EFFICIENT GROUP SEQUENTIAL TESTS FOR SUPERIORITY AND
NON-INFERIORITY HYPOTHESES IN CLINICAL TRIALS

by

Tze Leung Lai
Mei-Chiung Shih
Guangrui Zhu

September 2004

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065
EFFICIENT GROUP SEQUENTIAL TESTS FOR SUPERIORITY AND
NON-INFERIORITY HYPOTHESES IN CLINICAL TRIALS

by

Tze Leung Lai
Department of Statistics
Stanford University

Mei-Chiung Shih
Department of Biostatistics
Harvard School of Public Health

Guangrui Zhu
Aventis Pharmaceuticals
Bridgewater, NJ

September 2004

This research was supported in part by National Science Foundation
grant DMS 0305749 and National Institute of Health grant 5R01-CA 088890

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
Efficient group sequential tests for superiority and non-inferiority hypotheses in clinical trials

Tze Leung Lai\textsuperscript{1}, Mei-Chiung Shih\textsuperscript{2,\ast,\dagger}, and Guangrui Zhu\textsuperscript{3}

\textsuperscript{1}Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.
\textsuperscript{2}Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, U.S.A.
\textsuperscript{3}Aventis Pharmaceuticals, 200 Crossing Boulevard, Bridgewater, NJ 08807, U.S.A.

SUMMARY

In designing an active controlled clinical trial, one sometimes has to choose between a superiority objective (to demonstrate that a new treatment is more effective than an active control therapy) and a non-inferiority objective (to demonstrate that it is no worse than the active control within some pre-specified non-inferiority margin). It is often difficult to decide which study objective should be undertaken at the planning stage when one does not have actual data on the comparative advantage of the new treatment. By making use of recent advances in the theory of efficient group sequential tests, we show how this difficulty can be resolved by a flexible group sequential design that can adaptively choose between the superiority and non-inferiority objectives during interim analyses. While maintaining the type I error probability at a pre-specified level, the proposed test is shown to have power advantage and/or sample size saving over fixed sample size tests for either only superiority or non-inferiority, and over other group sequential designs in the literature.

KEY WORDS: futility boundary; generalized likelihood ratio statistics; group sequential design; Kullback-Leibler information; non-inferiority; superiority.

Short Title: EFFICIENT GROUP SEQUENTIAL TESTS

\ast Correspondence to: Mei-Chiung Shih, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, U.S.A.
\dagger E-mail: mcshih@hsph.harvard.edu

Contract/grant sponsor: NIH; contract/grant number: 5R01-CA088890 and AI-24643
1. INTRODUCTION

In the design of controlled clinical trials comparing a new treatment with an active control, one often has to choose between two different study objectives: either a superiority or a non-inferiority hypothesis that the new treatment is more effective, or no worse (within certain indifference limits) than, the active control. The following example concerning the clinical trial design of a new antimicrobial drug that we studied recently illustrates some of the issues in the choice between these two study objectives, at the design stage when there is not enough information to decide on which objective has a better chance of success. Let \( p_1 \) and \( p_2 \) denote the response rates of the new and control drugs, respectively. The null hypothesis is \( H_0 : p_1 - p_2 \leq -\gamma \) for a non-inferiority trial, and is \( H'_0 : p_1 - p_2 \leq 0 \) for a superiority trial. The \( \gamma \) is chosen by the FDA that requires two trials to prove non-inferiority and only one trial to establish superiority. The pharmaceutical company developing the new drug does not have a good feel of the magnitude of \( p_1 - p_2 \) to decide whether it should perform two non-inferiority trials or one superiority trial, and would like to have a flexible design which can adapt to the information about \( p_1 - p_2 \) during interim analyses so that it can switch from the superiority to the non-inferiority objective, if needed. Section 4 describes the design we developed for this study by making use of recent advances in the theory of efficient group sequential tests reviewed in Section 2.

For the simpler problem of choosing between a superiority and only one (as opposed to two) non-inferiority group sequential tests after specifying \( H_1 : p_1 - p_2 > -\gamma \) (or \( H'_1 : p_1 - p_2 > 0 \)) as the alternative hypothesis of non-inferiority (or superiority), a different adaptive group sequential strategy has been proposed by Wang et al. [1]. Their approach, which is based on a conditional power criterion, is reviewed in Section 3 and compared with ours that is based on the theory of efficient group sequential tests in Section 2. As this theory has been developed in the general setting of exponential families which include Bernoulli populations as a special case, we develop our methodology in Section 3 for two-armed tests in exponential families and then specialize them to the Bernoulli case. Section 5 gives some concluding remarks and further discussion of our approach.

2. THEORY OF EFFICIENT GROUP SEQUENTIAL DESIGN

In standard clinical trial designs, the sample size is determined by the power at a given alternative. In practice, however, it is often difficult for investigators to specify a realistic alternative at which sample size determination can be based, especially for new treatments
about which there is little information concerning the magnitude of the treatment effect before actual data are collected. Moreover, the choice of the alternative to determine the sample size is usually guided not only by published results on the magnitude of the treatment difference to be expected but also by economic considerations related to funding and duration for the trial and by administrative considerations related to other trials that compete for patients and investigators. Although the protocol of a trial typically justifies its choice of sample size by stating some conventional level, such as 80% or 90%, of power at a plausible alternative, there are actually many other feasibility considerations that are difficult to quantify and much harder to explain than the precise but oversimplified statement of some prescribed power at the alternative.

Clearly efficiency of a group sequential test depends not only on the design of the stopping rule but also on the test statistics used. To fix the ideas, we focus on the generic special case of a one-parameter exponential family \( f_\theta(x) = e^{\theta x - \psi(\theta)} \) of densities with respect to some measure on the real line, and consider the problem of testing the one-sided hypothesis \( H_0 : \theta \leq \theta_0 \) at significance level \( \alpha \) and taking no more than \( M \) observations \( X_1, X_2, \ldots \).

Sufficient statistics are the sample means which are maximum likelihood estimators of \( \psi'(\theta) \), and the Kullback-Leibler information number is given by

\[
I(\theta, \lambda) = E_\theta[\log\{f_\theta(X_i)/f_\lambda(X_i)\}] = (\theta - \lambda)\psi'(\theta) - \{\psi(\theta) - \psi(\lambda)\}. \tag{1}
\]

Letting \( S_n = X_1 + \ldots + X_n \), the fixed sample size test that rejects \( H_0 \) if \( S_M \geq c_\alpha \) has maximal power at any alternative \( \theta > \theta_0 \), in particular at the alternative \( \theta(M) \) 'implied' by \( M \) (in the sense that \( M \) can be derived from the assumption that the above fixed sample test has some prescribed power \( 1 - \alpha \) at \( \theta(M) \)) when one does not have much information on which to base a realistic alternative. Under the constraint of \( M \) on the maximum sample size, it is desirable to adapt to the information on the actual \( \theta \) gathered during the course of the trial, allowing early stopping at times of interim analysis, so that the test has nearly optimal expected sample size under a wide range of alternatives but with small loss in power from the fixed sample size test.

To achieve these goals in a group sequential test with \( k \) groups and group sizes \( n_1, n_2 - n_1, \ldots, n_k - n_{k-1} \) so that \( n_k = M \), Lai and Shih [2] made use of the theory of optimal sequential tests, which is now relatively complete in the fully sequential framework (cf. Section 2 of [3]), and modified it for the group sequential setting. This theory leads to a group sequential test with rejection region of the form \( S_{n_k} \geq c \) at the \( k \)th analysis, where \( c > c_\alpha \) but \( c \) does not differ much from \( c_\alpha \). Let \( \bar{X}_n = S_n/n \). During the first \( k-1 \) analyses, the
test uses the maximum likelihood estimator $\hat{\theta}_{n_i} = (\psi')^{-1}(\bar{X}_{n_i})$ to estimate $\theta$ and a stopping rule of the form

$$\hat{\theta}_{n_i} > \theta_0 \text{ and } n_i I(\hat{\theta}_{n_i}, \theta_0) \geq b, \text{ or}$$

$$\hat{\theta}_{n_i} < \theta(M) \text{ and } n_i I(\hat{\theta}_{n_i}, \theta(M)) \geq \tilde{b},$$

for $1 \leq i \leq k - 1$. If (2a) holds, the test rejects $H_0$ upon stopping. If stopping occurs with (2b), it accepts $H_0$. The thresholds $b, \tilde{b}$ and $c$ are so chosen that $\text{pr}_{\theta_0}(\text{Test rejects } H_0) = \alpha$ and that the power of the test at $\theta(M)$ does not differ much from its upper bound $1 - \alpha$. A simple way of choosing $b, \tilde{b}$ and $c$ satisfying these properties is given in [2]. Let $0 < \epsilon < \frac{1}{2}$ and define $\tilde{b}$ by the equation

$$\text{pr}_{\theta(M)}(\hat{\theta}_{n_i} < \theta(M) \text{ and } n_i I(\hat{\theta}_{n_i}, \theta(M)) \geq \tilde{b} \text{ for some } 1 \leq i \leq k - 1) = \epsilon \alpha.$$

After determining $\tilde{b}$, define $b$ and then $c$ by the equations

$$\sum_{j=1}^{k-1} \text{pr}_{\theta_0}(\hat{\theta}_{n_j} > \theta_0 \text{ and } n_j I(\hat{\theta}_{n_j}, \theta_0) \geq b),$$

$$n_i I(\hat{\theta}_{n_i}, \theta_0) 1_{\{\hat{\theta}_{n_i} > \theta_0\}} < b \text{ and } n_i I(\hat{\theta}_{n_i}, \theta(M)) 1_{\{\hat{\theta}_{n_i} < \theta(M)\}} < \tilde{b} \text{ for } i < j = \epsilon \alpha,$$

$$\text{pr}_{\theta_0}(S_{n_k} \geq c, n_i I(\hat{\theta}_{n_i}, \theta_0) 1_{\{\hat{\theta}_{n_i} > \theta_0\}} < b \text{ and } n_i I(\hat{\theta}_{n_i}, \theta(M)) 1_{\{\hat{\theta}_{n_i} < \theta(M)\}} < \tilde{b} \text{ for } i < k$$

$$= (1 - \epsilon) \alpha.$$

Note that although the rejection region in favor of the treatment at each of the $k$ times is one-sided, involving the thresholds $b$ and $c$, there is also a "futility" boundary in (2b) with threshold $\tilde{b}$ that stops the trial when it becomes unlikely to demonstrate efficacy of the treatment within the resources allocated to the trial.

3. GROUP SEQUENTIAL TESTS FOR SUPERIORITY OR NON-INFERIORITY WITH PRESCRIBED TYPE I ERROR

In this section we first make use of the preceding theory of efficient group sequential designs to construct efficient group sequential tests that can adaptively choose, with prescribed type I error probability, between a superiority and non-inferiority alternative during interim analyses. We then review a different adaptive group sequential test proposed by Wang et al. [1] and compare its performance with ours.

To begin with, consider the application of the group sequential methodology in Section 2 to the problem of adaptively choosing between testing $H_0 : \theta \leq \theta_0$ and $H'_0 : \theta \leq \theta'_0$ with significance level $\alpha$, where $\theta_0 < \theta'_0$. Note that in the special case of a normal distribution
with mean $\theta$ and variance 1, the alternative hypotheses $H_1: \theta > 0$ and $H'_1: \theta > -\delta$ are often used as the superiority and non-inferiority alternatives for asymptotically normal test statistics. After considering the methodology in a univariate exponential family to conform to the notation of Section 2, we show how it can be extended to more general multi-armed and multiparameter settings which can be used to test treatment differences.

3.1 Adaptation and group sequential tests of $H_0$ or $H'_0$

The group sequential test of $H_0: \theta \leq \theta_0$ in Section 2 can be easily modified to allow concurrent testing of $H_0$ and the larger null hypothesis $H'_0: \theta \leq \theta'_0$, with $\theta'_0 > \theta_0$. Consider testing jointly (i) $H_0: \theta \leq \theta_0$ versus $H'_1: \theta \geq \theta'_0$ and (ii) $H'_0: \theta \leq \theta'_0$ versus $H_1: \theta > \theta_0$.

The group sequential test of $H'_0$ versus $H_1$ involves at most $k'$ interim analyses, with $k' < k$ and maximum sample size $M' = n_{k'}$. If $H'_0$ is not rejected during the first $k'$ stages, the test switches to testing $H_0$ versus $H'_1$ thereafter. During these $k'$ stages the test can also terminate with acceptance of $H_0$ if the lower futility boundary of $H_0$ versus $H'_1$ is crossed, besides early stopping with rejection of $H'_0$.

Specifically, the test uses the generalized likelihood ratio (GLR) statistics $n_i I(\hat{\theta}_{n_i}, \theta_0)$ and $n_i I(\hat{\theta}_{n_i}, \theta'_0)$ in conjunction with stopping boundaries $\bar{b}, \bar{b}', b, c$ and $c'$ that are determined as follows. Let $0 < \epsilon < \frac{1}{2}$ and define $\bar{b}, \bar{b}'$ and then $c'$ by

$$\text{pr}_{\theta_0}\{\hat{\theta}_{n_i} < \theta'_0 \text{ and } n_i I(\hat{\theta}_{n_i}, \theta'_0) \geq \bar{b} \text{ for some } i \leq k - 1\} = \epsilon \alpha,$$

$$\text{pr}_{\theta'_0}\{\hat{\theta}_{n_i} > \theta'_0 \text{ and } n_i I(\hat{\theta}_{n_i}, \theta'_0) \geq \bar{b}' \text{ for some } i \leq k' - 1\} = \epsilon \alpha,$$

$$\text{pr}_{\theta'_0}(\{S_{n_{k'}} \geq c'\} \cap A_{k'-1}) = (1 - \epsilon) \alpha,$$

where $A_{k'-1} = \{n_i I(\hat{\theta}_{n_i}, \theta'_0)1_{\{\hat{\theta}_{n_i} > \theta'_0\}} < \bar{b}' \text{ and } n_i I(\hat{\theta}_{n_i}, \theta'_0)1_{\{\hat{\theta}_{n_i} < \theta'_0\}} < \bar{b} \text{ for all } i \leq k - 1\}$. The thresholds $b'$ and $c'$ are related to rejection of $H'_0$ before or at the $k'$th interim analysis; see (8a, b) below. Testing of $H_0$ instead of $H'_0$ begins at the $k'$th interim analysis if $H'_0$ is not rejected by that time. Define $b$ and $c$ by the equations

$$\text{pr}_{\theta_0}\{\text{Test rejects } H'_0\} + \Sigma_{j=k'}^{k-1} \text{pr}_{\theta_0}(\{S_{n_{k'}} < c'\} \cap A_{k'-1} \cap \{\hat{\theta}_{n_j} > \theta_0 \text{ and } n_j I(\hat{\theta}_{n_j}, \theta_0) \geq b\},$$

$$n_i I(\hat{\theta}_{n_i}, \theta_0)1_{\{\hat{\theta}_{n_i} > \theta'_0\}} < \bar{b} \text{ and } n_i I(\hat{\theta}_{n_i}, \theta'_0)1_{\{\hat{\theta}_{n_i} < \theta'_0\}} < \bar{b} \text{ for } k' \leq i < j\} = \epsilon \alpha,$$

$$\text{pr}_{\theta_0}(\{S_{n_{k'}} < c'\} \cap A_{k'-1} \{S_{n_k} \geq c, \ n_i I(\hat{\theta}_{n_i}, \theta_0)1_{\{\hat{\theta}_{n_i} > \theta_0\}} < \bar{b} \text{ and }$$

$$n_i I(\hat{\theta}_{n_i}, \theta'_0)1_{\{\hat{\theta}_{n_i} < \theta'_0\}} < \bar{b} \text{ for } k' \leq i \leq k - 1\} = (1 - \epsilon) \alpha.$$

4
Analogous to (2a) and (2b), the stopping rule of the test has the form

\[
\hat{\theta}_{n_i} > \theta_0' \text{ and } n_i I(\hat{\theta}_{n_i}, \theta_0') \geq b' \text{ if } 1 \leq i \leq k' - 1,
\]

or \( S_{n_{k'}} \geq c' \), or \( \hat{\theta}_{n_{k'}} > \theta_0 \) and \( n_{k'} I(\hat{\theta}_{n_{k'}}, \theta_0) \geq b \),

or \( \hat{\theta}_{n_i} > \theta_0 \) and \( n_i I(\hat{\theta}_{n_i}, \theta_0) \geq b \) if \( k' < i < k \),

or \( \hat{\theta}_{n_i} < \theta_0' \) and \( n_i I(\hat{\theta}_{n_i}, \theta_0') \geq \tilde{b} \) if \( 1 \leq i < k \).

If (8a) holds, the test rejects \( H_0' \) upon stopping. If stopping occurs at the \( k' \)th interim analysis, the test rejects \( H_0' \) (in favor of the superiority alternative) when \( S_{n_{k'}} \geq c' \), and rejects \( H_0 \) (in favor of the non-inferiority alternative) when \( S_{n_{k'}} < c' \) and the other event in (8b) occurs. If stopping occurs with (8c), the test rejects \( H_0 \) (in favor of the non-inferiority alternative) upon stopping. Early stopping (due to futility) with acceptance of \( H_0 \) occurs with (8d). If stopping does not occur during the first \( k - 1 \) interim analyses, the test rejects \( H_0 \) when \( S_{n_k} \geq c \) at the \( k \)th interim analysis. In view of (4)-(7), \( \Pr_{\theta_0} (\text{Test rejects } H_0') = \alpha \) and \( \Pr_{\theta_0} (\text{Test rejects } H'_0 \text{ or } H_0) = \alpha \). By monotonicity, the test has overall significance level \( \alpha \) for concurrently testing \( H_0 \) and \( H'_0 \). The lower futility boundary in (8d) is chosen such that the power of the test of \( H_0 \) versus \( H_1' : \theta \geq \theta_0' \) does not differ much from \( 1 - \tilde{\alpha} \), which is the power of the fixed sample size test (with sample size \( M' \)). Note that we have not imposed a similar futility boundary for testing \( H'_0 \) versus \( H_1 \) that has maximum sample size \( M' = n_{k'} \) as we may switch to non-inferiority at the \( k' \)th interim analysis.

### 3.2 Extensions to multi-armed and multiparameter problems

As noted in Section 3.4 of [2], the efficient group sequential test with stopping rule (2) can be readily extended to the multiparameter exponential family and to multi-armed clinical trials by using appropriate GLR statistics in these problems. In a similar way described as follows, we can generalize the group sequential test of \( H_0 \) or \( H'_0 \) to multiparameter and multi-armed settings.

First consider the multiparameter exponential family \( f_\theta(x) = \exp\{\theta^T x - \psi(\theta)\} \) and let \( d \) be a continuously differentiable real-valued function on the natural parameter space. Consider the null hypothesis \( H_0 : d(\theta) \leq \delta_0 \) or \( H'_0 : d(\theta) \leq \delta'_0 \) with \( \delta_0 < \delta'_0 \). The GLR statistic for testing \( d(\theta) = \delta \) at the \( j \)th interim analysis has the form

\[
n_j \{\theta^T \hat{X}_{n_j} - \psi(\hat{\theta}_{n_j})\} - \sup_{d(\theta) = \delta} n_j \{\theta^T \hat{X}_{n_j} - \psi(\theta)\} = \inf_{d(\theta) = \delta} n_j I(\hat{\theta}_{n_j}, \theta),
\]
in which \( I(\theta, \lambda) \) is given by (1) with \( \psi' \) denoting the gradient vector \( \nabla \psi \) of partial derivatives of \( \psi \) with respect to the components of \( \theta \). By a general result on group sequential distribution theory involving efficient test statistics (cf. [4], [5], [6]), the signed root likelihood ratio statistic (obtained by multiplying the square root of (9) by \( \sqrt{2n_j} \text{ sgn} \{ d(\hat{\theta}_n) - \delta \} \)) is approximately \( N(0, n_j) \) and has independent increments under \( d(\theta) = \delta \), and therefore we can apply the test in Section 3.1 with \( nI(\hat{\theta}_n, \theta_0) \) replaced by \( \inf_{d(\theta) = \delta_0} nI(\hat{\theta}_n, \theta) \), and \( nI(\hat{\theta}_n, \theta'_0) \) replaced by \( \inf_{d(\theta) = \delta'_0} nI(\hat{\theta}_n, \theta) \), and using the normal random walk approximation to the signed root likelihood ratio statistics in determining the thresholds \( \tilde{b}, \tilde{b}', \tilde{c}', b \) and \( c \) via (3)-(7).

We next consider the case of multi-armed clinical trials involving \( m \) independent populations having density functions \( \exp\{\theta_i x - \psi(\theta_i)\} \) and with different numbers of patients assigned to different populations. Let \( n_{ij} \) be the total number of observations \( X_{i,n} \) from the \( i \)th population up to the time of the \( j \)th interim analysis, and let \( \bar{X}_{i,n_{ij}} \) and \( \hat{\theta}_{i,n_{ij}} \) denote the sample mean and the maximum likelihood estimate of \( \theta_i \), respectively, based on these observations. Consider testing \( H_0 : d(\theta_1, \ldots, \theta_m) \leq \delta_0 \) or \( H_0' : d(\theta_1, \ldots, \theta_m) \leq \delta'_0 \) with \( \delta_0 < \delta'_0 \). The GLR statistic for testing \( d(\theta_1, \ldots, \theta_m) = \delta \) at the \( j \)th interim analysis has the form

\[
\inf_{d(\theta_1, \ldots, \theta_m) = \delta} \sum_{i=1}^{m} n_{ij} I(\hat{\theta}_{i,n_{ij}}, \theta_i)
\]

with \( I(\theta, \lambda) \) given by (1), and we can therefore proceed again as in the preceding paragraph. Details are provided in the special case of \( m = 2 \) Bernoulli populations below.

3.3 Comparison with an adaptive design based on conditional power

Suppose there are \( m = 2 \) treatment groups and the responses \( X_{i,n} \) are Bernoulli random variables with \( \text{pr}\{X_{i,n} = 1\} = p_i = 1 - \text{pr}\{X_{i,n} = 0\}, \ i = 1, 2 \). In this case, \( \theta_i = \log\{p_i/(1 - p_i)\} \) and

\[
I(\theta, \theta') = p \log(p/p') + (1 - p) \log((1 - p)/(1 - p')).
\]

Letting \( p_1 \) denote the response probability of the experimental treatment and \( p_2 \) that of the control treatment, testing for superiority (resp. non-inferiority) of the experimental treatment involves the null hypothesis \( H_0' : d \leq 0 \) (resp. \( H_0 : d \leq -\gamma \)), where \( d = p_1 - p_2 \) and \( \gamma > 0 \) denotes a prescribed non-inferiority margin. In this case, the GLR statistic for testing \( p_1 - p_2 = \delta \) at the \( j \)th interim analysis can be expressed explicitly as

\[
\Lambda_j(\delta) = \sum_{i=1}^{2} n_{ij} \{ \hat{p}_{i,j} \log(\hat{p}_{i,j}/\bar{p}_{i,j}(\delta)) + (1 - \hat{p}_{i,j}) \log((1 - \hat{p}_{i,j})/(1 - \bar{p}_{i,j}(\delta))) \},
\]
where \( \hat{p}_{i,j} = \tilde{X}_{1,n_{ij}} \), \( \bar{p}_{i,j}(\delta) = p + \delta \) and \( \bar{p}_{2,j}(\delta) = p \), in which \( p \) is the minimizer of (12) (over such values of \( \bar{p}_{1,j}(\delta) \) and \( \bar{p}_{2,j}(\delta) \)). In particular, \( \bar{p}_{1,j}(0) = \bar{p}_{2,j}(0) = (\sum_{i=1}^{n_{ij}} \bar{X}_{1,n_{ij}})/(n_{1j} + n_{2j}) \). For \( \delta = -\gamma \), the minimization problem leads to a nonlinear equation in \( p \). Since \( \gamma \) is typically small, we can replace it by the linear approximation

\[
\bar{p}_{2,j}(-\gamma) = \{n_{2j} \tilde{X}_{2,n_{2j}} + n_{1j}(\tilde{X}_{1,n_{1j}} + \gamma)\}/(n_{2j} + n_{1j}).
\] (13)

The GLR statistics (12) with \( \delta = 0 \), \(-\gamma \) can be applied to the group sequential test in the last paragraph of Section 3.2, in which we set \( \delta_0' = 0 \) (for superiority) and \( \delta_0 = -\gamma \) (for non-inferiority).

Instead of using GLR statistics, Wang et al. [1] used the Studentized statistics \( Z_j = \tilde{\Delta}_j / \hat{\sigma}_j \), where \( \tilde{\Delta}_j = \hat{p}_{i,j} - \bar{p}_{2,j} \) and

\[
\hat{\sigma}_j^2 = \bar{p}_{1,j}(1 - \bar{p}_{1,j})/n_{1j} + \bar{p}_{2,j}(1 - \bar{p}_{2,j})/n_{2j}.
\] (14)

Moreover, in place of the stopping boundaries introduced in [2], they used the O'’Brien-Fleming error spending function (cf. [6]). The major difference between their approach and ours, however, lies in how they switch from the superiority to the non-inferiority alternative during the course of the trial. Whereas our procedure makes the switch at the \( k' \)th interim analysis, where \( 2M' = 2n_{k'} \) is the maximum sample size for testing the superiority alternative \( p_1 - p_2 = \gamma > 0 \) at which the fixed sample size (FSS) GLR test attains some prescribed power, they make the switch at the first interim analysis at which the conditional power in favor of the non-inferiority alternative \( p_1 - p_2 > -\gamma \) exceeds that of the superiority alternative \( p_1 - p_2 > 0 \). Specifically, at the \( j \)th interim analysis with \( j < k \), consider the conditional probability \( CP_S(\Delta) \) at \( p_1 - p_2 = \Delta \), given \( (\hat{p}_{1j}, \hat{p}_{2j}) \), that the FSS test with sample size \( 2M \) rejects \( H'_0 : p_1 - p_2 \leq 0 \). Also compute the conditional probability \( CP_{NI}(\Delta) \) at \( p_1 - p_2 = \Delta \), given \( (\hat{p}_{1j}, \hat{p}_{2j}) \), that the FSS test rejects \( H_0 : p_1 - p_2 \leq -\gamma \) but accepts \( H'_0 \). The conditional power approach computes these conditional probabilities at \( \Delta = \tilde{\Delta}_j \), and the adaptive strategy in [1] switches from the superiority alternative (with maximum sample size \( 2M' \)) to the non-inferiority alternative (with maximum sample size \( 2M \)) when \( CP_S(\tilde{\Delta}_j) < CP_{NI}(\tilde{\Delta}_j) \). To circumvent the possibility of an inflated type I error probability due to such data-dependent switch, Wang et al. [1] proposed to modify both the times of interim analyses and the test statistics after the interim analysis at which such switch is made. Making use of previous work of Cui et al. [7], they showed that this modification indeed yields a type I error probability that is asymptotically no larger than \( \alpha \).
Wang et al. [1] reported a simulation study demonstrating the advantages of their adaptive strategy over the FSS and other group sequential methods that they considered earlier in [8]. The superiority alternative in their study is at \( p_2 = 0.25, p_1 = 0.35 \), at which the level \( \alpha = 0.025 \) FSS test of \( H'_0 : p_1 - p_2 \leq 0 \) with power 0.8 requires a sample size of \( M' = 330 \) from each population. The non-inferiority alternative is at \( p_1 = p_2 = 0.25 \), at which the level \( \alpha = 0.025 \) FSS test of \( H_0 : p_1 - p_2 \leq -0.05 \) with power 0.8 requires a sample size of \( M = 1200 \) from each population. Their group sequential tests involve \( k = 5 \) interim analyses, and of particular interest is their adaptive group sequential design that uses the above conditional power criterion to switch from the superiority to non-inferiority objective.

The simulation results of Wang et al. [1] show that at the prespecified superiority alternative \( (p_2 = 0.25, p_1 = 0.35) \), their adaptive procedure has power 0.947 and an expected sample size of 428 (which exceeds 330 for the FSS test) from each population. For comparison, we have computed corresponding operating characteristics of our group sequential test by Monte Carlo involving 50,000 simulations. Our test (described in the first paragraph of this section with \( \epsilon = 1/3 \)) also uses \( k = 5 \) analyses, with \( k' = 3, n_j - n_{j-1} = 110 \) (for \( j \leq 3 \)) or 435 (for \( j = 4, 5 \)). It has type I error probability 0.024 of falsely rejecting \( H'_0 \) at \( p_1 = p_2 = 0.25 \), which is close to the corresponding value of 0.0258 for the test of Wang et al.. At the superiority alternative \( p_2 = 0.25, p_1 = 0.35 \), it has power 0.785 (which is close to the target power 0.8) whereas the adaptive test of Wang et al. is substantially over-powered) and expected sample size of 299 from each population.

At the prespecified non-inferiority alternative \( (p_1 = p_2 = 0.25) \), the adaptive test of Wang et al. has power 0.768 and expected sample size 1144 (from each population), whereas ours has power 0.784 and expected sample size 1013. The type I error probability of falsely rejecting \( H_0 : p_1 - p_2 \leq -0.05 \) at \( p_2 = 0.25, p_1 = 0.2 \) is 0.024 for our test and 0.0253 for theirs.

Besides the particular configurations \( (p_1, p_2) = (0.25, 0.25), (0.35, 0.25) \), at which the power of the FSS test of \( H_0 \) or \( H'_0 \) is evaluated to determine the sample size, we also consider a variety of other parameter configurations in the superiority and non-inferiority regions. Table I gives the expected sample size \( E(T) \) from each population, and the probabilities \( \text{pr}(S) \) of claiming superiority, \( \text{pr}(\text{NI}) \) of claiming non-inferiority (and not superiority), and \( \text{pr}(+) = \text{pr}(S) + \text{pr}(\text{NI}) \) of a positive claim for the experimental treatment, for our group sequential test at these parameter configurations. Corresponding values of the FSS test for superiority, denoted by FSS(330) to indicate the sample size, and of the FSS test for
non-inferiority, denoted by FSS(1200), are also given for comparison.

INSERT TABLE I ABOUT HERE

4. A GROUP SEQUENTIAL APPROACH TO ADAPTIVE CHOICE BETWEEN ONE SUPERIORITY AND TWO NON-INFERIORITY TRIALS

In this section we show how the group sequential design of Section 3 that allows switching from superiority to non-inferiority objectives during the first $k'$ interim analyses can be modified for the clinical trial that has been mentioned in the Introduction. We begin by describing the background of the trial design. A pharmaceutical company that developed a new antimicrobial drug had planned to conduct two independent non-inferiority trials, as required by the FDA, with the same (fixed) sample size to demonstrate the drug’s non-inferiority relative to an active control. This plan had to be reconsidered when the FDA required a substantially smaller non-inferiority margin than what was assumed in the plan. A major issue was whether the increased total sample size for the two non-inferiority trials due to the decreased non-inferior margin would already suffice to establish superiority of the drug since the FDA only required a single trial to demonstrate superiority. Following the notation of Section 3.3, the null hypothesis is $H_0 : p_1 - p_2 \leq -\gamma$ to test for non-inferiority, and is $H_0' : p_1 - p_2 \leq 0$ test for superiority, where $p_2$ is the response probability of the active control and $p_1$ is that of the new drug. The narrower non-inferiority margin required by the FDA was $\gamma = 0.1$, and the response probability $p_2$ of the active control was estimated to be 0.7 from previous studies. Thus, to have 90% power at the alternative $p_1 = p_2 (= 0.7)$, a level $\alpha = 0.025$ test of $H_0$ requires a sample size of 882 (i.e., 441 per treatment arm), which means a total sample size of 1764 for two non-inferiority trials.

Because it could only make rough a priori guesses of $p_1$ and was also concerned that the estimate 0.7 might differ substantially from the actual $p_2$, the pharmaceutical company was unable to decide whether it should perform two non-inferiority trials or one superiority trial with the same number of subjects. It would prefer to make that decision during the course of the trial when accumulating data would provide information on the feasibility of demonstrating non-inferiority or superiority of the new drug at the end of the trial. How to do this without inflating the overall type I error was the question it posed to us. Moreover, 1764 was already considered to be somewhat too large for the total sample size because of the eligibility criterion that made it difficult to enroll subjects. The company, therefore, would also like to be able to terminate the study if interim analysis of the data would suggest
“futility” of a trial with a maximum sample size of 1764.

We now modify Section 3.3 to come up with a group sequential design that can “self-tune” to the unknown \((\hat{p}_1, \hat{p}_2)\) and thereby choose adaptively among testing for superiority at level \(\alpha\) (with no more than 1764 subjects), testing for non-inferiority (with two independent trials each of level \(\alpha\), as required by the FDA), and early termination due to futility. The test statistics are the GLR statistics \(\Lambda_j(\delta)\) in (12), with \(\delta = 0\) for superiority and \(\delta = -0.1\) for non-inferiority. The sequential design involves \(k = 5\) groups with \(n_1 = 300\) (or 150 patients per arm), \(n_2 = 600\), \(n_3 = 882\) (which is the fixed sample size of the non-inferiority trial), \(n_4 = 1320\), \(n_5 = 1764\). The group sequential trial terminates early at the \(j\)th analysis with the superiority claim (rejecting \(H_0^s\)) for \(j \leq 4\) if

\[
\hat{p}_{1,j} > \hat{p}_{2,j} \text{ and } \Lambda_j(0) \geq b'.
\] (15)

It can also terminate during the first two interim analyses due to futility (accepting \(H_0\)) if for \(j = 1\) or 2,

\[
\hat{p}_{1,j} < \hat{p}_{2,j} \text{ and } \Lambda_j(0) \geq \tilde{b}.
\] (16)

At the 3rd interim analysis (with total sample size 882, assuming termination has not occurred before), if \(H_0^s\) is not rejected, continue the trial if

\[
\hat{p}_{1,3} - \hat{p}_{2,3} > -0.1 \text{ and } \Lambda_3(-0.1) \geq b,
\] (17)

otherwise accept \(H_0\) and stop (for futility). At the final analysis \((j = 5)\), reject \(H_0^s\) (claiming superiority) if

\[
(\hat{p}_{1,5} - \hat{p}_{2,5})/\hat{\sigma}_5 \geq c',
\] (18)

where \(\hat{\sigma}_j^2\) is defined in (14), otherwise reject \(H_0\) (claiming non-inferiority) only if

\[
(\hat{p}_1 - \hat{p}_2 + 0.1)/\hat{\sigma} \geq c,
\] (19)

where \(\hat{p}_1 = \sum_{i=883}^{882+n_D} X_{1i}/n_D\), \(\hat{p}_2 = \sum_{i=883}^{882+n_C} X_{2i}/n_C\), \(\hat{\sigma}^2 = \hat{p}_1(1 - \hat{p}_1)/n_D + \hat{p}_2(1 - \hat{p}_2)/n_C\), and \(n_D(n_C)\) denotes the number of subjects receiving the new drug (active control) between the 3rd and 5th analyses (representing the second non-inferiority trial required by the FDA) so that \(n_D + n_C = 882\). Letting \(0 < \epsilon < \frac{1}{2}\), the thresholds \(b', \tilde{b}\) and then \(b, c'\) and \(c\) are determined by the equations

\[
\Pr_0\{\text{(15) holds for some } j \leq 4\} = \epsilon \alpha,
\] (20a)
\[ \text{pr}_0\{(16) \text{ holds for some } j \leq 2\} = \varepsilon \alpha, \] (20b)

\[ \text{pr}_{-0.1}\{\text{Test terminates at } j\text{th analysis with (15) for some } j \leq 3,} \] (20c)

or continues at 3rd analysis with (17)\} = \alpha,

\[ \text{pr}_0\{\text{Test terminates at the 5th analysis with (18)}\} = (1 - \varepsilon)\alpha, \] (20d)

\[ \text{pr}_{-0.1}\{(19) \text{ holds}\} + \text{pr}_{-0.1}\{(15) \text{ holds for } j = 4\} = \alpha. \] (20e)

The major difference between this group sequential design and that in Section 3.3, where only one trial is needed to establish non-inferiority, is that the design has to allow the option of two independent non-inferiority trials required by the FDA for a non-inferiority claim, under the maximum sample size constraint of 1764. In this connection, note that (19) represents the rejection region of the second non-inferiority trial involving a new set of 882 subjects accrued after the third interim analysis. Since superiority testing (which is based on all subjects that have entered the trial) can still proceed after the third interim analysis, (20e) involves the sum of the probability of falsely claiming non-inferiority because of (19) and the probability of falsely claiming superiority at the 4th interim analysis because of (15) at \( j = 4 \), for which the threshold \( b' \) is already determined by (20a).

Table II gives the results of a simulation study on this group sequential design with \( \alpha = 0.025 \) and \( \varepsilon = 1/3 \). Each result is based on 50,000 simulations. The expected sample size \( E(T) \) and the probabilities \( \text{pr}(S) \) of claiming superiority, \( \text{pr}(\text{NI}) \) of claiming non-inferiority (but not superiority), and \( \text{pr}(+) = \text{pr}(S) + \text{pr}(\text{NI}) \) of a positive claim for the new drug are given for a variety of parameter configurations. Note that \( \text{pr}(\text{NI}) = \text{pr}\{\text{Test terminates at the 5th analysis with rejection of } H_0 \text{ because of (19)}\} \), so the non-inferiority claim is supported by the first set of 882 subjects who result in (17) at the 3rd analysis and by a second set of 882 subjects yielding (19) at the 5th analysis. Also given for comparison in Table II are (a) the power \( \text{pr}(S) \) of a level-\( \alpha \) FSS test of \( H_0^* \) with sample size 1764, denoted by FSS\(_1\), and (b) the power \( \text{pr}(\text{NI}^2) \) of two independent level-\( \alpha \) FSS tests of \( H_0 \), denoted by FSS\(_2\), with sample size 882 for each test so that \( \text{NI}^2 \) represents non-inferiority claims for both tests. Table II shows that our group sequential design has markedly smaller expected sample size than 1764 while having better power than FSS\(_1\) or FSS\(_2\). Although \( \text{pr}(+) \) and \( \text{pr}(\text{NI}^2) \) seem to differ little, note that \( \text{pr}(+) = \text{pr}(S) + \text{pr}(\text{NI}) \) and that the probability \( \text{pr}(S) \) of a superiority claim by the group sequential test can be quite high, whereas FSS\(_2\) can only claim non-inferiority with probability \( \text{pr}(\text{NI}^2) \).

**INSERT TABLE II ABOUT HERE**

11
5. DISCUSSION

As the actual benefits of a new treatment are often unclear at the planning stage of a controlled clinical trial that compares it with an active control therapy, one encounters different formulations of the null hypothesis for the comparison. This paper considers in particular the one-sided null hypothesis $H'_0$ (versus the superiority alternative) and the smaller null hypothesis $H_0$ (versus the non-inferiority alternative) and shows that a suitably chosen group sequential design can choose efficiently which hypothesis to test, with a prescribed overall significance level, during interim analyses by using the information accumulated so far to estimate the treatment differences. The construction of the group sequential design is based on the theory of efficient group sequential tests in multiparameter exponential families developed in [2]. These group sequential tests do not involve error spending functions and conditional power that are used in the group sequential procedure of Wang et al. [2]. Their basic underlying idea is to estimate the unknown parameter $\theta$ of the exponential family by maximum likelihood during the course of the trial and to use the estimate to replace $\theta$ in an approximately optimal sequential test that assumes known $\theta$. As pointed out in Section 5 of [2], these group sequential tests can be conveniently extended to much more complex settings including time-sequential tests comparing the failure times between two treatment groups since they do not require pre-specification of the group sizes nor estimation of the “maximum information” of the trial in the error spending approach (cf. [6, Chapter 7], [9]). To circumvent the difficulties for time-to-event endpoints in the group sequential approach with error spending functions, Shih et al. [10] recently proposed to use a two-stage adaptive design instead of the more efficient group sequential strategy of Wang et al. [1] for choosing between the superiority and non-inferiority objectives.

ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health under grants 5R01-CA088890 to Stanford University and AI-24643 to Harvard University.
REFERENCES


<table>
<thead>
<tr>
<th>$p_2$</th>
<th>$p_1$</th>
<th>pr(NI)</th>
<th>pr(S)</th>
<th>pr(+)</th>
<th>FSS(330) pr(NI)</th>
<th>pr(S)</th>
<th>pr(+)</th>
<th>FSS(1200) pr(NI)</th>
<th>pr(S)</th>
<th>pr(+)</th>
<th>Group Sequential pr(NI)</th>
<th>pr(S)</th>
<th>pr(+)</th>
<th>E(T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.25</td>
<td>0.291</td>
<td><strong>0.025</strong></td>
<td>0.316</td>
<td>0.783</td>
<td>0.025</td>
<td>0.808</td>
<td>0.760</td>
<td><strong>0.024</strong></td>
<td>0.784</td>
<td>1013</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>0.025</td>
<td>0.000</td>
<td>0.025</td>
<td></td>
<td><strong>0.026</strong></td>
<td>0.000</td>
<td>0.026</td>
<td>0.023</td>
<td>0.001</td>
<td><strong>0.024</strong></td>
<td>973</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.27</td>
<td>0.452</td>
<td>0.087</td>
<td>0.539</td>
<td></td>
<td>0.773</td>
<td>0.201</td>
<td>0.975</td>
<td>0.888</td>
<td>0.080</td>
<td>0.968</td>
<td>815</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>0.517</td>
<td>0.302</td>
<td>0.819</td>
<td></td>
<td>0.218</td>
<td>0.782</td>
<td>1.000</td>
<td>0.714</td>
<td>0.286</td>
<td>1.000</td>
<td>536</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.32</td>
<td>0.413</td>
<td>0.514</td>
<td>0.927</td>
<td></td>
<td>0.032</td>
<td>0.968</td>
<td>1.000</td>
<td>0.507</td>
<td>0.493</td>
<td>1.000</td>
<td>416</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>0.187</td>
<td>0.802</td>
<td>0.989</td>
<td></td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.216</td>
<td>0.784</td>
<td>1.000</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>0.20</td>
<td>0.334</td>
<td><strong>0.025</strong></td>
<td>0.359</td>
<td>0.840</td>
<td>0.025</td>
<td>0.865</td>
<td>0.822</td>
<td><strong>0.025</strong></td>
<td>0.847</td>
<td>973</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>0.026</td>
<td>0.000</td>
<td>0.026</td>
<td></td>
<td><strong>0.025</strong></td>
<td>0.000</td>
<td>0.025</td>
<td>0.023</td>
<td>0.000</td>
<td><strong>0.023</strong></td>
<td>918</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.22</td>
<td>0.506</td>
<td>0.094</td>
<td>0.600</td>
<td></td>
<td>0.762</td>
<td>0.225</td>
<td>0.987</td>
<td>0.897</td>
<td>0.088</td>
<td>0.985</td>
<td>755</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.533</td>
<td>0.335</td>
<td>0.868</td>
<td></td>
<td>0.165</td>
<td>0.835</td>
<td>1.000</td>
<td>0.680</td>
<td>0.319</td>
<td>0.999</td>
<td>491</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.27</td>
<td>0.389</td>
<td>0.565</td>
<td>0.954</td>
<td></td>
<td>0.018</td>
<td>0.982</td>
<td>1.000</td>
<td>0.456</td>
<td>0.554</td>
<td>1.000</td>
<td>381</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>0.148</td>
<td>0.846</td>
<td>0.994</td>
<td></td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.169</td>
<td>0.831</td>
<td>1.000</td>
<td>278</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>0.30</td>
<td>0.263</td>
<td><strong>0.026</strong></td>
<td>0.289</td>
<td>0.736</td>
<td>0.025</td>
<td>0.761</td>
<td>0.716</td>
<td><strong>0.025</strong></td>
<td>0.741</td>
<td>1035</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.026</td>
<td>0.000</td>
<td>0.026</td>
<td></td>
<td><strong>0.026</strong></td>
<td>0.000</td>
<td>0.026</td>
<td>0.023</td>
<td>0.001</td>
<td><strong>0.024</strong></td>
<td>1006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.32</td>
<td>0.418</td>
<td>0.081</td>
<td>0.499</td>
<td></td>
<td>0.776</td>
<td>0.184</td>
<td>0.960</td>
<td>0.874</td>
<td>0.078</td>
<td>0.952</td>
<td>851</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>0.504</td>
<td>0.279</td>
<td>0.783</td>
<td></td>
<td>0.257</td>
<td>0.743</td>
<td>1.000</td>
<td>0.734</td>
<td>0.264</td>
<td>0.998</td>
<td>573</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.37</td>
<td>0.423</td>
<td>0.482</td>
<td>0.905</td>
<td></td>
<td>0.046</td>
<td>0.954</td>
<td>1.000</td>
<td>0.543</td>
<td>0.457</td>
<td>1.000</td>
<td>445</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.40</td>
<td>0.213</td>
<td>0.769</td>
<td>0.982</td>
<td></td>
<td>0.001</td>
<td>0.999</td>
<td>1.000</td>
<td>0.247</td>
<td>0.753</td>
<td>1.000</td>
<td>316</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE II. Type I error (in boldface) and power of fixed sample size and group sequential tests

<table>
<thead>
<tr>
<th>$p_2$</th>
<th>$p_1$</th>
<th>pr (NI)</th>
<th>pr (S)</th>
<th>pr (+)</th>
<th>E (T)</th>
<th>FSS$_1$</th>
<th>FSS$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.70</td>
<td>0.70</td>
<td>0.786</td>
<td><strong>0.025</strong></td>
<td>0.811</td>
<td>1664</td>
<td><strong>0.025</strong></td>
<td>0.810</td>
</tr>
<tr>
<td>0.60</td>
<td>0.001</td>
<td>0.000</td>
<td><strong>0.001</strong></td>
<td>0.000</td>
<td>693</td>
<td>0.000</td>
<td><strong>0.001</strong></td>
</tr>
<tr>
<td>0.73</td>
<td>0.706</td>
<td>0.275</td>
<td>0.981</td>
<td>1680</td>
<td>0.287</td>
<td>0.980</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.363</td>
<td>0.634</td>
<td>0.997</td>
<td>1529</td>
<td>0.653</td>
<td>0.998</td>
<td></td>
</tr>
<tr>
<td>0.77</td>
<td>0.094</td>
<td>0.906</td>
<td>1.000</td>
<td>1251</td>
<td>0.916</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>0.80</td>
<td>0.002</td>
<td>0.998</td>
<td>1.000</td>
<td>792</td>
<td>0.998</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>0.60</td>
<td>0.709</td>
<td><strong>0.025</strong></td>
<td>0.734</td>
<td>1626</td>
<td><strong>0.025</strong></td>
<td>0.736</td>
</tr>
<tr>
<td>0.50</td>
<td>0.001</td>
<td>0.000</td>
<td><strong>0.001</strong></td>
<td>0.000</td>
<td>713</td>
<td>0.000</td>
<td><strong>0.001</strong></td>
</tr>
<tr>
<td>0.63</td>
<td>0.712</td>
<td>0.244</td>
<td>0.956</td>
<td>1679</td>
<td>0.253</td>
<td>0.956</td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>0.429</td>
<td>0.563</td>
<td>0.992</td>
<td>1569</td>
<td>0.584</td>
<td>0.992</td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>0.150</td>
<td>0.849</td>
<td>0.999</td>
<td>1339</td>
<td>0.864</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>0.70</td>
<td>0.008</td>
<td>0.992</td>
<td>1.000</td>
<td>911</td>
<td>0.993</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>0.80</td>
<td>0.80</td>
<td>0.895</td>
<td><strong>0.025</strong></td>
<td>0.920</td>
<td>1715</td>
<td><strong>0.025</strong></td>
<td>0.922</td>
</tr>
<tr>
<td>0.70</td>
<td>0.001</td>
<td>0.000</td>
<td><strong>0.001</strong></td>
<td>0.000</td>
<td>646</td>
<td>0.000</td>
<td><strong>0.001</strong></td>
</tr>
<tr>
<td>0.83</td>
<td>0.645</td>
<td>0.351</td>
<td>0.997</td>
<td>1662</td>
<td>0.368</td>
<td>0.997</td>
<td></td>
</tr>
<tr>
<td>0.85</td>
<td>0.225</td>
<td>0.775</td>
<td>1.000</td>
<td>1425</td>
<td>0.791</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>0.87</td>
<td>0.025</td>
<td>0.975</td>
<td>1.000</td>
<td>1039</td>
<td>0.978</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>0.90</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>584</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>