Conditional estimation methods for the analysis of a group sequential experiment

Xiaoyin (Frank) Fan, Ph.D.
Conditional Estimation Methods for the Analysis of a Group Sequential Experiment

Xiaoyin (Frank) Fan

August 11, 2000

SUMMARY. Test statistics from a group sequential experiment can often be approximated as obtained from a Brownian motion with some drift parameter. In this paper we propose a new methodology to estimate the drift parameter. Although traditional method could provide little biased estimators or exact confidence intervals, when looking at individual stopping stage those estimators may still have very prominent bias and confidence intervals provide very poor coverage probability. Conditional methods to obtain conditional bias reduced estimators and exact conditional confidence intervals as proposed in Fan, DeMets & Lan (2000) and Fan (2000) are summarized in this paper as well as some simulation results comparing the conditional and some unconditional methods. The proposed methods always provide “future independent” solutions. Furthermore, a recent metoprolol clinical trial MERIT-HF is used as an example of applying the new methods.

KEY WORDS: Group sequential analysis; Brownian motion; Stopping time; Conditional estimation: Bias; Exact confidence interval; Maximum likelihood estimation; Future independent.
1 Introduction and Background

In phase III clinical trials with interim analyses, group sequential procedures have often been used to protect patients from undue harm or unnecessary exposure to an inferior drug, procedure or device. Because of the possibility of stopping the trial before the scheduled end, the group sequential design improves from the traditional fixed sample design in the aspect of economy as well as ethics. Early methods required fixed number of analyses to be specified in advance or equal numbers of subjects or events between analyses. However, in the group sequential procedures which might allow for unscheduled interim analysis, a flexible Lan-DeMets method using $\alpha$-spending function can be used to construct sequential boundaries while still protecting the overall Type I error (Lan & DeMets (1983)). Examples of the trials using group sequential design include $\beta$-Blocker Heart Attack Trial (BHAT), Vesmarinone Trial (VEST), PROMISE study and MERIT-HF trial, etc.

Different test statistics have been developed to test hypotheses or assess treatment effects such as the Wald and score statistics in the semi or full parametric models and the rank statistics in non-parametric models. When sequentially observed, those statistics have asymptotic independent increment structure and in large sample case behave approximately as a Brownian motion with some drift parameter (Tsiatis (1982), Sellke & Siegmund (1983), Slud (1984)). It is well known that because of the dependency between the drift parameter and the stopping time, the intuitive MLE and na"ive confidence interval by pretending the trial as a fixed sample design are generally biased (e.g. Kim(1989)). So by applying the group sequential boundary we certainly gain the convenience and flexibility in hypothesis testing. But meanwhile we lose the unbiasedness of the intuitive estimator and consequently that raises a lot of difficulty and complication in estimating treatment parameter as we will discuss later.

A number of authors have addressed the issue of estimation following the termination of group sequential test. By using different ordering of the sample space, many statisticians and researchers

It's interesting to notice that the issues of bias and confidence level those researchers studied are the overall bias of the MLE and overall coverage probability which do not distinguish different individual stopping stages. Knowing that the parameter estimation is only needed after the trial is stopped, we feel that at that time it is more relevant to consider the conditional bias and confidence level (upon the stopping time) as the subject than the overall ones. Surprisingly, by knowing the stopping stage, those overall exact confidence intervals tend to have significantly different coverage probabilities than the nominal level. Similarly, the conditional biases of the earlier proposed estimators are still remarkable in spite of the little or even zero overall bias.

In a paper by Fan, DeMets and Lan (2000) they proposed a conditional bias adjusting method and it was followed by another paper. Fan (2000), in which conditional confidence interval methods are proposed which generate confidence intervals with exact coverage probabilities at any stopping stage as well as in overall. The conditional point estimates can greatly reduce the bias at each stopping stage especially at very early ones. This paper summarizes the results from the above two papers. After the theoretical part is derived, simulation results and the comparisons between the conditional methods and some unconditional ones will also be presented. It's well known that one of the strengths of using the Lan-DeMets $\alpha$-spending function in group sequential design is that we don't have to decide the number or timing of the interim analyses in advance. Based on the same reason, we don't want our estimates or confidence intervals to depend on the time or boundaries of those interim after the stopping stage has been reached, because they may not have been determined.
yet at that time. Furthermore, intuitively those planned but not happening interim are irrelevant to the current observations and thus should not affect our estimation. This desirable property is termed as “future independent” and will be further addressed in this paper. Finally, a recent cardiovascular clinical trial (MERIT-HF) will be analyzed as an example of applying the proposed conditional methods.

2 Brownian Motion Notation

Since as mentioned before, many statistics obtained from group sequential designed trials can be approximated as from a Brownian motion at certain discrete time, the first part of this paper will be in the context of Brownian motion. To ease the reading and avoid some confusion of details, we’ll explain some notation used in this paper as the following:

- \( B(t), \ t \in [0, 1] \) denotes a standard Brownian motion and \( S_\theta(t) = B(t) + \theta t \) is the family of Brownian motion processes with the drift parameter \( \theta \).

- \( 0 = t_0 < t_1 < \cdots < t_M = 1 \) are the time at which interim analyses are to be performed and \( M \) is the number of the pre-planned interim.

- \( S_k := S_\theta(t_k), \ k = 0, 1, \ldots, M. \)

- \( X_k := S_k - S_{k-1}, \ k = 1, 2, \ldots, M. \)

- \( (b^*_k, c^*_k) \) is the acceptance region for \( S_k, \ k = 1, 2, \cdots, M, \) where the endpoints can be \( \pm \infty \).

- \( A_k = (b^*_k, c^*_k) \) if \( k < M \) and the empty set \( \phi \) if \( k = M \) is the continuation region for \( S_k \).

- \( \eta := \min\{k \leq M : S_k \not\in A_k\} \) where \( t_\eta \) is the time to stop the process \( S_\theta(t) \).

- \( \hat{\theta} := S_\eta/t_\eta, \) the usual maximum likelihood estimator of the true \( \theta \).
• $b(\theta) := E_{\hat{\theta}}(\hat{\theta}) - \theta$ is the bias of $\hat{\theta}$.

• similarly $b(\theta|\eta) := E_{\hat{\theta}(\hat{\theta}|\eta)} - \theta$ is the bias of $\hat{\theta}$ conditional on the stopping stage $\eta$ or, precisely, the event $\{X_\eta \notin A_\eta, \ X_k \in A_k, \ k < \eta\}$.

Some theoretical notes:

1. By the nature of $S_\theta(t)$, $X_1, X_2, \ldots, X_M$ is an independent sequence.

2. The total observations are $\{\eta, S_1, S_2, \ldots, S_\eta\}$ or, equivalently, $\{\eta, X_1, X_2, \ldots, X_\eta\}$

3. Suppose a spending function $\alpha(t), \ t \in [0, 1]$ is used which is increasing with $\alpha(0) = 0$ and $\alpha(1) = \alpha$. $\{(b_k^*, c_k^*), \ k = 1, 2, \ldots, M\}$ can be calculated numerically by solving sequential equations:

$$P_{\theta} [S_1 \in (b_1^*, c_1^*), \ldots, S_{k-1} \in (b_{k-1}^*, c_{k-1}^*), \ S_k \notin (b_k^*, c_k^*)|H_0] = \alpha(t_k) - \alpha(t_{k-1}).$$

Usually we require $0 \in (b_k^*, c_k^*)$ for $k = 1, 2, \ldots, M$.

3 Conditional Point Estimations

It’s known that the usual MLE $\hat{\theta}$ tends to exaggerate the true value of $\theta$ on average. However, that is not always true for every stopping stage. As a matter of fact, the conditional bias of $\hat{\theta}$ can be extraordinarily large when the trial is stopped very early. On the other hand, it also could be negative which means $\hat{\theta}$ will underestimate the true $\theta$ and sometime seriously at certain stages, say the trial is stopped very late or the treatment effect is huge.

As an example, a two-sided level $\alpha = 0.05$ O'Brien-Fleming(OFB) type group sequential boundary is considered which includes five interim looks to monitor $S_\theta(t)$, a standard Brownian motion
with drift parameter \( \theta \), at time 0.2, 0.4, 0.6, 0.8 and 1.0. Figure A.1 shows the conditional biases \( b(\hat{\theta}|\eta = k) \), \( k=1,2,3,4,5 \) and the marginal bias of \( \hat{\theta} \).

Because the symmetric boundary is used and thus the biases are also symmetric at \( \theta = 0 \), only the part where \( \theta \geq 0 \) is shown in the individual plots. Practically we're mostly interested in the areas where \( \theta \) is less than 5 or 6. The larger values of \( \theta \) are demonstrated for theoretical interest. From those graphs we can see that although the overall bias only ranges from 0 to 0.45 depending on the true value of the \( \theta \), the conditional bias could be as high as 9 or as low as -5 (not shown). The bias graphs \( b(\hat{\theta}|\eta) \)'s differ dramatically as the trial is stopped at different interim stages. Compared with the conditional biases, the overall bias \( b(\hat{\theta}) \) in the lower right corner does not look impressive at all. In fact, \( b(\hat{\theta}) = \sum_k b(\theta|\eta = k)P(\eta = k) \) which means that the overall bias is simply a weighted average of all the conditional biases with the stopping probability at each interim stage as the weight function. In addition, we can show that it is a general result under the symmetric sequential boundary that the bias is always negative at the last stage and positive at the first stage.

To minimize the conditional bias at each possible stopping stage, a proposed conditional estimator \( \hat{\theta} \) has three equivalent forms of conditional maximum likelihood estimate (CMLE), conditional moment estimate (CME) and conditional bias reduction estimate (CBRE) as defined next.
Conditional maximum likelihood estimate (CMLE)

Denoted by $\hat{\theta}_1$, the CMLE maximizes the likelihood of the statistics $(X_1, \ldots, X_\eta)$ conditional on the stopping stage $\eta = n$. Mathematically,

$$
\hat{\theta}_1 = \text{arg max}_\theta f_\theta(X_1, \ldots, X_\eta | \eta = n)
= \text{arg max}_\theta \frac{f_\theta(X_1, \ldots, X_n)}{P_\theta(\eta = n)}
= \text{arg max}_\theta L_{\text{cond}}.
$$

(3.1)

where $L_{\text{cond}} = \log (f_\theta (X_1, \ldots, X_n)) - \log (P_\theta (\eta = n))$ and $f_\theta$ represents the density function.

Equivalently, $\hat{\theta}_1$ is a solution to the following conditional likelihood equation:

$$
\frac{\partial}{\partial \theta} L_{\text{cond}} = \frac{\partial}{\partial \theta} \log (f_\theta (X_1, \ldots, X_n)) - \frac{\partial}{\partial \theta} \frac{P_\theta (\eta = n)}{P_\theta (\eta = n)} = 0
$$

(3.2)

Conditional moment estimate (CME)

Denoted by $\hat{\theta}_2$, the CME is the solution of $\theta$ to a conditional moment estimate equation:

$$
E_\theta[S_\eta | \eta = n] = s_n
$$

(3.3)

or equivalently,

$$
E_\theta[\theta | \eta = n] = \hat{\theta}_n
$$
Conditional bias reduction estimate (CBRE)

Denoted by $\hat{\theta}_3$, CBRE is obtained by subtracting an estimate of the conditional bias $b(\theta|\eta)$ from MLE $\hat{\theta}$. More specifically $\hat{\theta}_3$ is the solution to the following equation (in $\theta$):

$$\theta = \hat{\theta} - b(\theta|\eta)$$  \hspace{1cm} (3.4)

It’s obvious that equation (3.3) and (3.4) are equivalent. Based on Whitehead (1986), Pinheiro and DeMets (1995) provided a useful formula showing that under Gaussian independent increment structure (GIIS) the drifted Brownian motion $S(t) = B(t) + \theta t$ satisfies:

$$\frac{\partial}{\partial \theta} E_{\theta} [g(S(t_{\eta}), t_{\eta})] = E_{\theta} [(S(t_{\eta}) - \theta t_{\eta}) g(S(t_{\eta}), t_{\eta})]$$  \hspace{1cm} (3.5)

where $\eta$ is stopping time as defined before and $g$ is a real function such that $E_{\theta}[|g(S(t_{\eta}), t_{\eta})|] < \infty$ for any $\theta$.

By applying the above formula with $g$ selected to be the indicator function $I_{\{\eta = n\}}$, it can be shown that the equation (3.3) and (3.2) are also equivalent to (3.4). After showing that the equation (3.3) has a unique solution (see Fan, DeMets & Lan (2000)), we conclude that the three conditional estimators CME, CMLE and CBRE are identical and uniquely exist and thus they will be denoted by just $\hat{\theta}$ hereafter. That also provides us convenience of using any of above equations to solve for the $\hat{\theta}$.

To further investigate the conditional estimator $\hat{\theta}$, a simulation study is performed in the same scenario as the last example where the O'Brien-Fleming type group sequential boundary is used with five interim looks at time 0.2, 0.4, 0.6, 0.8 and 1.0 and the overall type I error rate $\alpha = .05$. The Whitehead’s bias adjusted estimator (1986) which is denoted by $\bar{\theta}$ and defined as the solution to the equation (in $\theta$) $\theta = \hat{\theta} - b(\theta)$ is also studied for comparison. In addition, two modification
of the \( \hat{\theta} \) are included in the simulation. The first one, denoted by \( \hat{\theta}' \), takes the value of \( \hat{\theta} \) or \( \bar{\theta} \), whichever is closer to zero. By doing that, we don’t inflate the MLE \( \hat{\theta} \) but do shrink it if necessary. Precisely,

\[
\hat{\theta}' = \begin{cases} 
\bar{\theta}, & \text{when } |\bar{\theta}| > |\hat{\theta}| \\
\hat{\theta}, & \text{otherwise}
\end{cases}
\]

The other modified estimator, denoted by \( \hat{\theta}'' \), differs from \( \hat{\theta}' \) only where \( |\bar{\theta}| > |\hat{\theta}| \) and when early stopping happens. Suppose the upper boundary is early crossed, i.e. \( s_\eta > c^*_\eta(> 0) \) and \( \eta < M \), let \( \theta_{\eta 0} \) be the solution to the equation (in \( \theta \)): \( E_\theta[\theta_{\eta}]=c^*_\eta \). Let \( \theta_{\eta 1} = \max(0, \theta_{\eta 0}) \) and \( \theta_{\eta 2}(> \theta_{\eta 1}) \) satisfy \( E_{\theta_{\eta 2}}[\bar{\theta}|\eta]=\theta_{\eta 2} \) if it exists. Then for the case \( s_\eta > 0 \),

\[
\hat{\theta}'' = \begin{cases} 
\hat{\theta}, & \text{if } \eta = M \text{ or } \hat{\theta} > \theta_{\eta e} \\
\theta_{\eta e} - \frac{(\theta_{\eta e} - \hat{\theta})(\theta_{\eta e} - \theta_{\eta m})}{\theta_{\eta e} - c^*_\eta / \eta}, & \text{if } 1 < \eta < M \text{ and } \hat{\theta} < \theta_{\eta e} \\
\hat{\theta}, & \text{if } \eta = 1
\end{cases}
\]

When \( s_\eta < 0 \), the other part of \( \hat{\theta}'' \) can be similarly defined.

Table A.1 shows the result of the bias, variance and mean squared error of different estimators from an 100,000 run simulation study where the true \( \theta = 3 \) which gives power of 83.9% to reject the null hypothesis. The stopping probability at each stage is also listed. We can see that in this case the conditional estimator \( \hat{\theta} \) has smaller bias than the usual MLE and Whitehead’s estimator in all five interim stages and even overall. It performs extremely well when the trial is stopped at early trial. It’s also noticed that the price \( \hat{\theta} \) paid for the small bias is its enlarged variance. However, by incorporating the unconditional MLE \( \hat{\theta} \) into \( \bar{\theta} \), the two modified versions significantly reduce the variance from \( \hat{\theta} \) without changing the bias much. Similar phenomena are observed when different
values of the true $\theta$ is used for the simulation. In Table A.2 and A.3 the $\theta$ is chosen to be 1 and 5 respectively. The first case has a small power with the rejection probability of only 16.5%. The power in the second case is much higher which is about 99.9%. Although the Whitehead’s estimator $\hat{\theta}$ may have the smallest overall bias, its bias at every stopping stage is much more significant than the three conditional estimators. In addition, it always worsens the underestimation caused by the usual MLE $\hat{\theta}$ at the last stopping stage. So looking at the overall bias alone might be misleading. As to the modified estimators, $\hat{\theta}'$ tends to have both small bias and variance when the true $\theta$ is small, i.e. small power case. In general, the $\hat{\theta}''$ has the most satisfactory performance at all scenarios in term of the mean squared error. Emerson (1993) proposed an unbiased estimator of $\theta$ which is in the form of $E[\hat{\theta}_1|\eta, S_\eta]$. It’s not studied in the simulation because of its computational complexity. However, since it doesn’t adjust from the $\hat{\theta}$ when the trial is stopped at the first stage (i.e. $\eta = 1$), we expect it to have the same problem of the prominent conditional biases in spite of the zero overall.

From the definition, the conditional estimator $\hat{\theta}$ depends only on the information of the interim stages at the stopping time or earlier. We call a parameter estimator future independent if it does not depend on the number, timing and boundary values of those interim stages which were planned but not used after the experiment is early stopped. Intuitively it is a very desirable feature because those future yet happened interim analyses have nothing to do with the trial going on so far and thus should not affect the estimation at current stage. Furthermore, in flexible trials the future interim stages may not be specifically determined at first. They could be changed in either the number or the timing and thus the dependence of estimation upon the future information is difficult and unnatural to explain. Both the $\hat{\theta}$ and the MLE $\hat{\theta}$ are future independent and so are $\hat{\theta}'$ and $\hat{\theta}''$ because they are functions of only $\hat{\theta}$ and $\hat{\theta}$. The Whitehead’s estimator $\hat{\theta}$ is not because it targets on the overall bias. Also because of the future independence property, the $\hat{\theta}$ is computationally
easier to calculate than $\tilde{\theta}$ when early stopping happens.

4 Conditional Exact Confidence Interval

Similar to the biased estimator $\hat{\theta}$, the naive $100(1 - 2\alpha)\%$ level confidence interval of $\hat{\theta} \pm Z_\alpha / \sqrt{T_\eta}$ is also biased, i.e. the coverage probability is different from $1 - 2\alpha$. By using the same idea of conditioning on the stopping time, a level $100(1 - 2\alpha)\%$ conditional exact confidence interval (CECI) $(\theta^L_\eta, \theta^U_\eta)$ can be defined as:

\begin{align*}
\theta^L_\eta &= \sup \{ \theta : P_{\theta}(S_\eta > s_\eta|\eta) \leq \alpha \} \\
\theta^U_\eta &= \inf \{ \theta : P_{\theta}(S_\eta > s_\eta|\eta) \geq 1 - \alpha \}
\end{align*} (4.1) (4.2)

Claim 4.1 Notation as before. The definition 4.2 and 4.2 uniquely define an exact confidence interval $(\theta^L_\eta, \theta^U_\eta)$ of the drift parameter $\theta$ such that for any stopping stage $n \leq M$

\[ P_{\theta} (\theta \in (\theta^L_\eta, \theta^U_\eta) | \eta = n) = 1 - 2\alpha \]

and $P_{\theta} (\theta \in (\theta^L_\eta, \theta^U_\eta)) = 1 - 2\alpha$

Proof: See Fan (2000).

So the CECI not only has overall coverage probability of exactly $1 - 2\alpha$, but also has the same coverage probability at any individual stopping stage. Besides, from the proof we know that $S_\eta$ is stochastically ordered conditional on the stopping stage $\eta$ and the CECI is uniquely defined. The endpoints $\theta^L_\eta$ and $\theta^U_\eta$ can be solved from the following equations

\[ P_{\theta^L_\eta}(S_\eta > s_\eta|\eta) = \alpha \]

and $P_{\theta^U_\eta}(S_\eta > s_\eta|\eta) = 1 - \alpha$
Conventionally an exact confidence set or interval following a sequential design is formulated by inverting a hypothesis test of the same level to preserve the confidence level, i.e., if

\[ A(\theta_0) = \{(n, s) : \alpha < P_{\theta_0}(g_{\theta_0}(\eta, S_\eta) > g_{\theta_0}(n, s)) < 1 - \alpha \} \]

is an acceptance region of a size $2\alpha$ test for $H_0 : \theta = \theta_0$ vs. $H_A : \theta \neq \theta_0$ where $g_{\theta_0}$ is a function of the statistics defining an order in the two-dimensional space $(1, 2, \ldots , M) \times \mathcal{R}$, then the $100(1 - 2\alpha)\%$ level confidence interval is defined as

\[ C(\eta = n, S_\eta = s) = \{ \theta : (\eta = n, S_\eta = s) \in A(\theta) \} = \{ \theta : \alpha < P_{\theta}(g_{\theta}(\eta, S_\eta) > g_{\theta}(n, s)) < 1 - \alpha \} \]  \hspace{1cm} (4.3)

Since the sufficient statistic $(\eta, S_\eta)$ is a two-dimensional variable, different types of the ordering of $(\eta, S_\eta)$, or equivalently different function $g_{\theta}$ will result in different confidence set. They include the intuitive ordering like in Siegmund (1978), likelihood ratio ordering in Rosner & Tsiatis (1988) and sample mean ordering in Emerson & Fleming (1990). Examples of the $g_{\theta}$ for those ordering are:

intuitive ordering : $g_{\theta}(\eta, S) = (M - \eta) \cdot \text{sign}(S) + \frac{1}{1 + e^{-S}}$.

likelihood ratio ordering : $g_{\theta}(\eta, S) = \sqrt{t_\eta} \left( \frac{S}{t_\eta} - \theta \right)$.

sample mean ordering : $g_{\theta}(\eta, S) = \frac{S}{t_\eta}$.

Unfortunately, there's still no general agreement on which type of the above ordering is most reasonable or which consequent confidence set is the best one. The CECI can also be written in the following form which is similar to (4.3)

\[ (\theta_L, \theta_U) = \{ \theta : \alpha_1 \leq P_{\theta}(S_\eta > s_\eta|\eta = n) \leq 1 - \alpha_2 \}. \] \hspace{1cm} (4.4)
It’s easy to see that if using conditional probability in the equation (4.3) with any of the three $g_p$’s described before, it becomes equivalent to the equation (4.4). Although those three exact unconditional methods are different, they will merge into the same CECI when conditional on stopping time.

One of the advantages of using the CECI method over those unconditional ones is that by conditioning on the stopping time we circumvent the confusion from defining an order in the two-dimensional sample space. Because of the conditional stochastic ordering property, the conditional method guarantees a confidence interval as result. The likelihood ratio ordering doesn’t have this property and several instances were detected though very rare that the confidence set was not interval (Emerson & Fleming (1990)).

The earlier simulation study is continued here to compare the performance of the different confidence intervals discussed before. In addition, another type of confidence interval called repeated confidence interval (RCI) proposed by Jennison and Turnbull (1989) is also tested. The RCI is constructed by using $S_k/t_k \pm d_k/\sqrt{t_k}$ as the $k$-th RCI of level $100(1-2\alpha)\%$ where $d_k$, $k = 1, \ldots, M$ are the standardized symmetric boundary values (usually the O’Brien-Fleming type) of a size $2\alpha$ sequential test. The nominal confidence level is set at 90%. Again, each 100,000-run simulation is generated for each case of $\theta = 1.3$ and 5 and the results are listed in Table A.4, A.5 and A.6.

From those tables we notice that in overall the RCI is unnecessarily conservative. The naive CI does not have the right coverage probability but it’s not too much off the 90% target. Those various exact confidence intervals have the right 90% overall coverage probability. However, when looking at individual stopping stage, only the CECI still maintains the nominal level no matter the true $\theta$. Others could have the conditional coverage probability anywhere from 0% to 100%. Generally, their coverage probability is too low (i.e. interval too narrow or off the target) at the less likely stopping stages but too high (i.e. interval too wide) at the more likely stopping stages.
The CECI is the best confidence interval in terms of the conditional coverage probability.

If we use $f_{k, \theta}(s)$ to denote the density of $S_k$, it's true (e.g., Emerson & Fleming (1990)) that the $f_{k, \theta}$ satisfies:

$$f_{k, \theta}(s) = f_{k, 0}(s) \exp(s\theta - t_k \theta^2/2).$$

So the conditional density of $S_\eta$ given $\eta$ satisfies

$$f_\theta(s_\eta|\eta) \propto f_{\eta, 0}(s_\eta) \exp(s_\eta \theta)$$

and it has a monotone likelihood ratio in $S_\eta$. Using a corollary of Lehmann (1986, p91) in the conditional context, the boundary values of the CECI discussed above $\theta^L_\eta$ and $\theta^U_\eta$ are the conditional uniformly most accurate (CUMA) lower and upper confidence bound respectively at the confidence level $100(1 - \alpha)\%$.

In addition, by applying Lehmann (1986, p220) to the conditional distribution of $S_\eta$ given $\eta$, a level $100(1 - 2\alpha)\%$ conditional uniformly most accurate unbiased (CUMAU) confidence interval can be defined as

$$C_{L, \eta}^{-1}(S_\eta) \leq \theta \leq C_{U, \eta}^{-1}(S_\eta) \quad (4.5)$$

where the $C_{L, \eta} < C_{U, \eta}$ are strictly increasing functions: $\theta \rightarrow S_\eta$ such that

$$P_\theta[S_\eta < C_{L, \eta}(\theta)|\eta] + P_\theta[S_\eta > C_{U, \eta}(\theta)|\eta] = 1 - 2\alpha$$

and $E_\theta[S_\eta \cdot (I_{S_\eta < C_{L, \eta}(\theta)} + I_{S_\eta > C_{U, \eta}(\theta)})|\eta] = E_\theta[S_\eta|\eta] \cdot 2\alpha.$

The CUMAU has the same coverage probability as the CECI at each stopping time and in overall and both of them are "future independent". However, it’ll be more difficult to compute the
So far we have discussed the conditional methods to obtain the point estimates and confidence intervals in group sequential experiment. They have many advantages over the traditional unconditional methods. Furthermore, it also needs to be pointed out that the conditional methods are not restricted to the symmetric or two-sided sequential boundary case which gives the methods much flexibility. They can be applied to asymmetric or even one-sided sequential boundaries too and all of the nice properties previously discussed still hold. The methods can also be applied to survival analysis without much difficulty.

5 Analysis of MERIT-HF Trial

Metoprolol is a lipophilic $\beta_1$-selective antagonist with no intrinsic sympathomimetic activity. Earlier results suggested that $\beta$-blocker including metoprolol could increase survival in patients with chronic heart failure. The MERIT-HF was a large-scale double-blinded randomized placebo-controlled trial to investigate whether metoprolol controlled release/extended release (CR/XL) once daily added to optimum standard therapy lowers mortality in patients with decreased ejection fraction and symptoms of heart failure.

A total of 3991 patients were enrolled in the MERIT-HF trial between February 1997 and April 1998. 1990 of those patients were randomly assigned into the treatment (metoprolol CR/XL) group and the other 2001 patients into the placebo group. The study was designed based on a significance level of $\alpha = 0.04$ for all-cause mortality and a power of at least 80%. Safety was monitored by an independent safety committee during the study. The predefined stopping rule for efficacy was based on all-cause mortality, analyzed by intention to treat, with three predetermined interim analyses, done when 25%, 50% and 75% of the expected total deaths had occurred. An asymmetric group sequential procedure was used to monitor the primary outcomes. A Haybittle-Peto type boundary
was used for monitoring the positive trend. The four critical Z-values at the four planned analyses are 3.04, 2.98, 2.93 and 2.05 with the cumulative crossing probabilities 0.0012, 0.0024, 0.0036 and 0.0215. The uniform type Lan-DeMets α-spending function was used to calculate the lower boundary Z-values for negative trend. They are -2.58, -2.49, -2.41 and -2.34 with the cumulative crossing probabilities 0.005, 0.010, 0.015 and 0.020. The upper boundary is more conservative than the lower boundary in the first three planned looks. The result of the MERIT-HF turned out to be that the trial was stopped on Oct 31, 1998 after the second pre-planned interim analysis where the upper boundary was exceeded ($Z_2 = 3.92$ vs. the bound of 2.98). The boundary values and observed standardized log-rank statistics are plotted in Figure A.2 and the Kaplan-Meier for the cumulative mortality rates of the two arms are plotted in Figure A.3.

The trial benefited from the group sequential design because it was stopped at only the half way to the pre-scheduled end with significantly positive results. The actual average patient follow-up time was 1 year vs. the planned 2.4 years shows that much time and resources have been saved. Now how to estimate the treatment effect of metoprolol becomes an intriguing question. Naively, we can pretend the trial is a fixed sample experiment and ignore the interim boundary. In that case by using the Cox’s proportional hazards model the estimate of the relative risk in mortality is 0.659 with the P-value of $8.9 \times 10^{-5}$ and the 95% confidence interval (0.812, 0.535). However, we know that because of the nature of sequential design those estimates are biased or incorrect. So was the treatment effect really that big and that significant?

Under the Cox’s proportional hazard model, let $\beta$ denote the logarithm of the hazard ratio of the treatment vs. placebo. Sellke and Siegmund (1983) showed that under certain regularity condition $I(t, \hat{\beta}(t)) \cdot (\hat{\beta}(t) - \beta)$ is an approximately scaled mean zero Brownian motion, where $t$ is calender time. $I(t, \beta)$ is the observed Fisher's information at the time $t$ under $\beta$ and $\hat{\beta}(t)$ is the maximum partial likelihood estimator of $\beta$ at the time $t$. To convert the test statistics into
Brownian framework as presented before, we need to rescale the time from the calendar time $t$ to the information time $\tau = I(t, \hat{\beta}(t))/I(t_M, \hat{\beta}(t_M))$ which is estimated by the ratio of the current observed number of death and the total expected number of deaths. In the MERIT-HF trial the new estimated information time at the first two interim are $\tau_1 = 0.24$ and $\tau_2 = 0.62$. Also, the observed rescaled Brownian motion process can be defined as $S_k = Z_k \cdot \sqrt{\tau_k}$ where $Z_k$ is the standardized log-rank statistic at interim $k$. The boundary value of $S_k$ will shrink accordingly from the one of $Z_k$ by the factor $\sqrt{\tau_k}$, too. Hence, $S_k$’s can be treated as obtained from a standard Brownian motion with drift parameter $\theta$. Different estimators and confidence intervals of the $\theta$ then can be calculated as discussed before. The log hazard ratio $\beta$ and the drift parameter $\theta$ will have a linear relationship:

$$\theta = \beta \sqrt{I(t_M, \hat{\beta}(t_M))}$$

Although the linear coefficient $\sqrt{I(t_M, \hat{\beta}(t_M))}$ is a random quantity and even may not be observed, it is a constant throughout the process in the sense that it does not depend on $t$ or $\tau$. Since $\hat{\theta} = S_k / \tau_k = Z_k / \sqrt{\tau_k}$ at the stopping stage and $\hat{\beta}$ can be readily obtained from the Cox’s model, the ratio $\hat{\beta}/\hat{\theta}$ is used as the linear factor to convert the estimates of the $\theta$ back to the estimates of the $\beta$. In the MERIT-HF trial, the $\hat{\theta} = 4.97$ and $\hat{\beta} = 0.417$ which corresponds to a hazard ratio of 0.659 or risk reduction of 34.1%. Several different estimation methods are applied, the results of the point estimates of the risk reduction in mortality and the corresponding confidence intervals of level 90% and 95% are listed in Table 5.1.

The point estimate and the 95% CI from the naive method agrees with the unadjusted P-value given in the article of the MERIT-HF trial report. The unconditionally adjusted method uses the Whitehead’s method to calculate the point estimate and Fairbanks & Madsen’s intuitive ordering
<table>
<thead>
<tr>
<th>method</th>
<th>estimated risk reduction</th>
<th>90% CI (%)</th>
<th>95% CI (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>naive</td>
<td>34.1%</td>
<td>(21.5, 44.7)</td>
<td>(18.8, 46.5)</td>
</tr>
<tr>
<td>unconditionally</td>
<td>31.3%</td>
<td>(17.7, 43.1)</td>
<td>(15.2, 46.0)</td>
</tr>
<tr>
<td>adjusted</td>
<td>32.8%</td>
<td>(8.6, 47.8)</td>
<td>(4.1, 49.9)</td>
</tr>
</tbody>
</table>

Table 5.1: Estimated risk reduction in mortality by using metoprolol in the MERIT-HF trial

To calculate P-value and (asymptotic) exact confidence intervals. Both the unconditional adjusted estimate (31.3%) and the conditional MLE 32.8% shrink from the naive estimate (34.1%) but their difference is very small in this case. The two modified conditional methods give the estimate of 32.8% and 32.0% which are almost no difference from others. However, the conditional adjusted P-value 0.0095 and the unconditional one 0.0024, though still highly significant, are much greater than the naive 0.000089 which means that given the stopping rule the outcome was not as significant as it appears. The conditional exact 90% and 95% confidence intervals for the risk reduction of the drug are (8.6%, 47.8%) and (4.1%, 49.9%) respectively which are wider than both of the naive and unconditional intervals. Since it's been demonstrated that the conditional intervals have the exact coverage probability when stopping at the second interim (or other possible ones) and in this case they cover the other two types of intervals entirely, we know the naive or the unconditional adjusted confidence intervals have lower than the nominal level coverage probabilities. The numbers of 8.6% (4.1%) and 47.8% (49.9%) are also the conditional uniformly most accurate lower and upper bounds at the 90% (95%) level for the mortality risk reduction of metoprolol.

6 Conclusions

We have proposed and discussed a new methodology to estimate the parameter in a group sequential experiment based on the conditional inference. The conditional methods improve from the earlier estimation methods in the accuracy of both point estimate and confidence interval at any given
stopping stage. The solutions from the conditional methods are future independent which are not only desirable but also computationally simpler than many other methods especially when very early stopping occurs. It’s easy to see that the conditional methods are natural extension from the fixed sample design case. When there’s no interim analysis, they are just same as the conventional method. The conditional methods can also be adapted into survival analysis and applied to practice without much difficulty.

Reference


A Figures and Tables

List of Figures

A.1 Plots of the conditional bias functions $b(\theta|\eta)$ and the marginal bias function $b(\theta)$ for OBF type boundary at $T=(.2,.4,.6,.8,1)$. .............................................. 23
A.2 MERIT-HF Monitoring Bounds for Mortality and the Observed Statistics .................. 24
A.3 Kaplan-Meier Curves of the Cumulative Mortality Rate. ...................................... 25

List of Tables

5.1 Estimated risk reduction in mortality by using metoprolol in the MERIT-HF trial 18
A.1 Table of the bias, variance and MSE of different estimators of $\theta$ when an OBF type boundary is used with $T=(.2,.4,.6,.8,1)$ and $\theta = 3$ (power $\approx 83.9\%$) ................... 26
A.2 Table of the bias, variance and MSE of different estimators of $\theta$ when an OBF type boundary is used with $T=(.2,.4,.6,.8,1)$ and $\theta = 1$ (power $\approx 16.5\%$) ................... 26
A.3 Table of the bias, variance and MSE of different estimators of $\theta$ when an OBF type boundary is used with $T=(.2,.4,.6,.8,1)$ and $\theta = 5$ (power $\approx 99.9\%$) ................... 27
A.4 Coverage probabilities of different 90% level confidence intervals when an OBF boundary is used with $T=(.2,.4,.6,.8,1)$ and $\theta = 5$ (power $\approx 99.9\%$) ................... 27
A.5 Coverage probabilities of different 90% level confidence intervals when an OBF boundary is used with $T=(.2,.4,.6,.8,1)$ and $\theta = 3$ (power $\approx 83.9\%$) ................... 28
A.6 Coverage probabilities of different 90% level confidence intervals when an OBF boundary is used with $T=(.2,.4,.6,.8,1)$ and $\theta = 1$ (power $\approx 16.3\%$) ................... 28
Figure A.1: Plots of the conditional bias functions $b(\theta|\eta)$ and the marginal bias function $b(\theta)$ for OBF type boundary at $T=(2,4,6,8,1)$.
† The $X$ axis is the true drift parameter $\theta$
Figure A.2: MERIT-HF Monitoring Bounds for Mortality and the Observed Statistics
Figure A.3: Kaplan-Meier Curves of the Cumulative Mortality Rate.

A: Metoprolol CR/XL and B: Placebo
<table>
<thead>
<tr>
<th>$\eta$ =</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\theta}$</td>
<td>7.45</td>
<td>2.62</td>
<td>0.98</td>
<td>0.08</td>
<td>-0.99</td>
<td>0.31</td>
</tr>
<tr>
<td>$\ddot{\theta}$</td>
<td>7.02</td>
<td>2.32</td>
<td>0.65</td>
<td>-0.21</td>
<td>-1.17</td>
<td>0.04</td>
</tr>
<tr>
<td>bias $\hat{\theta}$</td>
<td>0.50</td>
<td>-0.25</td>
<td>-0.19</td>
<td>-0.15</td>
<td>0.28</td>
<td>-0.03</td>
</tr>
<tr>
<td>$\ddot{\theta}$</td>
<td>0.50</td>
<td>-0.25</td>
<td>-0.47</td>
<td>-0.81</td>
<td>-0.99</td>
<td>-0.70</td>
</tr>
<tr>
<td>$\dddot{\theta}$</td>
<td>0.50</td>
<td>2.10</td>
<td>0.71</td>
<td>-0.15</td>
<td>-0.99</td>
<td>0.10</td>
</tr>
<tr>
<td>$\hat{\theta}$</td>
<td>0.43</td>
<td>0.42</td>
<td>0.28</td>
<td>0.17</td>
<td>0.42</td>
<td>1.70</td>
</tr>
<tr>
<td>$\ddot{\theta}$</td>
<td>0.49</td>
<td>0.44</td>
<td>0.28</td>
<td>0.15</td>
<td>0.34</td>
<td>1.55</td>
</tr>
<tr>
<td>variance $\hat{\theta}$</td>
<td>11.30</td>
<td>4.56</td>
<td>4.39</td>
<td>4.42</td>
<td>2.44</td>
<td>3.85</td>
</tr>
<tr>
<td>$\ddot{\theta}$</td>
<td>11.30</td>
<td>4.47</td>
<td>2.70</td>
<td>1.48</td>
<td>0.42</td>
<td>1.94</td>
</tr>
<tr>
<td>$\dddot{\theta}$</td>
<td>11.30</td>
<td>0.73</td>
<td>0.56</td>
<td>0.43</td>
<td>0.42</td>
<td>1.50</td>
</tr>
<tr>
<td>$\hat{\theta}$</td>
<td>56.0</td>
<td>7.31</td>
<td>1.24</td>
<td>0.18</td>
<td>1.41</td>
<td>1.80</td>
</tr>
<tr>
<td>$\ddot{\theta}$</td>
<td>49.8</td>
<td>5.84</td>
<td>0.70</td>
<td>0.19</td>
<td>1.70</td>
<td>1.55</td>
</tr>
<tr>
<td>MSE $\hat{\theta}$</td>
<td>11.6</td>
<td>4.62</td>
<td>4.43</td>
<td>4.45</td>
<td>2.52</td>
<td>3.85</td>
</tr>
<tr>
<td>$\ddot{\theta}$</td>
<td>11.6</td>
<td>4.54</td>
<td>2.92</td>
<td>2.14</td>
<td>1.41</td>
<td>2.43</td>
</tr>
<tr>
<td>$\dddot{\theta}$</td>
<td>11.6</td>
<td>5.14</td>
<td>1.07</td>
<td>0.45</td>
<td>1.41</td>
<td>1.51</td>
</tr>
<tr>
<td>Prob($\eta$)</td>
<td>0.001</td>
<td>0.118</td>
<td>0.296</td>
<td>0.262</td>
<td>0.322</td>
<td></td>
</tr>
</tbody>
</table>

Table A.1: Table of the bias, variance and MSE of different estimators of $\theta$ when an OBF type boundary is used with $T=(2,4,6,8,1)$ and $\theta=3$ (power$\approx$83.9%) 
† based on 100,000 simulations

<table>
<thead>
<tr>
<th>$\eta$ =</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\theta}$</td>
<td>9.84</td>
<td>4.28</td>
<td>2.64</td>
<td>1.82</td>
<td>-0.16</td>
<td>0.07</td>
</tr>
<tr>
<td>$\ddot{\theta}$</td>
<td>9.42</td>
<td>3.97</td>
<td>2.32</td>
<td>1.55</td>
<td>-0.23</td>
<td>-0.02</td>
</tr>
<tr>
<td>bias $\hat{\theta}$</td>
<td>5.26</td>
<td>0.89</td>
<td>0.80</td>
<td>0.78</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>$\ddot{\theta}$</td>
<td>5.26</td>
<td>0.89</td>
<td>0.72</td>
<td>0.54</td>
<td>-0.16</td>
<td>-0.09</td>
</tr>
<tr>
<td>$\dddot{\theta}$</td>
<td>5.26</td>
<td>3.68</td>
<td>2.27</td>
<td>1.46</td>
<td>-0.16</td>
<td>0.03</td>
</tr>
<tr>
<td>$\hat{\theta}$</td>
<td>0.32</td>
<td>1.16</td>
<td>0.83</td>
<td>0.41</td>
<td>0.77</td>
<td>1.31</td>
</tr>
<tr>
<td>$\ddot{\theta}$</td>
<td>0.35</td>
<td>1.06</td>
<td>0.71</td>
<td>0.34</td>
<td>0.65</td>
<td>1.10</td>
</tr>
<tr>
<td>Var $\hat{\theta}$</td>
<td>13.0</td>
<td>2.54</td>
<td>2.56</td>
<td>2.54</td>
<td>1.73</td>
<td>1.84</td>
</tr>
<tr>
<td>$\ddot{\theta}$</td>
<td>13.0</td>
<td>2.53</td>
<td>1.94</td>
<td>1.21</td>
<td>0.77</td>
<td>0.90</td>
</tr>
<tr>
<td>$\dddot{\theta}$</td>
<td>13.0</td>
<td>1.09</td>
<td>0.90</td>
<td>0.56</td>
<td>0.77</td>
<td>1.16</td>
</tr>
<tr>
<td>$\hat{\theta}$</td>
<td>97.1</td>
<td>19.4</td>
<td>7.81</td>
<td>3.72</td>
<td>0.80</td>
<td>1.32</td>
</tr>
<tr>
<td>$\ddot{\theta}$</td>
<td>89.1</td>
<td>16.8</td>
<td>6.10</td>
<td>2.75</td>
<td>0.70</td>
<td>1.10</td>
</tr>
<tr>
<td>MSE $\hat{\theta}$</td>
<td>40.6</td>
<td>3.3</td>
<td>3.20</td>
<td>3.14</td>
<td>1.75</td>
<td>1.89</td>
</tr>
<tr>
<td>$\ddot{\theta}$</td>
<td>40.6</td>
<td>3.3</td>
<td>2.45</td>
<td>1.49</td>
<td>0.80</td>
<td>0.91</td>
</tr>
<tr>
<td>$\dddot{\theta}$</td>
<td>40.6</td>
<td>14.6</td>
<td>6.03</td>
<td>2.68</td>
<td>0.80</td>
<td>1.17</td>
</tr>
<tr>
<td>Prob($\eta$)</td>
<td>$4 \times 10^{-5}$</td>
<td>0.007</td>
<td>0.033</td>
<td>0.056</td>
<td>0.904</td>
<td></td>
</tr>
</tbody>
</table>

Table A.2: Table of the bias, variance and MSE of different estimators of $\theta$ when an OBF type boundary is used with $T=(2,4,6,8,1)$ and $\theta=1$ (power$\approx$16.5%)  
† based on 100,000 simulations
<table>
<thead>
<tr>
<th>$\eta$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\theta}$</td>
<td>5.60</td>
<td>1.12</td>
<td>-0.56</td>
<td>-1.56</td>
<td>-2.30</td>
<td>0.31</td>
</tr>
<tr>
<td>$\hat{\delta}$</td>
<td>5.18</td>
<td>0.82</td>
<td>-0.89</td>
<td>-1.87</td>
<td>-2.55</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>bias</strong> $\hat{\theta}$</td>
<td>-0.77</td>
<td>-0.78</td>
<td>-0.48</td>
<td>-0.41</td>
<td>0.24</td>
<td>-0.63</td>
</tr>
<tr>
<td>$\hat{\delta}'$</td>
<td>-0.77</td>
<td>-0.83</td>
<td>-1.31</td>
<td>-1.99</td>
<td>-2.30</td>
<td>-1.12</td>
</tr>
<tr>
<td>$\hat{\delta}''$</td>
<td>-0.77</td>
<td>0.74</td>
<td>-0.69</td>
<td>-1.66</td>
<td>-2.30</td>
<td>-0.04</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.55</td>
<td>0.86</td>
<td>0.49</td>
<td>0.28</td>
<td>0.30</td>
<td>2.08</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.63</td>
<td>0.86</td>
<td>0.50</td>
<td>0.26</td>
<td>0.25</td>
<td>2.08</td>
</tr>
<tr>
<td><strong>Var</strong> $\hat{\theta}$</td>
<td>13.1</td>
<td>6.72</td>
<td>6.24</td>
<td>6.25</td>
<td>3.37</td>
<td>6.60</td>
</tr>
<tr>
<td>$\hat{\delta}'$</td>
<td>13.1</td>
<td>6.28</td>
<td>2.73</td>
<td>1.31</td>
<td>0.30</td>
<td>4.73</td>
</tr>
<tr>
<td>$\hat{\delta}''$</td>
<td>13.1</td>
<td>1.37</td>
<td>0.75</td>
<td>0.47</td>
<td>0.30</td>
<td>1.98</td>
</tr>
<tr>
<td>$\hat{\theta}$</td>
<td>31.9</td>
<td>2.11</td>
<td>0.80</td>
<td>2.71</td>
<td>5.60</td>
<td>2.18</td>
</tr>
<tr>
<td>$\hat{\delta}$</td>
<td>27.4</td>
<td>1.53</td>
<td>1.29</td>
<td>3.74</td>
<td>6.75</td>
<td>2.08</td>
</tr>
<tr>
<td><strong>MSE</strong> $\hat{\theta}$</td>
<td>13.7</td>
<td>7.33</td>
<td>6.47</td>
<td>6.42</td>
<td>3.43</td>
<td>6.99</td>
</tr>
<tr>
<td>$\hat{\delta}'$</td>
<td>13.7</td>
<td>6.98</td>
<td>4.44</td>
<td>5.27</td>
<td>5.60</td>
<td>5.98</td>
</tr>
<tr>
<td>$\hat{\delta}''$</td>
<td>13.7</td>
<td>1.91</td>
<td>1.22</td>
<td>3.21</td>
<td>5.60</td>
<td>1.99</td>
</tr>
</tbody>
</table>

| Prob($\eta$) | 0.016 | 0.520 | 0.373 | 0.079 | 0.012 |

Table A.3: Table of the bias, variance and MSE of different estimators of $\theta$ when an OBF type boundary is used with $T=(.2,.4,.6,.8,1)$ and $\theta = 5$ (power≈99.9%)

† based on 100,000 simulations

<table>
<thead>
<tr>
<th>Ordering or Method</th>
<th>Stopping interim stage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>naive CI</td>
<td>0</td>
</tr>
<tr>
<td>RCI</td>
<td>0.998</td>
</tr>
<tr>
<td>intuitive ordering</td>
<td>0</td>
</tr>
<tr>
<td>likelihood ratio</td>
<td>0</td>
</tr>
<tr>
<td>sample mean</td>
<td>0</td>
</tr>
<tr>
<td><strong>CECI</strong></td>
<td><strong>0.893</strong></td>
</tr>
</tbody>
</table>

Table A.4: Coverage probabilities of different 90% level confidence intervals when an OBF boundary is used with $T=(.2,.4,.6,.8,1)$ and $\theta = 5$ (power≈99.9%)

Stop frequency

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>1212</th>
<th>100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1567</td>
<td>52040</td>
<td>37270</td>
<td>7911</td>
<td>1212</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

27
<table>
<thead>
<tr>
<th>Ordering or Method</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>naive CI</td>
<td>0</td>
<td>0.590</td>
<td>0.965</td>
<td>1.000</td>
<td>0.846</td>
<td>0.890</td>
</tr>
<tr>
<td>RCI</td>
<td>0.976</td>
<td>0.987</td>
<td>0.998</td>
<td>1</td>
<td>0.872</td>
<td>0.957</td>
</tr>
<tr>
<td>intuitive ordering</td>
<td>0</td>
<td>0.594</td>
<td>1</td>
<td>1</td>
<td>0.842</td>
<td>0.900</td>
</tr>
<tr>
<td>likelihood ratio</td>
<td>0</td>
<td>0.657</td>
<td>0.975</td>
<td>1.000</td>
<td>0.843</td>
<td>0.900</td>
</tr>
<tr>
<td>sample mean</td>
<td>0</td>
<td>0.609</td>
<td>0.993</td>
<td>1</td>
<td>0.843</td>
<td>0.900</td>
</tr>
<tr>
<td><strong>CCI</strong></td>
<td><strong>0.912</strong></td>
<td><strong>0.902</strong></td>
<td><strong>0.901</strong></td>
<td><strong>0.902</strong></td>
<td><strong>0.900</strong></td>
<td><strong>0.901</strong></td>
</tr>
<tr>
<td>Stop frequency</td>
<td>125</td>
<td>11783</td>
<td>29644</td>
<td>26201</td>
<td>32247</td>
<td>100.000</td>
</tr>
</tbody>
</table>

Table A.5: Coverage probabilities of different 90% level confidence intervals when an OBF boundary is used with $T=(.2,.4,.6,.8,1)$ and $\theta = 3$ (power≈83.9%)