prevalence of Exposure 0.10** | | | | | | | | | | | | | |
| 1 | 1.3 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.054 | 0.048 | 0.054 | 0.052 | 0.054 | 0.056 | 0.054 |
| 1 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.058 | 0.060 | 0.058 | 0.062 | 0.060 | 0.058 | 0.066 |
| 1 | 1.3 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.054 | 0.054 | 0.052 | 0.056 | 0.056 | 0.054 | 0.058 |
| 1 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.050 | 0.062 | 0.050 | 0.052 | 0.052 | 0.062 | 0.058 |
| 1 | 1.5 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.062 | 0.056 | 0.058 | 0.058 | 0.058 | 0.052 | 0.058 |
| 1 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.052 | 0.054 | 0.050 | 0.050 | 0.050 | 0.052 | 0.052 |
| 1 | 1.5 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.048 | 0.060 | 0.048 | 0.048 | 0.048 | 0.062 | **0.070** |
| 1 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.052 | 0.056 | 0.052 | 0.052 | 0.052 | 0.062 | 0.058 |
| **Prevalence of Exposure 0.50** | | | | | | | | | | | | | |
| 1 | 1.3 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.054 | 0.056 | 0.054 | 0.056 | 0.054 | 0.056 | 0.056 |
| 1 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.040 | 0.042 | 0.040 | 0.040 | 0.038 | 0.042 | 0.042 |
| 1 | 1.3 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.042 | 0.042 | 0.042 | 0.038 | 0.038 | 0.042 | 0.042 |
| 1 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.048 | 0.054 | 0.044 | 0.046 | 0.046 | 0.054 | 0.052 |
| 1 | 1.5 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.048 | 0.050 | 0.048 | 0.050 | 0.050 | 0.052 | 0.052 |
| 1 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.056 | 0.060 | 0.056 | 0.062 | 0.060 | 0.058 | 0.054 |
| 1 | 1.5 | 1.2 | 1.2 | 1.5 | 2 | 2 | **0.024** | **0.028** | **0.024** | **0.024** | **0.024** | 0.036 | **0.030** |
| 1 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.036 | 0.036 | 0.034 | 0.036 | 0.036 | 0.040 | 0.036 |

*Bold indicates Type I error 0.05 +/- 0.02

Continuous T = Continuous Time (no aggregation), Continuous Z = Continuous Age (no 5 year age categorization), Aggregate T = De-identify data by creating one-week study time categories, Aggregate Z = Deidentify data by creating 5-year age categorization, Distributed = Only send summary statistics across sites instead of raw data, Cox = Cox PH not stratified by site, Strat. Cox = Cox PH stratified by site, and MH Strat. = Send separate site adjusted Cox PH model LR test statistics and variance to a central location and calculate a Mantel-Haenszel type estimate.

**Table 2. Power for two site confounders and age using Lan Demets Boundary Formation, for outcome frequency=0.05 and RR=1.5 for association between exposure and outcome**

| RR (Y\|X, Z, and Site) | | | | RR (X\|Z and Site) | | | POWER | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Continuous T Continuous Z | | Aggregate T Continuous Z | Continuous T Aggregate Z | Aggregate T Aggregate Z | | Distributed |
| X | Z | Site 1 | Site 2 | Z | Site 1 | Site 2 | Cox | Strat. Cox | Cox | Cox | Cox | Strat. Cox | MH Strat. |
| **Prevalence of Exposure 0.10** | | | | | | | | | | | | | |
| 1.5 | 1.3 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.874 | 0.856 | 0.872 | 0.882 | 0.878 | 0.864 | 0.856 |
| 1.5 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.860 | 0.874 | 0.858 | 0.866 | 0.868 | 0.874 | 0.866 |
| 1.5 | 1.3 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.870 | 0.860 | 0.872 | 0.872 | 0.872 | 0.870 | 0.862 |
| 1.5 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.870 | 0.870 | 0.870 | 0.872 | 0.872 | 0.882 | 0.866 |
| 1.5 | 1.5 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.860 | 0.856 | 0.858 | 0.858 | 0.858 | 0.856 | 0.856 |
| 1.5 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.876 | 0.878 | 0.872 | 0.878 | 0.876 | 0.878 | 0.882 |
| 1.5 | 1.5 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.866 | 0.852 | 0.864 | 0.870 | 0.870 | 0.852 | 0.852 |
| 1.5 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.894 | 0.884 | 0.894 | 0.900 | 0.900 | 0.892 | 0.892 |
| **Prevalence of Exposure 0.50** | | | | | | | | | | | | | |
| 1.5 | 1.3 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.984 | 0.988 | 0.984 | 0.986 | 0.986 | 0.988 | 0.988 |
| 1.5 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 | 0.994 |
| 1.5 | 1.3 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.982 | 0.982 | 0.982 | 0.982 | 0.982 | 0.982 | 0.982 |
| 1.5 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.992 | 0.992 | 0.992 | 0.994 | 0.994 | 0.992 | 0.988 |
| 1.5 | 1.5 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.992 | 0.992 | 0.992 | 0.992 | 0.992 | 0.992 | 0.992 |
| 1.5 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 |
| 1.5 | 1.5 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 | 0.990 |
| 1.5 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.990 | 0.990 | 0.990 | 0.992 | 0.992 | 0.992 | 0.990 |

Continuous T = Continuous Time (no aggregation), Continuous Z = Continuous Age (no 5 year age categorization), Aggregate T = De-identify data by creating one-week study time categories, Aggregate Z = Deidentify data by creating 5-year age categorization, Distributed = Only send summary statistics across sites instead of raw data, Cox = Cox PH not stratified by site, Strat. Cox = Cox PH stratified by site, and MH Strat. = Send separate site adjusted Cox PH model LR test statistics and variance to a central location and calculate a Mantel Haenszel type estimate.

**Table 3. Time to Study End (in days) for two site confounders and age using Lan Demets Boundary Formation, for outcome frequency=0.05 and RR=1.5 for association between exposure and outcome**

| RR (Y\|X, Z, and Site) | | | | RR (X\|Z and Site) | | | TIME TO STUDY END | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Continuous T Continuous Z | | Aggregate T Continuous Z | Continuous T Aggregate Z | Aggregate T Aggregate Z | | Distributed |
| X | Z | Site 1 | Site 2 | Z | Site 1 | Site 2 | Cox | Strat. Cox | Cox | Cox | Cox | Strat. Cox | MH Strat. |
| **Prevalence of Exposure 0.10** | | | | | | | | | | | | | |
| 1.5 | 1.3 | 1.2 | 1.2 | 1.2 | 2 | 2 | 380 | 382 | 381 | 379 | 382 | 384 | 382 |
| 1.5 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 381 | 378 | 382 | 379 | 378 | 376 | 377 |
| 1.5 | 1.3 | 1.2 | 1.2 | 1.5 | 2 | 2 | 384 | 388 | 385 | 383 | 383 | 385 | 387 |
| 1.5 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 388 | 389 | 388 | 388 | 388 | 386 | 391 |
| 1.5 | 1.5 | 1.2 | 1.2 | 1.2 | 2 | 2 | 386 | 382 | 388 | 388 | 389 | 382 | 379 |
| 1.5 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 377 | 380 | 378 | 378 | 379 | 380 | 374 |
| 1.5 | 1.5 | 1.2 | 1.2 | 1.5 | 2 | 2 | 394 | 398 | 395 | 392 | 393 | 397 | 398 |
| 1.5 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 373 | 371 | 373 | 368 | 368 | 368 | 371 |
| **Prevalence of Exposure 0.50** | | | | | | | | | | | | | |
| 1.5 | 1.3 | 1.2 | 1.2 | 1.2 | 2 | 2 | 289 | 285 | 289 | 288 | 289 | 285 | 284 |
| 1.5 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 267 | 265 | 267 | 265 | 265 | 263 | 264 |
| 1.5 | 1.3 | 1.2 | 1.2 | 1.5 | 2 | 2 | 280 | 277 | 280 | 278 | 279 | 275 | 277 |
| 1.5 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 275 | 275 | 275 | 273 | 273 | 273 | 273 |
| 1.5 | 1.5 | 1.2 | 1.2 | 1.2 | 2 | 2 | 269 | 268 | 269 | 269 | 269 | 268 | 268 |
| 1.5 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 281 | 280 | 282 | 280 | 280 | 281 | 281 |
| 1.5 | 1.5 | 1.2 | 1.2 | 1.5 | 2 | 2 | 289 | 285 | 289 | 286 | 287 | 284 | 285 |
| 1.5 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 284 | 281 | 284 | 279 | 280 | 280 | 282 |

Continuous T = Continuous Time (no aggregation), Continuous Z = Continuous Age (no 5 year age categorization), Aggregate T = De-identify data by creating one-week study time categories, Aggregate Z = Deidentify data by creating 5-year age categorization, Distributed = Only send summary statistics across sites instead of raw data, Cox = Cox PH not stratified by site, Strat. Cox = Cox PH stratified by site, and MH Strat. = Send separate site adjusted Cox PH model LR test statistics and variance to a central location and calculate a Mantel Haenszel type estimate.

**Table 4. Type I error for two site confounders and age using Lan Demets Boundary Formation, for outcome frequency=0.01 and no association between exposure and outcome**

| RR (Y\|X, Z, and Site) | | | | RR (X\|Z and Site) | | | TYPE I ERROR | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Continuous T Continuous Z | | Aggregate T Continuous Z | Continuous T Aggregate Z | Aggregate T Aggregate Z | | Distributed |
| X | Z | Site 1 | Site 2 | Z | Site 1 | Site 2 | Cox | Strat. Cox | Cox | Cox | Cox | Strat. Cox | MH Strat. |
| **Prevalence of Exposure 0.10** | | | | | | | | | | | | | |
| 1 | 1.3 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.068 | **0.082** | 0.066 | 0.066 | 0.064 | **0.084** | **0.130** |
| 1 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | **0.090** | **0.084** | **0.090** | **0.092** | **0.092** | **0.084** | **0.148** |
| 1 | 1.3 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.062 | **0.076** | 0.062 | 0.064 | 0.066 | **0.076** | **0.122** |
| 1 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | **0.092** | **0.088** | **0.092** | **0.090** | **0.092** | **0.090** | **0.140** |
| 1 | 1.5 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.058 | 0.066 | 0.058 | 0.066 | 0.066 | 0.068 | **0.114** |
| 1 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | **0.088** | **0.114** | **0.090** | **0.088** | **0.088** | **0.112** | **0.142** |
| 1 | 1.5 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.062 | 0.062 | 0.062 | 0.058 | 0.058 | 0.066 | **0.114** |
| 1 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | **0.074** | **0.078** | **0.074** | **0.084** | **0.082** | **0.082** | **0.126** |
| **Prevalence of Exposure 0.50** | | | | | | | | | | | | | |
| 1 | 1.3 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.044 | 0.040 | 0.044 | 0.042 | 0.042 | 0.040 | 0.048 |
| 1 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.038 | 0.032 | 0.038 | 0.040 | 0.042 | 0.038 | 0.034 |
| 1 | 1.3 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.040 | 0.034 | 0.040 | 0.040 | 0.040 | 0.034 | 0.036 |
| 1 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.032 | **0.028** | 0.032 | **0.030** | **0.030** | **0.028** | 0.034 |
| 1 | 1.5 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.038 | 0.036 | 0.040 | 0.038 | 0.038 | 0.036 | 0.038 |
| 1 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.036 | **0.028** | 0.036 | 0.038 | 0.036 | **0.028** | **0.026** |
| 1 | 1.5 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.048 | 0.048 | 0.048 | 0.050 | 0.052 | 0.046 | 0.054 |
| 1 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.040 | 0.042 | 0.038 | 0.038 | 0.038 | 0.040 | 0.054 |

*Bold indicates Type I error 0.05 +/- 0.02

Continuous T = Continuous Time (no aggregation), Continuous Z = Continuous Age (no 5 year age categorization), Aggregate T = De-identify data by creating one-week study time categories, Aggregate Z = Deidentify data by creating 5-year age categorization, Distributed = Only send summary statistics across sites instead of raw data, Cox = Cox PH not stratified by site, Strat. Cox = Cox PH stratified by site, and MH Strat. = Send separate site adjusted Cox PH model LR test statistics and variance to a central location and calculate a Mantel Haenszel type estimate.

**Table 5. Power for two site confounders and age using Lan Demets Boundary Formation, for outcome frequency=0.01 and RR=2 for association between exposure and outcome**

| RR (Y\|X, Z, and Site) | | | | RR (X\|Z and Site) | | | POWER | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Continuous T Continuous Z | | Aggregate T Continuous Z | Continuous T Aggregate Z | Aggregate T Aggregate Z | | Distributed |
| X | Z | Site 1 | Site 2 | Z | Site 1 | Site 2 | Cox | Strat. Cox | Cox | Cox | Cox | Strat. Cox | MH Strat. |
| **Prevalence of Exposure 0.10** | | | | | | | | | | | | | |
| 2 | 1.3 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.730 | 0.734 | 0.730 | 0.734 | 0.736 | 0.736 | 0.772 |
| 2 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.752 | 0.762 | 0.752 | 0.752 | 0.752 | 0.764 | 0.776 |
| 2 | 1.3 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.770 | 0.778 | 0.770 | 0.766 | 0.768 | 0.786 | 0.808 |
| 2 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.738 | 0.744 | 0.736 | 0.742 | 0.742 | 0.746 | 0.760 |
| 2 | 1.5 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.782 | 0.774 | 0.782 | 0.782 | 0.782 | 0.772 | 0.792 |
| 2 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.762 | 0.766 | 0.758 | 0.768 | 0.766 | 0.770 | 0.796 |
| 2 | 1.5 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.804 | 0.800 | 0.804 | 0.806 | 0.806 | 0.802 | 0.834 |
| 2 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.766 | 0.764 | 0.766 | 0.774 | 0.774 | 0.766 | 0.794 |
| **Prevalence of Exposure 0.50** | | | | | | | | | | | | | |
| 2 | 1.3 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.918 | 0.926 | 0.918 | 0.918 | 0.918 | 0.928 | 0.924 |
| 2 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.894 | 0.880 | 0.892 | 0.894 | 0.894 | 0.880 | 0.880 |
| 2 | 1.3 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.898 | 0.902 | 0.898 | 0.902 | 0.902 | 0.904 | 0.900 |
| 2 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.898 | 0.892 | 0.898 | 0.898 | 0.898 | 0.890 | 0.890 |
| 2 | 1.5 | 1.2 | 1.2 | 1.2 | 2 | 2 | 0.878 | 0.884 | 0.878 | 0.882 | 0.882 | 0.892 | 0.892 |
| 2 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.918 | 0.926 | 0.918 | 0.918 | 0.918 | 0.926 | 0.924 |
| 2 | 1.5 | 1.2 | 1.2 | 1.5 | 2 | 2 | 0.876 | 0.874 | 0.876 | 0.880 | 0.880 | 0.880 | 0.876 |
| 2 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.894 | 0.892 | 0.894 | 0.896 | 0.896 | 0.894 | 0.890 |

Continuous T = Continuous Time (no aggregation), Continuous Z = Continuous Age (no 5 year age categorization), Aggregate T = De-identify data by creating one-week study time categories, Aggregate Z = Deidentify data by creating 5-year age categorization, Distributed = Only send summary statistics across sites instead of raw data, Cox = Cox PH not stratified by site, Strat. Cox = Cox PH stratified by site, and MH Strat. = Send separate site adjusted Cox PH model LR test statistics and variance to a central location and calculate a Mantel Haenszel type estimate.

**Table 6. Time to study end (in days) for two site confounders and age using Lan Demets Boundary Formation, for outcome frequency=0.01 and RR=2 for association between exposure and outcome**

| RR (Y\|X, Z, and Site) | | | | RR (X\|Z and Site) | | | TIME TO STUDY END | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Continuous T Continuous Z | | Aggregate T Continuous Z | Continuous T Aggregate Z | Aggregate T Aggregate Z | | Distributed |
| X | Z | Site 1 | Site 2 | Z | Site 1 | Site 2 | Cox | Strat. Cox | Cox | Cox | Cox | Strat. Cox | MH Strat. |
| **Prevalence of Exposure 0.10** | | | | | | | | | | | | | |
| 2 | 1.3 | 1.2 | 1.2 | 1.2 | 2 | 2 | 447 | 444 | 447 | 445 | 445 | 444 | 407 |
| 2 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 437 | 437 | 436 | 437 | 437 | 437 | 408 |
| 2 | 1.3 | 1.2 | 1.2 | 1.5 | 2 | 2 | 437 | 439 | 437 | 436 | 436 | 438 | 407 |
| 2 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 439 | 440 | 439 | 436 | 437 | 440 | 413 |
| 2 | 1.5 | 1.2 | 1.2 | 1.2 | 2 | 2 | 435 | 435 | 435 | 435 | 435 | 435 | 400 |
| 2 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 434 | 437 | 434 | 432 | 433 | 437 | 400 |
| 2 | 1.5 | 1.2 | 1.2 | 1.5 | 2 | 2 | 412 | 416 | 413 | 409 | 411 | 411 | 381 |
| 2 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 426 | 431 | 426 | 425 | 424 | 429 | 398 |
| **Prevalence of Exposure 0.50** | | | | | | | | | | | | | |
| 2 | 1.3 | 1.2 | 1.2 | 1.2 | 2 | 2 | 368 | 391 | 368 | 369 | 368 | 392 | 383 |
| 2 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 395 | 417 | 395 | 391 | 391 | 415 | 406 |
| 2 | 1.3 | 1.2 | 1.2 | 1.5 | 2 | 2 | 386 | 408 | 386 | 384 | 384 | 408 | 404 |
| 2 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 376 | 404 | 376 | 373 | 374 | 403 | 396 |
| 2 | 1.5 | 1.2 | 1.2 | 1.2 | 2 | 2 | 382 | 402 | 382 | 381 | 382 | 402 | 395 |
| 2 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 379 | 399 | 379 | 377 | 378 | 398 | 394 |
| 2 | 1.5 | 1.2 | 1.2 | 1.5 | 2 | 2 | 391 | 411 | 391 | 389 | 389 | 411 | 404 |
| 2 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 378 | 407 | 378 | 375 | 375 | 405 | 398 |

Continuous T = Continuous Time (no aggregation), Continuous Z = Continuous Age (no 5 year age categorization), Aggregate T = De-identify data by creating one-week study time categories, Aggregate Z = De identify data by creating 5-year age categorization, Distributed = Only send summary statistics across sites instead of raw data, Cox = Cox PH not stratified by site, Strat. Cox = Cox PH stratified by site, and MH Strat. = Send separate site adjusted Cox PH model LR test statistics and variance to a central location and calculate a Mantel Haenszel type estimate.

**Table 7. Type I error for two site confounders and age using Exact Boundary Formation (P(Y)=0.01)**

| RR (Y\|X, Z, and Site) | | | | RR (X\|Z and Site) | | | TYPE I ERROR | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | **Continuous T Continuous Z** | **Aggregate T Aggregate Z** | **Distributed** |
| X | Z | Site 1 | Site 2 | Z | Site 1 | Site 2 | Cox | Cox | MH Strat. |
| **Prevalence of Exposure 0.10** | | | | | | | | | |
| 1 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.054 | 0.054 | **0.110** |
| 1 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.048 | **0.072** | **0.080** |
| 1 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.044 | 0.050 | **0.096** |
| 1 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.044 | **0.104** | 0.068 |
| **Prevalence of Exposure 0.50** | | | | | | | | | |
| 1 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.026 | 0.050 | 0.040 |
| 1 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.046 | **0.082** | 0.048 |
| 1 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.022 | 0.050 | 0.054 |
| 1 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.026 | **0.126** | 0.040 |

*Bold indicates Type I error 0.05+-0.02

Continuous T = Continuous Time (no aggregation), Continuous Z = Continuous Age (no 5 year age categorization), Aggregate T = De-identify data by creating one-week study time categories, Aggregate Z = De identify data by creating 5-year age categorization, Distributed = Only send summary statistics across sites instead of raw data, Cox = Cox PH not stratified by site, Strat. Cox = Cox PH stratified by site, and MH Strat. = Send separate site adjusted Cox PH model LR test statistics and variance to a central location and calculate a Mantel Haenszel type estimate.

**Table 8. Power for two site confounders and age using Exact Boundary Formation(P(Y)=0.01)**

| RR (Y\|Covariate) | | | | RR (X\|Covariate) | | | Power | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Continuous T Continuous Z | Aggregate T Aggregate Z | Distributed |
| X | Z | Site 1 | Site 2 | Z | Site 1 | Site 2 | Cox | Cox | MH Strat. |
| **Prevalence of Exposure 0.10** | | | | | | | | | |
| 2 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.696 | 0.746 | 0.724 |
| 2 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.734 | 0.794 | 0.706 |
| 2 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.682 | 0.728 | 0.712 |
| 2 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.678 | 0.792 | 0.712 |
| **Prevalence of Exposure 0.50** | | | | | | | | | |
| 2 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.892 | 0.924 | 0.912 |
| 2 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.892 | 0.958 | 0.884 |
| 2 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 0.890 | 0.928 | 0.906 |
| 2 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 0.908 | 0.970 | 0.892 |

Continuous T = Continuous Time (no aggregation), Continuous Z = Continuous Age (no 5 year age categorization), Aggregate T = De-identify data by creating one-week study time categories, Aggregate Z = De identify data by creating 5-year age categorization, Distributed = Only send summary statistics across sites instead of raw data, Cox = Cox PH not stratified by site, Strat. Cox = Cox PH stratified by site, and MH Strat. = Send separate site adjusted Cox PH model LR test statistics and variance to a central location and calculate a Mantel Haenszel type estimate.

**Table 9. Time to study end for two site confounders and age using Exact Boundary Formation (P(Y)=0.01)**

| RR (Y\|Covariate) | | | | RR (X\|Covariate) | | | Power | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Continuous T Continuous Z | Aggregate T Aggregate Z | Distributed |
| X | Z | Site 1 | Site 2 | Z | Site 1 | Site 2 | Cox | Cox | MH Strat. |
| **Prevalence of Exposure 0.10** | | | | | | | | | |
| 2 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 509 | 491 | 482 |
| 2 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 492 | 459 | 491 |
| 2 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 496 | 478 | 483 |
| 2 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 509 | 459 | 491 |
| **Prevalence of Exposure 0.50** | | | | | | | | | |
| 2 | 1.3 | 1.2 | 0.8 | 1.2 | 2 | 2 | 436 | 411 | 454 |
| 2 | 1.3 | 1.2 | 0.8 | 1.5 | 2 | 2 | 431 | 381 | 473 |
| 2 | 1.5 | 1.2 | 0.8 | 1.2 | 2 | 2 | 434 | 405 | 461 |
| 2 | 1.5 | 1.2 | 0.8 | 1.5 | 2 | 2 | 432 | 362 | 469 |

Continuous T = Continuous Time (no aggregation), Continuous Z = Continuous Age (no 5 year age categorization), Aggregate T = De-identify data by creating one-week study time categories, Aggregate Z = De identify data by creating 5-year age categorization, Distributed = Only send summary statistics across sites instead of raw data, Cox = Cox PH not stratified by site, Strat. Cox = Cox PH stratified by site, and MH Strat. = Send separate site adjusted Cox PH model LR test statistics and variance to a central location and calculate a Mantel Haenszel type estimate.

# IV. DATA STRUCTURE NEEDED TO CONDUCT ANALYSES IN MINI-SENTINEL

We will now outline the type of data needed to conduct the statistical methods described in Section II.B.1, standard Cox PH with de-identified data, and Section II.B.2, Mantel-Haenszel type test statistic in the distributed data setting. We will use the specific example outlined in the simulation evaluation, but scenarios with more confounders and longer study time can be easily generalized.

## A. STANDARD COX PH WITH DE-IDENTIFIED DATA

We will first define the individual level dataset that will be aggregated at each site to be shared across sites. First, divide the assumed two-year surveillance time into quarters, and categorize each participant's start day into the quarter in which that person first enters the study. Specifically, assume that surveillance started on January 1, 2012. Then, any study participants who initially entered the study from January 1, 2012 through March 31, 2012 (e.g. date participant started taking the exposure or comparator medical product and met enrollment criteria) is assigned to study quarter 1. Participants who entered the study from April 1, 2012 through June 30, 2012 are assigned to study quarter 2, and so on, up through study quarter 8.

Each participant has exposure status, *X*, and covariates such as site of enrollment (Site = 1, 2, or 3) and Age Category (Age (years) = 35-39, 40-44, 45-49, 50-54, 55-59, 60-65) at study entry. At the specified analysis time *a,* they have the outcome indicator $\delta_{si}^{c}(a) = I(E_{si} < C_{si} \cap E_{si} < T_{si}^{c}(a))$ which indicates if the participant experienced an outcome before they were censored or the current analysis time ended. At analysis time *a,* they also have the time to event or censoring variable ($T_{si}^{c}(a)$), defined as the minimum of the time to event, censoring, or analyses time, categorized into weekly categories. We will now walk through a test example of 10 participants at site 1 with 4 on comparator and 6 on exposure of interest, and will demonstrate how the dataset is created at analysis time June 30, 2012.

**Table 10. Example individual-level dataset at a site**

| Enrollment Date | Site | Age | Exposure | Date of Outcome | Date of Censoring | Outcome $\delta_{si}(a)$ | Outcome Time $T_{si}(a)$ |
|---|---|---|---|---|---|---|---|
| Jan 10, 2012 | 1 | 47 | 0 | . | . | 0 | 172 |
| Feb 1, 2012 | 1 | 55 | 1 | . | Mar 20, 2012 | 0 | 48 |
| Feb 20, 2012 | 1 | 60 | 0 | Apr 10, 2012 | . | 1 | 50 |
| Mar 12, 2012 | 1 | 64 | 0 | . | . | 0 | 110 |
| Mar 31, 2012 | 1 | 58 | 1 | . | Apr 18, 2012 | 0 | 18 |
| Apr 25, 2012 | 1 | 46 | 1 | May 1, 2012 | . | 1 | 6 |
| May 30, 2012 | 1 | 42 | 1 | Jun 12, 2012 | . | 1 | 13 |
| Jun 3, 2012 | 1 | 64 | 0 | . | . | 0 | 27 |
| Jun 10, 2012 | 1 | 38 | 1 | . | . | 0 | 20 |
| June 29, 2012 | 1 | 39 | 1 | . | . | 0 | 1 |

The first step is to deidentify the individual-level data in Table 10 by creating categories for study quarter and age, and to calculate weeks from study start for Outcome Time, $T_{si}^{c}(a)$as follows:

**Table 11. Example individual-level deidentified dataset at a site**

| Study Qtr | Site | Age Cat | Exposure | Outcome $\delta_{si}^{c}(a)$ | Outcome Time $T_{si}^{c}(a)$ in weeks |
|---|---|---|---|---|---|
| 1 | 1 | 3 | 0 | 0 | 25 |
| 1 | 1 | 5 | 1 | 0 | 7 |
| 1 | 1 | 6 | 0 | 1 | 8 |
| 1 | 1 | 6 | 0 | 0 | 16 |
| 1 | 1 | 5 | 1 | 0 | 3 |
| 2 | 1 | 3 | 1 | 1 | 1 |
| 2 | 1 | 2 | 1 | 1 | 2 |
| 2 | 1 | 6 | 0 | 0 | 4 |
| 2 | 1 | 1 | 1 | 0 | 3 |
| 2 | 1 | 1 | 1 | 0 | 1 |

The next step is to aggregate the individual-level data so that several participants can be represented in each row, to provide deidentification and the smallest number of data rows possible. To do this we propose the following aggregate dataset:

**Table 12. Example deidentified aggregate dataset at site**

| Study Qtr | Site | Age Cat | *N* | $N_x$ | *Y* | $Y_x$ | $E_1^0$ | $E_2^0$ | … | $E_8^0$ | … | $E_{25}^0$ | $C_1^0$ | $C_2^0$ | $C_3^0$ | $C_4^0$ | … | $C_{16}^0$ | … | $C_{25}^0$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | | 0 | | 0 | 0 | 0 | 0 | 0 | | 0 | | 1 |
| 1 | 1 | 5 | 2 | 2 | 0 | 0 | 0 | 0 | | 0 | | 0 | 0 | 0 | 0 | 0 | | 0 | | 0 |
| 1 | 1 | 6 | 2 | 0 | 1 | 0 | 0 | 0 | | 1 | | 0 | 0 | 0 | 0 | 0 | | 1 | | 0 |
| 2 | 1 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | | 0 | | 0 | 0 | 0 | 0 | 0 | | 0 | | 0 |
| 2 | 1 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | | 0 | | 0 | 0 | 0 | 0 | 0 | | 0 | | 0 |
| 2 | 1 | 3 | 1 | 1 | 1 | 1 | 0 | 0 | | 0 | | 0 | 0 | 0 | 0 | 0 | | 0 | | 0 |
| 2 | 1 | 6 | 1 | 0 | 0 | 0 | 0 | 0 | | 0 | | 0 | 0 | 0 | 0 | 1 | | 0 | | 0 |

| $E_1^1$ | $E_2^1$ | $E_3^1$ | $E_4^1$ | … | $E_{25}^1$ | $C_1^1$ | $C_2^1$ | $C_3^1$ | … | $C_7^1$ | … | $C_{25}^1$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 1 | | 1 | | 0 |
| 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | | 0 |
| 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | | 0 |
| 0 | 0 | 0 | 0 | | 0 | 1 | 0 | 1 | | 0 | | 0 |
| 0 | 1 | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | | 0 |
| 1 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | | 0 |
| 0 | 0 | 0 | 0 | | 0 | 0 | 0 | 0 | | 0 | | 0 |

where within each row defining study quarter and confounder stratum, we define the following counts: *N* is total number, $N_x$ is the number exposed, *Y* is the total number of outcomes, $Y_x$ is the number of exposed outcomes, $E_W^0$ is the number of outcomes in the comparator group observed at $T_{si}^c(a)$=*w*, $C_W^0$ is the number censored in the comparator group observed at $T_{si}^c(a)$=*w*, $E_W^1$ is the number of outcomes in the exposed group observed at $T_{si}^c(a)$=*w*, and $C_W^1$ is the number censored in the exposed group observed at $T_{si}^c(a)$=*w*. The number of rows in the dataset will be at most the number of study quarters times the number of confounder categories. As the sample size increases, the number of rows in the dataset will not increase beyond this maximum. This dataset can be securely sent to the coordinating center where the data can be de-aggregated to form the dataset needed to conduct the analysis.

## B. MANTEL-HAENSZEL TYPE ESTIMATE DATASET

This method is designed to send summary information across sites. Specifically, to conduct the primary observed test statistic the only information necessary to send across sites is a set of analyses time specific sample size, adjusted log rank test statistics, adjusted HR, and variance of adjusted HR. Further, it would be preferable to also submit Table 1 type information which includes by exposure and confounder categories the sample size, number of outcomes and total follow-up time. This would be the only information necessary when using normal approximation boundaries.

When doing the exact boundary formation additional information further needs to be shared which is a set of analysis time specific log rank calculations from permuted datasets under Ho: no effect of exposure. For better performance of the permutation approach and computational efficiency it may be preferable to instead submit a limited de-identified dataset which only includes by participant an indicator of outcome up to analysis time *a*, time of the outcome or censoring at analysis time *a*, and at each analysis times the calculated summary information, $\widehat{\boldsymbol{\beta}}_z^{(0)}\boldsymbol{Z}_{kl}$, where $\widehat{\boldsymbol{\beta}}_z^{(0)}$is the estimated confounder coefficient under Ho no effect of exposure (i.e. fit the site-specific Cox PH model only

including *Z* in the model without *X*). Including this richer dataset would allow us more flexibility to conduct the boundary calculation along with describing the unadjusted outcome data such as Kaplan-Meier curves.

# V. DISCUSSION AND FUTURE WORK

This workgroup has presented different postmarket surveillance methods applicable to the distributed data setting with rare outcomes. We have conducted a simulation study to assess performance of these new approaches compared to the non-distributed data setting. We found that the de-identified methods performed well in most settings using normal approximation boundary approaches except when there was both very rare outcomes and low exposure prevalence. In this setting in particular all methods including standard non-distributed methods did not perform well indicated by inflated type I error. We then conducted evaluations using the exact boundary formation and showed that using exact boundaries the type I error was held in the situations in which the normal approximation boundaries were not applicable except for in very extreme strength of confounding scenarios with low outcome and exposure prevalence. Therefore exact boundary methods should be utilized as outcome prevalence and exposure prevalence decrease. In the next survival task order we will assess if we can suggest other methods for situations with very strong confounding such as stratification or adjustment for propensity score categories to reduce the dimensionality of the confounders which may be causing some of the issues.

This task order developed two Cox PH methods using regression for active postmarket surveillance in which the data is distributed. Future work to assess other type of confounding control methods, such as the use of propensity scores via stratification or regression may improve upon the ability to conduct active surveillance evaluations for Mini-Sentinel.

# VI. REFERENCES


1. Mini-Sentinel Methods Development: Statistical methods for estimating causal risk differences in the distributed data setting for postmarket safety outcomes. 2012. (Accessed at http://www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_PRISM_Statistical-Methods-for-Estimating-Causal-Risk-Differences.pdf.)
2. Mini-Sentinel Methods Development: Sequential Testing Working Group Report. 2011. (Accessed at http://www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_Sequential-Testing-Report.pdf.)
3. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society, Series B 1972;34:187-220.
4. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719-48.
5. Lan KKG, Demets DL. Discrete Sequential Boundaries for Clinical-Trials. Biometrika 1983;70:659-63.
6. Pocock SJ. Interim Analyses for randomized clinical-trials - The Group Sequential Approach. Biometrics 1982;38:153-62.
7. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics 1979;35:549-56.
8. Kulldorff M, Davis RL, Kolczakâr 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.
9. Cook AJ, Wellman RD, Nelson JC, Jackson LA, Tiwari RC. Group sequential method for observational data using generalized estimating equations: Application Vaccine Safety Datalink (VSD). JRSS-C 2014;epub: doi:10.1111/rssc.12076.
10. Wang SK, Tsiatis AA. Approximately optimal one-parameter boundaries for group sequential trials. Biometrics 1987;43:193-9.
11. Kittelson JM, Emerson SS. A unifying family of group sequential test designs. Biometrics 1999;55:874-82.