

Design and Analysis of Thorough QT Studies

Alex Dmitrienko, PhD

Research Advisor, Global Statistics, Eli Lilly and Company

Charles Beasley, MD

Distinguished Lilly Scholar, Global Patient Safety, Eli Lilly and Company

Malcolm Mitchell, MD

Medical Director, Early-Phase Biopharmaceuticals, Eli Lilly and Company

April 29, 2008

This white paper has been written to facilitate an open discussion of issues related to the design and analysis of thorough QT studies. The correct bibliographic citation for this paper is as follows:

Dmitrienko A, Beasley C, Mitchell M. Design and Analysis of Thorough QT Studies. Biopharmaceutical Network. Report 2008-04-29.

A free copy of the paper can be downloaded from Biopharmaceutical Network's Cardiac Safety page at

<http://www.biopharmnet.com/cardiac.html>

Readers are encouraged to share their comments, thoughts and suggestions by e-mailing them to

ecg@biopharmnet.com

Please include the report's title (Design and Analysis of Thorough QT Studies) in the subject line of the e-mail.

1. Introduction

Assessment of cardiac liability with respect to life threatening ventricular tachydysrhythmias presumably due to delayed ventricular repolarization (e.g., Torsade de Pointes [TdP]) of new compounds is becoming an increasingly important component of pre-clinical and clinical drug development. Lengthening of QT interval corrected for heart rate (QTc interval), representing the duration of ventricular depolarization and subsequent repolarization on a 12-lead surface electrocardiogram (ECG) in clinical trials, is commonly used as a surrogate biomarker for an increased risk of TdP during clinical use of the compound.

The International Conference on Harmonization (ICH) published a guidance document (ICH E14, Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs) to describe strategies for the evaluation of cardiac safety of drugs in clinical development. This document introduced a new approach to the assessment of proarrhythmic potential of new, non-cardiac drugs with systemic bioavailability (thorough QT study). The objective of this study is “to determine whether the drug has a threshold pharmacologic effect on cardiac repolarization, as detected by QT/QTc prolongation” (ICH E14, Section 2.2). Although ICH E14 specifies requirements for a “negative study” (one demonstrating absence of effect on ventricular polarization), the document provides little guidance on the details of design of a thorough QT study or analysis of data collected in such a study.

This white paper discusses design and analysis considerations in thorough QT studies. Section 2 briefly summarizes main difficulties with the measurement and analysis of QT interval and provides a high-level description of key features and outcomes of thorough QT studies. Section 3 reviews design elements in thorough QT studies that impact the probability of a negative thorough QT study. Section 4 gives an overview of common approaches to the analysis of QT interval data collected in these studies that also influence the probability of a negative outcome under the assumption that the test drug, as administered, in fact has no effect on ventricular repolarization. It is worth noting that, as in other clinical trials, design considerations and analysis strategies are closely intertwined in thorough QT studies and certain design topics could have been included in the analysis section and vice versa.

2. Thorough QT studies

2.1. Difficulties with QT interval measurement and analysis

To understand the rationale behind and key features of thorough QT studies, it is helpful to review important properties of the QT interval:

- The most challenging problem faced by clinical trial sponsors and regulators is that it is generally quite difficult to measure with precision the length of QT interval. The precision of measurement problem was recognized 50 years ago (Lepeschkin and Surawicz, 1955) when QT intervals were measured using paper ECG recordings and remains unsolved in the era of electronic ECG recordings. The primary reason for this difficulty in precise measurement is that the offset (end) of the QT interval (and to a lesser extent the onset) is on a curve rather than being a distinct, easily identified, point. The problem is essentially equivalent to that of attempting to measure the “side” of a circle, which, of course, has no side.
- Analysis of drug-related QT interval changes is also complicated by the fact that its length depends on the heart rate. We still have a poor understanding of the precise nature of this relationship, not only across populations but even within an individual (Malik, 2001).
- Finally, an inherent variability of QT interval poses significant problems in the analysis of the QT interval. Even with stable heart rate as well as the best QT measurement techniques and QT correction methods, we continue to see a fair amount of variability over very brief time intervals, essentially on a beat-to-beat basis. The variability may be as much as 25 msec over 10 consecutive complexes and yet clinical trial sponsors are interested in detecting differences that may be as small as 5 msec (Malik and Camm, 2001).

2.2. Key features of thorough QT studies

Given the difficulties associated with the measurement and analysis of the QT interval (Section 2.1), regular monitoring of ECG data in clinical trials may not provide a reliable means of detecting small changes in QTc duration that are nevertheless predictive of some degree of risk. Conversely, changes may be observed in clinical trials that are due to benign, random variability and/or measurement/collection artifact. Thorough QT studies were introduced in ICH E14 in an attempt to create a more rigorous framework for the assessment of drug-related QTc prolongation. This is achieved by including the following elements in the design of thorough QT studies:

- Evaluation of the drug’s effect on QTc interval in a strictly controlled environment (healthy subjects without comorbidities are typically enrolled in thorough QT studies).
- Frequent ECG recordings to perform a comprehensive characterization of drug-related QTc changes.

- The QTc effect is evaluated at exposure levels that exceed the anticipated maximum therapeutic exposure (a substantial multiple of the therapeutic dose is typically used).

2.3. Classification of outcomes in thorough QT studies

The outcome of a thorough QT study is classified as negative (no evidence of QTc prolongation) if the upper limit of a one-sided 95% confidence interval (or a two-sided 90% confidence interval) for the largest time-matched mean difference between the test drug and placebo is below 10 msec (ICH E14, Section 2.2.4). It is pointed out in the document that “This definition is chosen to provide reasonable assurance that the mean effect of the study drug on the QT/QTc interval is not greater than around 5 ms.”

This definition serves as a useful starting point in classifying potential outcomes in thorough QT studies. To help individualize the design of a thorough QT study to the molecule of interest, it is helpful to define the following hierarchy of desired outcomes:

- Type 0: The mean treatment difference and upper confidence limit are less than 5 msec.
- Type 1: The mean treatment difference is less than 5 msec and upper confidence limit is between 5 and 10 msec.
- Type 2: The mean treatment difference and upper confidence limit are between 5 and 10 msec.
- Type 3: The mean treatment difference is between 5 and 10 msec and upper confidence limit is greater than 10 msec.
- Type 4: The mean treatment difference and upper confidence limit are greater than 10 msec.

The Type 0 outcome gives the clinical trial sponsor great certainty that the test drug has no impact on QTc prolongation and, as one moves higher in the continuum of outcomes, there is increasingly more evidence that the drug induces QT prolongation. Type 0, 1 and 2 outcomes are consistent with the ICH E14 definition of a negative study and Type 3 and 4 outcomes will indicate that the drug affects ventricular repolarization. The interpretation of the results of a thorough QT study is likely to depend on the population in which the drug is intended to be used and the condition for which the drug is intended to be used. Specifically, a Type 2 outcome might result in adverse labeling and Type 3 and 4 outcomes will be considered unacceptable for a drug intended to treat a relatively benign condition. Thus, a thorough QT study for a drug intended for wide chronic use in a non-life threatening condition, especially if intended for use in children, may need to be customized to ensure a high probability of detecting a Type 0 or a Type 1 outcome (under the assumption that the drug truly does not affect ventricular repolarization) in order to achieve commercially viable labeling or to even achieve approval in some regulatory venues. On the other hand, a thorough QT study with a Type 3 or a Type 4 outcome may have little impact on commercial opportunities for an oncology drug that is of substantial benefit over existing therapies.

To illustrate the impact of the study's outcome on the product label, consider the thorough QT studies of tadalafil (Beasley et al, 2005) and vardenafil (Morganroth et al, 2004). The first study produced a Type 0 outcome (mean difference, 2.8 msec; 90% CI, 1.2-4.4 msec) whereas the results of the other study were consistent with a Type 2 outcome (mean difference, 6 msec; 90% CI, 5-8 msec, at the suprathreshold dose, vardenafil 80 mg). The two studies were reviewed by the same division of the FDA but the label language for the two drugs was markedly different:

- The tadalafil label only includes a brief summary of the results of the thorough QT study in the Clinical Pharmacology section.
- The vardenafil label discusses the results of the thorough QT study in the Precautions section. It is stated in the label that the effects of the therapeutic and suprathreshold doses of vardenafil on QTc prolongation were similar to that of the positive control and patients on antiarrhythmic drugs should avoid using vardenafil.

QTc prolongation demonstrated in a thorough QT study does not necessarily prohibit a drug's ability to get approved. However, it is critical to design a thorough QT study to provide a comprehensive characterization of the QTc safety profile that will facilitate labeling and risk/benefit assessment. Any degree of QTc prolongation will be considered in the context of the target population, indication and efficacy of the drug (degree of positive impact on public health). As with other potential adverse events, QTc prolongation and the risk of TdP will be judged in the context of potential benefits to determine approvability and appropriate labeling.

2.4. Design and analysis considerations in thorough QT studies

In this short section we will outline main design and analysis considerations in a thorough QT study. Given the definition of a negative study (see Section 2.3), the sponsor can minimize the probability of false-positive and false-negative outcomes by focusing on the following objectives:

- Selecting design elements and study procedures (e.g., QT measurement procedures) that
 - Reduce the variability of point estimates of the treatment effect (standard deviation of point estimates needs to be minimized).
 - Reduce the systematic bias of point estimates of the treatment effect (e.g., if the test drug has no effect on ventricular repolarization, point estimates need to be close to 0).
- Selecting analysis methods with appropriate operating characteristics (adequate power and Type I error rate controlled at a nominal level).

The first objective (appropriate design elements) will be discussed in Section 3. Section 4 will focus on the second objective (appropriate analysis methods).

3. Design considerations in thorough QT studies

Given the difficulties in QT interval measurement and analysis described in Section 2, it is clear that one study design will not fit all molecules in clinical drug development. This is recognized in ICH E14 (Section 1.2): “The investigational approach used for a particular drug should be individualized, depending on the pharmacodynamic, pharmacokinetic, safety characteristics of the product, as well as on its proposed clinical use.” In this section we will go over key design features that may need to be customized in a thorough QT study to help achieve its goals.

3.1. Homogeneous and diverse subject populations

Within acceptable regulatory bounds, it is best to select a homogeneous subject population of healthy subjects¹. A homogeneous subject population will help reduce the within-subject (intra-subject) variability and, since within-subject variability is the primary determinate of the precision of the measured treatment difference, this will help improve the power of the study.

General exclusion criteria

To achieve a more homogeneous subject population, the trial’s sponsor needs to remove factors that directly contribute to higher degrees of inherent QTc variability, a higher risk of QTc prolongation due to factors other than drug exposure or increase the variability of drug exposure concentrations. For example, the following exclusion criteria can be considered in a thorough QT study:

- Subjects with personal or family history of long QT syndrome, heart failure or hypokalemia.
- Subjects with ECG abnormalities, including abnormal QTc interval and conduction abnormalities that may affect QTc analysis.
- Subjects who smoke.
- Subjects with larger BMI values.

Gender-specific exclusion criteria

The sponsor can also consider restricting the study population to one gender, e.g., males, in order to increase homogeneity; however, this is a fairly complex issue. There are at least four matters that might also interact with one another, several not completely resolved, that bear on the decision to include or exclude females from thorough QT studies. These matters are first enumerated and then discussed in more detail below. These matters are:

¹ A population of patients with the condition intended as the therapeutic indication may be used in a thorough QT study if the sponsor expects safety concerns such as with a cytotoxic compound intended as an oncolytic. See ICH E14 (Section 2.3) for more information.

- Females have longer QT intervals than males. Although this difference exists, this difference is not a reason to require that females be included in thorough QT studies. QTc duration at baseline is not known to be a factor influencing the likelihood of experiencing QTc prolongation or the magnitude of QTc prolongation when exposed to a QTc prolonging drug. Although a comparable magnitude of QTc prolongation, relative to males, from a greater baseline value, among females, might place females at a greater risk of TdP (because the absolute magnitude of QTc interval influences the degree of TdP risk); a thorough QT study evaluates magnitude of QTc changes rather than the incidence of TdP.
- Females are more likely to develop TdP when exposed to predisposing factor or develop a predisposing factor for TdP relative to males. Some regulators might believe that this is a reason to include females in a thorough QT study. However, this belief represents flawed logic; this is *not* a reason requiring female inclusion. A thorough QT study does not include the incidence of TdP as an analyzable endpoint (assuming TdP is not observed in any thorough QT study). The real endpoint is a drug-related QTc change and thus excluding females with a greater risk of TdP in the context of equal prolongation of QTc interval does not limit the sponsor's ability to detect QTc changes.
- Females might display within-subject, systematic variability over time (over the period of the menstrual cycle) that is not displayed by males. Although this has not been definitively established, it is probably the case that females do evidence systematic changes in QTc interval across the menstrual cycle due to hormonal changes as discussed below. This is a clear reason to exclude females from thorough QT studies. This systematic change would introduce variability and would potentially create bias unless appropriate controls were introduced. These controls include the requirement that each female begin each study treatment at the same point within their individual menstrual cycle (in a cross-over study) and a matching of females for points in their menstrual cycles across treatment arms (in a parallel study). Menstrual cycles of different lengths would further complicate the matter. Note also that, not only might there be systematic change in basal QTc interval with the menstrual cycle, but, if females do show greater sensitivity to QTc prolongation (point discussed below), this degree of differential sensitivity, relative to males, may fluctuate over the course of the menstrual cycle.
- Females might or might not display greater QTc prolongation when exposed to a QTc prolonging drug relative to males at the same exposure level (e.g., identical maximum value and area-under-the-curve of plasma drug concentration). It is unclear as to whether females are more sensitive to QTc prolonging effects (greater individual magnitude of prolongation and greater incidence of prolongation as a group with given exposure) as will be discussed below. If they are more sensitive to QTc prolonging effects, this would argue for inclusion of females. However this consideration must be balanced by consideration of the potential for undesirable variability that inclusion of females might bring. Differential sensitivity may well be accounted for and overcome, if the phenomenon exists, through suprathreshold dosing of males.

As noted above, premenopausal females may add more complexity in that their systematic QT variability may be linked to the phases of their menstrual cycle and/or interaction of the phase of menstrual cycle with autonomic tone (Burke et al, 1997; Hulot et al, 2003; Kadish et al, 2004; Nakagawa et al, 2006). Although these data are not completely consistent, when blockade of sympathetic and parasympathetic activity is accomplished, the relationship between this variability and the phase of the menstrual cycle appears clear (Burke et al, 1998). Even more importantly, the response to Ikr blockers, with respect to QTc length may change in females across phases of the menstrual cycle (Rodriguez et al, 2001). Such systematic, temporally linked within-subject variability in QTc interval is a strong reason to advise against the inclusion of female subjects.

However, as noted above, if females would be more sensitive to the effects of drugs that delay or prolong ventricular repolarization than men, then this gender difference would argue for their inclusion lest a positive finding be possibly missed in a relevant target population. QT/QTc assessment data (Benton et al, 2000; Drici et al, 1998; Drici et al, 2001; Ebert et al, 1998; El-Eraky and Thomas, 2003; Johansson and Carlsson, 2001; Lu et al, 2000; Rong et al, 2001; Shin et al, 2006) suggest that females might show a greater increase in QTc interval than males at a given concentration of a QT prolonging agent. However, it is important to note that

- The data are not consistent.
- Pure Ikr blockers may not show this difference across genders.
- The difference might be accounted for more by changes in QRS length than JT length.
- The magnitude of this difference is small and might be on the order of what might be observed within females with changes in phase of the menstrual cycle.

If there is a difference in the magnitude of QTc changes associated with known QT prolonging agents between males and females, it varies at various stages of the menstrual cycle and ranges from 15 to 44% (Benton et al, 2000; Rodriguez et al, 2003). For example, if a drug were to prolong QTc interval in a male by 10 msec, it might prolong QTc in a female by 11.5 to 14.4 msec, provided there is differential prolongation and depending on the point in the menstrual cycle. Thus, QTc changes associated with the phase of the menstrual cycle might well be greater than QTc changes that would be considered of potential clinical significance in a thorough QT study. On balance, inclusion of females in a thorough QT study would increase undesirable variability.

The bottom line with respect to inclusion of females in thorough QT studies is that this is likely to increase variability and the probability of false positive or false negative findings. As an alternative to including females, with a possible increase in the sensitivity of the assay but with the accompanying possibility of introduction of bias, increasing the dose of the test drug would serve to increase the sensitivity of the study. The increased sensitivity of females, if such an increased sensitivity does exist, is likely due to a decreased tissue concentration of the cardiac ion channels that, when blocked by

a given concentration of the test drug, lead to a delay in ventricular repolarization and in increase in QTc interval.

Genetic screening

Finally, to achieve a homogeneous subject population, the trial sponsor can theoretically consider genetic screening mentioned in ICH E14 (Section 4.3); however, it is a very expensive option. In general, it is recommended that the trial sponsor work with the regulatory agency that will be reviewing the study protocol to define an acceptable subject population consistent with the drug's indication.

3.2. Parallel and cross-over designs

The choice of the study design (parallel versus cross-over) is a fairly straightforward consideration. As in other early-phase clinical trials, this decision is driven almost exclusively by pharmacokinetic properties of the test drug. A cross-over design is recommended unless the washout period (the length of which is typically greater than 5 times the drug's half-life) is excessively long. For example, it may be difficult to retain subjects in a cross-over design with three treatment periods when the washout period is longer than 3 weeks. The length of each treatment period is also an important consideration since it directly contributes to the length of the study. Examples of thorough QT studies with longer treatment periods include dose-titration designs, e.g., duloxetine QT study (Zhang et al, 2007) or designs aimed at evaluating steady-state QTc effects, e.g., tolterodine QT study (Malhotra et al, 2007).

When a cross-over design is employed in a thorough QT study, it is important to ensure that a balanced Williams design is used (Williams, 1949). In a Williams design, each treatment precedes every other treatment, excluding itself, equally often and thus a balance for carry-over effects is achieved. As an illustration, consider a thorough QT study with four treatment groups (placebo, therapeutic and suprathreshold doses of the test drug and positive control labeled P, T, S and C, respectively). In this case, the number of treatment groups is even and only one square is needed. The four-period, four-sequence Williams design for this thorough QT study is displayed in Table 1. It is instructive to compare this design to the standard Latin square design which is also shown in Table 1. The Latin square design is not balanced, e.g., placebo always precedes the therapeutic dose. Further, to maintain the blinding of the test drug and placebo, the actual treatment randomization sequences should not be defined in the study protocol.

Table 1. Williams and standard Latin square designs for cross-over thorough QT studies with four treatment groups (P, placebo; T, therapeutic dose; S, suprathapeutic dose; C, positive control).

Sequence	Williams design	Standard Latin square design
Sequence 1	P T S C	P T S C
Sequence 2	T C P S	T S C P
Sequence 3	S P C T	S C P T
Sequence 4	C S T P	C P T S

When the number of treatment groups is odd, two squares need to be used to construct a balanced Williams design, i.e., the number of sequences is twice the number of treatment groups. For example, Table 2 shows the Williams design for a thorough QT study with three treatment groups (placebo, therapeutic dose of the drug and positive control labeled P, T and C, respectively).

Table 2. Williams design for cross-over thorough QT studies with three treatment groups (P, placebo; S, suprathapeutic dose; C, positive control).

Sequence	Williams design
Sequence 1	P S C
Sequence 2	S C P
Sequence 3	C P S
Sequence 4	C S P
Sequence 5	P C S
Sequence 6	S P C

In addition, unbalanced/incomplete cross-over designs can also be considered (e.g., placebo as the first treatment in all sequences). Although these designs translate into less time per individual subject and less total time for the study (e.g., one wash-out period can be eliminated when placebo is the first treatment arm), such designs are not recommended. Such unbalanced designs would not compensate for potential, systematic period effects or carry-over effects. At this time, such designs are likely to receive regulatory authority resistance because of this potential for period effects.

If the study sponsor has to use a parallel-group design, less time commitment from subjects will be required. However, since the analysis of QTc effects in parallel-group studies relies on between-subject comparisons (rather than within-subject comparisons in cross-over studies), the sample size in a parallel-group study will be considerably larger compared to a cross-over study with similar characteristics. As shown in Malik et al (2004), Zhang and Smith (2007) and Zhang, Dmitrienko and Luta (2008), the sample size in thorough QT studies with a parallel-group design can exceed 100 subjects per group whereas the total size in thorough QT studies with a cross-over design is typically less than 100 subjects (unless demonstration of a Type 0 or Type 1 outcome [see Section 2.3] is the goal of the study in which case a cross-over study may require as many as 100 subjects).

Lastly, it is also worth noting that hybrid designs have been discussed in the literature, e.g., Malik et al (2004) described studies with several cross-over groups that are run in parallel. However, it is unclear to what extent these designs will be acceptable from a regulatory perspective.

3.3. Single-dose and steady-state designs

It is important to ensure that adequate plasma concentrations can be achieved in thorough QT study. The doses used should be designed to cover the “worst-case” plasma concentrations that are likely to be achieved as a result of either intrinsic or extrinsic factors. The ability to achieve this by supratherapeutic concentrations should result in acceptable labeling. Potential options to achieve this include

- Administering the test drug as a single dose (tadalafil QT study, Beasley et al, 2005). This approach is generally recommended for well-tolerated drugs with no active metabolites.
- Multiple administrations of the drug to achieve a steady state level (darifenacin QT study, Serra et al, 2005). This approach is relevant in the presence of one or more metabolites and/or when the drug is poorly tolerated and requires up-titration to prevent subject discontinuation.
- Combining evaluation of steady state QT effects with determination of a tolerable dose (duloxetine QT study, Zhang et al, 2007). This approach is more suitable when tolerability at higher doses is unknown. In this case, the dose can be selected via within-subject dose-titration (note that the dose-titration component also helps assess the dose-response relationship for QTc effects)².

The choice of the number of doses and dose levels in a thorough QT study is driven by several considerations. Although ICH E14 does not require that multiple dose levels be evaluated in a thorough QT study, two dose levels (a therapeutic dose and a

² Such a design has the inherent disadvantage of being an unbalanced design (a lower dose arm always precedes a higher dose arm) but may be a practical necessity with certain poorly tolerated drugs. The alternative with poorly tolerated drugs, or drugs otherwise requiring titration and where it is desirable to study several doses, would be up-titration with evaluation *only* at one specific dose level.

supratherapeutic dose³) are often included in QT studies. This can be done to better understand the dose-response relationship or QTc-exposure relationship (note that it normally takes more than two doses to adequately characterize a dose-response function) if any such relationship exists. In addition, the therapeutic dose can serve as a fallback strategy when there is evidence of QTc prolongation at the supratherapeutic dose. However, if clinically important QTc effects are observed at the supratherapeutic dose, language to this effect will likely be included in the label. In general, if the test drug is unlikely to have a QTc prolongation effect, it is difficult to justify (from an efficiency perspective) an additional, lower dose level (therapeutic dose) in a thorough QT study.

The supratherapeutic dose level is intended to achieve the highest exposure levels anticipated in clinical practice, e.g., exposure levels caused by drug-drug interactions and/or genetic polymorphisms affecting drug metabolism. ICH E14 (Section 2.2.2) states that “if not precluded by considerations of safety or tolerability due to adverse effects, the drug should be tested at substantial multiples of the anticipated maximum therapeutic exposure.” The substantial multiple is typically defined as a 5-fold increase over the recommended therapeutic dose (for instance, a 5-fold multiple was used in the tadalafil QT study, Beasley et al, 2005). Exceptions include cases when tolerability issues are expected at higher doses⁴. For example, the supratherapeutic dose examined in the tolterodine QT study (Malhotra et al, 2007) was only 2-fold higher than the therapeutic dose (2 mg versus 4 mg). A higher dose could not be tested in this study because of dose-limiting anticholinergic side effects observed in previously conducted tolterodine trials.

Other methods that might be employed to achieve supratherapeutic concentrations include

- Administration of an additional drug that inhibits the metabolism of the test drug (if the test drug is metabolized through an enzymatic [typically a cytochrome P450 enzyme] pathway for which an inhibitor is available for human administration). This method is described in ICH E14 (Section 2.2.2).
- Use of subjects who are genetically predisposed to poor metabolism of the test drug due to genetic deficiency of an enzyme in the metabolic pathway for the test drug.

We strongly recommend against use of a metabolic inhibitor. Many drugs that do inhibit various cytochrome P450 enzymes are known to have an influence on ventricular repolarization and their administration would therefore confound the observed results of the study (Morganroth, 2005).

³ Examples include the alfuzosin QT study (Extramiana et al, 2005), darifenacin QT study (Serra et al, 2005), tiotropium QT study (Ring et al, 2006), tolterodine QT study (Malhotra et al, 2007).

⁴ It is worth noting that, when supratherapeutic exposures are not expected to be achieved in a thorough QT study, regulatory agencies may reserve the right to change the definition of a negative study outcome. Specifically, all available data (including the time course of QTc changes and QTc-exposure relationship) rather than the ICH E14 definition based on a 10-ms threshold may be used to define a negative thorough QT study.

3.4. Choice of positive control

A positive control is included in thorough QT studies to assess the sensitivity of the study in detecting drug-related QT effects (ICH E14, Section 2.2.1). Specifically, the assay sensitivity is established if the control is shown to induce QTc prolongation. A positive control is generally expected to have a moderate effect on QTc interval, in the range of 5-10 msec (greater than the effect observed with placebo), and this effect is expected to be statistically significantly different from the effect of placebo (see Section 4.2 for a detailed discussion of statistical criteria used in the assay sensitivity analysis).

Orally administered moxifloxacin 400 mg is currently the most commonly used positive control. It has been used in the overwhelming majority of thorough QT studies (the tadalafil QT study was the only known exception) mainly because its effect on QTc interval has been well characterized (Demolis et al, 2000). The peak plasma concentration for a single 400-mg dose of moxifloxacin occurs 2 hours after the dose and the maximum QTc effect is achieved slightly after 2 hours post-dose. As reported in recently published summaries of thorough QT studies, the mean maximum effect of this positive control is variable and ranges between 6 and 10 msec⁵.

Ibutilide, administered intravenously at a 0.002 mg/kg rate infusion for 10 min, was the only other positive control used in published thorough QT studies. One of the advantages of ibutilide is that it is relatively easy to control its effect on QTc interval. Unlike moxifloxacin, ibutilide does not produce a delayed response, its maximum QTc effect is close to the end of infusion, QTc interval begins to decline at about 10 minutes post infusion, and prolongation is resolved by about 30 min after the end of infusion (Beasley et al, 2005). However, since it is administered intravenously, ibutilide can complicate blinding of the positive control period or arm.

Blinding of the positive control may be accomplished through the use of the so-called double dummy method. This has been a recent requirement on the part of some regulators but is now being relaxed. The important requirement is for blinding of the individual or individuals who actually perform the measurements that will be used for analysis. Blinding of the positive control *might* serve to reduce the possibility of an unexpected magnitude of change with the positive control but there is no empirical validation of this theoretical utility. Overall, our experience is that blinding of the positive control is an unuseful burden on the sponsor and the site staff conducting the work with the subjects.

3.5. Control of QT variability

It was indicated in Section 2.4 that one of the main goals of a thorough QT study is the reduction of

- QT measurement variability (due to difficulties in precision of measurement).

⁵ The maximum effect of a single 400-mg dose of moxifloxacin can exceed 10msec (tolterodine QT study, Malhotra et al, 2007).

- Inherent QT variability (due to a large variety of influences, some well known, i.e., circadian changes in QTc length, food effect, and others that are occult).

This can be accomplished by controlling/removing known sources of QT variability and lowering the background QT measurement variability via signal-averaging.

Controlled environment

To help remove known sources of QT variability, steps need to be taken to ensure that the relationship between all temporal activity (eating, drinking, sleeping, exercise timing) and the ECG collection schedule be identical within subjects (across treatment periods or arms) and across subjects.

ECG recordings need to be taken in a quiet controlled environment. It is generally advisable that subjects rest in a supine position for approximately 10 minutes prior to ECG recordings. Additional factors that need to be taken into account include food intake. Consumption of a meal has been shown to lengthen QTc interval (Nagy et al, 1997). To prevent the confounding effect of meal consumption, subjects may be required to fast overnight prior to receiving the test drug and continue fasting until the lunch meal or until after the pharmacokinetic t_{max} has been achieved.

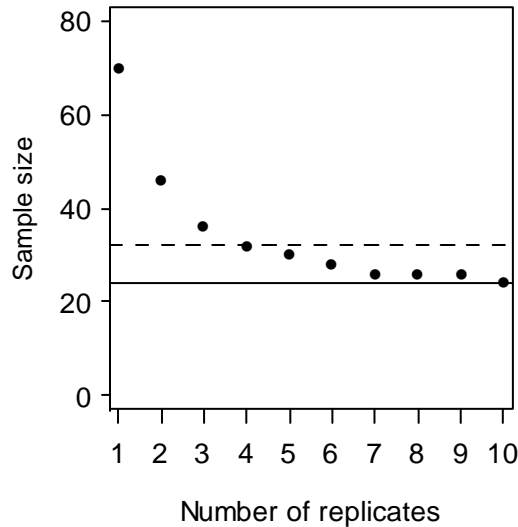
Signal-averaging

Signal-averaging is used in all thorough QT studies to strengthen the signal by averaging QTc values collected every 30 seconds to 2 minutes around each point of ECG acquisition. The number of replicate ECG recording at each time point directly influences the precision of the results and the power of a thorough QT study (or, equivalently, the sample size required to achieve a specified power level). In recently conducted thorough QT studies this number ranged between 3 and 10⁶. In general, the incremental improvement in the precision of measuring QT interval beyond 6-7 replicates is very small and thus 3-5 replicates are generally recommended for signal averaging purposes in thorough QT studies. The larger number of replicates is advised when a Type 0 or Type 1 outcome is the goal of the study (see Section 2.3).

When selecting the number of replicate ECG recordings in a thorough QT study, it is important to account for key study features such as the subject population and definition of the treatment difference (popular treatment difference definitions are discussed in Section 4.3). As shown in Zhang, Dmitrienko and Luta (2008), the subject population affects the variance components related to hour-to-hour, day-to-day and other sources of variability in QTc interval and, along with the treatment difference definition, has a direct impact on the number of replicate ECGs. As an illustration, consider a thorough QT study with a cross-over design and suppose the study is designed to achieve 95% power when the true mean treatment difference is 5 msec. Using Definition D2 of the treatment difference introduced in Section 4.3, Figure 1 depicts the relationship between the sample size and number of replicates in a mixed population (males and females). With an

⁶ For example, 3 replicates (tolterodine QT study, Malhotra et al, 2007), 4 replicates (duloxetine QT study, Zhang et al, 2007), 6 replicates (vardenafil QT study, Morganroth et al, 2004) and 10 replicates (tadalafil QT study, Beasley et al, 2005).

Figure 1. Relationship between the sample size to achieve 95% power in a cross-over design and number of replicate ECG recordings at each schedule time point (mean treatment difference is 5 ms). The solid line represents the sample size corresponding to an infinite number of replicate ECGs and the dashed line is the upper bound of a 30% window.



infinite number of replicate ECGs, 24 subjects are required to achieve the desired power level. To find the largest feasible number of replicates, it is reasonable to set up a window around this sample size, e.g., a 30% window. The upper bound of a 30% window corresponds to 32 subjects, which translates into 4 replicates. Note that the design with 4 replicates provides a considerable reduction (54%) in the required sample size compared to the design with a single ECG recording at each time point. Similar arguments can be used to determine the number of replicate ECGs for other treatment difference definitions and subject populations.

The calculations and results above emphasize the precision of treatment effect estimates. By reducing the background QT measurement variability, the sponsor can also address the second design objective given in Section 2.4, i.e., ensure that the observed estimates of the treatment difference are close to their true values (e.g., the estimates should be near 0 when a placebo is compared to a placebo in a thorough QT study). The bias reduction considerations can therefore guide the selection of the number of replicate ECG recordings, especially in a post-hoc analysis of the data collected in thorough QT studies.

Lastly, when selecting the number of replicates, it is worth keeping in mind that, as the length of the time frame within which replicate ECG recordings are collected extends, it becomes more difficult to control other factors influencing QT intervals such as circadian variability.

3.6. ECG acquisition schedule

The number and timing of ECG recordings taken after each dose of the test drug are driven mainly by the pharmacokinetic and pharmacodynamic properties of the drug.

Ideally, the time points for ECG acquisition should be chosen based on the expected time course of drug-related QTc changes, e.g., before, around and after the pharmacodynamic t_{\max} (the time point associated with the peak QTc effect). However, the pharmacodynamic t_{\max} and time course of QTc changes are difficult to predict precisely and pharmacokinetic parameters, e.g., the pharmacokinetic t_{\max} , are used to determine the number and timing of post-dose ECG recordings. It is worth noting that this approach relies on the assumption of a direct QTc-concentration relationship (direct and lag-phase [hysteresis] QTc-concentration relationships are discussed in Section 4.4).

Without major metabolites, at least four post-dose ECG recording time points need to be considered to adequately characterize the time course of QTc changes: before the pharmacokinetic t_{\max} , around t_{\max} , “slightly” after t_{\max} and at an anticipated trough point. The time points around t_{\max} must account for both individual differences in pharmacokinetics and hysteresis. The pharmacokinetic t_{\max} tends to be highly variable and thus multiple post-dose ECG recordings are generally included around the pharmacokinetic t_{\max} . In addition, when metabolites are present, ECGs need to be taken around the metabolite-specific t_{\max} points as well.

In general, it is clear that the more post-dose ECG recordings are taken the better the time course of the QTc effect will be estimated. However, as explained in Section 4.2.2, the multiplicity burden increases with the number of post-dose ECGs, i.e., including too many post-dose ECGs increases the likelihood of incorrectly concluding that the test drug induces QTc prolongation and reduces the statistical power of the study. The trial sponsor needs to find a balance between the requirement to accurately characterize the time course of QTc changes and the study power.

Lastly, as stated in Section 3.4, the positive control is typically required to be administered in a blinded fashion. Therefore, it is important to use the same ECG collection schedule across the treatment periods (in cross-over studies) or treatment arms (in parallel-group studies).

4. Analysis strategies in thorough QT studies

4.1. Correction of QT interval for heart rate

It was pointed out in Section 2 that the assessment of drug-related changes in QT interval is complicated by the fact that its duration depends on a large number of physiological factors. The most important of these factors is the heart rate. The length of QT interval is inversely proportional to heart rate or, equivalently, proportional to RR interval. To remove the confounding effect of heart rate or RR interval on the length of QT interval, a corrected QT interval (QTc interval) is commonly introduced. QTc interval is defined as the QT value that would have been observed if heart rate had been 60 bpm or RR interval had been 1 sec. Popular approaches to QT interval correction are described in this section.

4.1.1. Fixed QT correction methods

The first two widely used QT correction formulas were proposed by Bazett (1920) and Fridericia (1920):

$$QTc = QT / \sqrt{RR} \text{ (Bazett square-root correction),}$$

$$QTc = QT / \sqrt[3]{RR} \text{ (Fridericia cube-root correction).}$$

These and other QT correction formulas proposed in the literature (for example, Framingham formula, Sagie et al, 1992) are known as fixed or historical population QT corrections. They perform well only if a large number of restrictive conditions are satisfied, e.g., the actual QT-RR relationship for all subjects is close to the one assumed by the correction formula, the heart rate does not change over time, etc. If the test drug affects the heart rate, the corrections become unreliable, i.e., they are likely to underestimate or overestimate the true QTc effect (Dmitrienko and Smith, 2002, 2003). The Fridericia QT correction tends to perform better than the Bazett correction in this case but it still needs to be used with caution⁷.

⁷ To illustrate the performance of the Bazett and Fridericia corrections in the presence of drug-induced heart changes, consider the alfuzosin and tolterodine QT studies. Extramiana et al (2005) examined the Bazett and Fridericia corrections in a study that evaluated proarrhythmic potential of alfuzosin, an α_1 -blocker known to increase heart rate. Even through the suprathreshold dose of alfuzosin (40 mg) induced a fairly modest heart rate increase (mean change, 3.7 bpm), the use of fixed QT corrections led to a substantial probable overestimation of QTc effect. The treatment differences (95% CI) for the Bazett and Fridericia corrections were 10.8 (7.2, 14.4) msec and 6.9 (4.2, 9.5) msec, respectively. In comparison, the treatment difference and upper limit of a 95% CI were around 2.5 and 4.7 msec, respectively, for QT correction methods (based on Holter monitoring) that were developed specifically for the analysis of QTc effect for compounds that induce heart changes. Further, Malhotra et al (2007) reported the results of the tolterodine QT study. The mean increase in heart rate associated with a suprathreshold dose of tolterodine was 6.3 bpm. The use of the Bazett correction led to an probable overestimation of the true treatment difference in QTc interval (mean difference, 11.9 msec; 90% CI, 8.1-15.8 msec) and the Fridericia correction produced a much lower estimate of the true treatment difference (mean difference, 5.6 msec; 90% CI, 1.5-9.8 msec). To assess the performance of the Fridericia correction in this study, it is instructive to compare these results to the estimated treatment difference based on a more reliable QT

Despite this limitation, fixed QT corrections are widely used in clinical practice and are generally required to be included in thorough QT studies to provide supportive evidence. ICH E14 (Section 3.1) emphasizes that “QT interval data corrected using Bazett’s and Fridericia’s corrections should be submitted in all applications, in addition to QT interval data corrected using any other formulae.” However, since the Bazett QT correction is considered less reliable than the Fridericia QT correction, there has been a trend to put less emphasis on the former in thorough QT studies. The Bazett correction was not used in the vardenafil QT study (Morganroth et al, 2004) and was treated as the least reliable correction method in all other thorough QT studies.

4.1.2. Data-driven QT correction methods

Data-driven QT correction methods are becoming increasingly popular in clinical trials including routine ECG assessments as well as thorough QT studies. Unlike fixed corrections, such as the Bazett or Fridericia corrections, data-driven corrections take into account key features of the patient population (such as gender and age distribution) studied in each individual trial and thus lead to a better understanding of the QT-RR relationship and drug-related QTc effects (Dmitrienko et al, 2005). Hollister and Montague (2005) recommended that data-driven QT correction methods, including study population, individual and model-based, be used in the analysis of QT data in regular clinical studies and thorough QT studies.

Study population QT correction method

The most basic form of a data-driven QT correction method is a population-based QT correction specific to an individual study. A study population QT correction formula is computed from drug-free (off-treatment) ECG recordings using the methodology similar to the one utilized by Sagie et al (1992). A linear or log-linear regression model (other models could be used as well) is fitted to pooled off-treatment ECG recordings,

$$QT = a + bRR \text{ (linear regression),}$$

$$\log QT = c + d \log RR \text{ (log-linear regression),}$$

and the estimated slopes (b or d) are used to define linear or log-linear QT correction formulas,

$$QTc = QT + b(1 - RR) \text{ (linear population QT correction),}$$

$$QTc = QT / RR^d \text{ (log-linear population QT correction).}$$

Such population QT corrections rely on a number of assumptions, including the assumption of a constant QT-RR relationship across subjects and time points within each subject (Hollister and Montague, 2005), and tend to perform poorly if the conditions are not met.

correction method developed for this study (population-based correction). The treatment difference (90% CI) for this correction method were 8.3 (4.5, 12.2) msec, which suggests that the Fridericia correction likely underestimated the true QTc effect.

Study population QT corrections were utilized in several thorough QT studies as a secondary correction method; see, for instance, tolterodine QT study (Malhotra et al, 2007).

Individual QT correction method

Study population QT corrections assume that the QT-RR relationship is the same for all subjects enrolled in a thorough QT study and that a single model best describes this relationship across all subjects. However, there is increasing amount of evidence that the relationship varies substantially from individual to individual (Malik, 2001; Malik et al, 2002) and thus individual (subject-specific) QT correction formulas are likely to improve the assessment of QTc effects compared to a population-based QT correction. In fact, the individual QT correction method is recommended as “most suitable for the ‘thorough QT/QTc study’ and early clinical studies” in ICH E14 (Section 3.1.2).

Within the individual correction framework, a QT correction formula is derived for each subject using his or her off-treatment ECG recordings. For example, QT and RR measurements collected during four run-in placebo periods as well as placebo treatment period in a four-period cross-over design were used to compute a QT correction formula for each subject in the alfuzosin QT study (Extramiana et al, 2005)⁸.

Both linear and log-linear regression models (as well as other models) can be used to calculate individual QT corrections. This process is similar to the one used in the derivation of population-based QT corrections, i.e., QTc interval for each subject and each time point can be defined as

$$QTc = QT + b(1 - RR) \text{ (linear individual QT correction),}$$

$$QTc = QT / RR^d \text{ (log-linear individual QT correction),}$$

where b and d are the subject-specific correction factors (estimated slopes) obtained from linear and log-linear models, respectively.

The approach described above uses a fixed QT-RR regression model to derive individual QT correction formulas for all subjects. Alternatively, one can consider a more flexible individual correction method that selects the best regression model for each subject from several candidate models. Malik et al (2004) utilized this method to calculate individual QT corrections based on the best subject-specific models selected from 12 candidate models. The flexible individual QT correction method led to an improvement compared to the standard correction method (in terms of the precision of an estimated treatment effect and required sample size).

⁸ Further, a recent advance (Malik, 2008, personal communication) in the computation of the individual correction factor has involved the use of an average of multiple RR intervals surrounding an “individual” QT measurement (based on the standard “measurement” from a 10-second ECG [approximately 12 actual QRS-T complexes from 12 leads]). This can be accomplished by averaging RR intervals across all replicate ECGs collected around a time point of acquisition and using this common, average RR to correct replicate QT intervals. Most commonly, 3 to 7 RR intervals are averaged together and used to compute 3 to 7 distinct QTc values. More RR intervals can be employed if continuous 12-lead ECGs are used for data acquisition.

A sufficiently large number of ECG recordings per subject is typically required to compute reliable individual QT correction formulas, e.g., Hollister and Montague (2005) gave the following rule of thumb, “20 to 50 off-treatment observations per subject are needed for use of an individual correction, with more being better.” The actual number of recordings depends on a number of factors, including the range of heart rate values for each subject. If QT variability is low, the range of heart rates is fairly wide and heart rate values are distributed evenly over this range, a smaller number of recordings may be sufficient. However, in general, the individual QT correction approach becomes unreliable with less than 10 recordings per subject, especially if the test drug affects heart rate. A model-based approach can be applied to facilitate the estimation of individual QT-RR relationships and derivation of individual QT corrections with a limited number of ECG recordings per subject (Ma et al, 2008).

The individual QT correction approach was utilized as a primary or secondary correction method in all thorough QT studies.

Model-based QT correction method

Statistical modeling can improve the performance of popular QT correction methods by taking into account the correlations among repeated ECG recordings for each subject and drug-induced heart rate changes. For example, the model-based QT correction method proposed by Dmitrienko and Smith (2003) relies on modeling changes in QT interval as a function of changes in RR interval to determine a heart rate-independent estimate of the treatment difference (this estimate corresponds to the zero change in RR interval). This QT correction method enables the trial’s sponsor to perform a reliable analysis of QT changes independent of RR changes when the test drug affects heart rate. In addition, the method performs as well as (and sometimes better than) population QT corrections when no heart rate effect is present. Other model-based approaches to analyzing the QT-RR relationship were discussed by Shah and Hajian (2003).

To facilitate the comparison between the population and model-based QT correction methods, note that the population correction is reliable only over the range of RR values (or heart rate values) observed on off-treatment days. The population correction may be suspect and may not perform well when the test drug produces a marked change in heart rate and thus the RR range shifts to the left or the right on the treatment days compared to the off-treatment days. However, for the model-based QT correction method to be valid, one needs to assume that, as the RR range is shifted, the shape of the QT-RR relationship remains constant, i.e., the test drug does not affect the slope of the QT-RR relationship. It is generally accepted that the constancy of this relationship holds true, at least without significant alterations in the autonomic tone. However, it is theoretically possible that this relationship can be altered by the test drug and there is some evidence to support the possibility of a change over time (Morganroth et al, 1991; Extramiana et al, 1999). If this does occur, it may be impossible to conduct a valid thorough QT study because an altered QT-RR relationship may break the link between changes in QTc interval and changes in the ventricular repolarization rate. In particular, observed QTc changes may become an artifact of changes in the slope of the QT-RR relationship and lose their clinical relevance.

The model-based correction method developed by Dmitrienko and Smith (2003) was used as a key secondary correction method in the tadalafil QT study (Beasley et al, 2005) and duloxetine QT study (Zhang et al, 2007).

4.2. Comparison of test drug and positive control to placebo

4.2.1. Two-fold objective

Thorough QT studies are designed to evaluate QTc effects of a test drug, placebo and a positive control. The main objective of thorough QT studies is two-fold:

- Demonstrate that the test drug does not prolong QTc interval compared to placebo (non-inferiority test).
- Demonstrate that the study is capable of detecting prolongation of the QTc interval by a positive control compared to placebo (classical superiority test). This is referred to as the assay sensitivity analysis.

Patterson, Jones and Zariffa (2005) pointed out that other comparisons can be performed in thorough QT studies, for example, a direct comparison between the test drug and the positive control; however, it is uncommon to do this. A head-to-head comparison of this type was performed in the tadalafil QT study (Beasley et al, 2005) to demonstrate that the magnitude of QTc prolongation for the test drug was significantly lower than for the positive control (a “secondary” test of lack of effect).

The analysis corresponding to the first objective is clearly defined in ICH E14 (Section 2.2.4). The test drug is said to lack a QTc prolongation potential (be non-inferior to placebo) if the upper bound of the one-sided 95% confidence interval for the largest time-matched mean difference between the drug and placebo excludes 10 msec⁹ (see Section 2.3 for a discussion of the possible need to use more rigorous definitions of success in thorough QT studies). By contrast, the second objective is defined in a more ambiguous manner. It is stated in ICH E14 (Section 2.2.1) that assay sensitivity will be established if the mean difference between the positive control and placebo is close to 5 msec (this value is chosen because it represents the threshold of regulatory concern). The following interpretations of this definition have been proposed:

- Assay sensitivity will be established if the lower bound of the one-sided 95% confidence interval for the mean difference between the positive control and placebo is greater than a pre-specified threshold such as 5 msec at one or more post-dose time points. This criterion, by itself, does not appear to be consistent with the ICH E14 statement regarding the requirement for demonstrating assay sensitivity. In fact, with this only criterion it is easy to demonstrate assay sensitivity simply by selecting a positive control with a large QTc effect, which will defeat the purpose of including a positive control in a thorough QT study (Morganroth, 2005). The FDA at the present time emphasizes this approach to the assay sensitivity analysis.

⁹ This statement is commonly interpreted as follows: There is no evidence of drug-induced QTc effects if the upper bound of the one-sided 95% confidence interval is below 10 msec at all post-dose time points. Other interpretations/approaches to the analysis of the peak QTc effect are discussed in Section 4.2.2.

- Other regulatory bodies, e.g., Canadian regulatory authorities, rely on alternative interpretations. They are becoming concerned about the use of positive controls with excessively large QTc effects because they miss the 5-msec target that should be used in the assay sensitivity analysis. As a result, assay sensitivity is defined as the requirement to show that the positive control is statistically different from placebo at one or more post-dose time points *and* the mean difference between the positive control and placebo is close to 5 msec, e.g., it is less than 10 msec.

4.2.2. Multiplicity issues

The comparisons of the test drug and positive control to placebo are performed based on multiple post-dose ECG recordings, for example, the test drug was compared to placebo at 3, 4, 6 and 24 hours after dosing in the tadalafil QT study¹⁰ (Beasley et al, 2005). The multiple comparisons lead to a higher probability of incorrectly concluding that the test drug induces QTc prolongation (lower power) and a higher probability of incorrectly establishing assay sensitivity. These multiplicity issues are discussed below.

Comparison of the test drug to placebo

As was stated in Section 4.2.1, the popular interpretation of the definition of a negative thorough QT study relies on the so-called intersection-union test (Berger, 1982). Specifically, the comparison between the test drug and placebo requires no adjustment for multiplicity and thus standard one-sided 95% confidence intervals are used at all post-dose time points. In this case, the trial's sponsor encounters the effect known as reverse multiplicity (Offen et al, 2007). The probability of incorrectly detecting a significant QTc effect (failing to detect lack of effect) is higher than that associated with a single comparison. In fact, this probability increases (or, equivalently, the statistical power decreases) with the number of post-dose time points (Patterson et al, 2005).

As an illustration, consider a thorough QT study that employs a cross-over design and assume that the treatment difference is computed using Definition D2 introduced in Section 4.3. Further, suppose the study is powered at 95% when the true mean treatment difference is 5 msec. Using the sample size methodology developed in Zhang, Dmitrienko and Luta (2008), 35 subjects are required if the comparison between the test drug and placebo is performed at a single post-dose time point. If 5 and 10 post-dose time points are included in the study design, the sample size needs to be increased by 43% (50 subjects) and 60% (56 subjects), respectively.

Several authors discussed other interpretations of the definition of a negative thorough QT study, including those that account for multiplicity issues arising in the comparison of the test drug to placebo (Eaton et al, 2006; Boos et al, 2007). These issues can be addressed by employing a statistical method that accounts for the number of post-dose time points and correlations among the estimated treatment differences at these time points. The method results in an adjusted confidence interval for the largest time-matched treatment difference and, as a consequence, maintains the power of the study at the desired level. The width of adjusted interval depends on a number of factors. In

¹⁰ These analyses were performed along with a comparison based on the QTc difference associated with the individual highest plasma concentration at the post-dose time points.

general, the interval becomes narrower (and thus the power improves) as the number of time points is increased or weaker correlations among the treatment differences are observed at the post-dose time points¹¹.

To the best of our knowledge, the methodology based on multiplicity-adjusted confidence intervals has not yet been accepted by regulators. At this time, the only acceptable means of addressing this issue is to increase the sample size in order to achieve the desired power, based on the number of comparisons that can be made (number of post-dose time points at which ECGs are collected). However, we believe this is an important issue and further efforts will be required to find an appropriate solution to this multiplicity problem in thorough QT studies.

Comparison of the positive control to placebo

The assay sensitivity analysis (comparison between a positive control and placebo) leads to a different type of a multiplicity problem. As was explained earlier in this section that assay sensitivity is usually defined in terms of a statistically significant difference between the positive control and placebo at one or more post-dose time points. Due to the multiple comparisons, the probability of demonstrating assay sensitivity is inflated. In order to avoid the inflation, the clinical trial sponsor can consider the following two options:

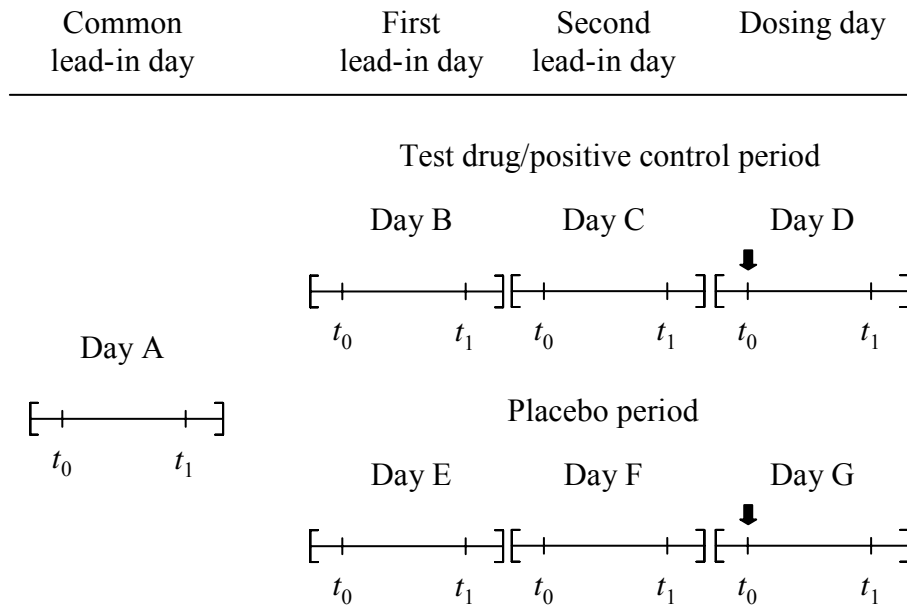
- Perform the assay sensitivity analysis at a fewer post-dose time points. Since the effect of a positive control on QTc interval is generally well understood, it is reasonable to restrict the control-placebo comparisons to the time points when the QTc effect of the positive control is most pronounced. For example, if moxifloxacin 400 mg serves as the control, significant QT interval prolongation is likely to occur during the 2-4-hour window after the dose and the sponsor can consider excluding the post-dose ECG recordings collected after 10 hours post-dose from the assay sensitivity analysis.
- Perform a multiplicity adjustment. When performing this adjustment, it is important to utilize a multiple testing procedure that takes into account correlations among the estimated treatment differences at post-dose time points (e.g., resampling-based multiplicity adjustments, Westfall and Young, 1993). Basic multiple tests such as the Bonferroni test need to be avoided because they tend to be very conservative in multiplicity problems of this kind.

4.3. Definition of the treatment difference for QTc interval

The outcome of a thorough QT study depends heavily on the definition of the treatment difference between the test drug/positive control and placebo. An “optimal” definition of

¹¹ Using a cross-over thorough QT study with 4 periods and 9 post-dose time points as an example, Boos et al (2007) demonstrated that the adjusted confidence interval for the largest time-matched treatment difference is about 30% narrower than the commonly used naïve confidence interval. For instance, if the largest mean time-matched treatment difference between the test drug and placebo is 5.5 msec, the upper limit of the adjusted upper confidence limit obtained using four different methods proposed in this paper varies between 9.2 and 9.4 msec whereas the naïve upper confidence limit is 11.0 msec.

Figure 2. ECG collection points in a thorough QT study (the arrow indicates the time point when the test drug/positive control or placebo are administered).



the treatment difference is expected to minimize the standard error of treatment effect estimates (or, equivalently, minimize the width of confidence intervals for the true treatment effect) in order to improve the power of a thorough QT study. Additionally, this “optimal” definition will minimize the bias of treatment effect estimates (i.e., result in an observed point estimate that approaches the true mean difference).

To introduce several popular definitions of the treatment difference, consider a single-dose thorough QT study described in Section 3.3 with a cross-over design. The study includes a common lead-in day at the beginning of the study, two lead-in days at the beginning of the test drug, positive control and placebo periods and a single dosing day for each of the three experimental treatments (see Figure 2). Day A represents the common lead-in day at the beginning of the study. Days B and C are the lead-in days at the beginning of the test drug/positive control periods and, similarly, Days E and F are the lead-in days at the beginning of the placebo period. Days D and G are the dosing days (test drug/positive control and placebo, respectively). The test drug/positive control and placebo are administered at the time point t_0 on the dosing day. ECG recordings are taken immediately prior to administration of treatment at t_0 and at multiple post-dose time points on all lead-in and dosing days. For the sake of simplicity, only one post-dose time point (t_1) is shown in Figure 2 (in actual studies, ECG data will be collected at multiple post-dose time points t_1, \dots, t_n).

The following definitions of the treatment difference can be used in the thorough QT study described above¹². To simplify the notation, let A0 denote QTc interval at the pre-

¹² Similar treatment difference definitions can be utilized in multiple-dose designs described in Section 3.3. Note that it is important to use QTc intervals measured on the lead-in day (or the average of QTc intervals

dose time point t_0 on Day A, A_1 denote QTc interval at the post-dose time point t_1 on Day A, etc.

Single-delta definitions

- Definition S1. The analysis is based on the treatment difference D_1-G_1 .

Similar definitions can be constructed by introducing covariates, e.g., one can perform the analysis of QTc interval based on the treatment difference D_1-G_1 adjusted for:

- D_0 and G_0 .
- C_1 and F_1 .
- $(B_1+C_1)/2$ and $(E_1+F_1)/2$.
- A_1 .

Double-delta definitions

- Definition D1. The analysis is based on the treatment difference $(D_1-A_1)-(G_1-A_1)$ [note that it reduces to D_1-G_1 and is therefore equivalent to Definition S1].
- Definition D2. The analysis is based on the treatment difference $(D_1-D_0)-(G_1-G_0)$.
- Definition D3. The analysis is based on the treatment difference $(D_1-C_1)-(G_1-F_1)$.
- Definition D4. The analysis is based on the treatment difference $(D_1-(B_1+C_1)/2)-(G_1-(E_1+F_1)/2)$.

As in the case of single-delta definitions, one can consider treatment difference definitions adjusted for covariates.

Triple-delta definitions

- Definition T1. The analysis is based on the treatment difference $((D_1-D_0)-(C_1-C_0))-((G_1-G_0)-(F_1-F_0))$.
- Definition T2. The analysis is based on the treatment difference $((D_1-D_0)-((B_1+C_1)/2-(B_0+C_0)/2))-((G_1-G_0)-((E_1+F_1)/2-(E_0+F_0)/2))$.

Again, one can also introduce covariate-adjusted triple-delta definitions of the treatment difference.

Some of the definitions on this list have never been used in practice and others have attracted more attention; however, all of them may need to be examined when the trial's sponsor is assessing pros and cons of available options. The following definitions have been used in recently conducted thorough QT studies:

- Definition T1 was used in the tolterodine QT study (Malhotra et al, 2007); it was referred to as the placebo-subtracted time-matched change from baseline.

on two lead-in days) or pre-dose QTc interval on the first dosing day as the baseline values in multiple-dose designs. Pre-dose QTc interval on the last dosing day is not a valid baseline value.

- Definition T2 was used in the tadalafil QT study (Beasley et al, 2005).

Other definitions

An important property of the definitions of the treatment difference discussed so far is that they are robust to systematic changes in QTc interval, e.g., circadian changes in QTc duration. The definitions rely on time-matched comparisons and produce the same result regardless of the pattern of systematic circadian changes as long as the pattern is constant across subjects and study days.

If circadian changes are treated as random noise (i.e., QTc interval is not expected to change throughout the day in a predictable pattern), other definitions of the treatment difference can be considered. For example, if multiple time points of QTc acquisition are used on the lead-in days (e.g., B1-Bn, C1-Cn, E1-En, F1-Fn rather than just B1, C1, E1, F1), the corresponding QTc values can be averaged and the averages can be used to define the treatment difference. To illustrate, let $C1^*$ and $F1^*$ denote the average QTc values on Days C and F, then Definition D3 can be modified as follows:

- Definition D3*. The analysis is based on the treatment difference $(D1-C1^*)-(G1-F1^*)$.

It is clear that this approach moves away from the concept of time-matched comparisons inherent in Definition D3. The advantage of this alternative definition is that, under the assumption of random circadian changes, Definition D3* would provide richer signal averaging compared to Definition D3.

Comparison of treatment difference definitions

Although a comprehensive comparison of the definitions given above is yet to be performed, the performance of selected definitions has been evaluated in terms of the precision of the treatment effect estimate. Zhang, Dmitrienko and Luta (2008) computed the standard errors of the estimated treatment difference for Definitions S1, D2, D3 and T1 and utilized this information to perform sample size calculations in cross-over and parallel-group thorough QT studies. Using the QTc data collected in four thorough QT studies conducted at Eli Lilly and Company, Zhang et al showed that Definitions S1 and D2 are associated with the most precise estimates of the treatment difference. Definitions of the treatment difference that use a lead-in day at the beginning of each period (e.g., Definitions D3 and T1) tend to increase the variability of the treatment difference because a lead-in day essentially provides a double control.

In addition, as with selection of the number of replicates for signal averaging in Section 3.5, the choice of a treatment difference definition can be driven by other criteria, e.g., the bias of the resulting estimate of the treatment effect. Ideally, if a placebo is compared to a placebo in a thorough QT study, the point estimate of the treatment difference for an “optimal” definition should be 0 (assuming that other aspects of the study design and analytical procedures are optimized).

4.4. QTc-exposure analysis

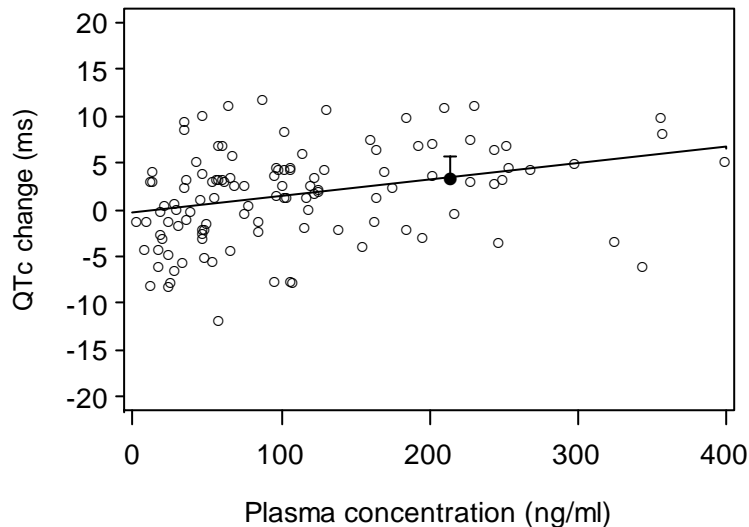
Although ICH E14 does not provide much information about QTc-exposure analysis¹³, it is commonly recognized that characterization of the exposure-response relationship for QTc changes is a key component of cardiac safety assessment. The exposure-response relationship is examined to understand QTc effects at exposure levels that are substantially higher than those associated with the therapeutic doses. To facilitate the evaluation of the exposure-response curve, PK samples need to be collected at the time of ECG recordings¹⁴.

The analysis of the relationship between drug exposure and changes in QTc duration is based on modelling a direct effect of the test drug's plasma concentration on QTc prolongation (separate analysis of the parent compound and metabolites may be required when metabolites are present). Specifically, changes in QTc interval are plotted against plasma concentrations and an appropriate regression model (linear or sigmoid) is used to estimate the slope of the exposure-response relationship and predict the maximum QTc effect. This effect is evaluated at the point corresponding to the mean (or median) subject-specific maximum plasma concentration. The mean treatment difference and upper bound of the associated one-sided 95% confidence interval are typically used to summarize the maximum effect of the test drug on QTc interval (Garnett et al, 2008). As an illustration, Figure 3 displays a QTc-concentration plot with the estimated mean treatment difference and one-sided 95% confidence limit.

¹³ It is stated in ICH E14 (Section 3.2.3) that this area is currently under active investigation.

¹⁴ Note that it is critical to draw blood samples after taking ECG recordings to avoid undesirable effects on ECG outcomes and their quality.

Figure 3. The mean treatment difference (3.4 ms) and one-sided 95% confidence limit (5.7 ms) evaluated at the mean subject-specific maximum plasma concentration (215 ng/ml).



The advantage of this approach to defining the maximum QTc effect is that, unlike the standard analysis of drug-related QTc changes described in Section 2.2.4 of ICH E14, it is not affected by reverse multiplicity (see Section 4.2.2) and takes into account exposure information. Although the QTc-exposure analysis is unlikely to be viewed as a definitive method for the assessment of drug-related QTc effects, it can serve as a reasonable supportive analysis with caveats discussed below.

An important underlying assumption for this analysis is that the QTc-concentration relationship is a direct one. In other words, the QTc effect is immediately associated with the plasma concentration and thus the maximum physiological effect (QTc prolongation) is achieved when the plasma concentration reaches its peak level. The maximum QTc effect may occur after the pharmacokinetic t_{\max} , due to a delay in the distribution from plasma to cardiac tissue. In this case, the QTc-concentration relationship has a lag-phase form (hysteresis of the QTc-concentration relationship is observed). Unless prior information is available to support the hypothesis of a direct effect, it appears unreasonable to rule out the possibility of a lag-phase QTc-concentration relationship. In fact, publications in this area suggest that drugs with a QTc prolongation potential are likely to exhibit a delayed QTc effect. Examples include

- Anti-arrhythmic drugs. Intravenous administration of dofetilide, a class III antiarrhythmic drug, was shown to produce a hysteresis of the concentration-QTc relationship (Le Coz et al, 1995).
- Antipsychotic drugs. A counter-clockwise hysteresis loop was observed with the maximum QTc effect delayed by 3 hours compared to the pharmacokinetic t_{\max} after a single oral dose of 50 mg of thioridazine (Salih et al, 2007).
- Antibiotics. Moxifloxacin, commonly used as a positive control in thorough QT studies, produces a 1-2-hour delay in the maximum QTc effect (Malhotra et al, 2007).

In order to account for a delayed QTc effect, an indirect-response approach based on an exposure-response model with a pharmacological effect compartment can be used (see, for example, Shi et al, 1995). This approach enables the trial sponsor to predict the concentration range associated with the target QT effect (Hollister and Montague, 2005). It is worth noting that further research is required to better characterize statistical properties of this approach (Type I error rate, robustness, etc).

Other important considerations that need to be taken into account when the QTc-exposure analysis is performed include:

- It is critical to account for circadian changes in QTc duration in the QTc-exposure analysis, e.g., define the treatment effect by comparing QTc changes in the test drug period (cross-over studies) or test drug arm (parallel-group studies) to time-matched QTc changes in the placebo period or arm, or explicitly model the circadian component (Piotrovsky, 2005).
- Since the parent compound and metabolites may have a synergistic effect on QTc interval (see, for example, Salih et al, 2007), QTc-exposure analysis can be performed for major metabolites.

Acknowledgements

The authors would like to thank Ms. Corina Loghin and Ms. Lu Zhang for their valuable comments.

References

- Bazett HC. An analysis of time relation of electrocardiograms. *Heart*. 1920; 20:174-195.
- Beasley CM, Mitchell MI, Dmitrienko AA, Emmick JT, Shen W, Costigan TM, Bedding AW, Turik MA, Bakhtyari A, Warner MR, Ruskin JN, Cantilena LR, Kloner RA. The combined use of ibutilide as an active control with intensive ECG sampling and signal averaging as a sensitive method to assess the effects of tadalafil on the human QT interval. *Journal of American College of Cardiology*. 2005; 46: 678-687.
- Benton RE, Sale M, Flockhart DA, Woosley RL. Greater quinidine-induced QTc interval prolongation in women. *Clinical Pharmacology and Therapeutics*. 2000; 67: 413-418.
- Berger RL. Multiparameter hypothesis testing and acceptance sampling. *Technometrics*. 1982; 24: 295-300.
- Boos D, Hoffman D, Kringle R, Zhang J. New confidence bounds for QT studies. *Statistics in Medicine*. 2007; 26: 3801-3817.
- Burke JH, Ehlert FA, Kruse JT, Parker MA, Goldberger JJ, Kadish AH. Gender-specific differences in the QT interval and the effect of autonomic tone and menstrual cycle in healthy adults. *American Journal of Cardiology*. 1997; 79: 178-181.
- Demolis J-L, Kubitzka D, Tenneze L et al. Effect of a single oral dose of moxifloxacin (400 mg and 800 mg) on ventricular repolarization in healthy subjects. *Clinical Pharmacology and Therapeutics*. 2000; 68: 658-666.

- Drici M-D, Clement N. Is gender a risk factor for adverse drug reactions? *Drug Safety*. 2001; 24: 575-585.
- Drici M-D, Knollmann BC, Wang W-X, Woosley RL. Cardiac actions of erythromycin. *Journal of the American Medical Association*. 1998; 280: 1774-1776.
- Dmitrienko A, Smith B. Analysis of QT interval in clinical trials. *Drug Information Journal*. 2002; 36: 269-279.
- Dmitrienko A, Smith B. Repeated-measures models in the analysis of QT interval. *Pharmaceutical Statistics*. 2003; 2: 175-190.
- Dmitrienko A, Sides G, Winters K, Kovacs R, Eisenberg P, Groh W. Electrocardiogram reference ranges derived from a standardized clinical trial population. *Drug Information Journal*. 2005; 39: 395-406.
- Eaton ML, Muirhead RJ, Mancuso JY, Lolluri S. A confidence interval for the maximal mean QT interval change due drug effect. *Drug Information Journal*. 2006; 40: 267-271.
- Ebert SN, Liu X-K, Woosley RL. Female gender as a risk factor for drug-induced cardiac arrhythmias: evaluation of clinical and experimental evidence. *Journal of Women's Health*. 1998; 7:547-557.
- El-Eraky H, Thomas SHL. Effects of sex on the pharmacokinetic and pharmacodynamic properties of quinidine. *British Journal of Clinical Pharmacology*. 2003; 56: 198-204.
- Extramiana F, Maison-Blanche P, Badilini F, Pinoteau J, Deseo T, Coumel P. Circadian modulation of QT rate dependence in healthy volunteers. *Journal of Electrocardiology*. 1999; 32: 33-43.
- Extramiana F, Maison-Blanche P, Cabanis MJ, Ortemann-Renon C, Beaufile P, Leenhardt A. Clinical assessment of drug-induced QT prolongation in association with heart rate changes. *Clinical Pharmacology and Therapeutics*. 2005; 77: 247-258.
- Fridericia LS. Die Systolendauer im Elektrokardiogramm bei normalen Menschen und bei Herzkranken. *Acta Medica Scandinavia*. 1920; 53: 469-486 [English translation: Fridericia LS. The duration of systole in an electrocardiogram in normal humans and in patients with heart disease. *Annals of Noninvasive Electrocardiology*. 2003; 8: 343-351]
- Garnett CE, Beasley N, Bhattaram VA, Jadhav P, Madabushi R, Stockbridge N, Tornoe CW, Wang Y, Zhu H, Gobburu JV. Concentration-QTc relationships play a key role in the evaluation of pro-arrhythmic risk during regulatory review. *The Journal of Clinical Pharmacology*. 2008; 48: 13-18.
- Hollister AS, Montague TH. Statistical analysis plans for ECG Data: Controlling the intrinsic and extrinsic variability in QT data. *Cardiac Safety of Noncardiac Drugs*. Morganroth J, Gussak I (Editors). Humana Press, 2005: 239-257.
- Hulot J-S, Demolis J-L, Riviere R, Strabach S, Christin-Maitre S, Funck-Brentano C. Influence of endogenous oestrogens on QT interval duration. *European Heart Journal*. 2003; 24:1663-1667.

- ICH E14. Clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs. 2005.
- Johansson M, Carlsson L. Female gender does not influence the magnitude of ibutilide-induced repolarization delay and incidence of torsades de pointes in an in vivo rabbit model of acquired long QT syndrome. *Journal of Cardiovascular Pharmacology and Therapeutics*. 2001; 6: 247-254.
- Kadish AH, Greenland P, Limacher MC, Frishman WH, Daugherty SA, Schwartz JB. Estrogen and progestin use and the QT interval in postmenopausal women. *A.N.E.* 2004; 9: 366-374.
- Le Coz F, Funck-Brentano C, Morell T, Ghadanfar MM, Jaillon P. Pharmacokinetic and pharmacodynamic modeling of the effects of oral and intravenous administrations of dofetilide on ventricular repolarization. *Clinical Pharmacology and Therapeutics*. 1995; 57: 533-542.
- Lepeschkin E, Surawicz B. The measurement of the Q-T interval of the electrocardiogram. *Circulation*. 1952; 6: 378-388.
- Lu HR, Marien R, Saels A, De Clerk F, Are there sex-specific differences in ventricular repolarization or in drug induced early afterdepolarizations in isolated rabbit purkinje fibers? *Journal of Cardiovascular Pharmacology*. 2000; 36: 132-139.
- Lu HR, Remeysen P, Somers K, Saels A, De Clerk F. Female gender is a risk factor for drug-induced long QT and cardiac arrhythmias in an in vivo rabbit model. *Journal of Cardiovascular Electrophysiology*. 2001; 12: 538-545.
- Ma H, Smith B, Dmitrienko A. Statistical analysis methods for QT/QTc prolongation. *Journal of Biopharmaceutical Statistics*. 2008; 18: 553-563.
- Malhotra BK, Glue1 P, Sweeney K, Anziano R, Mancuso J, Wicker P. Thorough QT study with recommended and suprathreshold doses of tolterodine. *Clinical Pharmacology and Therapeutics*. 2007; 81: 377-385.
- Malik M, Camm AJ. Evaluation of drug-induced QT interval prolongation: Implications for drug approval and labelling. *Drug Safety*. 2001; 24: 323-351.
- Malik M. Problems of heart rate corrections in assessment of drug-induced QT interval prolongation. *Journal of Cardiovascular Electrophysiology*. 2001; 12: 411-420.
- Malik M, Farbom P, Batchvarov V, Hnatkova K, Camm AJ. Relation between QT and RR intervals is highly individual among healthy subjects: implications for heart rate correction of the QT interval. *Heart*. 2002; 87: 220-228.
- Malik M, Hnatkova K, Batchvarov V, Gang Y, Smetana P, Camm J. Sample size, power calculations, and their implications for the cost of thorough studies of drug induced QT interval prolongation. *Pacing and Clinical Electrophysiology*. 2004; 27: 1659-1669.
- Morganroth J, Brozovich FV, McDonald JT, Jacobs RA. Variability of the QT measurement in healthy men, with implications for selection of an abnormal QT value to predict drug toxicity and proarrhythmia. *American Journal of Cardiology*. 1991; 67: 774-776.

- Morganroth J, Ilson BE, Shaddinger BC, Dabiri GA, Patel BR, Boyle DA, Sethuraman VS, Montague TH. Evaluation of vardenafil and sildenafil on cardiac repolarization. *American Journal of Cardiology*. 2004; 93:1378-1383.
- Morganroth J. Design and conduct of the thorough Phase I ECG trial for new bioactive drugs. *Cardiac Safety of Noncardiac Drugs*. Morganroth J, Gussak I (Editors). Humana Press, 2005: 205-222.
- Nagy D, DeMeersman R, Gallagher D, Pietrobelli A, Zion AS, Daly D, Heymsfield SB. QTc interval (cardiac repolarization): lengthening after meals. *Obesity Research*. 1997; 5: 531-537.
- Nakagawa M, Ooie T, Takahashi N, Taniguchi Y, Anan F, Yonemochi H, Saikawa T. Influence of menstrual cycle on QT interval dynamics. *PACE*. 2006; 29: 607-613.
- Offen W, Chuang-Stein C, Dmitrienko A, Littman G, Maca J, Meyerson L, Muirhead R, Stryszak P, Boddy A, Chen K, Copley-Merriman K, Dere W, Givens S, Hall D, Henry D, Jackson JD, Krishen A, Liu T, Ryder S, Sankoh AJ, Wang J, Yeh CH. Multiple co-primary endpoints: Medical and statistical solutions. A report from the Multiple Endpoints Expert Team of the Pharmaceutical Research and Manufacturers of America. *Drug Information Journal*. 2007; 41: 31-46.
- Patterson S, Agin M, Anziano R, Burgess T, Chuang-Stein C, Dmitrienko A, Ferber G, Geraldes M, Ghosh K, Menton R, Natarajan J, Offen W, Saoud J, Smith B, Suresh R, Zariffa N. 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*. 2005; 39: 243-266.
- Patterson S, Jones B, Zariffa N. Modeling and interpreting QTc prolongation in clinical pharmacology studies. *Drug Information Journal*. 2005; 39: 437-445.
- Piotrovsky V. Pharmacokinetic-pharmacodynamic modeling in the data analysis and interpretation of drug-induced QT/QTc prolongation. *American Association of Pharmaceutical Scientists Journal*. 2005; 7: E609-E624.
- Ring A, Koenen-Bergmann M, Ritzhaupt A, Gehlhar B, Platz J. Some properties of different analyses in thorough QT studies compared using a QT study of tiotropium. Poster at Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie 2006.
- Rodriguez I, Kilboron MJ, Liu X-K, Pezzullo JC, Woosley RL. Drug-induced QT prolongation in women during the menstrual cycle. *Journal of the American Medical Association*. 2001; 285: 1322-1326.
- Sagie A, Larson MG, Goldberg RJ, Bengtson JR, Levy D. An improved method for adjusting the QT interval for Heart Rate (the Framingham Heart Study). *American Journal of Cardiology*. 1992; 70: 797-801.
- Salih ISM, Thanacoody RHK, McKay GA, Thomas SHL. Comparison of the effects of thioridazine and mesoridazine on the QT interval in healthy adults after single oral doses. *Clinical Pharmacology and Therapeutics*. 2007; 82: 548-554.

Serra DB, Affrime MB, Bedigian MP, Greig G, Milosavljev S, Skerjanec A, Wang Y. QT and QTc interval with standard and suprathapeutic doses of darifenacin, a muscarinic M3 selective receptor antagonist for the treatment of overactive bladder. *Journal of Clinical Pharmacology*. 2005; 45: 1038-1047.

Shah A, Hajian G. A maximum likelihood approach for estimating the QT correction factor using mixed effects model. *Statistics in Medicine*. 2003; 22: 1901-1909.

Shi J, Lasser T, Koziol T, Hinderling PH. Kinetics and dynamics of sematilide. *Therapeutic drug monitoring*. 1995; 17: 437-444.

Shin J-G, Kang W-k, Shon J-H, Arefayene M, Yoon Y-R, Kim K-A, Kim D-I, Kim D-S, Cho K-H, Woosley RL, Flockhart DA. Possible interethnic differences in quinidine-induced QT prolongation between healthy Caucasian and Korean subjects. *British Journal of Clinical Pharmacology*. 2006; 1-10.

Westfall PH, Young SS. *Resampling-based multiple testing: examples and methods for p-value adjustment*. Wiley, New York, 1993.

Williams EJ. Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research, Series A*. 1949; 2: 149-168.

Zhang L, Chappell J, Gonzales C, Small D, Knadler M, Callaghan JT, Francis JL, Desai D, Leibowitz M, Ereshefsky L, Hoelscher D, Leese PT, Derby M. QT effects of duloxetine at suprathapeutic doses: A placebo and positive controlled study. *Journal of Cardiovascular Pharmacology*. 2007; 49: 146-153.

Zhang L, Smith B. Sex differences in QT interval variability and implication on sample size of thorough QT studies. *Drug Information Journal*. 2007; 5: 619-628.

Zhang L, Dmitrienko A, Luta G. Sample size calculations in thorough QT studies. *Journal of Biopharmaceutical Statistics*. 2008; 18: 468-482.