Nonparametric Smoothing Method for Assessing Effect Modification by a Post-Randomization Biomarker with Application to Vaccine Efficacy Trials

Fred Hutch Technical Report

Michal Juraska and Peter B. Gilbert

February 10, 2016

1 Notation

We consider a randomized placebo-controlled vaccine efficacy trial that measures the first occurrence of a clinical endpoint event (e.g., a protocol-specified dengue disease endpoint) during a pre-specified follow-up period. Let Y denote the indicator of the first occurrence of the clinical endpoint. Let S be a continuous biomarker that measures, at a fixed time t_0 after randomization (e.g., at Month 13), an immune response to prevent the clinical endpoint event. It is of interest to evaluate S as a correlate of protection against the clinical endpoint; therefore, we restrict the analysis to subjects who are endpoint-free at t_0 and denote this status as $V = 1$. The biomarker S is observable only in subjects with $V = 1$; otherwise S is unobserved. In placebo recipients with $V = 1$, S is allowed to vary at t_0. We consider a two-phase outcome-dependent case-cohort sampling design with a vector of continuous and/or discrete baseline covariate(s) X measured in everyone (phase 1) and the biomarker S only measured in all cases with $V = 1$ and an unstratified random subcohort of controls (phase 2). We assume that baseline values, S_b, of the biomarker are measured in a subset of those with S measured. Let Z be the vaccine group indicator ($Z = 1$, vaccine; $Z = 0$, placebo).

2 Estimand of Interest

Let $Y(z)$, $S(z)$ and $V(z)$ be the observations of Y, S and V under assignment to treatment group $Z = z$. Define

$$\text{risk}_{(i)}(s_1, s_0, x) = P(Y(i) = 1 | S(1) = s_1, S(0) = s_0, X = x, V(1) = 1), \quad i = 0, 1. \tag{1}$$

For simplicity of exposition, henceforth all probability statements involving $S(Z)$ implicitly condition on $V(Z) = 1$. Under the assumption that $P(Y(1) \neq Y(0)) = 0$, the causal estimand of interest is

$$VE(s_1) = 1 - \text{risk}_{(1)}(s_1)/\text{risk}_{(0)}(s_1).$$

We assume that

$$\text{risk}_{(0)}(s_1, s_0, x) = \text{risk}_{(0)}(s_0, x) \quad \forall (s_0, x), \tag{2}$$

i.e., for placebo recipients, the risk of registering the study endpoint is assumed to be conditionally independent of the value of S under assignment to vaccine given the values of S and X under assignment to placebo. Assumption (2) and the Bayes Theorem yield

$$\begin{aligned}\text{risk}_{(0)}(s_1) &= \iint \text{risk}_{(0)}(s_1, s_0, x) \frac{f(s_1|S(0) = s_0, X = x) g(s_0|X = x) h(x)}{p(s_1)} \, \mathrm{d}s_0 \mathrm{d}x \\ &= \frac{\iint \text{risk}_{(0)}(s_0, x) f(s_1|S(0) = s_0, X = x) g(s_0|X = x) h(x) \, \mathrm{d}s_0 \mathrm{d}x}{\iint f(s_1|S(0) = s_0, X = x) g(s_0|X = x) h(x) \, \mathrm{d}s_0 \mathrm{d}x} \,. \end{aligned} \tag{3}$$

3 Vaccine Efficacy Curve Estimation Method

To estimate $VE(s_1)$, we first estimate separately the probabilities $\text{risk}_{(1)}(s_1)$ and $\text{risk}_{(0)}(s_1)$. To estimate $\text{risk}_{(1)}(s_1)$, we fit an inverse probability weighted (IPW) logistic regression model for two-phase designs to group $Z = 1$ data (implemented, e.g., in the R `osDesign` package). Under assumption (2), we estimate $\text{risk}_{(0)}(s_1)$ by estimating each component in (3). Analogously to $\text{risk}_{(1)}(s_1)$ estimation, we estimate the probability $\text{risk}_{(0)}(s_0, x)$ by fitting an IPW logistic regression model for two-phase designs to group $Z = 0$ data.

Further, to estimate the conditional density $f(s_1|S(0) = s_0, X = x)$, we adopt the assumption that

$$f(s_1|S(0) = s_0, X = x) = f(s_1|S_b = s_0, X = x) \quad \forall (s_0, x). \tag{4}$$

Assumption (4) may be plausible as both $S(0)$ and S_b measure the pre-existing/natural immunity level, only $S(0)$ is measured t_0 time units later. If diagnostics indicate that assumption (4) is plausible, we will directly assess $f(s_1|S_b = s_0, X = x)$ based on group $Z = 1$ data. If diagnostics indicate that assumption (4) might be violated, we will first examine the association of $S(0)$ and S_b in group $Z = 0$, and, if there is a noticeable trend, an offset will be used to calibrate $f(s_1|S_b = s_0, X = x)$. To estimate $f(s_1|S_b = s_0, X = x)$, we use a nonparametric kernel density estimator based on a a subset of data in group $Z = 1$ with an available S_b measurement (using the generalized product kernel method as described by Hall, Racine, and Li (2004) and implemented in the R `np` package). To obtain unbiased kernel density estimates, the two-phase sampling design needs to be accounted for. One approach, considered in the CYD14/15 analyses and yielding unbiased estimates, is to randomly delete a subset of cases in the immunogenicity set to attain the same case:control ratio as is present in the target population (in CYD14/15, the ITT set at-risk at month 13 and with no observed dengue endpoint between Month 0 and Month 13). Another, more powerful, approach would use weighting, however, we conjecture that, in the CYD14/15 analyses, minimal efficiency gain would be achieved given that most of the data originate from the controls.

Similarly, we use a nonparametric kernel density estimator for the conditional density $g(s_0|X = x)$ based on data in group $Z = 0$ after adjustment of the case:control ratio by random deletion of a subset of cases. We estimate the multivariate density $h(x)$ by a nonparametric kernel density estimator using all available data in the target population. In CYD14/15, all components of X are discrete, and therefore we estimate $P(X = x)$ for each level of x by a sample proportion in the target population.

Confidence intervals (CI) are obtained by assessing $VE(s_1)$ based on 10^3 bootstrap samples from the ITT set at-risk at Month 13 and with no prior infection, where cases and controls were sampled separately yielding a fixed number of cases and controls in each bootstrap sample. Throughout the

bootstrap procedure, the kernel conditional density estimates use optimal bandwidths estimated based on the original data-set. Wald bootstrap 95% pointwise and simultaneous CI for $VE(s_1)$ are reported.

3.1 Simultaneous Confidence Interval for the Vaccine Efficacy Curve

To obtain the Wald simultaneous CI for the $VE(s_1)$ curve, we denote $RR(s_1) = 1 - VE(s_1)$,

$$U = \sup_{s_1} \left| \{\log \widehat{RR}(s_1) - \log RR(s_1)\} / SE(\log \widehat{RR}(s_1)) \right|,$$

and, for a fixed $\alpha \in (0, 1)$, we define c_α as the solution to the equation

$$P\left(U \leq c_\alpha\right) = 1 - \alpha.$$

Further, we denote $\widehat{RR}^{(b)}(s_1)$ as the estimator for $RR(s_1)$ based on the b-th bootstrap sample, $b = 1, \ldots, B$, and $SE^*(\log \widehat{RR}(s_1))$ as the sample standard deviation of the bootstrap estimates $\log \widehat{RR}^{(1)}(s_1), \ldots, \log \widehat{RR}^{(B)}(s_1)$. Let $U^{(b)} = \sup_{s_1} \left| \{\log \widehat{RR}^{(b)}(s_1) - \log \widehat{RR}(s_1)\} / SE^*(\log \widehat{RR}(s_1)) \right|$. Because the distributions of U and $U^{(b)}$ are asymptotically equivalent, we estimate c_α by c^*_α defined as the empirical quantile in the bootstrap sample $U^{(1)}, \ldots, U^{(B)}$ at the probability level $1 - \alpha$. Finally, the Wald bootstrap $(1 - \alpha) \times 100\%$ simultaneous CI for the $VE(s_1)$ curve is obtained by transformation of the bounds

$$(l_\alpha(s_1), u_\alpha(s_1)) = \log \widehat{RR}(s_1) \mp c^*_\alpha SE^*(\log \widehat{RR}(s_1)).$$

Note that the width of the simultaneous CI depends on the range of marker values s_1.

4 Vaccine Efficacy Curve Inferential Methods

It is of interest to evaluate, separately, the following three null hypotheses, each against a general alternative hypothesis:

(i) $H_0 : VE(s_1) = VE$ for all $s_1 \in [s_{\min}, s_{\max}] =: S$,

(ii) $H_0 : VE_x(s_1) = VE_y(s_1)$ for $s_1 \in [s_l, s_u] \subseteq S$, where, for serotypes x and y, VE_x and VE_y denote serotype-specific vaccine efficacies as functions of homologous titers, and

(iii) $H_0 : VE_{\text{CYD14}}(s_1) = VE_{\text{CYD15}}(s_1)$ for $s_1 \in [s_l, s_u] \subseteq S$.

We employ the method of inverting simultaneous confidence intervals to obtain p-values for tests of the null hypotheses (i)–(iii).

For testing (i), we obtain the two-sided p-value as the value α solving the equation

$$\inf_{s_1 \in S} u_\alpha(s_1) = \sup_{s_1 \in S} l_\alpha(s_1).$$

For testing (ii), denote $RR_x(s_1) = 1 - VE_x(s_1)$ and $RR_y(s_1) = 1 - VE_y(s_1)$. Analogously to Section 3.1, we first construct the Wald bootstrap simultaneous CI for the $\log RR_x(s_1) - \log RR_y(s_1)$

curve, and then invert the simultaneous CI to obtain the p-value. Let $SE^*(\log \widehat{RR}_x(s_1) - \log \widehat{RR}_y(s_1))$ denote the sample standard deviation of the bootstrap estimates $\log \widehat{RR}_x^{(1)}(s_1) - \log \widehat{RR}_y^{(1)}(s_1), \ldots, \log \widehat{RR}_x^{(B)}(s_1) - \log \widehat{RR}_y^{(B)}(s_1)$. Let

$$U_d^{(b)} = \sup_{s_1 \in [s_l, s_u]} \frac{\left| \log \widehat{RR}_x^{(b)}(s_1) - \log \widehat{RR}_y^{(b)}(s_1) - \left(\log \widehat{RR}_x(s_1) - \log \widehat{RR}_y(s_1) \right) \right|}{SE^*(\log \widehat{RR}_x(s_1) - \log \widehat{RR}_y(s_1))}.$$

We define c_α^{*d} as the empirical quantile in the bootstrap sample $U_d^{(1)}, \ldots, U_d^{(B)}$ at the probability level $1-\alpha$. Subsequently, the Wald bootstrap $(1-\alpha)\times 100\%$ simultaneous CI for the $\log RR_x(s_1) - \log RR_y(s_1)$ curve is

$$\left(l_\alpha^d(s_1), u_\alpha^d(s_1) \right) = \log \widehat{RR}_x(s_1) - \log \widehat{RR}_y(s_1) \mp c_\alpha^{*d} SE^*(\log \widehat{RR}_x(s_1) - \log \widehat{RR}_y(s_1)).$$

The two-sided p-value for the test of (ii) is defined as the minimum of α_1 and α_2 that satisfy

$$\inf_{s_1 \in [s_l, s_u]} u_{\alpha_1}^d(s_1) = 0, \qquad \sup_{s_1 \in [s_l, s_u]} l_{\alpha_1}^d(s_1) \le 0,$$
$$\sup_{s_1 \in [s_l, s_u]} l_{\alpha_2}^d(s_1) = 0, \qquad \inf_{s_1 \in [s_l, s_u]} u_{\alpha_2}^d(s_1) \ge 0.$$

Note that at least one of α_1 and α_2 always exists.

For testing (iii), we replace $RR_x(s_1)$ and $RR_y(s_1)$ in the test of (ii) with $RR_{\text{CYD14}}(s_1)$ and $RR_{\text{CYD15}}(s_1)$ and, due to independence, we obtain $SE^*(\log \widehat{RR}_{\text{CYD14}}(s_1) - \log \widehat{RR}_{\text{CYD15}}(s_1))$ as

$$\left[SE^{*^2}(\log \widehat{RR}_{\text{CYD14}}(s_1)) + SE^{*^2}(\log \widehat{RR}_{\text{CYD15}}(s_1)) \right]^{1/2}.$$

The Wald bootstrap $(1-\alpha)\times 100\%$ simultaneous CI for the $\log RR_{\text{CYD14}}(s_1) - \log RR_{\text{CYD15}}(s_1)$ curve is

$$\left(l_\alpha^D(s_1), u_\alpha^D(s_1) \right) = \log \widehat{RR}_{\text{CYD14}}(s_1) - \log \widehat{RR}_{\text{CYD15}}(s_1) \mp c_\alpha^{*D} SE^*(\log \widehat{RR}_{\text{CYD14}}(s_1) - \log \widehat{RR}_{\text{CYD15}}(s_1)).$$

The two-sided p-value for the test of (iii) is defined as the minimum of α_1 and α_2 that satisfy

$$\inf_{s_1 \in [s_l, s_u]} u_{\alpha_1}^D(s_1) = 0, \qquad \sup_{s_1 \in [s_l, s_u]} l_{\alpha_1}^D(s_1) \le 0,$$
$$\sup_{s_1 \in [s_l, s_u]} l_{\alpha_2}^D(s_1) = 0, \qquad \inf_{s_1 \in [s_l, s_u]} u_{\alpha_2}^D(s_1) \ge 0.$$

5 Vaccine Efficacy Curve Estimation and Inference in the Dengue CYD Vaccine Trials

The analysis of the combined CYD14/15 trial data is restricted to the age range of 9–16 years. The probabilities $\text{risk}_{(1)}(s_1)$ and $\text{risk}_{(0)}(s_1)$ are assessed using logistic change point models (with a 'hinge') if supported by goodness-of-fit assessments, and using logistic linear models otherwise. The threshold in the hinge point model is estimated using a grid search method (Fong, Huang, and Gilbert, 2015) using the R `chngpt` package, which is hosted by the Comprehensive R Archive

Network. The combined CYD14/15 VE curve analyses in 9–16 year olds control for the categorical baseline covariates sex, age category (9–11 vs. 12–16 years), and country. The ITT set at-risk at Month 13 and without a prior dengue endpoint is treated as the Phase I cohort of interest in the two-phase sampling design framework. In the overall (serotype-specific) analyses, controls are defined as participants who never registered the overall (serotype-specific) endpoint between Months 0–25 and who were at-risk at Month 13, and cases are defined as participants who did not register the overall (serotype-specific) endpoint between Months 0–13 but did so between Months 13–25.

To assess the validity of assumption (4) in $VE(s_1)$ estimation, we examine the association of $S(0)$ and S_b in arm $Z = 0$ for each of the six biomarkers by fitting a robust linear regression model of Yohai (1987) and a locally-weighted polynomial regression model, and by estimating Spearman's correlation coefficient, r.

To test the null hypotheses (ii) and (iii) in Section 4, we need to specify the interval $[s_l, s_u]$ of marker values. In the combined CYD14/15 analysis of 9–16 year olds, it is of interest to test whether serotype-specific $VE(s_1)$ curves in (ii) and trial-specific $VE(s_1)$ curves in (iii) vary in the tails of the Month 13 titer distribution. For consistency with the definition of the neutralization response categories, we define the $[s_l, s_u]$ intervals by using percentiles of the titer readout distribution pooling over the vaccine and placebo groups in both trials and the four serotypes, whose unbiased estimation requires accounting for the case-cohort marker sampling design. More specifically, we use the first tertile estimate, 2.1, as a cut-off for defining the left tail (identical to the 'low' response category), i.e., $[s_l, s_u] = [\log_{10}(5), 2.1]$, and the 80^{th} percentile estimate, 3.1, as a cut-off for defining the right tail, i.e., $[s_l, s_u] = [3.1, 5.0]$.

We will conduct 5 tests of (i) (i.e., for AUC-MB and the four serotypes pooling over both trials), 12 tests of (ii) (i.e., for 6 serotype pairs and 2 marker intervals pooling over both trials), and 10 tests of (iii) (i.e., for AUC-MB and the four serotypes, and 2 marker intervals comparing the two trials). All tests involving the four serotypes are treated as multiple comparison, i.e., 4 tests of (i), 12 tests of (ii), and 8 tests of (iii). For each set of multiple comparisons, we will report unadjusted p-values as well as Holm-adjusted p-values controlling the familywise error rate and q-values controlling the false discovery rate. We define statistical significance as a q-value ≤ 0.2 for all multiply compared serotypes, and as an unadjusted p-value ≤ 0.05 for AUC-MB.

References

Fong, Y., Huang, Y., and Gilbert, P. (2015), "Estimation and Hypothesis Testing under Four Types of Change Point Models Using chngpt," *Department of Biostatistics, University of Washington Technical Report.*

Hall, P., Racine, J., and Li, Q. (2004), "Cross-validation and the estimation of conditional probability densities," *Journal of the American Statistical Association*, 99, 1015–1026.

Yohai, V. (1987), "High breakdown-point and high-efficiency robust estimates for regression," *Annals of Statistics*, 15, 642–656.