Journal of Biopharmaceutical Statistics, 18: 494–501, 2008

ISSN: 1054-3406 print/1520-5711 online
DOI: 10.1080/10543400802020987

Taylor & Francis
Taylor & Francis Group

STATISTICAL ASSESSMENT OF QT/QTc PROLONGATION BASED ON MAXIMUM OF CORRELATED NORMAL RANDOM VARIABLES

Bin Cheng[1], Shein-Chung Chow[2], David Burt[3], and Dennis Cosmatos[3]

[1]*Department of Biostatistics, Columbia University, New York, New York, USA*
[2]*Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, USA*
[3]*Biostatistics for Translational Development, Wyeth Research, Collegeville, Pennsylvania, USA*

To establish noninferiority in QT\|QTc prolongation of a test drug with respect to either a placebo or an active control, a thorough QT\|QTc study is recommended by ICH (ICH E14, 2005) which concerns statistical inference on the maximal time-matched drug effect. The existing statistical methods for assessing such effects suffer either power loss or parameter restriction. In this paper, we propose a new asymptotic test with small sample correction based on distribution of maximum of correlated random variables under both a parallel-group design and a crossover design. Simulations indicate that our proposed test has adequate powers.

Key Words: Crossover design; Heteroscedasticity; Maximum of normal variables; Parallel-group design; Small sample correction; Thorough QT/QTc studies.

1. INTRODUCTION

In clinical practice, it is recognized that the prolongation of QT/QTc intervals is related to an increased risk of cardiotoxicity such as a life-threatening arrhythmia (Temple, 2003). Thus it is suggested that a careful evaluation of potential QT/QTc prolongation be performed for potential drug-induced cardiotoxicity. As a result, a draft guidance on the clinical evaluation of QT/QTc interval prolongation and proarrythmic potential for non-antiarrythmic drugs has been prepared by the ICH (ICH E14, 2005). This draft guidance calls for a placebo-controlled study in normal healthy volunteers with a positive control to assess cardiotoxicity by examining QT/QTc prolongation. Under a valid study design (e.g., a parallel-group design or a crossover design), ECG readings will be collected at baseline and at several time intervals posttreatment for each subject. And at each time interval, it is suggested that 3 to 5 ECG recording replicates within a 2- to 5-minute period at each

Received February 14, 2008; Accepted February 28, 2008
Address correspondence to Bin Cheng, Department of Biostatistics, Columbia University, New York, NY 10032, USA; E-mail: bc2159@columbia.edu

time interval be collected for a more accurate and reliable assessment of QT/QTc prolongation (Malik and Camm, 2001). The ICH E14 guidance recommends a thorough QT/QTc study to decide whether the drug induces QT/QTc prolongation as is evidenced if the upper bound of the 95% confidence interval of the mean drug effect on QTc exceeds 10 ms. Statistical methods for thorough QT/QTc study have been proposed by Patterson et al. (2005b) under linear mixed models and by Eaton et al. (2006) using a confidence interval approach. Hosmane and Locke (2005) examined the power in thorough QT/QTc studies via a simulation study. For a review of statistical design and analysis in QT/QTc studies, see Patterson et al. (2005a).

The testing method proposed in Patterson et al. (2005b) was essentially an intersection-union method, which is typically conservative. To address this issue, Eaton et al. (2006) constructed a confidence interval, via δ-method, for a parameter which sufficiently approximates the parameter of interest. However, this method technically assumes that mean QT/QTc differences between drug and placebo are positive at all time intervals, which is both restrictive and unverifiable. Furthermore, when applying to a function (although smooth) that approximates a nonsmooth function (i.e., maximum function), the δ-method may yield a confidence interval with an actual coverage considerably different from the nominal one, particularly when the sample size is moderate. To address these limitations, we propose a new testing method based on the maximum of correlated normal random variables.

The remainder of this article is organized as follows. In the next section, the hypotheses associated with thorough QT/QTc study are formulated under a parallel-group design. An asymptotic test for assessment of QT/QTc prolongation is derived in Section 2.1. A small sample correction of this test is described in Section 2.2. The extension of the results in Section 2 to the crossover design case is briefly discussed in Section 3. In Section 4, simulation results are provided to investigate the finite sample performance of the proposed method. An example is also discussed as an illustration of the proposed method. Discussions and conclusions are provided in Section 5.

2. TEST OF QT/QTc PROLONGATION UNDER PARALLEL-GROUP DESIGN

As indicated in Patterson et al. (2005a), a typical study design for thorough QT/QTc studies is either a parallel-group design or a crossover design. Under a parallel-group design, qualified subjects will be randomly assigned to receive either treatment A or treatment B. ECGs will be collected at baseline and at several time intervals posttreatment. Subjects usually fast at least 3 hours and rest at least 10 minutes prior to scheduled ECG measurements. Identical lead-placement and the same ECG machine will be used for all measurements. As recommended by Malik and Camm (2001), 3 to 5 recording replicate ECGs at each time interval should be obtained within a 2- to 5-minute period. In practice, the time intervals are about 2 hours apart. For simplicity of exposition, in this paper we will model the averaged QTc measures over the recording replicates at each time interval.

Let y_{ijk} be the average QTc response (possibly adjusted for baseline) over the recording replicates at the kth time interval of the jth subject receiving the ith

treatment, where $i = 1$ indicates test drug and $i = 2$ indicates placebo (or active control); $j = 1, \ldots, n$; and $k = 1, \ldots, m$. Under a parallel-group design, y_{ijk} can be described by the following model:

$$y_{ijk} = \mu_{ik} + e_{ij} + \epsilon_{ijk}, \tag{1}$$

where μ_{ik} is the mean of the ith treatment at the kth time interval, e_{ij}'s are independent and normally distributed subject random effects with mean 0 and variance σ_1^2, and ϵ_{ijk}'s are independent and identically distributed normal random errors with mean 0 and variance σ^2. Define $\delta_k = \mu_{1k} - \mu_{2k}$, and $\theta = \max_{1 \le k \le m} \delta_k$; then, a thorough QT/QTc study is equivalent to test the following hypotheses:

$$H_0 : \theta \ge 10, \quad \text{versus} \quad H_1 : \theta < 10. \tag{2}$$

In fact, suppose the noninferiority in QTc prolongation is claimed via a 95% confidence upper bound based on a statistic U; then, according to the ICH E14 guidance, this means that $U + z_{0.05}\ \mathrm{SE}(U) < 10$, or equivalently, $\frac{U-10}{\mathrm{SE}(U)} < -z_{0.05}$, which rejects H_0 in (2) at level 0.05. Here, $\mathrm{SE}(U)$ denotes the estimated standard error of U.

2.1. Asymptotic Testing Procedure

Define $W_k = \bar{y}_{1.k} - \bar{y}_{2.k}$, where $\bar{y}_{i.k}$ is the sample mean for the ith treatment at the kth time interval. We first establish the following asymptotic result.

Theorem 1. *Let* $T = \max_{1 \le k \le m} W_k$, *and* θ *be defined in* (2); *then*

$$\sqrt{n}(T - \theta) \to_d N\big(0, 2(\sigma_1^2 + \sigma^2)\big),$$

where $\to_d$ *means convergence in distribution.*

Proof. The random vector $\mathbf{W} = (W_1, \ldots, W_m)'$ is normally distributed with mean $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)'$ and variance $\boldsymbol{\Sigma} = \frac{2\sigma_1^2}{n}\mathbf{U}_m + \frac{2\sigma^2}{n}\mathbf{I}_m$, where $\mathbf{U}$ is the $m \times m$ matrix of one's and $\mathbf{I}_m$ is the $m \times m$ identity matrix. By Afonja (1972), the moment-generating function of T is

$$M_T(t) = \sum_{k=1}^{m} e^{\delta_k t + (\sigma_1^2 + \sigma^2)\frac{t^2}{n}} \Phi_{m-1}(\mathbf{d}_k; \mathbf{R}_{-k}),$$

where

$$\mathbf{d}_k = \{d_{kl}\}_{l \ne k}, \quad d_{kl} = \frac{(\delta_l - \delta_k)}{2\sigma}\sqrt{n} - \frac{\sigma t}{\sqrt{n}},$$

and $\Phi_{m-1}(\mathbf{d}_k; \mathbf{R}_{-k}) = \int_{\mathbf{d}_k}^{\infty} \phi_{m-1}(\mathbf{z}, \mathbf{R}_{-k})$ is the survival function of an $(m-1)$ dimensional, mean $\mathbf{0}$, normal random vector whose variance is the correlation

matrix of the $(m-1)$ dimensional random vector $\mathbf{W}_{-k} = \{W_k - W_l\}$, $l \neq k$. Then, the moment-generating function of $\sqrt{n}(T-\theta)$ is

$$\begin{aligned} M_{\sqrt{n}(T-\theta)}(t) &= e^{-t\sqrt{n}} \sum_{k=1}^{m} e^{\delta_k t\sqrt{n} + (\sigma_1^2+\sigma^2)t^2} \Phi_{m-1}(\mathbf{d}_k; \mathbf{R}_{-k}) \\ &= e^{-t\sqrt{n}} \sum_{k=1}^{m} e^{\delta_k t\sqrt{n} + (\sigma_1^2+\sigma^2)t^2} \mathrm{I}(\delta_k = \theta)(1 + \mathrm{o}(1)) \\ &= e^{(\sigma_1^2+\sigma^2)t^2}(1 + \mathrm{o}(1)), \end{aligned}$$

which implies the claim.

By Theorem 1, an asymptotic α level test rejects H_0 in (2) if and only if

$$\frac{T-10}{\sqrt{2(\hat{\sigma}_1^2 + \hat{\sigma}^2)/n}} < -z_\alpha, \tag{3}$$

where

$$\hat{\sigma}_1^2 + \hat{\sigma}^2 = \sum_{i=1}^{2} \sum_{k=1}^{m} \sum_{j=1}^{n} \big(y_{ijk} - \bar{y}_{i.k}\big)^2 / \big(2m(n-1)\big).$$

2.2. Small Sample Correction

When the number of patients n per treatment group is small, the normal approximation of distribution of T as suggested in Theorem 1 may not work well. In this section, we propose a small sample correction of the distribution of T and illustrate how to modify the test described in the previous section.

Define $\mathbf{a}_k = \{a_{kl}\}$, $a_{kl} = \frac{(\delta_l - \delta_k)}{2\sigma}\sqrt{n}$ for $k \neq l$ and $a_{kk} = -\infty$, and define $\mathbf{R}_k = \{r_{k,ll'}\}_{l,l'=1}^m$, $r_{k,ll'} = \sigma^2/(\sigma_1^2 + \sigma^2)$ for $l \neq l'$ and $r_{k,ll} = 1$. Let k_0 be such that $\delta_{k_0} = \max_{1 \leq k \leq m} \delta_k = \theta$; then, by Afonja (1972),

$$\begin{aligned} E(T) &= \sum_{k=1}^{m} \delta_k \int_{\mathbf{a}_k}^{\infty} \phi_m(\mathbf{z}, \mathbf{R}_k) + \sqrt{\frac{2(\sigma_1^2 + \sigma^2)}{n}} \sum_{k=1}^{m} \int_{\mathbf{a}_k}^{\infty} z_k \phi_m(\mathbf{z}, \mathbf{R}_k) \\ &= \theta \int_{\mathbf{a}_{k_0}}^{\infty} \phi_m(\mathbf{z}, \mathbf{R}_{k_0}) + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right) \\ &= \theta\varrho + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right); \end{aligned}$$

thus,

$$E(T) \approx \theta\varrho, \quad \varrho = \int_{\mathbf{a}_{k_0}}^{\infty} \phi_m(\mathbf{z}, \mathbf{R}_{k_0}). \tag{4}$$

Similarly, because

$$E(T^2) = \sum_{k=1}^{m} \delta_k^2 \int_{\mathbf{a}_k}^{\infty} \phi_m(\mathbf{z}, \mathbf{R}_k) + \sqrt{\frac{2(\sigma_1^2 + \sigma^2)}{n}} \sum_{k=1}^{m} \int_{\mathbf{a}_k}^{\infty} z_k \phi_m(\mathbf{z}, \mathbf{R}_k) + \frac{2(\sigma_1^2 + \sigma^2)}{n} \sum_{k=1}^{m} \int_{\mathbf{a}_k}^{\infty} z_k^2 \phi_m(\mathbf{z}, \mathbf{R}_k)$$
$$= \theta^2 \int_{\mathbf{a}_{k_0}}^{\infty} \phi_m(\mathbf{z}, \mathbf{R}_{k_0}) + \frac{2(\sigma_1^2 + \sigma^2)}{n} \int_{\mathbf{a}_{k_0}}^{\infty} z_{k_0}^2 \phi_m(\mathbf{z}, \mathbf{R}_{k_0}) + \mathrm{o}\left(\frac{1}{\sqrt{n}}\right),$$

we have

$$\mathrm{Var}(T) \approx \theta^2 \varrho(1 - \varrho) + \frac{2(\sigma_1^2 + \sigma^2)}{n} \gamma, \quad \gamma = \int_{\mathbf{a}_{k_0}}^{\infty} z_{k_0}^2 \phi_m(\mathbf{z}, \mathbf{R}_{k_0}). \tag{5}$$

Now, by replacing in (4) and (5) $k_0, \mathbf{a}_{k_0}, \sigma_1^2$, and σ^2 with their obvious estimators, we get $\hat{\varrho}$ and $\hat{\gamma}$. Then, a small sample corrected level α test rejects H_0 in (2) if and only if

$$\frac{T - 10\hat{\varrho}}{\sqrt{100\hat{\varrho}(1 - \hat{\varrho}) + 2(\hat{\sigma}_1^2 + \hat{\sigma}^2)\hat{\gamma}/n}} < -z_\alpha, \tag{6}$$

3. TEST OF QT/QTc PROLONGATION UNDER CROSSOVER DESIGN

Let A and B be the two treatments under investigation. Under a crossover design, qualified subjects will be randomly assigned to either sequence AB (sequence 1) or Sequence BA (sequence 2) and a sufficient length of washout period will be imposed between the two periods. As in the parallel-group design, for each treating period, ECGs will be collected at baseline and posttreatment at several 2-hour time intervals. Subjects usually fast at least 3 hours and rest at least 10 minutes prior to scheduled ECG measurements, and 3 to 5 recording replicate ECGs at each time interval will be obtained within a 2- to 5-minute period.

For simplicity, let y_{ijkl} be the average QTc responses (possibly adjusted for baseline) over the recording replicates at the lth time interval of the kth treating period for the jth subjects in ith sequence, where $i = 1, 2$; $j = 1, \ldots, n$; $k = 1, 2$; and $l = 1, \ldots, m$. Under a crossover design, treatment index u is a function of (i, k), denoted as $u = d(i, k)$ with $u = 1$ denoting test drug and $u = 2$ placebo or active control. We consider the following model:

$$y_{ijkl} = \mu + \alpha_k + \beta_{ul} + a_{ij} + b_{ijk} + \epsilon_{ijkl}, \tag{7}$$

where μ is the overall mean, α_k is the period effect, β_{ul} is the treatment effect at the lth time interval, a_{ij} is the subject random effect, b_{ijk} is the period random effect nested in the jth subject in the ith sequence, and ϵ_{ijkl} is the random error. We assume that $a_{ij} \sim N(0, \sigma_2^2)$, $b_{ijk} \sim N(0, \sigma_1^2)$, $\epsilon_{ijkl} \sim N(0, \sigma^2)$, a_{ij}, b_{ijk}, and ϵ_{ijkl}'s are independent.

Under model (7), the treatment effect at the lth time interval is $\delta_l = \beta_{1l} - \beta_{2l}$. Let $\theta = \max_{1 \le l \le m} \delta_l$; then, the hypotheses of QTc prolongation in a thorough QT/QTc study under the crossover design are formulated the same as in (2). Define $W_l = (\bar{y}_{1.1l} - \bar{y}_{1.2l} + \bar{y}_{2.2l} - \bar{y}_{2.1l})/2$, $l = 1, \ldots, m$; then, W_l is an estimator of δ_l, and it is straightforward to show that the random vector $\mathbf{W} = (W_1, \ldots, W_m)'$ has the same distribution as described in Section 2.1. Therefore, a test similar to the one derived in the previous section can be constructed.

4. NUMERICAL STUDY

4.1. Simulation

We conduct a simulation to check the performance of the asymptotic test proposed in Section 2.1. For ease of comparison, we adopt a similar setup as in Eaton et al. (2006), although we acknowledge that the confidence intervals constructed in their paper were around $\theta = 5$; therefore, the estimated coverage probabilities reported in their simulations were not related to the power for testing our hypothesis (2).

Specifically, we consider 6 time intervals ($m = 6$) and consider $\sigma_1^2 + \sigma^2 = 100$; $\rho = \sigma_1^2/(\sigma_1^2 + \sigma^2) = 0.2, 0.4, 0.6, 0.8$; and $n = 40, 60, 80, 100$. The estimated size for $(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6) = (1, 1, 10, 1, 1, 1)$ is given in Table 1. The estimated power for $(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6) = (1, 2, 5, 2, 1, 1)$ is given in Table 2. All estimations are based on 5,000 simulation runs. From Table 1, we can seen that our proposed test has sizes very close to the 5% nominal level. Table 2 indicates that for a sample size as small as 40 to 60 subjects per group, the test has enough powers.

4.2. Example

To illustrate the proposed test procedure, we consider an example concerning a thorough QTc study with time-dependent recording replicates. Under the parallel-group design, 380 qualified subjects were randomly assigned to either a test treatment or an active control agent ($n = 190$). Subjects were at rest prior to the scheduled ECG. QT measurements were taken in recordings of 5 replicates within 2 minutes of one another. Five time intervals ($m = 5$) were considered 2 hours apart. The vector $\mathbf{W}$ defined in Section 2.1 was calculated as

$$\mathbf{W} = (8.98, 8.47, 7.96, 8.78, 10.05)', \quad T = 10.05.$$

Table 1 Estimated size under $(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6) = (1, 1, 10, 1, 1, 1)$

n	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
40	0.0452	0.0494	0.0482	0.0516
60	0.0524	0.0548	0.0520	0.0528
80	0.0486	0.0502	0.0496	0.0594
100	0.0478	0.0524	0.0514	0.0484

Table 2 Estimated power under $(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6) = (1, 2, 5, 2, 1, 1)$

n	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
40	0.6794	0.7054	0.7202	0.7286
60	0.8562	0.8570	0.8574	0.8650
80	0.9396	0.9370	0.9344	0.9350
100	0.9714	0.9714	0.9740	0.9684

And because $\hat{\sigma}_1^2 + \hat{\sigma}^2 = 229.78$, we have

$$\frac{T - 10}{\sqrt{2(\hat{\sigma}_1^2 + \hat{\sigma}^2)/n}} = \frac{10.05 - 10}{\sqrt{2 \times 229.78/190}} = 0.03 > -1.64 = -z_{0.05};$$

hence we do not reject H_0, implying there was no statistical evidence to claim the test drug's noninferiority to placebo in QTc prolongation.

5. DISCUSSION

In this paper, we have proposed a new test procedure based on maximum of correlated normal random variables. Although the proposed test was derived under a balanced design without covariates, the method we used can be generalized to allow for not only unbalancedness between the two treatment groups but also adjustment of important covariates such as baseline QTc measures and/or heart rates.

Note that in justifying our method, we essentially do not need to assume any specific form for the variance structure of the random vector $\mathbf{W}$. This implies that our proposed method will still be valid when covariance structures other than the model implied compound symmetric structure such as an AR(1) structure or heterogenous covariance (e.g., unequal variance at different time intervals) are assumed.

It should be noted that our formulation of hypotheses in (2) represents only one of the interpretations of QTc prolongation evidence assessment. Other definitions are worthy of consideration. For example, under a parallel-group design, we could define $\vartheta = \max_{1 \le k \le m} \mu_{1k} - \max_{1 \le k \le m} \mu_{2k}$ and consider testing the following hypotheses:

$$H_0 : \vartheta \ge 10, \quad \text{versus} \quad H_1 : \vartheta < 10.$$

The above hypotheses are relevant in an active-controlled QT/QTc study where the maximal prolongation of the two drugs occurs at different time intervals and where a global comparison rather than a time-matched comparison is desirable. Our proposed method can be easily modified to test these hypotheses.

REFERENCES

Afonja, B. (1972). The moments of the maximum of correlated normal and t-variates. *Journal of the Royal Statistical Society Ser. B* 34:251–262.

Eaton, M. L., Muirhead, R. J., Mancuso, J. Y., Kolluri, S. (2006). A confidence interval for the maximal mean QT interval change caused by drug effect. *Drug Information Journal* 40:267–271.

Hosmane, B., Locke, C. (2005). A simulation study of power in thorough QT/QTc studies and a normal approximation for planning purposes. *Drug Information Journal* 39:447–455.

ICH (2005). ICH E14 Guidance on The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrythmic Potential for Non-Antiarrythmic Drugs. Geneva, Switzerland: International Conference on Harmonisation, May 2005.

Malik, M., Camm, A. J. (2001). Evaluation of drug-induced QT interval prolongation. *Drug Safety* 24:323–351.

Patterson, S., Agin, M., Anziano, R. et al. (2005a). Investigating drug-induced QT and QTc prolongation in the clinic: A review of statistical design and analysis considerations: Report from the Pharmaceutical Research and Manufacturers of America QT Statistics Expert Team. *Drug Information Journal* 39:243–266.

Patterson, S. D., Jones, B., Zariffa, N. (2005b). Modelling and interpreting QTc prolongation in clinical pharmacology studies. *Drug Information Journal* 39:437–445.

Temple, R. (2003). Overview of the concept paper, history of the QT/TdP concern; regulatory implications of QT prolongation. Presentations at Drug Information Agency/FDA Workshop. Available at: www.diahome.org.