Published in final edited form as:

Stat Med. 2012 February 28; 31(5): 420–435. doi:10.1002/sim.4430.

Conditional Estimation after a Two-stage Diagnostic Biomarker Study that Allows Early Termination for Futility

Joseph S. Koopmeiners^{a,*}, Ziding Feng^b, and Margaret Sullivan Pepe^b

^aDivision of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, U.S.A.

^bDivision of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, U.S.A.

Abstract

Many biomarkers identified in marker discovery are shown to have inadequate performance in validation studies. This motivates the use of group sequential designs that allow early termination for futility. However, an option for early termination will lead to biased estimates for studies that reach full enrollment. We propose conditional estimators and confidence intervals that correct for this bias assuming that an unadjusted estimator exists that has an independent increments covariance structure. The proposed estimators and confidence intervals are applied to conditional estimation of the receiver operating characteristic (ROC) curve and the positive predictive value (PPV) curve after a two-stage study that allows early termination for futility and their performance is evaluated by simulation.

Keywords

Conditional Estimation; Diagnostic Biomarkers; Group Sequential Design; ROC Curve; PPV Curve

1. Introduction

Biomarkers are identified during marker discovery and their performance is evaluated in subsequent validation studies. The majority of candidate markers will not have adequate performance upon further evaluation. This, along with the need to conserve specimens and minimize cost, motivates the use of group sequential study designs that allow early termination for poorly performing markers.

Group sequential designs that allow early termination for futility lead to biased estimates of marker performance in studies that reach full enrollment. A candidate marker will have to show signs of promise at interim analyses in order to reach full enrollment, leading to optimistic estimates at study completion. Biased estimates at study completion have negative consequences even when the markers are studied further. These include: future studies wasted on inadequate markers and future studies that are underpowered. In contrast, biased estimates from studies that appropriately terminate early are not of great concern because the marker is deemed inadequate and estimation of its performance is not of interest. Due to these differences in the scientific implications we believe it is inappropriate

to average over stopping times when studying the statistical properties of estimators used after group sequential studies that allow early termination for futility. Instead, we focus on properties conditional on reaching full enrollment.

Group sequential designs for diagnostic biomarker studies have been discussed in the literature [1, 2, 3]. These methods are most appropriate in settings where markers already exist and the focus is on comparisons between markers and controlling the type-I error rate. However, our primary concern is estimation of the performance characteristic of a new biomarker or biomarker combination. Estimation after group sequential studies has been studied in the therapeutic clinical trials literature [4]. These methods were developed to maintain good statistical properties averaging over all possible stopping times whereas our interest lies in estimators with good statistical properties conditional on reaching full enrollment.

Pepe et al. considered conditional estimation of sensitivity and specificity for a dichotomous diagnostic biomarker [5]. They proposed the conditional UMVUE for a binomial probability after a two-stage study that allows early termination for futility. The conditional UMVUE performed well in simulation studies but its use is limited to evaluation of the performance of a dichotomous marker. In practice, many biomarkers are continuous. There is a need to extend these methods to accommodate the evaluation of continuous markers.

In Section 2, we discuss conditional estimation of a normal mean after a two-stage study that allows early termination for futility. In Section 3, we discuss how the results from Section 2 can be generalized to conditional estimation of any parameter when an unadjusted estimator exists that is asymptotically normal with an independent increments covariance structure. In Sections 4 and 5, we apply these results to conditional estimation of the receiver operating characteristic (ROC) curve and the positive predictive value (PPV) curve, two common summaries of performance for a continuous biomarker. Finally, we summarize our work and suggest possible extensions in Section 6.

2. A General Framework for Conditional Estimation

2.1. Notation and Study Design

Consider a two-stage study that allows early termination for futility. Let $x_1, \dots, x_n$ be *i.i.d.* normal random variables with mean, β , and known variance, σ^2 , where the first n_{stg1} observations are drawn in stage one and the remaining $n - n_{stg1}$ observations are drawn in stage 2. Furthermore, let $\widehat{\beta}_{stg1}$ be the sample mean for the first n_{stg1} observations, $\widehat{\beta}_{all}$ be the sample sample mean for all n observations and define,

$$\sigma_{stg1}^2 \equiv \frac{\sigma^2}{n_{stg1}} \quad \text{and} \quad \sigma_{all}^2 \equiv \frac{\sigma^2}{n}$$

where σ_{stg1}^2 and σ_{all}^2 are the variances of $\widehat{\beta}_{stg1}$ and $\widehat{\beta}_{all}$, respectively. $(\widehat{\beta}_{stg1}, \widehat{\beta}_{all})$ follows a bivariate normal distribution with an independent increments covariance structure. That is,

$$\begin{pmatrix} \widehat{\beta}_{stg1} \\ \widehat{\beta}_{all} \end{pmatrix} \sim N \left(\begin{pmatrix} \beta \\ \beta \end{pmatrix}, \begin{pmatrix} \sigma_{stg1}^2 & \sigma_{all}^2 \\ \sigma_{all}^2 & \sigma_{all}^2 \end{pmatrix} \right),$$

where $\frac{\widehat{\beta}_{stg1}}{\sigma_{stg1}^2}$ and $\frac{\widehat{\beta}_{all}}{\sigma_{all}^2}$ have independent increments. In the diagnostic testing setting, we will often consider estimators that follow this distribution approximately. We discuss conditional estimation when $(\widehat{\beta}_{stg1}, \widehat{\beta}_{all})$ are asymptotically normal in Section 3.

Our focus is on estimation after study completion and, rather than discussing the specifics of study design, we only assume that the study has continuation regions of the form

$$\begin{aligned} C=1 & \text{ if } \widehat{\beta}_{stg1} \geq h, \\ C=0 & \text{ if } \widehat{\beta}_{stg1} < h, \end{aligned}$$

where C is an indicator function taking the value 1 if the study continues to stage 2 and 0 if the study terminates after stage 1. We note that this implies that $\widehat{\beta}_{all}$ is undefined, or at the very least, unobserved, when $C=0$. This is a sufficiently broad class of stopping rules that should include any sensible two-stage design that allows early termination for futility.

2.2. Behavior of the Unadjusted Estimate Conditional on Study Completion

We begin by investigating the behavior of $\widehat{\beta}_{all}$ for studies that reach full enrollment. $\widehat{\beta}_{stg1}$ and $\widehat{\beta}_{all}$ follow a bivariate normal distribution in studies that do not allow early termination. In contrast, $\widehat{\beta}_{stg1}$ follows a truncated normal distribution conditional on reaching full enrollment in a study that allows for early termination. Therefore, the joint density of $\widehat{\beta}_{stg1}$ and $\widehat{\beta}_{all}$ conditional on reaching full enrollment is

$$f(\widehat{\beta}_{stg1}, \widehat{\beta}_{all} | \beta, C=1) = \frac{f(\widehat{\beta}_{stg1}, \widehat{\beta}_{all} | \beta)}{1 - \Phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)} I(\widehat{\beta}_{stg1} > h), \quad (1)$$

where $f(\widehat{\beta}_{stg1}, \widehat{\beta}_{all} | \beta)$ is the bivariate normal density and I is an indicator function that takes the value 0 if $\widehat{\beta}_{stg1}$ is less than h and 1 if $\widehat{\beta}_{stg1}$ is greater than h . Integrating (1) with respect to $\widehat{\beta}_{stg1}$ gives the conditional density of $\widehat{\beta}_{all}$

$$f(\widehat{\beta}_{all} | \beta, C=1) = \frac{1 - \Phi\left(\frac{h-\widehat{\beta}_{all}}{\sqrt{\sigma_{stg1}^2 - \sigma_{all}^2}}\right)}{1 - \Phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)} \frac{1}{\sqrt{2\pi\sigma_{all}^2}} e^{-\frac{1}{2}\left(\frac{\widehat{\beta}_{all}-\beta}{\sigma_{all}}\right)^2}. \quad (2)$$

The conditional density is essentially the marginal normal density multiplied by a term that changes the shape of the distribution depending on the probability of early termination. There will be large deviations from the normal density for lower values of β due to the high probability of early termination but very little deviation from the normal density for higher values of β when it is unlikely that the study terminates after stage 1.

We next consider the effect of early termination on the conditional mean and variance of $\widehat{\beta}_{all}$. The mean of $\widehat{\beta}_{all}$ conditional on reaching full enrollment is

$$E[\widehat{\beta}_{all}|C=1] = \beta + \left(\frac{\sigma_{all}^2}{\sigma_{stg1}^2} \right) \frac{\phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)}{1-\Phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)} \quad (3)$$

$$\equiv \beta + b(\beta|C=1).$$

To see this, note that $Cov(\widehat{\beta}_{stg1}/\sigma_{stg1}^2 - \widehat{\beta}_{all}/\sigma_{all}^2, \widehat{\beta}_{stg1}) = 0$, which implies that $\widehat{\beta}_{stg1}/\sigma_{stg1}^2 - \widehat{\beta}_{all}/\sigma_{all}^2$ is independent of C . Therefore,

$$E[\widehat{\beta}_{stg1}/\sigma_{stg1}^2 - \widehat{\beta}_{all}/\sigma_{all}^2 | C=1] = E[\widehat{\beta}_{stg1}/\sigma_{stg1}^2 - \widehat{\beta}_{all}/\sigma_{all}^2] = \beta(1/\sigma_{stg1}^2 - 1/\sigma_{all}^2).$$

Solving the previous equation for $E[\widehat{\beta}_{all}|C=1]$, we obtain

$$E[\widehat{\beta}_{all}|C=1] = \beta + \sigma_{all}^2 E[\widehat{\beta}_{stg1}/\sigma_{stg1}^2 - \beta/\sigma_{stg1}^2 | C=1] = \beta + (\sigma_{all}^2/\sigma_{stg1}^2) E[(\widehat{\beta}_{stg1} - \beta)/\sigma_{stg1} | C=1].$$

Taking the expectation of $E[(\widehat{\beta}_{stg1} - \beta)/\sigma_{stg1} | C=1]$ gives the desired result. From (3), we see that bias increases as the cut-off for continuing to full enrollment, h , and the ratio of the variance of $\widehat{\beta}_{all}$ to the standard deviation of $\widehat{\beta}_{stg1}$ increase. In contrast, the bias decreases as the true value of β increases. The conditional variance of $\widehat{\beta}_{all}$ is

$$Var[\widehat{\beta}_{all}|C=1] = \sigma_{all}^2 + \frac{\sigma_{all}^4}{\sigma_{stg1}^2} \left[\left(\frac{h-\beta}{\sigma_{stg1}} \right) \left(\frac{\phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)}{1-\Phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)} \right) - \left(\frac{\phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)}{1-\Phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)} \right)^2 \right]. \quad (4)$$

The conditional variance of $\widehat{\beta}_{all}$ will be used in the next section when we present bias adjusted conditional estimators. A detailed derivation of the conditional variance and an example illustrating the effect of early termination on the density and conditional expectation can be found in Appendix A.

2.3. Bias Corrected Conditional Estimation

In this section, we consider three biased corrected estimators and examine some of their properties. The first is motivated by Whitehead [6] who derived the bias of the MLE after a sequential trial and proposed a mean adjusted estimator to correct for this bias. We define a conditional mean adjusted estimator, $\widehat{\beta}_{mn}$, as the solution to the following equation

$$\widehat{\beta}_{all} = \beta_{mn} + b(\widehat{\beta}_{mn}|C=1). \quad (5)$$

The mean adjusted estimator is the value of β that results in a conditional mean equal to $\widehat{\beta}_{all}$. This estimator is equivalent to the K -th order bias adjusted estimator proposed by Troendle and Yu [7] and the conditional maximum likelihood estimator discussed by Liu et al [8].

Theorem 1 $\widehat{\beta}_{mn}$ is approximately conditionally unbiased with variance

$$V ar \left[\widehat{\beta}_{mn} | C=1 \right] \approx \left(\frac{1}{1+b'(\beta|C=1)} \right)^2 V ar \left[\widehat{\beta}_{all} | C=1 \right].$$

Proof Using a Taylor expansion, $\widehat{\beta}_{mn}$ can be approximated by

$$\widehat{\beta}_{all} \approx \widehat{\beta}_{mn} + b(\beta|C=1) + b'(\beta|C=1)(\widehat{\beta}_{mn} - \beta)$$

or

$$\widehat{\beta}_{mn} - \beta \approx \frac{\widehat{\beta}_{all} - \beta}{1+b'(\beta|C=1)} - \frac{b(\beta|C=1)}{1+b'(\beta|C=1)} \quad (6)$$

Taking the expectation and variance of the right side of (6) gives the result.

The second estimator is motivated by Whitehead's median adjusted estimator [9] that solves for the value of β that results in a marginal median equal to the observed value of $\widehat{\beta}_{all}$. Again, we extend this to the conditional case by using the conditional density. The conditional median adjusted estimator, $\widehat{\beta}_{md}$, is the solution to the following equation

$$.5 = \int_{-\infty}^{\widehat{\beta}_{all}} f(x|\widehat{\beta}_{md}, C=1) dx \equiv F(\widehat{\beta}_{all}|\widehat{\beta}_{md}, C=1), \quad (7)$$

, where $f(\cdot|\beta, C=1)$ is the conditional density from (2) and $F(\cdot|\beta, C=1)$ is the conditional CDF. $\widehat{\beta}_{md}$ is the value of β that results in a conditional median equal to $\widehat{\beta}_{all}$. Similar median adjusted for conditional estimation after adaptive clinical trials have been discussed in the literature [10,11].

Theorem 2 $\widehat{\beta}_{md}$ is approximately conditionally median unbiased and has conditional mean and variance equal to

$$E \left[\widehat{\beta}_{md} | C=1 \right] \approx \beta + \frac{.5 - F(E[\widehat{\beta}_{all}|C=1]|\beta, C=1)}{\frac{\partial F(E[\widehat{\beta}_{all}|C=1]|\beta, C=1)}{\partial \beta}}$$

and

$$V ar \left[\widehat{\beta}_{md} | C=1 \right] \approx \left[\frac{\partial}{\partial y} \frac{.5 - F(y|\beta, C=1)}{\frac{\partial F(y|\beta, C=1)}{\partial \beta}} \Big|_{y=E[\widehat{\beta}_{all}|C=1]} \right]^2 V ar \left[\widehat{\beta}_{all} | C=1 \right].$$

Proof Using a Taylor expansion, $\widehat{\beta}_{md}$ can be approximated by

$$.5 \approx F\left(\widehat{\beta}_{all} \middle| \beta, C=1\right) + \left(\widehat{\beta}_{md} - \beta\right) \frac{\partial F\left(\widehat{\beta}_{all} \middle| \beta, C=1\right)}{\partial \beta}$$

or

$$\widehat{\beta}_{md} - \beta \approx \frac{.5 - F\left(\widehat{\beta}_{all} \middle| \beta, C=1\right)}{\frac{\partial F\left(\widehat{\beta}_{all} \middle| \beta, C=1\right)}{\partial \beta}} \quad (8)$$

From (8) we can show that $\widehat{\beta}_{md}$ is approximately median unbiased

$$\begin{aligned} Pr\left[\widehat{\beta}_{md} < \beta \middle| \beta, C=1\right] &\approx Pr\left[\beta + \frac{.5 - F\left(\widehat{\beta}_{all} \middle| \beta, C=1\right)}{\frac{\partial F\left(\widehat{\beta}_{all} \middle| \beta, C=1\right)}{\partial \beta}} < \beta \middle| \beta, C=1\right] \\ &= Pr\left[F\left(\widehat{\beta}_{all} \middle| \beta, C=1\right) < .5 \middle| \beta, C=1\right] \quad \text{or} \quad Pr\left[F\left(\widehat{\beta}_{all} \middle| \beta, C=1\right) > .5 \middle| \beta, C=1\right] \end{aligned}$$

The mean and variance of $\widehat{\beta}_{md}$ can be found by applying the delta method to the right side of (8).

The third estimator is motivated by the conditional UMVUE for a binomial probability after a two-stage study that allows early termination for futility [5]. The conditional UMVUE for a binomial probability is the expectation of the sample proportion using only stage 2 data conditional on the sample proportion using all data and the study reaching full enrollment. We can follow the same approach to define the conditional UMVUE for a normal mean after a two-stage study that allows early termination for futility, which we call the Rao-Blackwell estimator, $\widehat{\beta}_{rb}$,

$$\begin{aligned} \widehat{\beta}_{rb} &= E\left[\widehat{\beta}_{stg2} \middle| \widehat{\beta}_{all}, C=1\right] \\ &= \widehat{\beta}_{all} - \frac{\sigma_{all}^2}{\sqrt{\sigma_{stg1}^2 - \sigma_{all}^2}} \frac{\phi\left(\frac{h - \widehat{\beta}_{all}}{\sqrt{\sigma_{stg1}^2 - \sigma_{all}^2}}\right)}{\Phi\left(\frac{h - \widehat{\beta}_{all}}{\sqrt{\sigma_{stg1}^2 - \sigma_{all}^2}}\right)} \quad (9) \\ &\equiv \widehat{\beta}_{all} - g\left(\widehat{\beta}_{all}\right), \end{aligned}$$

, where $\widehat{\beta}_{stg2}$ is the sample mean using only stage 2 data.

Theorem 3 $\widehat{\beta}_{rb}$ is the conditional UMVUE and has variance,

$$Var\left[\widehat{\beta}_{rb} \middle| C=1\right] \approx \left(1 - g'(\beta)\right)^2 Var\left[\widehat{\beta}_{all} \middle| C=1\right].$$

Proof Conditional on $C = 1$, the density of the X_i 's is a single parameter exponential family with complete sufficient statistic $\widehat{\beta}_{all}$ [12]. $\widehat{\beta}_{rb}$ is the conditional UMVUE by the Rao-Blackwell Theorem. The conditional variance of $\widehat{\beta}_{rb}$ is an application of the delta method.

Theorem 3 illustrates the advantage to considering the special case of estimation conditional on full enrollment in a two-stage study that allows early termination for futility. Estimators that consider conditional estimation more generally only provide approximate unbiasedness, whereas we are able to derive an estimator that is not only unbiased but optimal within the class of unbiased estimators.

While $\widehat{\beta}_{rb}$ is optimal for conditional estimation of a normal mean, this will not necessarily be the case when $\widehat{\beta}$ is asymptotically normal. For this reason, $\widehat{\beta}_{mn}$ and $\widehat{\beta}_{md}$ are still of interest and their performance will be compared to the performance of $\widehat{\beta}_{rb}$ when we consider conditional estimation of the ROC and PPV curves in Sections 4 and 5.

2.4. Small-sample Properties

The theoretical results presented in Section 2.3 show that $\widehat{\beta}_{rb}$ is optimal within the class of unbiased estimators but do not provide information about its relative merits compared to approximately unbiased estimators $\widehat{\beta}_{mn}$ and $\widehat{\beta}_{md}$. We completed a small simulation study to compare the performance of $\widehat{\beta}_{rb}$, $\widehat{\beta}_{mn}$ and $\widehat{\beta}_{md}$. Data were simulated from *i.i.d* normal distributions with mean β and known variance equal to 1. We considered sample sizes of 20, 40 and 60, with one interim analysis using 10, 20 and 30 observations, respectively. The study terminates for futility if the upper limit of the 95% confidence interval formed at the interim analysis is less than 0.8. All summaries are based only on studies that reached full enrollment. 10,000 simulations were considered for each scenario.

Table 1 presents simulation results comparing the performance of $\widehat{\beta}_{rb}$, $\widehat{\beta}_{mn}$ and $\widehat{\beta}_{md}$. Also included are summaries for the biased, naive estimator that considers all of the data and does not adjust for the possibility of early termination, $\widehat{\beta}_{all}$, and the estimator only using the stage 2 data, $\widehat{\beta}_{stg2}$, which represents the simplest conditionally unbiased estimator but is inefficient. We see that $\widehat{\beta}_{rb}$ is unbiased and has smaller standard error than $\widehat{\beta}_{stg2}$ but larger standard error than $\widehat{\beta}_{all}$. $\widehat{\beta}_{mn}$ and $\widehat{\beta}_{md}$ have similar standard errors to $\widehat{\beta}_{rb}$ (smaller, in some cases) but are biased, over-correcting for the bias in $\widehat{\beta}_{all}$. The bias decreases as sample size increases but a small amount of bias is still visible when $n = 60$. Finally, standard errors of $\widehat{\beta}_{rb}$, $\widehat{\beta}_{mn}$ and $\widehat{\beta}_{md}$ are larger than the standard error for $\widehat{\beta}_{all}$ when β is small and there is a high probability for early termination but there is little difference when β is larger and early termination is rare.

2.5. Conditional Confidence Intervals

Unadjusted confidence intervals for $\widehat{\beta}_{all}$ fail to provide correct coverage conditional on reaching full enrollment. The conditional coverage probability can be found by integrating (2) and ranges from 0 to the nominal rate depending on the probability of early termination. An example illustrating the effect of early termination on unadjusted confidence intervals is in Appendix A. Standard approaches to forming confidence intervals after group sequential clinical trials [13, 14, 4] also fail to provide uniformly correct conditional coverage [15]. This is not surprising as these confidence intervals were developed to provide correct coverage averaging over all stopping times. Different approaches are needed to provide correct coverage conditional on reaching full enrollment.

Ohman-Strickland and Casella [15] present conditional confidence intervals for a normal mean that provide adequate conditional coverage. Following their approach, we define (β_L, β_U) as a $1 - \alpha$ conditional confidence interval where

$$1 - \frac{\alpha}{2} = F(\widehat{\beta}_{all} | \beta_L, C=1)$$

and

$$\frac{\alpha}{2} = F(\widehat{\beta}_{all} | \beta_U, C=1).$$

These intervals provide correct conditional coverage, as we are simply pivoting the conditional CDF [12], and are the uniformly most accurate $1 - \alpha$ level conditional confidence interval for β [15].

3. Conditional Estimation when $\widehat{\beta}$ is Asymptotically Normal

In the diagnostic testing setting, we often work with estimators that are asymptotically normal with an independent increments covariance structure. For example, a binomial proportion, such as the true positive fraction (TPF) or the false positive fraction (FPF), when estimated using the sample proportion or the empirical estimate of a point on the receiver operating characteristic (ROC) curve are asymptotically normal with an independent increments covariance structure [16]. Furthermore, we must also account for the possibility that the variance of $\widehat{\beta}$ depends on the true value of β . The estimators and confidence intervals proposed in Section 2 can be adapted to these settings.

Let $\widehat{\beta}$ be an estimate of marker performance, β . Recall that $\widehat{\beta}_{stg1}$ is the estimate of β using only data from stage 1 and $\widehat{\beta}_{all}$ is the estimate of β at full enrollment. Assume that $(\widehat{\beta}_{stg1}, \widehat{\beta}_{all})$ is asymptotically normal with,

$$\begin{pmatrix} \widehat{\beta}_{stg1} \\ \widehat{\beta}_{all} \end{pmatrix} \sim N \left(\begin{pmatrix} \beta \\ \beta \end{pmatrix}, \begin{pmatrix} \sigma_{\beta, stg1}^2 & \sigma_{\beta, all}^2 \\ \sigma_{\beta, all}^2 & \sigma_{\beta, all}^2 \end{pmatrix} \right),$$

where $\sigma_{\beta, stg1}^2$ and $\sigma_{\beta, all}^2$ are the variances of $\widehat{\beta}_{stg1}$ and $\widehat{\beta}_{all}$, respectively, and $\frac{\widehat{\beta}_{stg1}}{\sigma_{\beta, stg1}^2}$ and $\frac{\widehat{\beta}_{all}}{\sigma_{\beta, all}^2}$ have independent increments. This is a common distributional assumption in the group-sequential testing literature [17]. We use the notation, $\sigma_{\beta, stg1}^2$ and $\sigma_{\beta, all}^2$, to indicate that the variance of $\widehat{\beta}$ is allowed to depend on β , which will often be the case in the diagnostic testing setting.

To calculate the mean adjusted estimator when $\widehat{\beta}$ is asymptotically normal, we let $b(\beta|C=1)$ depend on β through $\sigma_{\beta, stg1}^2$ and $\sigma_{\beta, all}^2$ and solve (5) numerically. Similarly, the median adjusted estimator is calculated by allowing $f(x|\beta, C=1)$ to depend on β through $\sigma_{\beta, stg1}^2$ and $\sigma_{\beta, all}^2$ and solve (7) numerically. We have applied these estimators with various mean-variance relationships, including the two found in Sections 4 and 5, and had no difficulty

solving these equations using the `nlm` function in R. When $\widehat{\beta}$ is asymptotically normal with β -dependent variance, the Rao-Blackwell estimator becomes,

$$\widehat{\beta}_{rb} = \widehat{\beta}_{all} - \frac{\sigma_{\beta,all}^2}{\sqrt{\sigma_{\beta,stg1}^2 - \sigma_{\beta,all}^2}} \frac{\phi\left(\frac{h - \widehat{\beta}_{all}}{\sqrt{\sigma_{\beta,stg1}^2 - \sigma_{\beta,all}^2}}\right)}{\Phi\left(\frac{h - \widehat{\beta}_{all}}{\sqrt{\sigma_{\beta,stg1}^2 - \sigma_{\beta,all}^2}}\right)}$$

and must be calculated using an iterative procedure. At the first iteration,

$$\widehat{\beta}_{rb}^{(1)} = \widehat{\beta}_{all} - \frac{\sigma_{\beta_{all},all}^2}{\sqrt{\sigma_{\beta_{all},stg1}^2 - \sigma_{\beta_{all},all}^2}} \frac{\phi\left(\frac{h - \widehat{\beta}_{all}}{\sqrt{\sigma_{\beta_{all},stg1}^2 - \sigma_{\beta_{all},all}^2}}\right)}{\Phi\left(\frac{h - \widehat{\beta}_{all}}{\sqrt{\sigma_{\beta_{all},stg1}^2 - \sigma_{\beta_{all},all}^2}}\right)}$$

and at subsequent iterations,

$$\widehat{\beta}_{rb}^{(i)} = \widehat{\beta}_{all} - \frac{\sigma_{\beta_{rb}^{(i-1)},all}^2}{\sqrt{\sigma_{\beta_{rb}^{(i-1)},stg1}^2 - \sigma_{\beta_{rb}^{(i-1)},all}^2}} \frac{\phi\left(\frac{h - \widehat{\beta}_{all}}{\sqrt{\sigma_{\beta_{rb}^{(i-1)},stg1}^2 - \sigma_{\beta_{rb}^{(i-1)},all}^2}}\right)}{\Phi\left(\frac{h - \widehat{\beta}_{all}}{\sqrt{\sigma_{\beta_{rb}^{(i-1)},stg1}^2 - \sigma_{\beta_{rb}^{(i-1)},all}^2}}\right)}.$$

We continue until the difference between $\widehat{\beta}_{rb}^{(i)}$ and $\widehat{\beta}_{rb}^{(i-1)}$ is small. Calculation of the conditional confidence intervals when $\widehat{\beta}$ is asymptotically normal is analogous to calculation of the median adjusted estimator.

It is often the case that $\sigma_{\beta,stg1}^2$ and $\sigma_{\beta,all}^2$ depend on unknown nuisance parameters that must be estimated. For our examples in Sections 4 and 5, unadjusted estimates of the nuisance parameters using all data available at the end of stage 2 were plugged into the variance formulas and we proceeded to calculate the conditional estimators and confidence intervals as if the nuisance parameters were known.

4. Conditional Estimation of ROC(t)

In this section, we apply our methods to conditional estimation of $ROC(t)$, a point on the ROC curve. Let X_D be a biomarker value for a case with distribution function $F_D(x)$, X_C be a biomarker value for a control with distribution function $F_C(x)$ and assume that larger biomarker values are more indicative of disease. If we consider a biomarker value greater than or equal to a threshold c to be a positive test, the performance of the biomarker can be summarized by the true positive fraction, $TPF(c) = P[X_D \geq c]$, and the false positive fraction,

$FPF(c) = P[X_C \geq c]$. The ROC curve is the set of all possible true and false positive fractions: $ROC(\cdot) = \{(TPF(c), FPF(c)), c \in (-\infty, \infty)\}$. The ROC curve can alternately be expressed as

$$ROC(t) = S_D \left(S_D^{-1}(t) \right), t \in (0, 1), \quad (10)$$

where $S_D(x)$ is the biomarker survival function for the cases, $S_D(x)$ is the biomarker survival function for the controls and t is the false positive fraction. $ROC(t)$ is the true positive fraction corresponding to a false positive fraction equal to t .

The ROC curve is often estimated empirically in order to avoid making distributional assumptions about $F_D(x)$ and $F_D(x)$. The sequential empirical estimate of $ROC(t)$ is defined by plugging the sequential empirical estimates of $S_D(x)$ and $S_D^{-1}(t)$ into (10)

$$\widehat{ROC}_{r_D, r_D}(t) = \widehat{S}_D, r_D \left(\widehat{S}_D^{-1}(t) \right),$$

where r_D and r_D are the proportion of cases and controls, respectively, observed at the interim analysis. The sequential empirical estimate of $ROC(t)$ is asymptotically normal with an independent increments covariance structure [16] and variance

$$\sigma_{\widehat{ROC}_{r_D, r_D}(t)}^2 = \frac{ROC(t)(1 - ROC(t))}{n_D r_D} + \left(\frac{f_D \left(S_D^{-1}(t) \right)}{f_D \left(S_D^{-1}(t) \right)} \right)^2 \frac{t(1-t)}{n_D r_D},$$

where n_D and n_D are the number of cases and controls, respectively, at full enrollment.

Therefore, if we let $\widehat{ROC}_{r_D, r_D, stg1}(t) = \widehat{\beta}_{stg1}$, $\widehat{ROC}_{1,1}(t) = \widehat{\beta}_{all}$, $\sigma_{\widehat{ROC}_{r_D, r_D, stg1}(t)}^2 = \sigma_{\widehat{\beta}_{stg1}}^2$ and $\sigma_{\widehat{ROC}_{1,1}(t)}^2 = \sigma_{\widehat{\beta}_{all}}^2$, we can use the estimators proposed in Section 3 to estimate $ROC(t)$ after a

two-stage study that allows early termination for futility. The variance of $\widehat{ROC}_{r_D, r_D}(t)$

depends on $f_D \left(S_D^{-1}(t) \right)$ and $f_D \left(S_D^{-1}(t) \right)$, which must be estimated in order to use the

estimators from Section 3. For our simulations, we estimated $f_D \left(S_D^{-1}(t) \right)$ and $f_D \left(S_D^{-1}(t) \right)$ using a normal kernel density estimator and proceeded as if they were known.

Although we expect the normality approximation for $\widehat{ROC}_{r_D, r_D}(t)$ to work well, transformed values might provide faster convergence to normality. For example, the logit transformation is preferable when t or $ROC(t)$ are close to 0 or 1 [18]. We note that the conditional estimators and confidence intervals presented in Section 3 apply to transformed values as well.

4.1. Small Sample Properties

We simulated 10,000 studies of 400 cases and 400 controls with an interim analysis using data from 200 cases and 200 controls. We note that, while not what would typically considered a small sample size, 400 cases and 400 controls is representative of phase 2 diagnostic biomarker studies found in the literature [19]. Studies were terminated if the upper limit of the two-sided 95% Wald confidence interval for $ROC(0.2)$ was less than 0.7.

Marker values for the controls were simulated from a standard normal distribution while marker values for the cases were simulated from a normal distribution with variance 1 and mean chosen to achieve the desired true value of $ROC(0.2)$. We are interested in the properties of our estimators and confidence intervals conditional on reaching full enrollment. Therefore, all summaries are based on only those studies that completed stage 2. If, for example, 75% of the studies terminated after stage 1, then summaries are based on the 2,500 studies that completed stage 2.

Table 2 presents results for the unadjusted estimator and the three bias-adjusted estimators.

Summaries for $\widehat{ROC}(0.2)_{stg2}$, the empirical estimator using only stage 2 data, are also included because it is the simplest estimator that is not biased by the possibility of early termination. However, $\widehat{ROC}(0.2)_{stg2}$ is inefficient because it does not use all of the observations. The bias of the unadjusted estimator is substantial and decreases as $ROC(0.2)$ increases. All three bias corrected estimators provide unbiased estimates of $ROC(0.2)$ and have smaller standard errors than $\widehat{ROC}(0.2)_{stg2}$. They have larger standard errors than $\widehat{ROC}(0.2)_{all}$ but the difference decreases as $ROC(0.2)$ increases and the difference is negligible when $ROC(0.2) > 0.7$. Table 3 presents coverage probabilities and average widths. The unadjusted Wald interval provides correct conditional coverage when $ROC(0.2) > 0.6$ but not when there is a high probability of early termination. The conditional intervals provide coverage that is slightly below the nominal level (approximately 0.94) but provide substantially better conditional coverage than the unadjusted intervals when $ROC(0.2) < 0.6$ and the study is likely to terminate early. The conditional confidence intervals are wider than the unadjusted intervals but the difference decreases as $ROC(0.2)$ increases. For example, the conditional intervals are 27% wider when $ROC(0.2) = 0.55$ but are nearly the same width when $ROC(0.2) > 0.7$.

5. Conditional Estimation of PPV(u)

The predictive accuracy of a binary marker can be evaluated by the positive predictive value (PPV) and negative predictive value (NPV). The PPV and NPV curves have been proposed as a method for evaluating the predictive accuracy of a continuous marker [20, 21]. Let D be the binary outcome, $\rho = \text{Prob}(D = 1)$ and let X denote the biomarker value with conditional distribution $F(x|D = 1) = F_D(x)$ and $F(x|D = 0) = F_{\bar{D}}(x)$. For a given threshold c , the positive predictive value is $P[D = 1|X > c]$ and the negative predictive value is $P[D = 0|X \leq c]$. The PPV and NPV curves are indexed by the proportion of the population that are classified as negative, u , and positive, $1 - u$, where $u = P(X \leq c)$. That is, the PPV and NPV curves are defined as $PPV(u) = P[D = 1|X > F^{-1}(u)]$ and $NPV(u) = P[D = 0|X \leq F^{-1}(u)]$ for all $u \in (0, 1)$. We can also write

$$PPV(u) = \frac{S_D(F^{-1}(u))\rho}{1-u} \quad \text{and} \quad NPV(u) = \frac{F_{\bar{D}}(F^{-1}(u))(1-\rho)}{u}, \quad (11)$$

where $F(x) = \rho F_D(x) + (1-\rho) F_{\bar{D}}(x)$ is the marginal distribution function for X in the entire population.

Consider conditional estimation of $PPV(u)$ under case-control sampling. Estimation for $NPV(u)$ is analogous. The sequential empirical estimator of $PPV(u)$ under case-control sampling is defined by plugging the sequential empirical estimates of $S_D(x)$ and $F^{-1}(u)$ into (11)

$$\widehat{PPV}_{r_D, r_{\bar{D}}} (u) = \frac{\widehat{S}_{D, r_D} \left(\widehat{F}_{r_D, r_{\bar{D}}}^{-1} (u) \right) \rho}{1 - \nu},$$

where r_D and $r_{\bar{D}}$ refer to the proportion of cases and controls, respectively, observed at a given time point and ρ is assumed known. The sequential empirical estimator of $PPV(u)$ is asymptotically normal with an independent increments covariance structure [16] and variance

$$\sigma_{\widehat{PPV}_{r_D, r_{\bar{D}}} (u)}^2 = \frac{\left(\frac{f_D(F^{-1}(u))}{f(F^{-1}(u))} (1-\rho) \right)^2 PPV(u) \left(\frac{\rho}{1-u} - PPV(u) \right)}{n_D r_D} + \frac{\left(\frac{f_D(F^{-1}(u))}{f(F^{-1}(u))} \rho \right)^2 (1-PPV(u)) \left(\frac{u-\rho}{1-u} + PPV(u) \right)}{n_{\bar{D}} r_{\bar{D}}}.$$

where n_D and $n_{\bar{D}}$ are the number of cases and controls, respectively, at full enrollment.

Therefore, if we let $\widehat{PPV}_{r_D, stg1, r_{\bar{D}}, stg1} (u) = \widehat{\beta}_{stg1}$, $\widehat{PPV}_{1,1} (u) = \widehat{\beta}_{all}$, $\sigma_{\widehat{PPV}_{r_D, stg1, r_{\bar{D}}, stg1} (u)}^2 = \sigma_{\widehat{\beta}_{stg1}}^2$ and $\sigma_{\widehat{PPV}_{1,1} (u)}^2 = \sigma_{\widehat{\beta}_{all}}^2$, we can use the estimators proposed in Section 3 to estimate $PPV(u)$ after a

two-stage study that allows early termination for futility. The variance of $\widehat{PPV}_{r_D, r_{\bar{D}}} (u)$

depends on $f_D(F^{-1}(u))$ and $f(F^{-1}(u))$, which must be estimated in order to use the estimators from Section 3. Again, we estimated $f_D(F^{-1}(u))$ and $f(F^{-1}(u))$ using a normal density estimator and proceeded as if they were kernel known.

5.1. Small Sample Properties

We simulated 10,000 studies with 200 cases and 200 controls allowing early termination after 100 cases and 100 controls. The parameter of interest was $PPV(0.9)$, the positive predictive value when biomarker values at the 90th percentile or above are considered positive. Early termination occurred if the upper limit of the two-sided 95% Wald confidence interval for $PPV(0.9)$ was less than 0.8. Marker values for the controls were simulated from a standard normal distribution while marker values for the cases were simulated from a normal distribution with variance 1 and mean chosen to achieve the desired true value of $PPV(0.9)$. Again, all summaries are based on only those studies that reached full enrollment.

Table 4 presents simulation results evaluating the bias and standard errors of the three bias adjusted estimators along with the unadjusted empirical estimator, $\widehat{PPV}(0.9)_{all}$, and the empirical estimator using only stage 2 data, $\widehat{PPV}(0.9)_{stg2}$. The unadjusted estimator has substantial bias that decreases as $PPV(0.9)$ increases. The three bias adjusted estimators are unbiased and have standard errors that are larger than $\widehat{PPV}(0.9)_{all}$ but smaller than $\widehat{PPV}(0.9)_{stg2}$. We see that the difference between the standard errors of $\widehat{PPV}(0.9)_{all}$ and the three bias adjusted estimators decreases as $PPV(0.9)$ increases. Table 5 presents conditional coverage probabilities and average widths for the unadjusted 95% Wald confidence interval and the conditional confidence interval. We see that the unadjusted interval provides correct conditional coverage when $PPV(0.9) \geq 0.65$ but not for smaller values of $PPV(0.9)$. Again, we see that the coverage probability for the conditional intervals are below the nominal rate

(between 0.92 and 0.93 in most cases) but the conditional intervals provide a clear improvement over the unadjusted intervals when $PPV(0.9) < 0.65$ and the study is likely to terminate early. The conditional intervals are wider than the unadjusted intervals but we see that the difference decreases as $PPV(0.9)$ increases.

6. Discussion

We present a general framework for conditional estimation after a two-stage study that allows early termination for futility. We show that the unadjusted estimator has substantial bias conditional on reaching full enrollment and propose three bias adjusted conditional estimators. The three bias adjusted estimators are shown to have good theoretical properties conditional on reaching full enrollment and perform well in two small simulation studies. All three estimators correct for the bias due to early termination with a minimal loss in efficiency when there is only a small chance of early termination.

Neither unadjusted confidence intervals, that do not account for the possibility of early termination, nor standard approaches to forming confidence intervals after group sequential studies provide adequate conditional coverage. Confidence intervals that provide correct conditional coverage for a normal mean have been discussed in the literature [15]. In

Sections 4 and 5, we apply these intervals to the case where $\hat{\beta}$ is asymptotically normal. The conditional intervals are wider than the unadjusted intervals but the difference is small for large values of the performance parameter of interest.

There are limitations to conditional estimation. Conditional estimation results in a loss of information compared to the unconditional approach and, for this reason, it is important to consider the scientific implications when deciding between the two. In the diagnostic testing setting, where high-quality resources are very scarce, we feel that the scientific implications (i.e. future studies of inadequate markers or underpowered future studies) support the use of conditional estimation.

We have proposed an estimator that is unbiased conditional on reaching full enrollment but we did not discuss conditional estimation for studies that terminated early for futility. Estimators that adjust for bias conditional on early termination exist but are not unbiased [7, 8]. Therefore, we are left to choose between an approach that is biased conditional on early termination, biased conditional on reaching full enrollment but unbiased unconditionally and an approach that is biased conditional on early termination, unbiased conditional on reaching full enrollment and biased unconditionally. The former is the standard approach in the clinical trials literature but, again, we feel that the scientific implications make the second approach more appropriate in the diagnostic testing setting.

Our simulation results illustrate that the conditional estimators and confidence intervals proposed in this paper can be applied when $\hat{\beta}$ is only asymptotically normal. This is a standard assumption in the group sequential testing literature. We chose to focus on conditional estimation of $ROC(t)$ and $PPV(u)$ where the independent increments structure holds and we expect the assumption to hold for other measures of discrimination such as AUC.

We framed our discussion in the context of a diagnostic biomarker study. However, the same methods can be used in a therapeutic study as well as long as conditional estimation is appropriate and an unadjusted estimator of treatment effect exists that meets the independent increments assumption.

The Early Detection Research Network has several biomarker validation studies underway that utilize two-stage designs that allow early termination for futility. This paper provides methodology for conditional estimation at study completion but further research is needed to identify strategies for designing diagnostic biomarker studies that utilize conditional estimation at study termination. The sample size used in stage 1, overall sample size and criteria for early termination should be determined in order to achieve desired type-I error rate and power while minimizing expected sample size.

We considered the case of a two-stage design that evaluates the performance of a single biomarker. Another possibility is group sequential designs that evaluate the performance of several markers at once. In this case, the stopping rule may be applied to one marker but unbiased estimates are desired for the other markers as well. Conditional estimation of a secondary parameter after a group sequential trial has been discussed in the literature [8]. This may provide a starting point for developing conditional estimators for several markers when a group sequential stopping rule is applied to only one marker.

Acknowledgments

The authors would like to thank the Associate Editor and two Referees for their helpful comments that resulted in a greatly improved manuscript. This work is partially supported by NIH grants P01-CA53996 and U01-CA86368.

A. Supplementary Materials for Section 2

A.1. Derivation of the conditional variance of $\widehat{\beta}_{all}$

The derivation of the conditional variance of $\widehat{\beta}_{all}$ relies on the iterated variance formula. Line 2 is a consequence of $\widehat{\beta}_{stg1}$ and $\widehat{\beta}_{all}$ following a bivariate normal distribution and line 4 is a results of $\widehat{\beta}_{all}$ following a truncated normal distribution conditional on $C = 1$.

$$\begin{aligned} \text{Var}[\widehat{\beta}_{all}|C=1] &= \text{Var}\left[E[\widehat{\beta}_{all}|\widehat{\beta}_{stg1}, C=1]|C=1\right] + E\left[\text{Var}[\widehat{\beta}_{all}|\widehat{\beta}_{stg1}, C=1]|C=1\right] \\ &= \text{Var}\left[\beta + \left(\frac{\sigma_{all}^2}{\sigma_{stg1}^2}\right)(\widehat{\beta}_{stg1} - \beta)|C=1\right] + E\left[\sigma_{all}^2\left(1 - \frac{\sigma_{all}^2}{\sigma_{stg1}^2}\right)|C=1\right] \\ &= \left(\frac{\sigma_{all}^2}{\sigma_{stg1}^2}\right)^2 \text{Var}[\widehat{\beta}_{stg1}|C=1] + \sigma_{all}^2\left(1 - \frac{\sigma_{all}^2}{\sigma_{stg1}^2}\right) \\ &= \frac{\sigma_{\beta,all}^4}{\sigma_{stg1}^2} \left[1 + \left(\frac{h-\beta}{\sigma_{stg1}}\right)\left(\frac{\phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)}{1-\Phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)}\right) - \left(\frac{\phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)}{1-\Phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)}\right)^2\right] + \sigma_{all}^2\left(1 - \frac{\sigma_{all}^2}{\sigma_{stg1}^2}\right) \\ &= \sigma_{all}^2 + \frac{\sigma_{all}^4}{\sigma_{stg1}^2} \left[\left(\frac{h-\beta}{\sigma_{stg1}}\right)\left(\frac{\phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)}{1-\Phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)}\right) - \left(\frac{\phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)}{1-\Phi\left(\frac{h-\beta}{\sigma_{stg1}}\right)}\right)^2\right]. \end{aligned}$$

A.2. Illustration of the effect of early termination on $\widehat{\beta}_{all}$

To illustrate the effect of early stopping we consider estimation of $ROC(t)$ after a two-stage study that allows early termination for futility. $ROC(t)$ is the sensitivity at a specificity of $1 - t$. In this example, we consider $ROC(0.2)$, the sensitivity at a specificity of 0.8. The empirical estimate of $ROC(0.2)$ is asymptotically normal and has an independent increments covariance structure [16]. Consider a two stage study with 400 cases and 400 controls, an interim analysis using the first 200 cases and 200 controls, where the study terminates if the upper limit of the 95% confidence interval formed at the interim analysis is less than 0.7. Figure 1 presents the conditional asymptotic density of the empirical estimate of $ROC(0.2)$ for different true values of $ROC(0.2)$. The solid line represents the conditional asymptotic

density, while the dashed line represents the usual unconditional asymptotic normal density for the empirical estimate of $ROC(0.2)$. There is considerable difference between the conditional and unconditional density when $ROC(0.2)$ is equal to 0.55 but very little difference when $ROC(0.2)$ equals 0.65 and the probability of early termination is low.

Figure 2 presents the conditional mean of $\widehat{ROC}(0.2)_{all}$ under different design scenarios. We consider a two-stage study designs with 400 cases and 400 controls, one interim analysis using the first 30%, 40% or 50% of the observations where the study terminates if the upper limit of the 95% confidence interval formed at the interim analysis is less than 0.70 or 0.80. There is substantial bias for lower true values of $ROC(0.2)$ and the bias decreases as the true value of $ROC(0.2)$ increases. In contrast, the bias increases as the cut-off for continuing to full enrollment and the number of subjects used in stage 1 increase as they both increase the probability of early termination.

Finally, we illustrate the effect of early termination on the coverage probability of an unadjusted confidence interval. Consider the usual Wald confidence interval for $ROC(t)$,

$$\left(\widehat{ROC}(t)_{all} - Z_{\alpha/2} * \sigma_{\widehat{ROC}(t)_{all}}, \widehat{ROC}(t)_{all} + Z_{\alpha/2} * \sigma_{\widehat{ROC}(t)_{all}} \right). \quad (12)$$

We can calculate the conditional coverage of the Wald confidence interval using the conditional density from Section 2. Figure 3 presents the conditional coverage probability of the Wald confidence interval for $ROC(0.2)$ under the same design scenarios as Figure 2. The unadjusted Wald intervals provide adequate coverage for higher true values of $ROC(0.2)$, but coverage decreases as the true value of $ROC(0.2)$ decreases. It is unlikely that the study will reach full enrollment if $ROC(0.2) = 0.2$, for example, where the conditional coverage is 0 but the Wald interval also fails to provide adequate conditional coverage for intermediate values of $ROC(0.2)$. This illustrates that the unadjusted Wald interval fails to provide uniformly adequate conditional coverage. We expect that other unadjusted confidence intervals that do not account for the possibility of early terminate will also fail to provide adequate coverage conditional on reaching full enrollment.

References

1. Mazumdar M. Group sequential design for comparative diagnostic accuracy studies: Implications and guidelines for practitioners. *Medical Decision Making* SEP-OCT. 2004; 24(5):525–533. doi: 10.1177/0272898X04269240.
2. Mazumdar M, Liu A. Group sequential design for comparative diagnostic accuracy studies. *Statistics in Medicine*. 2003; 22(5):727–739. [PubMed: 12587102]
3. Tang L, Emerson SS, Zhou XH. Nonparametric and Semiparametric Group Sequential Methods for Comparing Accuracy of Diagnostic Tests. *Biometrics*. 2008; 64(4):1137–1145. [PubMed: 18371124]
4. Emerson SS, Fleming TR. Parameter estimation following group sequential hypothesis testing. *Biometrika*. 1990; 77(4):875–892. doi:10.1093/biomet/77.4.875. URL <http://biomet.oxfordjournals.org/cgi/content/abstract/77/4/875>.
5. Pepe MS, Feng Z, Longton G, Koopmeiners J. Conditional estimation of sensitivity and specificity from a phase 2 biomarker study allowing early termination for futility. *Statistics in Medicine*. 2009; 28(5):762–779. [PubMed: 19097251]
6. Whitehead J. On the bias of maximum likelihood estimation following a sequential test. *Biometrika*. 1986; 73(3):573–581. doi:10.1093/biomet/73.3.573. URL <http://biomet.oxfordjournals.org/cgi/content/abstract/73/3/573>.
7. Troendle JF, Yu KF. Conditional estimation following a group sequential clinical trial. *Communications in Statistics: Theory and Methods*. 1999; 28:1617–1634.

8. Liu A, Troendle JF, Yu KF, Yuan VW. Conditional maximum likelihood estimation following a group sequential test. *Biometrical Journal*. 2004; 46(6):760–768.
9. Whitehead, J. The design and analysis of sequential clinical trials. Ellis Horwood Ltd; 1983.
10. James Hung HM, Wang SJ, O'Neill RT. Methodological issues with adaptation of clinical trial design. *Pharmaceutical Statistics*. 2006; 5(2):99–107. doi:10.1002/pst.219. URL <http://dx.doi.org/10.1002/pst.219>. [PubMed: 17080766]
11. Wang Y, Li G, Shih W. Estimation and confidence intervals for two-stage sample-size-flexible design with lsw likelihood approach. *Statistics in Biosciences*. 2010; 2:180–190. URL <http://dx.doi.org/10.1007/s12561-010-9023-0>, 10.1007/s12561-010-9023-0.
12. Casella, G.; Berger, RL. *Statistical Inference*. Duxbury Press; 2002.
13. Tsiatis AA, Rosner GL, Mehta CR. Exact confidence intervals following a group sequential test. *Biometrics*. 1984; 40:797–803. [PubMed: 6518248]
14. Jennison C, Turnbull BW. Interim analyses: The repeated confidence interval approach (C/R: p344–361). *Journal of the Royal Statistical Society, Series B, Methodological*. 1989; 51:305–334.
15. Ohman-Strickland PA, Casella G. Conditional inference following group sequential testing. *Biometrical Journal*. 2003; 45(5):515–526.
16. Koopmeiners, JS. PhD Thesis. University of Washington; Seattle, Washington: 2009. Methods for group sequential diagnostic biomarker studies.
17. Jennison, C.; Turnbull, BW. *Group sequential methods with applications to clinical trials*. CRC Press Inc.; 2000.
18. Pepe, MS. *The statistical evaluation of medical tests for classification and prediction*, Oxford Statistical Science Series. Vol. vol. 28. Oxford University Press; Oxford: 2003.
19. Marrero JA, Feng Z, Wang Y, Nguyen MH, Befeler AS, Roberts LR, Reddy KR, Harnois D, Llovet JM, Normolle D, et al. [alpha]-fetoprotein, des-[gamma] carboxyprothrombin, and lectin-bound [alpha]-fetoprotein in early hepatocellular carcinoma. *Gastroenterology*. 2009; 137(1):110–118. [PubMed: 19362088]
20. Moskowitz CS, Pepe MS. Quantifying and comparing the predictive accuracy of continuous prognostic factors for binary outcomes. *Biostatistics*. 2004; 5(1):113–127. [PubMed: 14744831]
21. Zheng Y, Cai T, Pepe MS, Levy WC. Time-Dependent Predictive Values of Prognostic Biomarkers With Failure Time Outcome. *Journal of the American Statistical Association*. 2008; 103(481):362–368. [PubMed: 19655041]

Figure 1.

Conditional (solid line) and unconditional (dashed line) density of $\widehat{ROC}_{all}(0.2)$ for different true values of $ROC(0.2)$ in a two-stage study with 400 cases and 400 controls, an interim analysis using the first 200 cases and 200 controls where the study terminates if the upper limit of the 95% confidence interval formed at the interim analysis is less than 0.7

Figure 2.

Conditional mean of $\widehat{ROC}(0.2)_{all}$ in a two-stage study with 400 cases and 400 controls. An interim analysis was completed using 30%, 40% or 50% of the subjects and continued to full enrollment if the 95% confidence interval for $ROC(0.2)$ formed at the interim analysis was less than 0.70 or 0.80.

Figure 3. Conditional coverage probability (dashed lines) of the unadjusted Wald confidence interval for $\widehat{ROC}(0.2)_{all}$ compared to the unconditional coverage probability (solid lines) in a two-stage study with 400 cases and 400 controls. An interim analysis was completed using 30%, 40% or 50% of the subjects and continued to full enrollment if the 95% confidence interval for $ROC(0.2)$ formed at the interim analysis was less than 0.70 or 0.80.

Table 1

Simulation results illustrating the small-sample properties of the estimators presented in Section 2 using a sample size of n and an interim analysis using the first n_{stg1} observations. The study terminates if the upper limit of the 95% confidence interval formed at the interim analysis is less than 0.8. Shown are the mean(se) of estimated β for studies that reached full enrollment. 10000 simulated studies per scenario

True β	% Early Stopping	$\hat{\beta}_{all}$	$\hat{\beta}_{stg2}$	$\hat{\beta}_{md}$	$\hat{\beta}_{mn}$	$\hat{\beta}_{rb}$
$n = 20, n_{stg1} = 10$						
0.0	0.714	0.189 (0.180)	-0.001 (0.318)	-0.010 (0.282)	-0.015 (0.281)	0.000 (0.285)
0.2	0.479	0.319 (0.182)	0.202 (0.316)	0.187 (0.265)	0.181 (0.264)	0.199 (0.267)
0.4	0.244	0.467 (0.197)	0.402 (0.319)	0.390 (0.258)	0.385 (0.259)	0.402 (0.257)
0.6	0.094	0.630 (0.211)	0.604 (0.316)	0.592 (0.249)	0.587 (0.251)	0.601 (0.245)
0.8	0.027	0.808 (0.217)	0.797 (0.320)	0.792 (0.236)	0.789 (0.238)	0.798 (0.231)
1.0	0.004	1.001 (0.223)	0.997 (0.315)	0.996 (0.230)	0.994 (0.232)	0.999 (0.227)
1.2	0.000	1.200 (0.223)	1.199 (0.316)	1.199 (0.225)	1.199 (0.225)	1.200 (0.223)
$n = 40, n_{stg1} = 20$						
0.0	0.947	0.228 (0.111)	0.001 (0.212)	-0.004 (0.193)	-0.007 (0.192)	0.000 (0.195)
0.2	0.761	0.351 (0.124)	0.206 (0.223)	0.201 (0.198)	0.198 (0.197)	0.208 (0.200)
0.4	0.439	0.477 (0.132)	0.399 (0.224)	0.391 (0.188)	0.387 (0.188)	0.400 (0.189)
0.6	0.146	0.628 (0.145)	0.598 (0.224)	0.591 (0.179)	0.587 (0.180)	0.598 (0.177)
0.8	0.027	0.805 (0.154)	0.798 (0.226)	0.794 (0.168)	0.792 (0.169)	0.798 (0.164)
1.0	0.001	1.001 (0.156)	1.002 (0.225)	0.999 (0.159)	0.998 (0.160)	1.000 (0.158)
1.2	0.000	1.199 (0.161)	1.200 (0.229)	1.199 (0.162)	1.199 (0.162)	1.199 (0.161)
$n = 60, n_{stg1} = 30$						
0.0	0.992	0.246 (0.095)	0.005 (0.186)	-0.011 (0.174)	-0.012 (0.174)	-0.008 (0.175)

True β	% Early Stopping	$\hat{\beta}_{all}$	$\hat{\beta}_{stg2}$	$\hat{\beta}_{md}$	$\hat{\beta}_{mn}$	$\hat{\beta}_{rb}$
0.2	0.906	0.370 (0.096)	0.214 (0.179)	0.206 (0.164)	0.204 (0.163)	0.211 (0.165)
0.4	0.587	0.487 (0.106)	0.398 (0.182)	0.393 (0.160)	0.390 (0.159)	0.400 (0.161)
0.6	0.196	0.631 (0.116)	0.600 (0.183)	0.593 (0.147)	0.590 (0.147)	0.600 (0.146)
0.8	0.025	0.804 (0.127)	0.798 (0.184)	0.795 (0.138)	0.793 (0.139)	0.799 (0.135)
1.0	0.002	0.999 (0.129)	0.997 (0.185)	0.998 (0.131)	0.997 (0.132)	0.999 (0.130)
1.2	0.000	1.199 (0.131)	1.198 (0.186)	1.199 (0.131)	1.199 (0.131)	1.199 (0.131)

Table 2

Simulation results for conditional estimation of $ROC(t)$ using 400 cases and 400 controls, an interim analysis using the first 200 cases and 200 controls where the study terminates if the upper limit of the 95% confidence interval formed at the interim analysis is less than 0.7. Shown are the mean(se) of estimated $ROC(0.2)$ for studies that reached full enrollment. 10000 simulated studies per scenario

True ROC(0.2)	% Early Stopping	$\widehat{ROC}(0.2)_{all}$	$\widehat{ROC}(0.2)_{stg2}$	$\widehat{ROC}(0.2)_{md}$	$\widehat{ROC}(0.2)_{mn}$	$\widehat{ROC}(0.2)_{rb}$
0.550	80.8	0.588 (0.031)	0.555 (0.053)	0.548 (0.050)	0.547 (0.050)	0.549 (0.051)
0.575	65.7	0.603 (0.031)	0.578 (0.053)	0.573 (0.048)	0.572 (0.048)	0.574 (0.049)
0.600	46.1	0.620 (0.031)	0.603 (0.052)	0.598 (0.045)	0.597 (0.045)	0.600 (0.046)
0.625	28.3	0.638 (0.031)	0.627 (0.051)	0.623 (0.043)	0.622 (0.043)	0.625 (0.043)
0.650	13.9	0.657 (0.032)	0.652 (0.050)	0.648 (0.040)	0.647 (0.041)	0.650 (0.040)
0.675	5.4	0.679 (0.033)	0.678 (0.048)	0.675 (0.038)	0.674 (0.038)	0.676 (0.037)
0.700	1.8	0.703 (0.033)	0.702 (0.047)	0.701 (0.035)	0.700 (0.036)	0.701 (0.035)
0.725	0.5	0.726 (0.032)	0.728 (0.046)	0.725 (0.033)	0.725 (0.034)	0.726 (0.033)
0.750	0.0	0.751 (0.032)	0.752 (0.044)	0.751 (0.032)	0.751 (0.032)	0.751 (0.032)

Table 3

Simulation results for conditional estimation of $ROC(t)$ using 400 cases and 400 controls, an interim analysis using the first 200 cases and 200 controls where the study terminates if the upper limit of the 95% confidence interval formed at the interim analysis is less than 0.7. Shown are the conditional coverage probabilities for studies that reached full enrollment. 10000 simulated studies per scenario

True ROC(0.2)	% Early Stop	Unadjusted Interval		Adjusted Interval		Width Ratio
		Cov	Ave Width	Cov	Ave Width	
0.550	80.8	0.850	0.143	0.933	0.182	1.27
0.575	65.7	0.898	0.142	0.938	0.176	1.24
0.600	46.1	0.934	0.140	0.940	0.169	1.21
0.625	28.3	0.950	0.139	0.942	0.162	1.17
0.650	13.9	0.953	0.136	0.940	0.154	1.13
0.675	5.4	0.951	0.134	0.943	0.145	1.08
0.700	1.8	0.951	0.130	0.947	0.136	1.05
0.725	0.5	0.942	0.126	0.943	0.129	1.02
0.750	0.0	0.939	0.122	0.942	0.123	1.01

Table 4

Simulation results for conditional estimation of $PPV(u)$ using 200 cases and 200 controls, an interim analysis using the first 100 cases and 100 controls where the study terminates if the upper limit of the 95% confidence interval formed at the interim analysis is less than 0.8. Shown are the mean(se) of estimated $PPV(0.9)$ for studies that reached full enrollment. 10000 simulated studies per scenario

True $PPV(0.9)$	% Early Stopping	$\widehat{PPV}(0.9)_{all}$	$\widehat{PPV}(0.9)_{stg2}$	$\widehat{PPV}(0.9)_{md}$	$\widehat{PPV}(0.9)_{mn}$	$\widehat{PPV}(0.9)_{rb}$
0.55	79.7	0.612 (0.057)	0.552 (0.095)	0.549 (0.089)	0.547 (0.089)	0.551 (0.090)
0.60	60.8	0.645 (0.057)	0.603 (0.095)	0.600 (0.084)	0.598 (0.084)	0.603 (0.084)
0.65	38.7	0.680 (0.057)	0.654 (0.092)	0.650 (0.078)	0.649 (0.078)	0.654 (0.078)
0.70	19.0	0.716 (0.059)	0.704 (0.092)	0.697 (0.074)	0.696 (0.075)	0.701 (0.074)
0.75	6.3	0.759 (0.060)	0.755 (0.090)	0.750 (0.070)	0.749 (0.070)	0.753 (0.069)
0.80	1.4	0.805 (0.059)	0.807 (0.086)	0.801 (0.064)	0.800 (0.065)	0.803 (0.063)
0.85	0.2	0.852 (0.056)	0.854 (0.080)	0.851 (0.058)	0.850 (0.059)	0.852 (0.058)

Table 5

Simulation results for conditional estimation of $PPV(u)$ using 200 cases and 200 controls, an interim analysis using the first 100 cases and 100 controls where the study terminates if the upper limit of the 95% confidence interval formed at the interim analysis is less than 0.8. Shown are the conditional coverage probabilities for studies that reached full enrollment. 10000 simulated studies per scenario

True PPV(0.9)	% Early Stop	Unadjusted Interval		Adjusted Interval		Width Ratio
		Cov	Ave Width	Cov	Ave Width	
0.55	79.7	0.835	0.245	0.919	0.312	1.27
0.60	60.8	0.879	0.246	0.926	0.303	1.23
0.65	38.7	0.945	0.245	0.934	0.292	1.19
0.70	19.0	0.946	0.242	0.924	0.280	1.15
0.75	6.3	0.943	0.236	0.925	0.263	1.11
0.80	1.4	0.935	0.226	0.929	0.246	1.09
0.85	0.2	0.934	0.211	0.941	0.226	1.07