Sample Size Calculations for Crossover Thorough QT Studies: Non-inferiority and Assay Sensitivity

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 #2621

Abstract

The cost involved in running a ‘thorough QT/QTc 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 an adequately powered thorough QT study. Current literature offers some help in designing such studies but these methods are limited and mostly apply only in the context of linear mixed models with compound symmetry covariance structure. It is not evident that such models can satisfactorily be employed to represent all kinds of QTc data, and the existing literature inadequately addresses if there is a change in sample size and power, in case one looks to more general covariance structures for the linear mixed models. We assess the use of some of the existing methods to design a thorough QT study, based on data arising from a GlaxoSmithKline (GSK) conducted thorough QT study, and explore newer models for sample size calculation. We also provide a new method to calculate the sample size required to detect assay sensitivity with a high power.

Keywords: Correlation structure; ICH E14; Monte Carlo simulation; Power and Sample size; QTc prolongation; thorough QT/QTc study

[*]Correspondence: Suraj Anand, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA; E-mail: suraj.anand@gmail.com

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 conducting 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.

The ICH E14 guidance document describes in detail the scope, conduct and implications of a thorough QT study. Recent developments in the area of analysis of a thorough QT study include articles by Eaton et al. (2006), Boos et al. (2007) and Anand and Ghosh (2009), though the standard way of analyzing QT data still remains to be the intersection-union-based test that compares the two-sided 90% upper confidence interval for the time-mached mean difference between drug and placebo at each time point with the regulatory threshold of 10 msec. A review of statistical design and analysis in thorough QT studies can be found in Patterson et al., 2005.

Measurements taken in a QT study are naturally in time order and hence multivariate methods are the most appropriate tools for analyzing such a data. Current literature offers some help in designing such studies but these methods are limited and apply only in the context of linear mixed models with rather restrictive assumptions on the covariance structure. Hosmane and Locke (2005) performed a simulation study to assess the impact of sample size on power for fixed covariance parameters, estimated from their own data, for a four-period crossover study. They used a multivariate model for the data and their results indicate that power may depend on the covariance structure and the components of the mean difference vector between drug and placebo. Recently, the Journal of Biopharmaceutical Statistics issued a special edition with a wide range of articles on statistical issues in design and analysis of thorough QT studies (Volume 18, Issue 3, 2008), some of which address the sample size calculations in commonly encountered settings.

Zhang and Machado (2008) provide a closed form sample size formula assuming independence between time points for a constant time-matched mean structure. They also present some simulation-based sample size estimates for a multivariate model under the compound symmetry covariance model. Dmitrienko et al. (2008) provide sample size estimates for some popular mixed model-based designs using data from thorough QT studies conducted in Eli Lilly. Cheng et al. (2008) assess the impact of the number of replicate measurements on sample size for linear mixed models. It is not known whether such models will fit all data from QT studies well and if there is a change in sample size and power, in case one looks beyond the commonly employed linear mixed models. We address the robustness of sample size and power calculations to departures from the linear mixed models by considering a compound symmetry correlation structure with a random period effect, a mixed model with variable correlation across time points, and a completely unstructured covariance matrix for the time-matched mean difference vector.

Another important aspect of a thorough QT study is the assay sensitivity of the trial. Zhang and Machado (2008) point out that the trial must be able to detect obvious differences, as indicated by a 'positive' result for the positive control. Zhang (2008) provides a local test using the Bonferoni adjustment for multiplicity, as well as a global test using the averaged time-matched mean difference that does not require a multiplicity correction. Both these tests provide reasonable methods for assessing assay sensitivity but they have their advantages and disadvantages. Tsong et al. (2008) provides a discussion on these two approaches for assessing assay sensitivity. We propose a slightly different approach that is less stringent than the Bonferoni correction and does not have the risk involved with the global test.

In Section 2, we present a motivating example and review the existing literature for sample size calculations. In Section 3, we provide a few plausible models for the QT data for crossover studies that extend the usual linear mixed models to a more general mixed model representation, and can be used to calculate power and sample size in more general settings. More specifically, we provide a methodology to calculate sample sizes for a design that accounts for variable correlation between the time points within any period. We provide a new method to calculate the minimum sample size required to detect assay sensitivity with a high power in Section 4. In Section 5, we use the example dataset to illustrate our methods and in Section 6 we present the conclusions and a discussion. We also provide sample size tables in the Appendix for a variety of crossover settings for the linear mixed effect model with and without a random period effect and they can be used by practitioners to design a crossover thorough QT study.

2 A Motivating Example

Zhang and Machado (2008) discuss a variety of designs suitable for a thorough QT study. We consider designing a thorough QT study for a new GSK drug based on data arising from an earlier GSK thorough QT study. This thorough QT 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 with 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 of dosing. 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 time points 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.

Suppose we use a mixed effect model for the change from baseline QTcF values, with baseline value, sequence, period, treatment, time, and a treatment by time interaction term as fixed effects, and subject as the random effect. This model inherently assumes that the covariance structure for the vector of baseline corrected QTcF measurements is compound symmetric; a mixed model analysis resulted in the between subject variability estimate $\sigma_s = 13$ and the intra-subject correlation $\rho = 0.8$. The mean difference vector between the baseline corrected QTcF values for the supratherapeutic dose and placebo for this thorough QT study was $(-1.64, -1.02, 2.29, 1.86, 0.79, 0.1, 1.95, 0.17, 0.03)$.

Suppose that we are planning to run a similar crossover study with ten time points and a full day of baseline measurements prior to dosing, and we wish to power against an alternative hypothesis with a hill structure, such that QTcF initially increases over time and then decreases after reaching the maximum, and $\theta = 3$ msec. Using the sample size estimates under independence given in Table 10 by Zhang and Machado (2008) one would require over 87 subjects to declare a thorough QT study successfully negative with 85% power. But this is a very conservative sample size estimate as it ignores the correlation between the time points and assumes a constant time-matched mean structure which may not be the case. The sample size required if we do not assume independence cannot be obtained in a straightforward manner and requires simulations. Zhang and Machado (2008) present some sample size estimates using simulations for a hill mean structure and a compound symmetry covariance structure with $\rho = 0.5$ for various values of σ_s, but these results are presented for $\theta = 5$ msec and hence cannot be used to design our study.

Assuming compound symmetry for the covariance matrix may be too simplistic and may not be appropriate, as it may be that observations between periods are less correlated than observations taken within the same period. One could also consider a variable correlation between the time points within any period given that the time points pertaining to the measurements taken on the same subject are not equispaced. For example, it is likely that the measurements taken within the first six hours are more correlated than those taken after that. It is of interest to assess the impact of the covariance structure on the planned sample size for the new study.

The time-matched mean difference vector between the baseline corrected QTcF values for the positive control moxifloxacin and placebo was $(0.69, 5.05, 9.85, 10.63, 11.16, 11.64, 8.07, 6.71, 5.25)$. To ensure trial sensitivity of the new study, the sample size also needs to be large enough to have a high power to detect a positive QT effect for the positive control.

3 A Multivariate Framework for Crossover Design

In this section, we provide a framework for crossover designs that can be used to conduct simulations to compute the power and sample size for a thorough QT study with varying assumptions about the structure of the variance-covariance matrix. 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 treatment 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 $\boldsymbol{Y}_{2i},\ i = 1, \dots n$ 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 for each treatment are assumed to be distributed identically and independently (iid), arising from corresponding p-variate normal distributions. However, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are not independent because these are observations taken on the same individuals. In matrix notation, this can be represented as follows:

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

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 covariances between data vectors corresponding to the 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 is the vector of the 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

$$\boldsymbol{D}_i \overset{iid}{\sim} \mathcal{N}_p(\boldsymbol{\Delta}, \boldsymbol{\Sigma}_{diff}), \quad i = 1, \ldots, n \tag{2}$$

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 sample time-matched mean difference by $\boldsymbol{d}$, $\boldsymbol{d} = \sum_i (\boldsymbol{y}_{2i} - \boldsymbol{y}_{1i})/n$, then

$$\boldsymbol{d} \sim \mathcal{N}_p(\boldsymbol{\Delta}, \boldsymbol{\Sigma}_d), \tag{3}$$

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

Each individual component of $\boldsymbol{d}$ follows a univariate normal distribution.

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

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 $d_k + z_{0.05}\sqrt{\sigma_{d(kk)}}$, 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

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

where power is calculated under any configuration of $\boldsymbol{\Delta}$ in the alternative space of non-inferiority H_1. To calculate power using equation (5) we use Monte Carlo simulations to repeatedly generate the 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 time points is below the regulatory threshold of 10 msec. To obtain the sample size for a 90% power, we start with a reasonable initial guess and change it in a stepwise manner until we exceed the desired power.

3.1 A Simple Linear Mixed Effect Model

We start by adopting the widely used linear mixed effect model for the crossover data, which can be written succinctly in a matrix notation for the ith subject as

$$\begin{pmatrix} \boldsymbol{Y}_{1i} \\ \boldsymbol{Y}_{2i} \end{pmatrix} = \boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i, i = 1, \ldots, n, \tag{6}$$

where $\boldsymbol{X}_i\boldsymbol{\beta}$ denotes the fixed effect part of the model with treatment, time, treatment by time interaction, period 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 in subject 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 $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{2p}(\boldsymbol{0}, \sigma_e^2\boldsymbol{I}_{2p})$, $\boldsymbol{\gamma}_i \sim \mathcal{N}_3\left(\boldsymbol{0}\ , \begin{pmatrix} \sigma_s^2 & 0 & 0 \\ 0 & \sigma_p^2 & 0 \\ 0 & 0 & \sigma_p^2 \end{pmatrix}\right)$ with all random effect components independent of each other. With $\boldsymbol{Z}_i = \begin{pmatrix} \boldsymbol{1}_p & \boldsymbol{1}_p & \boldsymbol{0}_p \\ \boldsymbol{1}_p & \boldsymbol{0}_p & \boldsymbol{1}_p \end{pmatrix}$,

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

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

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

where $\boldsymbol{J}_p = \boldsymbol{1}_p\boldsymbol{1}_p'$. Using the above results, we can calculate the variance of $\boldsymbol{D}_i$ as

$$Var(\boldsymbol{D}_i) = \Sigma_{diff} = \begin{pmatrix} -\boldsymbol{I} & \boldsymbol{I} \end{pmatrix} Var(\boldsymbol{Y}_i) \begin{pmatrix} -\boldsymbol{I} \\ \boldsymbol{I} \end{pmatrix} = 2\sigma^2\{(1-\rho_1)\boldsymbol{I}_p + (\rho_1 - \rho_2)\boldsymbol{1}_p\boldsymbol{1}_p'\}, \tag{7}$$

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

A special case of this model is one with a fixed period effect ($\sigma_p = 0$) as utilized by Zhang and Machado (2008), where all the observations taken on any subject follow a simple compound symmetry covariance model regardless of the period to which they belong, with the overall variability $\sigma = \sqrt{(\sigma_s^2 + \sigma_e^2)}$ and $\rho_1 = \rho_2 = \rho = \sigma_s^2/\sigma^2$.

Boos et al. (2007) showed that the variance for the difference in this linear mixed model with a random period effect can be represented in a compound symmetric form. The above model assumes a compound symmetric covariance structure for the observations taken in any one period. Such an assumption may not be appropriate given that the time points pertaining to the measurements taken on the same subject are not equispaced. For example, it is likely that the measurements taken within the first six hours are more correlated than those taken after that. We next propose a model to account for such a phenomenon.

3.2 A Mixed Model to Account for Variable Correlation Across Time Points

On lines similar to the above set up, we add one additional random component for timeband×period effect with its contribution to the overall variability denoted by σ_{tb}. Timeband is an indicator variable with two classes, with observations for the time points close together clubbed in one class and the remaining observations in the other class. This is equivalent to assuming that the observations taken on the first l time points $l < p$ in any period are more correlated than those taken after that. In addition, observations taken in the same period are more correlated than those taken in different periods. If we let $\sigma^2 = \sigma_s^2 + \sigma_{tb}^2 + \sigma_p^2 + \sigma_e^2$, $\rho_{11} = (\sigma_s^2 + \sigma_{tb}^2 + \sigma_p^2)/\sigma^2$, $\rho_{12} = (\sigma_s^2 + \sigma_p^2)/\sigma^2$, and $\rho_2 = \sigma_s^2/\sigma^2$, then the variance of $\boldsymbol{Y}_i$ is given by $Var(\boldsymbol{Y}_i) =$

$$\sigma^2 \begin{pmatrix} \begin{pmatrix} (1-\rho_{11})\boldsymbol{I}_l + \rho_{11}\boldsymbol{J}_l & \rho_{12}\boldsymbol{J}_{l\times(p-l)} \\ \rho_{12}\boldsymbol{J}_{(p-l)\times l} & (1-\rho_{12})\boldsymbol{I}_{p-l} + \rho_{12}\boldsymbol{J}_{p-l} \end{pmatrix} & \rho_2\boldsymbol{J}_p \\ \rho_2\boldsymbol{J}_p & \begin{pmatrix} (1-\rho_{11})\boldsymbol{I}_l + \rho_{11}\boldsymbol{J}_l & \rho_{12}\boldsymbol{J}_{l\times(p-l)} \\ \rho_{12}\boldsymbol{J}_{(p-l)\times l} & (1-\rho_{12})\boldsymbol{I}_{p-l} + \rho_{12}\boldsymbol{J}_{p-l} \end{pmatrix} \end{pmatrix},$$

where $\boldsymbol{J}_{r\times s} = \boldsymbol{1}_r\boldsymbol{1}_s^{'}$.

Doing similar algebra as in the previous subsection, we obtain

$$Var(\boldsymbol{D}_i) = 2\sigma^2 \begin{pmatrix} (1-\rho_{11})\boldsymbol{I}_l + (\rho_{11}-\rho_2)\boldsymbol{J}_l & (\rho_{12}-\rho_2)\boldsymbol{J}_{l\times(p-l)} \\ (\rho_{12}-\rho_2)\boldsymbol{J}_{(p-l)\times l} & (1-\rho_{12})\boldsymbol{I}_{p-l} + (\rho_{12}-\rho_2)\boldsymbol{J}_{p-l} \end{pmatrix}, \tag{8}$$

which can be used for the calculation of sample size and power as outlined above.

3.3 A Mixed Model With Completely Unstructured Covariance Matrix for the Difference

There is no evidence that any particular covariance structure fits QT data well. Even the existing literature does not favor any one model to represent the QT data. A compound symmetric model is usually invoked for the sake of simplicity. An unstructured covariance matrix for $Var(\boldsymbol{Y}_i)$ involves a large number of parameters. For example, if ten time points are used in a 2-way crossover study, then the completely unstructured design yields $2p(2p+1)/2 = 210$ covariance parameters. An alternative is to allow a completely unstructured covariance matrix for the difference vector, i.e., $Var(\boldsymbol{D}_i) = \Sigma_{diff} = ((\sigma_{kk'}))$, where $\sigma_{kk'}$ is the covariance between any two time points k and $k^{'}$. For a 2-way crossover study with ten time points, $Var(\boldsymbol{D}_i)$ will have $p(p+1)/2 = 55$ covariance parameters. Given an estimate for $Var(\boldsymbol{D}_i)$, sample size and power can be calculated as outlined above.

4 Sample Size Assessment for Assay Sensitivity

According to the ICH E14 guidelines, any thorough QT study should include a positive control to assess the assay sensitivity. The positive control should demonstrate a mean QT effect in excess of the regulatory threshold of 5msec. This gives reasonable assurance that the trial is sensitive enough to detect such an effect of the study drug. The hypotheses of interest can be set based on a testing procedure as outlined in Zhang and Machado (2008) as follows:

$$\begin{aligned} H_0 &: \bigcap_{k=1}^{q} \{\mu_{2(k)} - \mu_{1(k)}\} \leq 5 \\ H_1 &: \bigcup_{k=1}^{q} \{\mu_{2(k)} - \mu_{1(k)}\} > 5 \end{aligned} \tag{9}$$

where $\mu_{1(k)}$ and $\mu_{2(k)}$ are the baseline corrected QTc values for the placebo and positive control arms respectively at time point k, and q is the number of time points pre-specified to assess the assay sensitivity. Usually, the QT profile of the positive control is well known, and so one can examine the assay sensitivity at a subset of the total number of time points, $q < p$, where the positive control is most likely to show an effect larger than 5 msec. Assay sensitivity is established if at least one of the q lower bounds for two-sided 90% confidence intervals for the time-matched mean differences between the positive control and placebo is above 5 ms. The advantage of this approach is to reduce the extent for adjusting for multiplicity for a large number of time points. Zhang and Machado (2008) indicate a few references that provide methods to account for multiplicity in a general setting.

Zhang (2008) provides a local method using the Bonferoni adjustment for multiplicity, as well as a global method using the averaged time-matched mean difference that does not require a multiplicity correction. The paper also provides sample size calculations based on these methods. While both of these methods are reasonable methods for assessing assay sensitivity, the local test may be too stringent if the number of pre-specified time points is too large. On the other hand, the main risk with the global test is that the average time-matched mean difference may have dilution due to some time points with small differences even though there is a large difference at some time point. Tsong et al. (2008) provides a discussion on these two approaches for assessing assay sensitivity.

We take a somewhat different approach that is less stringent than the Bonferoni correction and does not suffer from the risk involved with the global test as proposed in Zhang (2008). More specifically, we use the Hailperin-Ruger method to assess for significance at a subset of the total number of time points. This approach was described by Abt (1991) for partially global hypotheses in the context of multiplicity problems in clinical trials.

For the assessment of assay sensitivity with the Hailperin-Ruger method, the global null hypothesis in (9) for the q pre-specified time points has contradiction with one-sided type I error when a pre-specified minimum number $q^{'} \leq q$ of the individual hypotheses within (9) for the respective time points have contradiction at the adjusted $\alpha^{*} = \frac{q^{'}}{q}\alpha$ one-sided significance level. These $q^{'}$ hypotheses, with each to have contradiction at the α^{*} significance level, may appear anywhere within the 'Hailperin-Ruger area' defined by the null hypothesis in (9), and the basis for their contradiction can be through $q^{'}$ of the corresponding two-sided 100(1-2α^{*})% confidence intervals having their lower bounds exceed 5 msec. A proof of how the Hailperin-Ruger method maintains type I error at one-sided α for the assessment pertaining to (9) is provided in the Appendix.

To obtain sample size for a given power, one can use a superiority test for a crossover design, based on a two-sample t-test with level α^{*} and a superiority margin of 5 using any standard software, e.g., PASS or n-Query. The mean squared error (MSE) estimate from an earlier study that is similar in design to the planned study can provide an estimate of the within subject variability (σ_w).

5 Application to the Real Dataset

We revisit the example given in Section 2. This GSK conducted thorough QTc study was a crossover trial with 4 treatment arms, with QTcF measurements taken at nine time points in each period. The model used to analyze the data was as described in Section 3.1 (Equation (6)) 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 3.1. Since period is treated as a fixed effect, $\sigma_p^2 = 0$. In this case, the individual components of the vector of differences can be easily shown to be independent with $\boldsymbol{\Sigma}_{diff} = 2\sigma_e^2 \boldsymbol{I}$.

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 description for the mean difference vector.

Suppose that we are planning to conduct a similar crossover study with nine time points and a full day of baseline measurements prior to dosing, and we wish to power against an alternative hypothesis with $\theta = 3$ msec. We compute sample sizes for three mean structures corresponding to this θ value. 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 time points. 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. We performed simulations based on the methods described in Section 3, using the formula for $Var(\boldsymbol{D}_i)$ provided in Section 3.1 (eq. (7)), with $\sigma^2 = 209.2$ and $\rho_1 = \rho_2 = 0.806$ to obtain the required sample sizes for 90% power. The first block of Table 1 summarizes these results.

It is clear from the results that the hill and the steady state mean models have a substantially lower sample size than the constant mean model; the sample sizes for the hill mean model are similar to those for the steady state mean model.

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

Table 1: Sample size calculations for the crossover design for 90% power for $\theta = 3$ for various covariance structures

Covariance Structure	Covariance Parameters	$\boldsymbol{\Delta}^*$	N
CS; Fixed period	$\rho_1 = \rho_2 = 0.806$	$(0, 1, 2, 2.5, 3, 2.5, 2, 1, 0)$	20
		$(0, 0.5, 1, 1.5, 2, 2.5, 3, 3, 3)$	22
		$(3, 3, 3, 3, 3, 3, 3, 3, 3)$	26
Random period	$\rho_1 = 0.841, \rho_2 = 0.786$	$(0, 1, 2, 2.5, 3, 2.5, 2, 1, 0)$	21
		$(0, 0.5, 1, 1.5, 2, 2.5, 3, 3, 3)$	23
		$(3, 3, 3, 3, 3, 3, 3, 3, 3)$	27
Timeband	$\rho_{11} = 0.845, \rho_{12} = 0.822, \rho_2 = 0.782$	$(0, 1, 2, 2.5, 3, 2.5, 2, 1, 0)$	22
		$(0, 0.5, 1, 1.5, 2, 2.5, 3, 3, 3)$	23
		$(3, 3, 3, 3, 3, 3, 3, 3, 3)$	27
Unstructured		$(0, 1, 2, 2.5, 3, 2.5, 2, 1, 0)$	41
		$(0, 0.5, 1, 1.5, 2, 2.5, 3, 3, 3)$	39
		$(3, 3, 3, 3, 3, 3, 3, 3, 3)$	50

* Results for hill, steady state, and constant mean structures.

0.841 and 0.786 respectively. The sample size results from simulations based on the methods described in Section 3, using the formula for $Var(\boldsymbol{D}_i)$ provided in Section 3.1 (eq. (8)), with $\sigma = 14.3$, $\rho_1 = 0.841$ and $\rho_2 = 0.786$ for the three mean structures are presented in the second block of Table 1. As with the earlier model, the sample sizes for the hill mean model are similar to those for the steady state mean model; the hill mean model and the steady state mean model have a substantially lower sample size than the constant mean model. It can also be seen that adding the random period effect to the model increases the sample size by only one in each case.

Next, we analyzed the data with an additional random component for timeband, so as to make observations taken during the first six hours in any period more correlated than those taken after that. The resulting covariance parameters obtained from this analysis were $\sigma^2 = 202.39$, $\sigma_s^2 = 159.05$, $\sigma_p^2 = 7.86$, $\sigma_{tb}^2 = 4.62$ and $\sigma_e^2 = 30.86$, so that $\sigma = 14.23$, $\rho_{11} = 0.848$ $\rho_{12} = 0.825$ and $\rho_2 = 0.786$. To obtain the required sample sizes for 90% power, we performed simulations based on the methods described in Section 3, using the formula for $Var(\boldsymbol{D}_i)$ provided in Section 3.2, with the above parameter estimates. The third block of Table 1 summarizes these results. We see that adding a random timeband effect in addition to the random period effect does not change the required sample substantially. This is not surprising given that the covariance parameter estimates ρ_{11} and ρ_{12} are very similar for this particular example. As with the earlier models, the sample sizes for the hill and the steady state mean models are similar and both these models have a substantially lower sample size than the constant mean model.

Finally, we allowed for a completely unstructured covariance structure for the data for the time-matched mean difference vector between drug and placebo. This is the most general model and can be invoked if one is not comfortable making any assumptions about the covariance structure for the original data. This model can also be tried to assess the sensitivity of the sample sizes to the covariance assumption. One should note, however, that a totally unstructured covariance structure might lead to a substantial loss of power because of the large number of parameters that need to be estimated. So, such a structure would be of some value only if the estimated unstructured covariance matrix was derived from a large study resulting in precise covariance parameter estimates. The sample size results for this model for the GSK data, from simulations based on the methods described in Section 3, using the formula for $Var(\boldsymbol{D}_i)$ provided in Section 3.3, are presented in the last block of Table 1. As expected, the sample size required to achieve 90% power is substantially high with the unstructured covariance structure.

All the analyses and simulations were performed using *SAS*, version 8. The sample size results were based on 1000 simulation runs since the corresponding span of a two-sided 95% confidence interval about 90% power is about $2 * 100\sqrt{(.9)(.1)/1000} \approx 2\%$. Any sample size thus calculated for a 90% nominal power is expected to have the actual power in the range of 88%-92%. It is evident from the above analysis that there isn't much change in power and sample size with a departure from the linear mixed model with a compound symmetric covariance structure unless one is confident about invoking a completely unstructured covariance matrix based on estimates from a large study. So, the commonly invoked compound symmetric covariance structure with fixed period effect or a linear mixed effect model with random period effect seems adequate. We provide some sample size results for these two popular models, for a variety of plausible parameter settings in the Appendix. We have given sample sizes under various mean and covariance structure parameter settings for a crossover design with ten time points, with fixed/random period effect. For the simplest model with a compound symmetry covariance structure for the data, we have provided results for two different parameterizations, one using variability estimates of the covariance parameters, and the other using the overall variability estimate and intra-subject correlation, so that depending on what information is available, one may use one or the other. These results are based on simulations similar to the ones described above and can be of great use for practicing statisticians in designing thorough QT studies for crossover designs. The details for these simulation results can be found in Anand, Murray and Koch (2008).

To assess the sample size required for trial sensitivity, we used a superiority test for a crossover design, based on a two-sample t-test with level α^*, with a superiority margin of 5 msec, as described in Section 4. We used the mean squared error (MSE) estimate from the model with a random period effect as an estimate of the within subject variability ($\sigma_w = \sqrt{\sigma_p^2 + \sigma_e^2} = 6.6$). The time-matched mean difference vector between the baseline corrected QTcF values for the positive control moxifloxacin and placebo was $(0.69, 5.05, 9.85, 10.63, 11.16, 11.64, 8.07, 6.71, 5.25)$. So, we used a value of 11.5 msec in the alternative space for the calculations. It is easy to see from the time-matched mean difference vector that the time points 2 hr, 3 hr, 4 hr, and 6 hr appear to have a higher chance of yielding differences greater that 5 msec, so we selected these time points for our partial null with $q = 4$. We chose $q^{'} = 2$, with one-sided $\alpha = 0.05$ that yields $\alpha^* = 0.025$ for our calculations. We used the sample size software PASS for our calculations. A sample size of 24 achieves 90% power to detect assay sensitivity, based on a superiority test using a one-sided t-test with a superiority margin of 5 msec. One would then choose the maximum of the two sample sizes computed above, one for the study drug safety and the other for sensitivity analysis, as the final sample size for the study. Assuming a hill or a steady state mean structure for the vector of the time-matched mean differences, a sample size of 24 subjects will result into a study that is adequately powered to show non-inferiority as well as establish the sensitivity of the trial based on the example data. We see that assay sensitivity drives the sample size for this study.

6 Conclusions and Discussion

Multivariate methods are the most appropriate tools for analyzing data from a thorough QT study. Zhang and Machado (2008) provide a sample size formula assuming independence between time points for a constant time-matched mean difference vector and they also present some simulation-based sample size estimates under the compound symmetry covariance structure for a hill-type time-matched mean vector. Dmitrienko et al. (2008) provide sample size estimates for a few popular mixed models, whereas, Cheng et al. (2008) assess the impact of number of replicate measurements on sample size for linear mixed models.

It is not evident that linear mixed effect models with compound symmetry can satisfactorily be employed to represent all kinds of QT data, and the existing literature inadequately addresses whether there is a change in sample size and power, if any, in case one looks beyond such models. We provide a way to compute sample sizes for a few mixed effect models that extend the sample size and power calculations to more general models, allowing different correlations between different set of time points. The results in this paper extend the results from Zhang and Machado (2008) to examine models with a variety of covariance structures.

This issue may be important if correlation is not constant within a period, given that the time points pertaining to the measurements taken on the same subject are not equispaced.

For example, it is likely that the measurements taken within the first six hours are more correlated than those taken after that. Our results show that there is no loss of power that would impact the sample size substantially unless the correlations for the different sets of time points within any period are largely different. A conservative sample size estimate can be obtained by assuming a compound symmetric covariance structure with the smallest correlation for different sets. We have provided a model with two sets of time points with different correlations but it can be easily extended to a model with more sets of time points. This might be useful if QTc measurements are taken at a larger number of time points.

We have also provided sample size results for linear mixed models with the commonly invoked compound symmetric covariance structure for the original data or a compound symmetric covariance structure for the difference vector, that is equivalent to using a random period effect. We have provided these results for a variety of plausible parameter settings in the Appendix; these can be of great use for practicing statisticians in designing thorough QT studies for crossover designs. We also compared the Akaike Information Criterion (AIC) for the above three models using our example dataset and found out that the AIC values for the model with a random period effect and the model with a random period effect and time-varying correlation were similar (12763.5 and 12764.9 respectively) and were significantly smaller than the AIC for the model with a fixed period effect (12962.6). This shows that for this particular dataset, the models with a random period effect explain the data in a better way than the models with a fixed period effect, with a minimum of free parameters. We have also given a general framework for simulations, with explicit formulas for the difference vector $Var(\boldsymbol{D}_i)$, which would allow one to conduct the simulations with a different set of operating characteristics, if desired. The SAS code for conducting the simulations to obtain sample size and power can be obtained by request from the first author.

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 imply 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 shorter 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.

Another important aspect of a thorough QT study is the sensitivity of the trial. Zhang and Machado (2008) point out that the trial must be able to detect obvious differences, as indicated by a 'positive' result for the positive control. We take a slightly different approach than the other recently proposed methods in the literature by the use of the Hailperin-Ruger method to test the corresponding global null hypothesis with multiplicity adjusted significance level for a subset of the total number of time points. Our approach is easy to use and can be used to compute a sample size for sensitivity using standard software (e.g., PASS) under the paradigm of a superiority test for a crossover design. The methods proposed in this paper provide a reasonable methodology for sample size determination that can be used to design an adequately powered crossover thorough QT study which is also sensitive enough to detect obvious QT prolongation signals. We observe that the sample size required to demonstrate assay sensitivity may drive the overall sample size if the regulators insist that assay sensitivity needs to be strictly a matter of statistical methodology with multiplicity control rather than a matter of clinical judgement with "beauty in the eye of the clinical beholder." The extension of our methods to parallel designs is straightforward and can be easily accomplished. Research of interest within the purview of our proposed methods will include the design of a parallel study with unequal sample sizes for the drug/placebo/postive control arm, that satisfies the power requirements for both non-inferiority and sensitivity in the context of thorough QT studies.

7 References

Anand, S. P. and Ghosh, S. K. (2009). A Bayesian approach for investigating the risk of QT prolongation, *Journal of Statistical Theory and Practice*, (To appear in **3(2)**).

Anand S. P., Murray S. C., Koch G. G. (2008). A multivariate approach to sample size calculations for Thorough QT studies, *Institute of Statistics Mimeo Series*, # **2608**. (`www.stat.ncsu.edu/library/papers/ISMS2608.pdf`)

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

Chow, S., Cheng, B., Cosmatos, D. (2008). On Power and Sample Size Calculation for QT Studies with Recording Replicates at Given Time Point, *Journal of Biopharmaceutical Statistics*, **18(3)**, 483-493.

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 Information Journal*, **40**, 267-271.

Hailperin, T. (1965). Best possible inequalities for the probability of a logical function of events, *American Mathematical Monthly*, **72**, 343-359.

Hosmane, B. and 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.

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`

Patterson, S., Agin, M., Anziano, R., et al. (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 Information Journal*, **9**, 243-266.

Rüger, B. (1978). Das maximale Signifikanzniveau des Tests: "Lehne H ab, wenn k unter n gegebenen Tests zur Ablehnung führen", *Metrika*, **25**, 171-178.

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

Zhang, L., Dmitrienko, A., Luta, G. (2008). Sample Size Calculations in Thorough QT Studies, *Journal of Biopharmaceutical Statistics*, **18(3)**, 468-482.

Zhang, J. and Machado, S. G. (2008). Statistical Issues Including Design and Sample Size Calculation in Thorough QT/QTc Studies, *Journal of Biopharmaceutical Statistics*, **18(3)**, 451-467.

8 Appendix

Proof of Hailperin-Ruger inequality. Let $p_i, i = 1, 2, \dots q$, denote the probability of detecting a significant result for the hypothesis H_{0i} in (9). Let $\alpha^* = \frac{q^{'}}{q}\alpha$, where $q^{'} \leq q$ and α is the global type I error rate for H_0.

$$\text{Suppose } \boldsymbol{X}_i = \begin{cases} 1 & \text{if } p_i \leq \alpha^* \\ 0 & \text{if otherwise.} \end{cases}$$

Note that the expected value of $\boldsymbol{X}_i$, $E(\boldsymbol{X}_i) = \alpha^*$. Let $\boldsymbol{S} = \sum_{i=1}^{q} \boldsymbol{X}_i$, then $E(\boldsymbol{S}) = q\alpha^*$.

$$\text{Now, let } \boldsymbol{Y} = \begin{cases} 1 & \text{if } \boldsymbol{S} \geq q^{'} \\ 0 & \text{if otherwise.} \end{cases}$$

Note that $\boldsymbol{Y} \leq \boldsymbol{S}/q^{'}, \because \boldsymbol{Y} = 0 \; \forall \; \boldsymbol{S} \leq q'$, and $\boldsymbol{Y} = 1 \; \forall \; \boldsymbol{S} \geq q'$, but for those values of $\boldsymbol{S}$, $\boldsymbol{S}/q^{'} \geq 1$.

Since $\boldsymbol{Y} \leq \boldsymbol{S}/q^{'} \; \forall \; \boldsymbol{S}$, $E(\boldsymbol{Y}) \leq E(\boldsymbol{S}/q^{'}) = \frac{q}{q^{'}}\alpha^* = \alpha$. But $E(\boldsymbol{Y}) = P[\boldsymbol{S} \geq q^{'}] = P[\sum_{i=1}^{q} \boldsymbol{X}_i \geq q^{'}]$. Thus, the probability that at least $q^{'}$ of the $\boldsymbol{X}_i$ have their $p_i \leq \alpha^*$ is less than or equal to α. □

Table 2: Sample size calculations for the crossover design for 90% power with no random period effect ($\sigma_p = 0$)

		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 3: Sample size calculations for the crossover design for 90% power with random period effect ($\sigma_p = 4$)

		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 4: Sample size calculations for the crossover design for 90% power given overall σ with $\rho_1 = \rho_2$

			Sample Size (n)		
θ	$\mathbf{\Delta}$	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