A Multivariate Approach to Sample Size Calculations for Thorough QT Studies

Suraj P. Anand,[1,*] Sharon C. Murray,[2] and Gary G. Koch[3]

[1]Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA

[2]Discovery Biometrics, Oncology, GlaxoSmithKline, Research Triangle Park, North Carolina, USA

[3]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Institute of Statistics Mimeo Series #2608

Abstract

Thorough QT studies are becoming a necessary part of the clinical profile of existing and investigational new drugs. The E14 guidelines mandate performing a thorough QT study on any non antiarrythmic drug to ensure that the drug does not produce a QT prolongation effect before it can be approved and marketed. The cost involved in running a thorough QT study is substantial and an adverse outcome of the study can be detrimental to the safety profile of the drug, hence sample size calculations play a very important role in ensuring a small but adequately powered thorough QT study. Measurements taken in a QT study are naturally in time order and hence correlated, so multivariate methods are the most appropriate tools for analyzing such data. However, there is no readily available methodology that makes use of a multivariate model for power and sample size calculations. Most of the current methods take a univariate approach that depends on a standard noninferiority test at each timepoint, assuming independence between the timepoints. In this paper, we propose a Monte Carlo simulation-based approach using a multivariate model that has several advantages over the univariate approach, namely, a) It makes use of the correlation structure prevalent in the data, b) Power can be evaluated for various structures of mean difference vector for a given maximum difference, and c) It generally results in a smaller sample size for fixed power. We also illustrate the use of our methodology by applying it to a real life dataset.

*Correspondence: Suraj Anand, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA; E-mail: spanand@ncsu.edu

Keywords: Power and Sample size ; ICH; Monte Carlo simulation; Correlation structure; QTc prolongation; Thorough QT/QTc study
Mathematics Subject Classification: 60E05; 62F03; 62H99; 65C05.

1 Introduction

Thorough QT studies are becoming a necessary part of the clinical profile of existing and investigational new drugs. The International Conference of Harmonization (ICH) E14 guidelines (available at `www.fda.gov/cder/guidance/`) mandate performing a thorough QT study on any non antiarrythmic drug to determine whether the drug has a threshold pharmacological effect on cardiac repolarization, as detected by QTc prolongation, before it can be approved and marketed. The cost involved in running a thorough QT study is substantial and an adverse outcome of the study can be detrimental to the safety profile of the drug, hence sample size calculations play a very important role in ensuring a small but adequately powered thorough QT study.

A thorough QT study is typically carried out in healthy volunteers (as opposed to individuals at increased risk of arrhythmias) and is used to determine whether or not the effect of a drug on the QTc interval in target patient populations should be studied intensively during later stages of drug development. The standard way of analyzing data from a thorough QT study is to construct a 90% two-sided (same as a 95% one-sided) confidence interval (CI), for the difference in mean QTc between drug and placebo at each time-point, and to conclude non-inferiority if the upper limits for all these CIs are less than a pre-specified constant (10 milliseconds (msec) according to E14 guidelines). A mixed model approach is used to model the baseline corrected QTc values with treatment, timepoint, and their interaction as fixed effects, and subject as a random effect, to construct the 90% two-sided CIs for the time-matched mean differences. Other study specific fixed and random effects are usually included according to the study design and objectives. A positive control that is known to have a QT prolonging effect (e.g. Moxifloxacin) is included to assess the sensitivity of the study in detecting the QT effects. If the positive control demonstrates an effect as compared to placebo, then the assay is considered reasonably sensitive to detection of QTc prolongation. Under the standard normality assumptions, this approach corresponds to both intersection-union test and the likelihood ratio test (Eaton et. al., 2006).

The ICH E14 guidance document defines a negative 'Thorough QT/QTc Study' as the one in which the upper bound of the 95% one-sided confidence

interval for the largest time-matched mean effect of the drug on the QTc interval excludes 10 msec. According to the guidance, this definition is chosen to provide reasonable assurance that the mean effect of the study drug on the QTc interval is not greater than 5 msec. When the largest time-matched mean difference exceeds the threshold, the study is termed positive. Hence, one alternative approach would be to base the non-inferiority inference on the largest time-matched difference between population mean QTc between drug and placebo. Coming up with an interval estimate for the maximum difference however is a non-trivial task. There have been few attempts (Eaton et. al., 2006; Boos et. al., 2007; Ghosh and Anand, 2008 (Submitted)) in this direction but these methods are still to come into serious consideration by Food and Drug Administration (FDA) and other regulatory agencies. Hence, the intersection-union test-based standard approach as described above is still the most widely used method for analyzing data arising from a thorough QT study. In this context, the standard approach is known to be very conservative (Patterson et. al., 2005) but still finds favor from the regulatory agencies since it does not compromise on the consumer's interests. Kong et. al. (2004) conducted comprehensive simulation studies to assess type I error and power for noninferiority/equivalence trial settings in the context of vaccine studies with multiple correlated antibody measures in sera and confirmed the conservative nature of the intersection-union test.

Measurements taken in a QT study are naturally in time order and hence multivariate methods are the most appropriate tools for planning and analyzing such a study. However, there is no readily available methodology that makes use of a multivariate model for the power and sample size calculations for a thorough QT study. Most of the current methods used for power and sample size calculations take a univariate approach that depends on a standard noninferiority test at each timepoint, assuming independence between the timepoints. One way to account for multiple timepoints would be to divide the target type II error rate by the total number of timepoints, and calculate the sample size by trial and error method to attain the target type II error rate (and hence power). Such methods may not be satisfactory since they do not take into account the correlation structure that may be present between the measurements taken at different timepoints on the same subject. Depending on the length of the intervals between the measurements the correlation may be of a small or large magnitude. The correlation structure between the QT measurements can have an impact on the sample size and power and therefore should be taken into account.

In this paper, we propose a Monte Carlo simulation-based approach using a multivariate model for power and sample size calculations for a thorough QT study. It has several advantages over the inadequate univariate approaches, namely, a) It makes use of the correlation structure prevalent in the data. b) Power can be evaluated for various types of mean difference vectors for a given maximum difference. For example, with a maximum difference of 3 msec and assuming 5 timepoints, one could look at an alternative hypothesis of constant mean structure (3,3,3,3,3), hill structure (1,2,3,2,1), or steady state structure (1,2,3,3,3). c) It generally results in a smaller sample size. We present a detailed multivariate framework for both parallel and crossover designs in Section 2. In Section 3, we provide the details of our simulation methods and results. We illustrate the use of our approach by applying it to real data arising form a GSK run thorough QT study in Section 4, and finally, in Section 5 we provide a short conclusion and discussion on the implications of the results.

2 Multivariate Framework for QTc Data

Thorough QT studies can be crossover or parallel with multiple arms that generally include a placebo, a positive control, and one or more dose levels of the study treatment. It is common to have at least one therapeutic and a supratherapeutic dose of the study treatment. For the purpose of calculating sample size, we will assume one placebo and one drug arm, and would then recommend the sample size for all arms of the study. The hypotheses of interest can be set based on a non-inferiority type of testing procedure as follows:

$$\begin{aligned} H_0 &: \bigcup_{k=1}^{p} \{\mu_{2(k)} - \mu_{1(k)}\} \geq \delta \\ H_1 &: \bigcap_{k=1}^{p} \{\mu_{2(k)} - \mu_{1(k)}\} < \delta \end{aligned} \tag{1}$$

where $\mu_{1(k)}$ and $\mu_{2(k)}$ are the baseline corrected QTc values for the placebo and drug arms respectively at timepoint k. The regulatory threshold is in terms of a two-sided 90% upper CI for the difference in $\mu_{2(k)}$ and $\mu_{1(k)}$ at each timepoint and is set to 10 msec. The type I error associated with this set of hypotheses would be exempting a drug incorrectly from the risk of QT prolongation, whereas, type II error will constitute falsely incriminating a drug

with the risk of QT prolongation.

2.1 Parallel Design

A parallel design comparing drug and placebo consists of two separate groups of subjects with each subject assigned to either placebo or the drug arm. Let $\boldsymbol{Y}_{1i},\ i = 1,\dots n_1$ and $\boldsymbol{Y}_{2j},\ j = 1,\dots n_2$ denote the vectors of p baseline corrected QTc measurements taken on ith and jth subjects in placebo and drug arms respectively, where n_1 and n_2 denote the number of subjects receiving placebo and drug, and p is the total number of time points. The measurements are assumed to be distributed identically and independently (iid), arising from two independent p-variate normal distributions (denoted by $\mathcal{N}_p$), and can be represented as follows:

$$\begin{aligned} \boldsymbol{Y}_{1i} &\overset{iid}{\sim} \mathcal{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \quad i = 1,\dots,n_1 \\ \boldsymbol{Y}_{2j} &\overset{iid}{\sim} \mathcal{N}_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), \quad j = 1,\dots,n_2 \end{aligned} \tag{2}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the population means for placebo and drug respectively, and the covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are assumed to be completely known, either on the basis of previous studies or a large sample size for the current study. Let $\boldsymbol{\Delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ be the p-dimensional vector of time-matched population mean differences between drug and placebo. By the Multivariate Central Limit Theorem, it is known that the maximum likelihood estimates (MLE) $\hat{\boldsymbol{\mu}}_1 = \bar{\boldsymbol{y}}_1$ and $\hat{\boldsymbol{\mu}}_2 = \bar{\boldsymbol{y}}_2$, where $\bar{\boldsymbol{y}}_1 = \sum_i \boldsymbol{y}_{1i}/n_1$ and $\bar{\boldsymbol{y}}_2 = \sum_j \boldsymbol{y}_{2j}/n_2$ denote the sample means, follow p-variate normal distributions. More specifically, $\bar{\boldsymbol{y}}_1 \sim \mathcal{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1/n_1)$, and $\bar{\boldsymbol{y}}_2 \sim \mathcal{N}_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2/n_2)$. If we denote by $\boldsymbol{d}$ the time-matched mean difference between the sample means $\bar{\boldsymbol{y}}_2$ and $\bar{\boldsymbol{y}}_1$, $\boldsymbol{d} = \bar{\boldsymbol{y}}_2 - \bar{\boldsymbol{y}}_1$, and the known covariance matrix of $\boldsymbol{d}$, $\boldsymbol{\Sigma}_d = \boldsymbol{\Sigma}_1/n_1 + \boldsymbol{\Sigma}_2/n_2$, then it follows that (Anderson, 1984),

$$\boldsymbol{\Sigma}_d^{-1/2}(\boldsymbol{d} - \boldsymbol{\Delta}) \xrightarrow{d} \mathcal{N}_p(\boldsymbol{0}, \boldsymbol{I}_p). \tag{3}$$

It is clear that each individual component of $\boldsymbol{d}$ follows a univariate normal distribution,

$$d_k \sim \mathcal{N}_1(\Delta_{(k)}, \sigma_{d(kk)}), k = 1,\dots p, \tag{4}$$

where $\Delta_{(k)}$ is the kth component of the mean difference vector $\boldsymbol{\Delta}$ and $\sigma_{d(kk)}$ is the kth diagonal element of the covariance matrix $\boldsymbol{\Sigma}_d$. Thus if we sample from the distribution of $\boldsymbol{d}$, we can very easily construct confidence intervals on the individual components Δ_k. A 95% one sided upper confidence limit for Δ_k,

k=1,... p, is given by $d_k + z_{0.05}\sqrt{\sigma_{d(kk)}}$, where $z_{0.05} = 1.645$ is the 95th quantile of the standard normal distribution. Redefining the standard test in terms of confidence intervals, we can set the power equation to be

$$power = P\{\bigcap_{k=1}^{p}(95\% \text{ upper confidence limit (UCL) for } \Delta_k) < 10|H_1\}, \quad (5)$$

which is equivalent to writing

$$power = P\{\max_{1\leq k\leq p}(95\% \text{ upper confidence limit (UCL) for } \Delta_k) < 10|H_1\}, \quad (6)$$

where power is calculated under any configuration of $\boldsymbol{\Delta}$ in the alternative space H_1. To avoid the complicated numerical integration in the multivariate alternative space, we use Monte Carlo simulations to repeatedly generate vectors of mean differences ($\boldsymbol{d}$), under some configuration in H_1. We then construct a 95% upper confidence limit for each Δ_k, k=1,... p, and calculate the proportion of times the largest 95% upper confidence limit across the p timepoints is below the regulatory threshold of 10 msec as an estimate for power.

2.1.1 Mixed Effect Model as a Special Case for Parallel Data

As a special case, we adopt the widely used mixed effect model for the parallel data, which can be written succinctly in a matrix notation for the ith subject as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i, i = 1, \ldots, n_1 + n_2, \quad (7)$$

where $\boldsymbol{X}_i\boldsymbol{\beta}$ denotes the fixed effect part of the model with treatment, time, treatment by time interaction and baseline QTc as the fixed effects, and $\boldsymbol{Z}_i\boldsymbol{\gamma}_i$ denotes the random effect part of the model, with random subject effect as the random component. The actual model may have more study specific covariates. We assume that $\boldsymbol{\epsilon}_i \sim \mathcal{N}_p(\boldsymbol{0}, \sigma_e^2\boldsymbol{I}_p)$, $\boldsymbol{\gamma}_i \sim \mathcal{N}(0, \sigma_s^2)$, with $\boldsymbol{\epsilon}_i$ independent of $\boldsymbol{\gamma}_i$. With $\boldsymbol{Z}_i = \boldsymbol{1}_p$,

$$Var(\boldsymbol{Y}_i) = \boldsymbol{Z}_i Var(\boldsymbol{\gamma}_i)\boldsymbol{Z}_i^{'} + \sigma_e^2\boldsymbol{I}_p = \sigma_s^2\boldsymbol{1}_p\boldsymbol{1}_p^{'} + \sigma_e^2\boldsymbol{I}_p.$$

If we let $\sigma^2 = \sigma_s^2 + \sigma_e^2$ and $\rho = \sigma_s^2/\sigma^2$, then

$$Var(\boldsymbol{Y}_i) = \sigma^2\{(1-\rho)\boldsymbol{I}_p + \rho\boldsymbol{1}_p\boldsymbol{1}_p^{'}\}.$$

Using the above results,

$$\boldsymbol{\Sigma}_d = (1/n_1 + 1/n_2)\sigma^2\{(1-\rho)\boldsymbol{I}_p + \rho\mathbf{1}_p\mathbf{1}_p^{'}\}, \tag{8}$$

Further assuming $n_1 = n_2 = n$, we can calculate the variance of $\boldsymbol{d}$ as

$$\boldsymbol{\Sigma}_d = (2/n)\sigma^2\{(1-\rho)\boldsymbol{I}_p + \rho\mathbf{1}_p\mathbf{1}_p^{'}\}, \tag{9}$$

and use it for the calculation of sample size and power as outlined above.

The advantage of using a mixed effect model for the parallel data lies in the special structure of the covariance matrix resulting from such modeling, that requires fewer parameter estimates, namely σ^2 and ρ. Estimates of these parameters can be calculated fairly easily using the estimated covariance parameters corresponding to random between subject and error effects obtained from the reports for earlier studies, or running a simple mixed effect model as envisioned above on available data.

2.2 Crossover Design

In a crossover design, subjects receive sequences of treatments (some (incomplete block) or all (complete block) of the study treatments) with the objective of studying differences between individual treatments (or sub-sequences of treatments) (Senn, S. J., 2002). Each individual in such a design serves as his own control and hence a crossover design requires a smaller sample size to achieve the same precision of the estimates than a parallel design. A 2x2 cross-over design refers to two treatments and two sequences (treatment orderings). For example, if the two treatments are placebo and drug then one sequence receives placebo in the first period followed by drug in the second period. The other sequence receives drug and placebo in a reverse ordering. The design includes a washout period between the two active periods to remove any carry-over effects. In an actual Thorough QT study, subjects would actually receive multiple treatments, including placebo, active control, and 2 or 3 drug arms. We will calculate sample size based on a single comparison of drug to placebo and then apply that sample size to the full study.

Let $\boldsymbol{Y}_{1i}$ and $\boldsymbol{Y}_{2i},\ i = 1, \ldots\ n$ denote the vectors of p baseline corrected QTc measurements taken on subject i on the placebo and drug arms respectively, where n denotes the total number of subjects in the study, and p is the total number of time points. The measurements are assumed to be distributed identically and independently (iid), arising from two p-variate normal distributions.

However, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are no more independent because these are observations taken on the same individuals. In matrix notation, this can be represented as follows:

$$\begin{pmatrix} \boldsymbol{Y}_{1i} \\ \boldsymbol{Y}_{2i} \end{pmatrix} \sim \mathcal{N}_{2p} \left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right), \quad i = 1, \ldots, n \tag{10}$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the population means for placebo and drug respectively, the covariance sub-matrices $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ represent the covariance structures for the data vectors corresponding to subjects in the placebo and the drug arms respectively, covariance sub-matrices $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^{'}$ represent the covariance between data vectors corresponding to same subject in period 1 and period 2 respectively. All the covariance sub-matrices are assumed to be completely known, either on the basis of previous studies or a large sample size for the current study. The parameter of interest still is the vector of time-matched population mean differences between drug and placebo $\boldsymbol{\Delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. Analogous to a univariate paired sample test, this problem can be reduced to a multivariate paired one sample setup. If we denote by $\boldsymbol{D}_i$ the difference between the subject matched data vectors for ith subject $\boldsymbol{D}_i = \boldsymbol{Y}_{2i} - \boldsymbol{Y}_{1i}$, then

$$\boldsymbol{D}_i \overset{iid}{\sim} \mathcal{N}_p(\boldsymbol{\Delta}, \boldsymbol{\Sigma}_{diff}), \quad i = 1, \ldots, n \tag{11}$$

where $\boldsymbol{\Sigma}_{diff} = \boldsymbol{\Sigma}_{11} + \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{12}^{'}$ is the completely known covariance structure for the difference. Now if we denote the time-matched mean difference by $\boldsymbol{d}$, $\boldsymbol{d} = \sum_i \boldsymbol{d}_i/n$, then

$$\boldsymbol{d} \sim \mathcal{N}_p(\boldsymbol{\Delta}, \boldsymbol{\Sigma}_d), \tag{12}$$

where $\boldsymbol{\Sigma}_d = \boldsymbol{\Sigma}_{diff}/n$.

Just like the parallel case, each individual component of $\boldsymbol{d}$ follows a univariate normal distribution.

$$\boldsymbol{d}_k \sim \mathcal{N}_1(\Delta_{(k)}, \sigma_{d(kk)}), k = 1, \ldots p, \tag{13}$$

where $\sigma_{d(kk)}$ is the kth diagonal element of the covariance matrix $\boldsymbol{\Sigma}_d$. Thus we can sample from the distribution of $\boldsymbol{d}$ and construct confidence intervals on the individual components of Δ_k. A 95% one sided upper confidence limit for Δ_k, k=1,... p, is given by $\boldsymbol{d}_k + z_{0.05}\sqrt{\sigma_{d(kk)}}$, where $z_{0.05} = 1.645$ is the 95th quantile of the standard normal distribution. To calculate power using equation 6 we use Monte Carlo simulations to repeatedly generate vector of mean differences ($\boldsymbol{d}$) under H_1. We then construct a 95% upper confidence limit for each Δ_k, and

estimate power by calculating the proportion of times the largest 95% upper confidence limit across the p timepoints is below the regulatory threshold of 10 msec.

Mixed Effect Model as a Special Case for Crossover Data

As a special case, we adopt the widely used mixed effect model for the crossover data, which can be written succinctly in a matrix notation for the ith subject as

$$\begin{pmatrix} \boldsymbol{Y}_{1i} \\ \boldsymbol{Y}_{2i} \end{pmatrix} = \boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i, i = 1, \ldots, n, \tag{14}$$

where $\boldsymbol{X}_i\boldsymbol{\beta}$ denotes the fixed effect part of the model with treatment, time, treatment by time interaction, baseline QTc, period and sequence as the fixed effects, and $\boldsymbol{Z}_i\boldsymbol{\gamma}_i$ denotes the random effect part of the model, with random subject effect and random period effect as the random components. The actual model may have more study specific covariates. The benefit of using a random period effect is that it allows for the correlation between two observations from the same period (ρ_1) to differ from correlations between two observations from different periods (ρ_2). We assume that $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{2p}(\boldsymbol{0}, \sigma_e^2\boldsymbol{I}_p)$, $\boldsymbol{\gamma}_i \sim \mathcal{N}_2\left(\boldsymbol{0} \quad , \quad \begin{pmatrix} \sigma_s^2 & 0 \\ 0 & \sigma_p^2 \end{pmatrix}\right)$ with all random effect components independent of each other. With $\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{1}_p & \boldsymbol{1}_p \\ \boldsymbol{1}_p & \boldsymbol{0}_p \end{pmatrix}$,

$$Var(\boldsymbol{Y}_i) = \boldsymbol{Z}_i Var(\boldsymbol{\gamma}_i)\boldsymbol{Z}_i^{'} + \sigma_e^2\boldsymbol{I}_{2p} = \begin{pmatrix} (\sigma_s^2 + \sigma_p^2)\boldsymbol{1}_p\boldsymbol{1}_p^{'} + \sigma_e^2\boldsymbol{I}_p & \sigma_s^2\boldsymbol{1}_p\boldsymbol{1}_p^{'} \\ \sigma_s^2\boldsymbol{1}_p\boldsymbol{1}_p^{'} & (\sigma_s^2 + \sigma_p^2)\boldsymbol{1}_p\boldsymbol{1}_p^{'} + \sigma_e^2\boldsymbol{I}_p \end{pmatrix}.$$

If we let $\sigma^2 = \sigma_s^2 + \sigma_p^2 + \sigma_e^2$, $\rho_1 = (\sigma_s^2 + \sigma_p^2)/\sigma^2$, and $\rho_2 = \sigma_s^2/\sigma^2$, then

$$Var(\boldsymbol{Y}_i) = \begin{pmatrix} \sigma^2\{(1-\rho_1)\boldsymbol{I}_p + \rho_1\boldsymbol{1}_p\boldsymbol{1}_p^{'}\} & \rho_2\sigma^2\boldsymbol{1}_p\boldsymbol{1}_p^{'} \\ \rho_2\sigma^2\boldsymbol{1}_p\boldsymbol{1}_p^{'} & \sigma^2\{(1-\rho_1)\boldsymbol{I}_p + \rho_1\boldsymbol{1}_p\boldsymbol{1}_p^{'}\} \end{pmatrix}.$$

Using the above results, we can calculate the variance of $\boldsymbol{D}_i$ as

$$Var(\boldsymbol{D}_i) = \Sigma_{diff} = \begin{pmatrix} -\boldsymbol{I} & \boldsymbol{I} \end{pmatrix} Var(\boldsymbol{Y}_i) \begin{pmatrix} -\boldsymbol{I} \\ \boldsymbol{I} \end{pmatrix} = 2\sigma^2\{(1-\rho_1)\boldsymbol{I}_p + (\rho_1 - \rho_2)\boldsymbol{1}_p\boldsymbol{1}_p^{'}\}, \tag{15}$$

and use it for the calculation of sample size and power as outlined above. An alternative representation of this variance in terms of random period and error

effects, obtained by simplifying equation 15 is as follows.

$$\Sigma_{diff} = 2\{\sigma_e^2 \boldsymbol{I}_p + \sigma_p^2 \mathbf{1}_p \mathbf{1}_p^{'}\} \tag{16}$$

The benefit of a using a mixed effect model for crossover data is that instead of requiring an estimate for the entire covariance matrix for the data, one can take advantage of the special structure of the covariance matrix resulting from a mixed effect modeling that is based on fewer parameters, namely σ^2, ρ_1 and ρ_2. Estimates of these parameters can be calculated fairly easily using the estimated covariance parameters corresponding to random between subject, period and error effects obtained from the reports for earlier studies, or running a simple mixed effect model as envisioned above on available data. If one or more of these estimates cannot be obtained from the available information, one has to make certain assumptions based on past experience or subjective knowledge of the process.

For the special case where the correlation between observations on a subject can be assumed to be the same regardless of whether the observations are taken in the same or a different period, the period effect can be taken to be fixed rather than random. In that case, $\sigma_p^2 = 0, \rho_1 = \rho_2$. Thus, although the observations taken on a given subject are correlated, the differences are independent under this model and the variance of $\boldsymbol{D}_i$ using equation 16 reduces to

$$\Sigma_{diff} = 2(\sigma_e^2 \boldsymbol{I}_p). \tag{17}$$

3 Simulation Details and Results

The distributions for the time matched mean difference vectors for both the parallel and the crossover designs are multivariate normal. Calculation of power in such cases requires integration in higher dimensions in the alternative space which is analytically intractable and is very complex even numerically as the dimension goes up. So we adopt a Monte Carlo simulation method to estimate power. To calculate power, we repeatedly generate vectors of mean differences ($\boldsymbol{d}$) under H_1 and construct a 95% upper confidence limit for each Δ_k, k=1,... p, as outlined in Section 2, and estimate power by calculating the proportion of times the largest 95% upper confidence limit across the p timepoints is below the regulatory threshold of 10 msec.

Our methods are applicable for arbitrary structures of the covariance matrices but researchers who design the studies seldom have such estimates available

to them. Most of the times, they have to use the estimates available from an analysis conducted on some earlier study that used a mixed effect model with a special covariance structure (mostly Compound Symmetry). We present simulation results based on a compound symmetry covariance structure for the parallel group design. For the crossover design, we present results for a slightly more general model in which the correlation between observations is greater for observations taken within the same period (ρ_1) than for observations taken between periods (ρ_2), and for which the compound symmetry covariance structure is a special case with $\rho_1 = \rho_2$.

3.1 Simulation Methods and Results for the Parallel Design

For a Parallel design , let $n_1 = n_2 = n$ denote the sample size for each study arm, with measurements taken on each subject at $p = 10$ timepoints. Instead of a subject level sampling from each study arm, we generate vectors from the distribution of time matched mean difference $\boldsymbol{d}$ under the assumption of a compound symmetric covariance structure as obtained in Section 2.1.1. We do the power calculations for the following configuration of the operating characteristics to accommodate the plausible values as expected in a parallel study.

- Significance level $\alpha = 0.05$
- Overall variation $\sigma = 15, 18$
- Correlation coefficient for a compound symmetric structure $\rho = 0, 0.5, 0.65, 0.8$
- Maximum time matched mean difference $\theta = 0, 1, 2$
- Time matched mean structure $\boldsymbol{\Delta}$ = Constant, Hill, Steady state

The three mean structures (Constant, Hill and Steady state) that we have used for the simulations are the most plausible ones. In the Constant mean model, the baseline corrected QTc is expected to remain constant across the timepoints. In the Hill mean model, the baseline corrected QTc is expected to initially increase and then decrease after reaching the maximum, whereas, in the Steady state mean model, it is expected to increase initially and then stabilize after reaching the maximum.

To calculate power using equation 6, we use Monte Carlo simulations to repeatedly generate mean difference vectors ($\boldsymbol{d}$) under H_1 with one of the three structures for $\boldsymbol{\Delta}$. We then construct a 95% upper confidence limit for each Δ_k,

			Sample Size (n) per treatment arm			
θ	$\boldsymbol{\Delta}$	σ	$\rho = 0$	$\rho = 0.5$	$\rho = 0.65$	$\rho = 0.8$
0	$(0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$	15	72	63	60	54
		18	103	91	86	78
1	$(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$	15	88	78	74	68
		18	127	113	107	98
	$(0, 0, 0, 1, 1, 1, 1, 0, 0, 0)$	15	81	73	69	64
		18	116	104	99	92
	$(0, 0, 0, 0, 0, 1, 1, 1, 1, 1)$	15	81	72	68	63
		18	115	102	98	91
2	$(2, 2, 2, 2, 2, 2, 2, 2, 2, 2)$	15	112	100	94	86
		18	161	144	134	124
	$(0, 0, 1, 1.5, 2, 2, 1.5, 1, 0, 0)$	15	94	85	81	76
		18	135	123	116	109
	$(0, 0, 0, 1, 1, 1.5, 1.5, 2, 2, 2)$	15	96	87	82	78
		18	138	126	118	112

Table 1: Sample size calculations for the parallel design

and estimate power by calculating the proportion of times the largest 95% upper confidence limit across the 10 timepoints is below the regulatory threshold of 10 msec. To obtain the sample size under any given configuration for a 90% power, we start with a reasonable initial guess and change it in a stepwise manner until we exceed the desired power. Table 1 summarizes the sample size calculations for a 90% power under different combinations of the operating characteristics. The results were based on 1000 simulation runs for each configuration. To use this table to design a parallel Thorough QT study, all one has to do is obtain estimates for σ and ρ from earlier studies of similar design, decide on a structure for the mean difference based on earlier analysis or pharmacokinetic profile of the drug, and look up the sample size entry corresponding to nearest such configuration in the table.

Our simulation results show that the sample size for parallel design (Table 1) decreases with an increase in the intrasubject correlation ρ and increases with an increase in the overall variability σ and maximum time-matched mean difference θ. Among the different mean structures explored, the hill mean model and the steady state mean model result into a substantial decrease in sample size over the constant mean model; the sample sizes for the hill mean model are similar to those for the steady state mean model.

We ran the simulations for different values of θ ranging from 0 to 5 but the results are presented only for $\theta = 0, 1, 2$, because the required sample size to attain a 90% power becomes unreasonably large for higher values of θ. Figure 1 presents a plot of power vs ρ with $n = 100$ per arm, $\sigma = 15$ and $\theta = 5$. It can be seen that despite a relatively large sample size the power remains low even with large values of ρ when $\theta = 5$.

[Figure 1 about here]

3.2 Simulation Methods and Results for the Crossover Design

For a Crossover design, let n denote the total sample size, with measurements taken on each subject at $p = 10$ timepoints in both periods. Instead of a subject level sampling from each study arm (period), we generate vectors from the distribution of time matched mean difference $\boldsymbol{d}$ under the assumption of a compound symmetric covariance structure as obtained in Section 2.2.1. It should be noted that we have given two representations for the variance of time matched difference vector $\boldsymbol{D}_i$ for subject i. Equation 13 and 14 lay out this variance in terms of σ_e^2 and σ_p^2, or in terms of σ_e^2 when σ_p^2 is assumed to be 0. For this variance representation, we do the power calculations for the following configuration of the operating characteristics to accommodate the plausible values as expected in a crossover study.

- Significance level $\alpha = 0.05$
- $\sigma_p = 0, 4$
- $\sigma_e = 7, 10, 12$
- Maximum time matched mean difference $\theta = 0, 1, 3, 5$
- Time matched mean structure $\boldsymbol{\Delta}$ = Constant, Hill, Steady state

To calculate the required sample size for 90% power, we again use Monte Carlo simulations for generation of mean difference vectors ($\boldsymbol{d}$) under H_1, calculate power, and vary the sample size until we obtain the desired power as expained in Section 3.1. Tables 2 and 3 summarize these results. To design a crossover Thorough QT study, one can make use of these tables by obtaining estimates for σ_e and σ_p from earlier studies that used a mixed effect model with random subject and period effect for analysis, and deciding on a plausible structure for the mean difference vector based on earlier analysis or the

		Sample Size (n)		
θ	$\boldsymbol{\Delta}$	$\sigma_e = 7$	$\sigma_e = 10$	$\sigma_e = 12$
0	$(0,0,0,0,0,0,0,0,0,0)$	16	32	45
1	$(1,1,1,1,1,1,1,1,1,1)$	20	40	57
	$(0,0,0,1,1,1,1,0,0,0)$	18	35	51
	$(0,0,0,0,0,1,1,1,1,1)$	18	37	53
3	$(3,3,3,3,3,3,3,3,3,3)$	32	65	95
	$(0,0,1,2,3,3,2,1,0,0)$	25	50	71
	$(0,0,0,1,1.5,2,2.5,3,3,3)$	27	54	77
5	$(5,5,5,5,5,5,5,5,5,5)$	63	127	184
	$(1,2,3,4,5,5,4,3,2,1)$	46	95	135
	$(0,1,2,2.5,3,3.5,4,4.5,5,5)$	48	98	141

Table 2: Sample size calculations for the crossover design with no random period effect ($\sigma_p = 0$)

pharmacokinetic profile of the drug. One can just look up the sample size entry corresponding to nearest such configuration in the table. Often, the mixed effect model used to analyze the QT data does not have a random period effect in the model. In the absence of any estimate for the random period effect, one could use results presented in Table 2 if it is reasonable to assume that $\sigma_p = 0$. Such an assumption might be reasonable if the two active periods are not separated by a long washout period.

The results presented in Tables 2 and 3 indicate that for crossover designs, the sample size increases with an increase in σ_e^2 and σ_p^2. For $\sigma_p^2 = 0$, the sample size increases by nearly three folds as σ_e^2 increases from 7 to 12. In contrast, the increase in sample size is only marginal as σ_p^2 increases from 0 to 4. As expected, the sample size also increases rapidly with an increase in the maximum time-matched mean difference θ. The hill mean model and the steady state mean model result into a substantial reduction in sample size over the constant mean model; the sample sizes for the hill mean model are marginally lower than that for the steady state mean model.

We ran the simulations for different values of θ ranging from 0 to 5 but present the results only for $\theta = 0, 1, 3, 5$. It is easy to see that the sample size increases rapidly with θ. Figure 2 presents a plot of power vs σ_e for a crossover design with $n = 50$, $\theta = 5$, and $\sigma_p = 0, 4$. Power appears to decrease very rapidly with σ_e even with a reasonably large sample size of 50.

[Figure 2 about here]

		Sample Size (n)		
θ	$\boldsymbol{\Delta}$	$\sigma_e = 7$	$\sigma_e = 10$	$\sigma_e = 12$
0	$(0,0,0,0,0,0,0,0,0,0)$	20	36	51
1	$(1,1,1,1,1,1,1,1,1,1)$	25	45	63
	$(0,0,0,1,1,1,1,0,0,0)$	23	41	56
	$(0,0,0,0,0,1,1,1,1,1)$	23	41	57
3	$(3,3,3,3,3,3,3,3,3,3)$	40	74	105
	$(0,0,1,2,3,3,2,1,0,0)$	31	57	80
	$(0,0,0,1,1.5,2,2.5,3,3,3)$	34	60	84
5	$(5,5,5,5,5,5,5,5,5,5)$	78	143	201
	$(1,2,3,4,5,5,4,3,2,1)$	58	107	149
	$(0,1,2,2.5,3,3.5,4,4.5,5,5)$	59	107	150

Table 3: Sample size calculations for the crossover design with no random period effect ($\sigma_p = 4$)

As pointed out earlier, we have given two representations for the variance of time matched difference vector $\boldsymbol{D}_i$ for subject i. Equation 13 lays out this variance in terms of σ_e^2 and σ_p^2 and offers an easy way to calculate the sample size if estimates of σ_e and σ_p are readily available. However, there might be instances when this is not the case. For example, if analysis results from a crossover study were not available or were obtained using a different model, making the estimation of σ_e and/or σ_p impossible, one could not use the results given above. The variance representation given by Equation 12 can be used in these circumstances if one can get a reasonable estimates of the overall variation σ, and the correlations ρ_1 and ρ_2. Such an estimate for σ could come from a parallel QT study or literature. Estimation of both ρ_1 and ρ_2 is not possible in the absence of a random period effect in the model, but if one were willing to make the assumption $\rho_1 = \rho_2 = \rho$, then an estimate for ρ could be obtained very easily from a parallel study. For this variance representation, we do the power calculations for the following configuration of the operating characteristics.

- Significance level $\alpha = 0.05$
- Overall variation $\sigma = 15, 18$
- Correlation coefficient $\rho_1 = \rho_2 = \rho = 0.5, 0.65, 0.8$
- Maximum time matched mean difference $\theta = 0, 1, 3, 5$
- Time matched mean structure $\boldsymbol{\Delta}$ = Constant, Hill, Steady state

			Sample Size (n)		
θ	$\boldsymbol{\Delta}$	σ	$\rho = 0.5$	$\rho = 0.65$	$\rho = 0.8$
0	$(0,0,0,0,0,0,0,0,0,0)$	15	36	25	15
		18	52	36	21
1	$(1,1,1,1,1,1,1,1,1,1)$	15	44	31	18
		18	64	45	26
	$(0,0,0,1,1,1,1,0,0,0)$	15	40	28	16
		18	57	40	23
	$(0,0,0,0,0,1,1,1,1,1)$	15	40	28	16
		18	58	41	24
3	$(3,3,3,3,3,3,3,3,3,3)$	15	73	51	30
		18	105	74	42
	$(0,0,1,2,3,3,2,1,0,0)$	15	55	39	22
		18	80	56	32
	$(0,0,0,1,1.5,2,2.5,3,3,3)$	15	60	42	24
		18	87	61	35
5	$(5,5,5,5,5,5,5,5,5,5)$	15	143	100	57
		18	205	144	83
	$(1,2,3,4,5,5,4,3,2,1)$	15	105	74	42
		18	151	106	61
	$(0,1,2,2.5,3,3.5,4,4.5,5,5)$	15	108	76	43
		18	155	109	62

Table 4: Sample size calculations for the crossover design given overall σ with $\rho_1 = \rho_2$

Using these configurations, we present the sample sizes for situations when at best we have estimates for σ and ρ, and the active periods are not separated by a long washout period making the assumption $\rho_1 = \rho_2 = \rho$ a reasonable one. Tables 4 summarizes these results. Under the above assumptions, sample size decreases with an increase in the intrasubject correlation ρ and increases with an increase in the overall variability σ and maximum time-matched mean difference θ. As in the earlier cases, the hill mean model and the steady state mean model result into a substantial decrease in sample size over the constant mean model while the sample sizes for the hill mean model are generally similar to that for the steady state mean model.

4 Real Data Example

We consider an example from a GlaxoSmithKline trial of a new drug. This thorough QTc study was conducted on healthy volunteers as a crossover trial with 4 treatment arms; placebo, moxifloxacin (positive control), and two doses of the new drug. Subjects were randomized to one of the four sequences defined using a William's square design (Senn, S. J., 2002), with a washout period of at least 14 days between treatments. During each period, subjects received the assigned treatment for 5 days and ECG readings were taken in triplicate at each time point for a full day prior to dosing and on the fifth day. Times on the baseline day were matched to the time points for the fifth day of the dosing period. Change from baseline values were calculated by subtracting the time-matched averaged baseline value from the average of the triplicate readings on the fifth day of dosing. The nine timepoints in hours (hr) used in this study were predose, 0.5 hr, 1 hr, 2 hr, 3 hr, 4 hr, 6 hr, 12 hr, and 23.25 hr. Several different corrections for QT interval were actually done in the study, but the outcome of interest that we will focus on here is Fridericia's corrected QT interval (QTcF). Sample size was planned to ensure that 40 subjects would complete the study with evaluable ECG readings. The achieved sample size was slightly larger.

The model used to analyze the data was as described in Section 2.2.1 (Equation 14) with $\boldsymbol{Y}_i$ as a $4p \times 1$ vector of change from baseline QTcF values from each of the 4 periods. Fixed effects included baseline value, sequence, period, treatment, time, and a treatment by time interaction term. This model inherently assumes that the data structure for the vector $\boldsymbol{Y}_i$ is compound symmetric with a given total variance $\sigma^2 = \sigma_s^2 + \sigma_e^2$ and $\rho = \sigma_s^2/\sigma^2$ as shown in Section 2.2.1. Since period is treated as a fixed effect, $\sigma_p^2 = 0$. In this case, the individual components of the vector of differences are independent with $\boldsymbol{\Sigma}_{diff} = 2\sigma_e^2\boldsymbol{I}$ as shown in Equation 17.

The covariance parameters obtained from this analysis were $\sigma^2 = 209.2$, $\sigma_s^2 = 168.5$, and $\sigma_e^2 = 40.7$, so that $\sigma = 14.5$ and $\sigma_e = 6.4$. This corresponds to the vector of change from baseline QTcF data, $\boldsymbol{Y}_i$, having compound symmetric structure with $\sigma = 14.5$ and $\rho = 0.806$. The mean difference vector between the baseline corrected QTcF values for the supratherapeutic dose and placebo was $(-1.64, -1.02, 2.29, 1.86, 0.79, 0.1, 1.95, 0.17, 0.03)$ and hence a hill mean structure seems to be a suitable fit for the mean difference vector.

Suppose that we are planning to run a similar crossover study with ten timepoints and a full day of baseline measurements prior to dosing, and we wish to power against an alternative hypothesis with a hill structure and $\theta = 3$

msec. Using the conservative value of 7 for σ_e, and $\sigma_p = 0$, we see from Table 2 that we would require 25 subjects for our study in order to have 90% power. Alternatively, we could use Table 4 with a value of $\sigma = 15$ and $\rho = 0.8$. Using this parameterization gives a sample size of 22 subjects. This value is of course similar to that obtained using Table 2 although not exactly the same because we had to extrapolate to use the tables. If we had run simulations using the exact values of $\sigma^2 = 209.2$, $\sigma_s^2 = 168.5$, $\sigma_e^2 = 40.7$ and $\rho = 0.806$, the results under both parameterizations would be the same. In fact, we did this and the resulting sample size was 21. It is not surprising that this number is slightly smaller than the table values because it is based on the exact estimates and not conservative extrapolation.

Rerunning the analysis with period as a random, rather than fixed effect, gives $\sigma^2 = 204.6$, $\sigma_s^2 = 160.9$, $\sigma_p^2 = 11.1$, $\sigma_e^2 = 32.5$, so that $\sigma = 14.3$, $\sigma_p = 3.3$ and $\sigma_e = 5.7$; ρ_1 and ρ_2 are 0.841 and 0.786 respectively. Determining sample size for 90% power against the alternative hill structure with $\theta = 3$, with conservative values of 7 for σ_e, and 4 for σ_p using Table 4, gives a conservative sample size of 31. Again, If we used the exact estimates and run the simulations using either representation, this number would be 21. This number is equal to the sample size obtained from the earlier model without a random period effect. It appears that the effect of the increase in σ_p from 0 to 3.3 was offset by the decrease in σ_e from 6.4 to 5.7. In this case, ρ_1 is close to ρ_2 and hence the assumption of compound symmetry appears to be sensible in this case. Adding the random period effect to the model does not change the required sample size by much. If the required sample size is very small, we would generally recommend an effective sample size of at least 30 so that large sample theory results used in the derivation of results are valid.

5 Conclusions and Discussion

Thorough QT studies by design generate multivariate data with moderate to high correlation between the QTc measurements taken at different timepoints. Hence the currently used power and sample size calculations that are based on univariate methods with some adjustments are overly conservative, and our method that uses a multivariate model allowing for correlation between the measurements across timepoints appears to achieve a substantial decrease in sample size in most of the cases. In our real life data example and several other data sets at GSK we saw high correlations ($\rho > 0.7$) and hence an independence assumption may not be reasonable for the sample size calculations.

We have given a Monte Carlo simulation-based general methodology that uses multivariate models and can be used for power and sample size calculations for both parallel and the crossover designs. We have given sample size under various mean and variance covariance structures for both designs. For the crossover design, we have provided two different parameterizations so that depending on what information is available, one may use one or the other. We have also given a general framework for simulations so that if one would like to alter the assumptions somewhat, one can use this framework to determine sample size for given power.

Our simulation results show that for both the parallel and the crossover designs, the sample size decreases with an increase in the intrasubject correlation ρ. For the crossover design, the sample size increases with an increase in σ_e^2, and σ_p^2, which is expected because a large period effect indicates lower correlations between meaurements taken in different periods. In reality, σ_p^2 may not be very large compared to the overall within subject variability $\sigma_e^2 + \sigma_p^2$. In that case, the assumption of coumpound symmetry with $\sigma_p^2 = 0$ may be a reasonable assumption and our simulations show that the increase in sample size in the presence of a small period effect is not substantial. Our results also show that the sample sizes for crossover design are much smaller than that for the parallel design. Crossover studies will always be preferred in terms of sample size, but may not always be logistically feasible. For example, if the half-life of a drug is very long, a long washout period would be required for the crossover design and it might not be feasible to expect subjects to return after such a long time away. Also, if the drug requires many days of dosing in order to reach a steady state concentration where QTc would be measured, a crossover study could be too long for consideration. If a parallel study is necessary, one would want to design the study so as to have the least possible total variability in change from baseline QTc measurements.

We have considered three different structures for the alternative hypothesis; constant, hill, and steady state. Of these, the constant structure is the least likely to be seen in reality unless there is no effect of the drug on QTc at all. If a pre-dose measure at a single time is used as the baseline measure, then the hill structure is likely even if there is no drug effect. This is due to the diurnal patterns that are observed in QTc measures. QTc is lowest in the morning, increases during the day, and is lower in the evening. Baseline measures taken in the morning will mean that a hill structure exists in the data. If the drug has a short-lived effect, then the hill structure is also an appropriate alternative. The QTc effect would increase as the concentration of the drug in the body increases

and then may decrease as the drug is eliminated. For a drug with a long half-life and QTc measures taken over a relatively short time-period compared with the half-life of the drug, a steady state alternative may be most appropriate. In our simulation study we found that the sample sizes for the hill mean model and the steady state mean model are much lower than the corresponding constant mean model and are generally similar and hence a multivariate model that allows for a specification of the mean structure based on past data and/or clinical judgement seems to be a more appropriate and attractive approach.

The ICH E14 guidance document defines a negative 'Thorough QT/QTc Study' as 'one in which the upper bound of the 95% one-sided confidence interval for the largest time-matched mean effect of the drug on the QTc interval excludes 10 msec. This definition is chosen to provide reasonable assurance that the mean effect of the study drug on the QTc interval is not greater than 5 msec' . This definition has been a subject of much debate because one might acheive a 95% one-sided upper confidence interval below 10 msec even with a mean effect larger than 5 msec by rigorously controlling the variability of QTc measurements. In our simulations, we found that with a parallel study, the sample size required to show noninferiority to 10 msec when the true maximum is 5 msec would be too large to be logistically feasible. However, for a crossover study this may not be the case. For example, even with a moderate sample size of 50 in a crossover trial, a power of over 90% can be obtained to establish non-inferiority with within subject variability as small as 7 msec for the hill and the steady state mean models (Figure 2), which can be obtained by employing rigorous control over study conditions to ensure precise measurement of QTc effects. With a larger, but potentially feasible variability, say a within subject variability of 10 msec with no period effect, it requires around 100 subjects in a crossover study to show non-inferiority with 90% power (Table 2) when $\theta = 5$ msec.

6 References

Anand, S. P. and Ghosh, S. K. (2008). A Bayesian methodology for investigating the risk of QT prolongation, *Journal of Statistical Theory and Practice*, (Submitted).

Boos, D., Hoffman, D., Kringle, R., Zhang, J. (2007). New confidence bounds for QT studies, *Statist. Med.*, **26**, 3801-3817.

Eaton, M. L., Muirhead, R. J., Mancuso, J. Y., Lolluri, S. (2006). A confidence interval for the maximal mean QT interval change due drug effect, *Drug*

Inf J., **40**, 267-271.

Gupta, A. K. and Nagar, D. K. (2000). *Matrix Variate Distributions*, Chapman and Hall, New York.

International Conference on Harmonisation (2005). Guidance for Industry: E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrythmic Potential for Non-Antiarrythmic Drugs.
Available at: `www.fda.gov/cber/gdlns/iche14qtc.htm`

Kong, L., Kohberger, R. C., Koch, G. G. (2004). Type I Error and Power in Noninferiority/Equivalence Trials with Correlated Multiple Endpoints: An Example from Vaccine Development Trials, *Journal of Biopharmaceutical Statistics*, **14:4**, 893-907.

Patterson, S. (2005). 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 Inf J.*, **9**, 243-266.

Patterson, S., Jones, B. Zariffa, N. (2005). Modeling and interpreting QTc prolongation in clinical pharmacology studies, *Drug Inf J.*, **39**, 437-445.

Patterson S. and Jones B. (2005) *Bioequivalence and Statistics in Clinical Pharmacology*, Chapman and Hall, CRC Press; London.

Senn, S. J. (2002) *Cross-over Trials in Clinical Research*, John Wiley, Chichester.

Temple, R. (2005). 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`

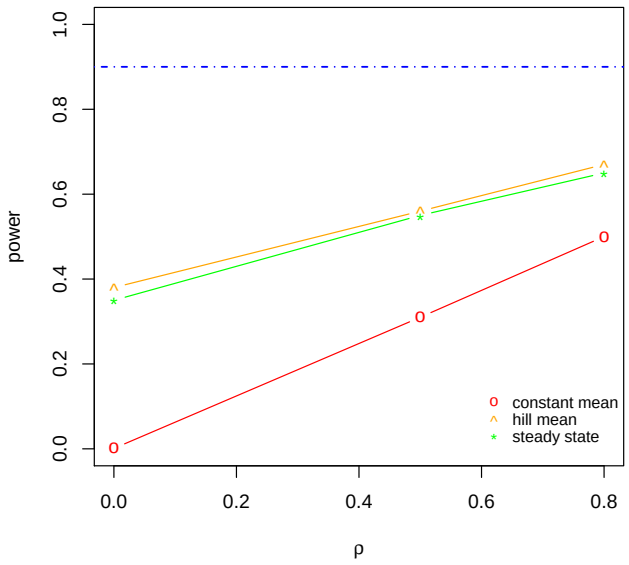

Figure 1: Plot of power vs ρ for the parallel design with $n = 100$ per arm, $\sigma = 15$ and $\theta = 5$

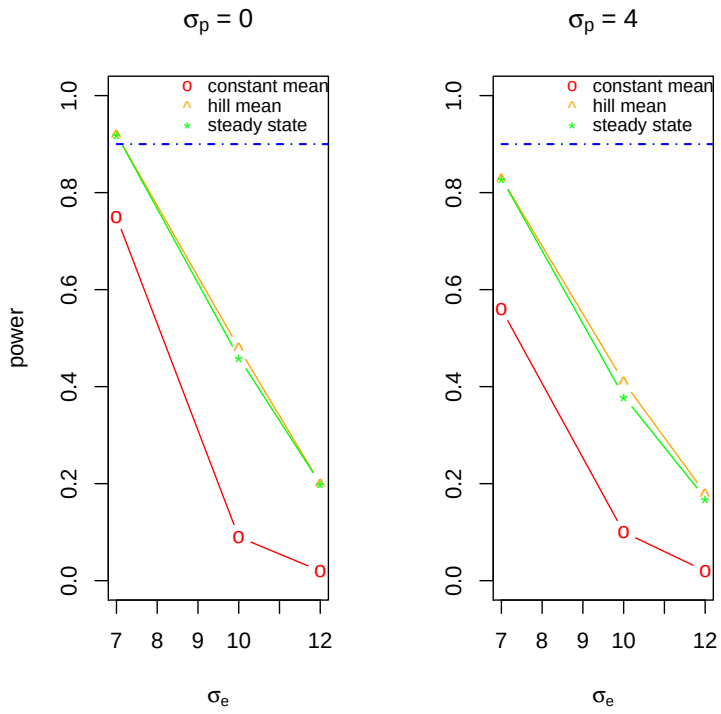

Figure 2: Plot of power vs σ_e for the Crossover design with $n = 50$, $\theta = 5$, and $\sigma_p = 0, 4$