14 Statistical Analysis Plans for ECG Data

Controlling the Intrinsic and Extrinsic Variability in QT Data

Alan S. Hollister, MD, PhD
and Timothy H Montague

CONTENTS

INTRODUCTION

The safety and regulatory needs to detect small drug-induced changes in the QT interval have created many challenges for the design and analysis of "thorough" QT studies. The measurement techniques available, the correlation between the RR interval and the QT interval, and the high variability in the QT interval have made the detection of changes in the QT interval difficult, and the verification of a lack of an effect on the QT interval even more difficult. The purpose of this chapter is to provide statistical and empirical rationales for key elements of study design, and statistical analysis that will control for sources of QT variability and will enhance study sensitivity. We will identify study design and statistical techniques to reduce QT variability, discuss the assumptions inherent in many of the choices available in study design, and recommend study designs based on these principles.

The QT interval, and its heart rate corrected value (QTc) varies widely throughout the day in normal individuals with reports ranging from 76 ms to 117 ms *(1,2)*. Numerous groups have reported the influence of meals, sleep, age, autonomic tone or balance,

From: *Cardiac Safety of Noncardiac Drugs:*
Practical Guidelines for Clinical Research and Drug Development
Edited by: J. Morganroth and I. Gussak © Humana Press Inc., Totowa, NJ

gender, body position, electrolyte abnormalities, exercise, and insulin levels, as well as the effects of drugs, disease, and genetic abnormalities on the QT interval *(2–11)*. Despite study designs that control for many of these factors, several factors may change during the course of a study, and a time or sequence effect may be present in both the QT variability as well as the absolute value of the QT. It is also important to understand that it is not known whether the diurnal variation in the QT is a pattern or rhythm that is reproducible day to day within individuals. In general, minute-to-minute variations in the QT interval are less than day-to-day variations, and much less than week-to-week variations *(12)*.

The sensitivity of a study is dependent on the ratio of the change in the QT to the variability of that change. The magnitude of the change in QT is dependent on the drug, dose, and occasionally the study population, but is often limited by drug tolerability and/or safety. Thus, study designs that minimize the variability of the QT and reliably measure the QT at the time of maximal drug-induced changes will most efficiently detect the change, and will require the smallest sample size.

STATISTICAL LIMITATION OF THE QT VARIABILITY

The simplest and easiest technique for reducing the minute-to-minute variability in the QT is to standardize which complexes are measured within an electrocardiogram (ECG), to measure two or more complexes, and to average values from two or more ECGs. Typically, the QT interval is measured on three consecutive complexes from the same ECG lead during a period of stable heart rate and rhythm, and the QT and its corrected value averaged from the three complexes. Duplicate or triplicate ECGs are obtained at 1-to-5 min intervals and the values of these are all averaged to estimate the QT and QTc values. In this fashion, the minute-to-minute variability in the QT is reduced by the square root of the number of complexes and the number of ECGs measured. The effect of this technique on reducing the variability (standard deviation) of the change in QT, as well as on the sample size, is shown in Fig. 1.

The bars represent the standard deviation for the change in QTc from a model data set where the true difference between baseline and on-drug QT values is 10 ms and the within-subject standard deviations at both baseline and on-drug are 15 ms. Note how the standard deviation decreases in an exponential fashion as the number of replicate ECGs increase. The numbers above the bars indicate the number of subjects required for an 80% probability of finding a $p < 0.05$ difference between baseline and on-drug QTc by the *t*-test. Multiple iterations of this model found a statistically significant reduction in the standard deviation between one and two ECGs per time point, and a $p = 0.07$ reduction between two and three replicate ECGs per time point. Thus, the QT and QTc averaged from two or more ECGs per time point is a simple and effective method for increasing study sensitivity and reducing sample size.

CHOICE OF BASELINE

The baseline against which the effect of drug is to be compared needs to be chosen carefully. Factors to consider in this choice are listed previously and include the time of day, meals, period of awakening, familiarity of the experimental surroundings, and the time interval between the baseline and on-drug measurements. The assumption present in the choice of baseline is that it should neither increase nor decrease the magnitude of

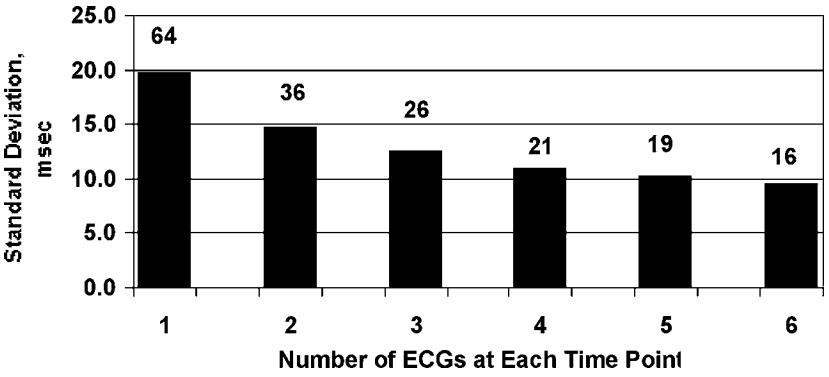

Fig. 1. Effect of the number of replicate ECGs per time point on the standard deviation of the change in QTc.

the QT change, and it should minimize the variance of the estimate of the QT change. There is some debate about what is the "best" baseline: a predose value of the QT or an on-placebo, time-matched QT value. The predose baseline may be collected immediately before dosing, or may be an average of baselines over several periods (as proposed in the original Canadian/FDA draft of the ICH document). The on-placebo baseline may be the QT obtained at the same time of day as the on-drug measurement (i.e., time-matched), or an average of QTs over the course of multiple measurements on a placebo day. Each of these options makes one or more assumptions about QT variability:

1. Predose QT as baseline: assumes time proximity to the drug-induced change minimizes the variability, no diurnal pattern of QT changes, no study protocol-induced QT changes.
2. Multiple predose baselines over several days or periods: assumes no difference in the absolute QT value or its variance between days (or weeks) compared to within day, no sequence effect, no difference in QT or its variability caused by diurnal changes.
3. Same time of day on placebo day: assumes a stable, reliable pattern in QT with time of day and study conditions, no sequence effect, stable day-to-day variability, stable QT–RR relationship between the placebo day and the active drug day, and no effect of time interval on magnitude of QT change.
4. Multiple ECGs averaged over a period of time on a baseline/placebo day: assumes stable and consistent QT variability from day to day, with limited within day variability.

In the experience of the authors, the within subject moment-to-moment QT variability (average SD = 6–9 ms) is slightly less than within day variability (SD = 9–10) which is less than day-to-day variability (SD = 9–13) and less than between week variability (SD = 10–15). A sequence effect has been noted in some studies, and the time-matched placebo as baseline has been criticized for a lack of sensitivity and reliability *(13)*. In addition, the apparent magnitude of the QT change increases with the interval between the baseline and on-drug measurements. Figure 2 illustrates the effect of time between baseline and on-drug measurement of the Fridericia's corrected QT (QTcF) change caused by a single 400 mg oral dose of moxifloxacin (data taken from three studies reported in the 4/27/01 Summary Basis for Approval for moxifloxacin *[14]*).

Figure 2 illustrates the change in QTcF at the time of drug Cmax when the time between baseline and on-drug measurements was 0 (predose on the same day as drug dosing), 1 d, 1 to 3 wk, and 1 to 5 wk. The diamond symbols are data from single ECG

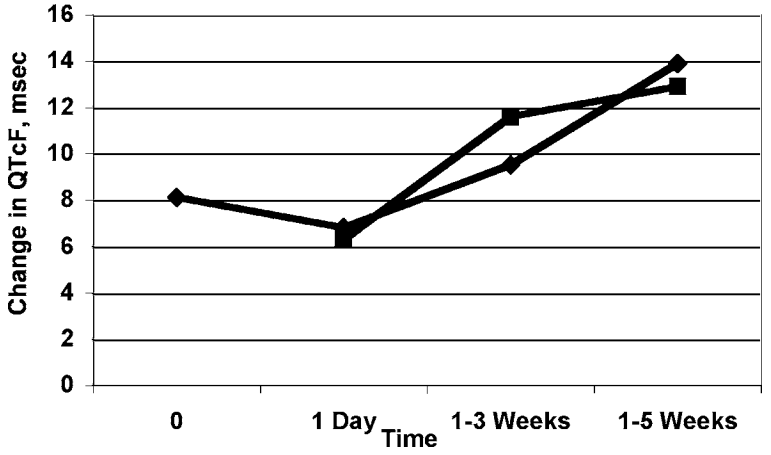

Fig. 2. Effect of time between baseline and on-drug measurement of the change in QTcF at Cmax after oral doses of 400 mg moxifloxacin.

determinations of baseline, and the squares are data from multiple ECG determinations of baseline. These results suggest that the greater the interval between the baseline and on-drug measurements of QTc, the greater the apparent effect of the drug. This phenomenon was observed in three studies designed *a priori* to test for this effect, and warrant consideration in the design of crossover studies or in parallel design studies requiring longer periods between predose baseline and drug steadystate concentration levels.

A similar effect was observed in the variability of the change in QTc in these moxifloxacin studies. Figure 3 illustrates the standard deviation for the change in QTcF at Cmax for moxifloxacin plotted against the time between the baseline and on-drug measurements. The diamond symbols are data from single determinations of the baseline QT, and the squares display data from multiple ECG determinations of baseline. Although the differences between single and multiple ECG measurements is clear, there is also a trend for greater variability in the group standard deviation when the time between baseline and on-drug measurements increase. This increase in variability as the time between baseline and on-drug measurements increases will impact the sample size (number of subjects) necessary to detect a drug effect in longer term studies.

In summary, whether the drug effect is expressed as a change from baseline or a change from placebo, there are multiple assumptions inherent in the choice. For data transparency, display of data as a change from the predose baseline for both drug and placebo will enhance the understanding of the drug effect as well as effects related to experimental conditions. As indicated by the experience with moxifloxacin, the closer in time the baseline is to the on-drug measurements of QT, the lower the magnitude of QT change, and the lower the variance of the change. There appears to be little difference between same day and preceding day baselines, but longer intervals have the potential to falsely elevate the magnitude of QT effect, increase the variability of the estimate of this effect, and will require larger sample sizes. This will be of particular concern in the design of "thorough" QT studies with drugs or metabolites that have long half-lives, or that require titration to reach the tested dose. Parallel designs that incorporate assessment of the interval and sequence effect may be the best approach for evaluation of these drugs.

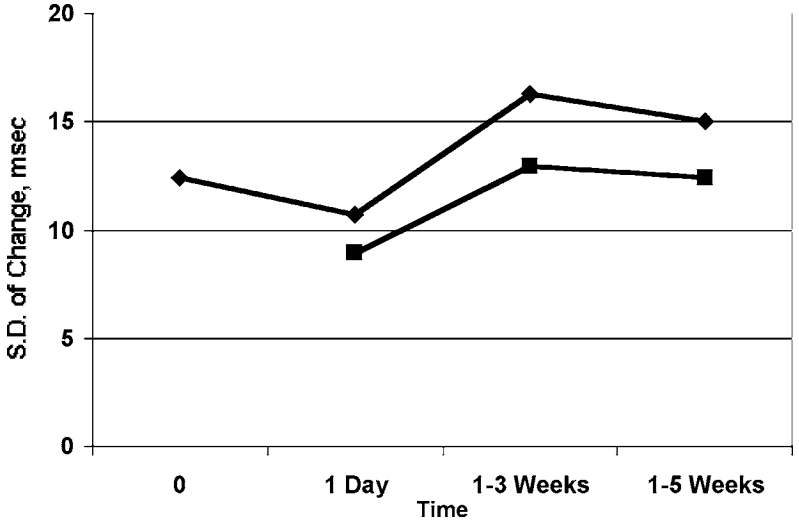

Fig. 3. Effect of time between baseline and on-drug measurement of QTcF on the standard deviation of the change in QTcF.

QT CORRECTION METHODS

The RR and QT intervals are highly correlated, with the QT interval increasing with increasing RR interval. As such, an observed increase in the absolute QT interval could be the results of changes in the RR interval rather than a drug effect. Several methods have been proposed to "correct" the QT interval with respect to the RR interval, such that the corrected QT interval (QTc) is independent of the RR interval *(15)*.

Figure 4 illustrates the relationship between the RR interval (*x*-axis) and the QT interval (*y*-axis), showing a clear trend for the QT interval to increase with increasing values of the RR interval. An appropriate correction method should show no trend in the data when the corrected QT (QTc) is plotted vs the RR interval.

All methods are based on defining the RR–QT relationship, and then standardizing the QT interval around an RR value of 1 s (equivalent to a heart rate of 60 bpm). In this chapter, "population" and "individual" correction methods will be reviewed, as well as a method that requires no correction. Finally, a brief overview of the use of Holter ECGs and their analyses will be provided.

Population corrections are the most common and historically used methods. The oldest of these is Bazett's, $QTcB = QT/(RR^{1/2})$, where RR is in seconds and QT in milliseconds. Another common correction, is Fridericia's, $QTcF = QT/(RR^{1/3})$. These methods assume a log-linear QT–RR relationship. The problem with these "fixed" corrections is that if the actual QT–RR relationship differs from the fixed relationship, then the estimate of treatment effects will be biased. In the case of Bazett's correction, it is widely recognized that Bazett's over-corrects for the RR interval at higher heart rates, resulting in an increase in false positive effects. Fridericia's typically performs a bit better, but also is susceptible to both over- and under-correcting, leading to both false positive and negative conclusions, respectively.

Figure 5 illustrates that there is an inverse relationship between the RR interval (*x*-axis) and the Bazett's corrected QT (QTcB) interval (*y*-axis). The QTcB interval decreases

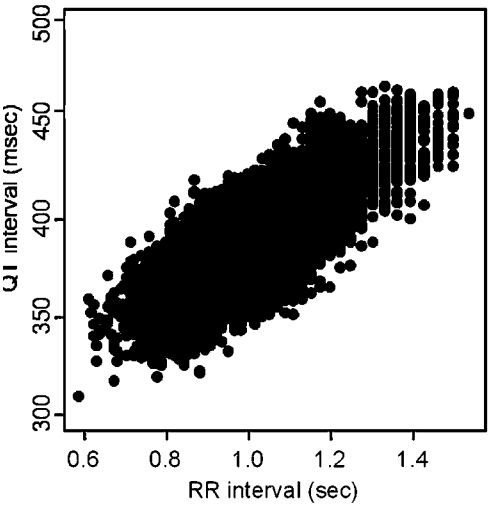

Fig. 4. QT–RR relationship.

Table 1
Linear and Log-Linear Models of QT–RR Relationship and Corresponding Formulae for Corrected QT (QTc)

	Model QT	*Calculation of QTc*
Linear	$QT = \alpha + \beta RR$	$QTc = QT + \beta(1–RR)$
Log–linear	$\log(QT) = \alpha + \beta\log(RR)$	$QTc = QT/(RR^{\beta})$

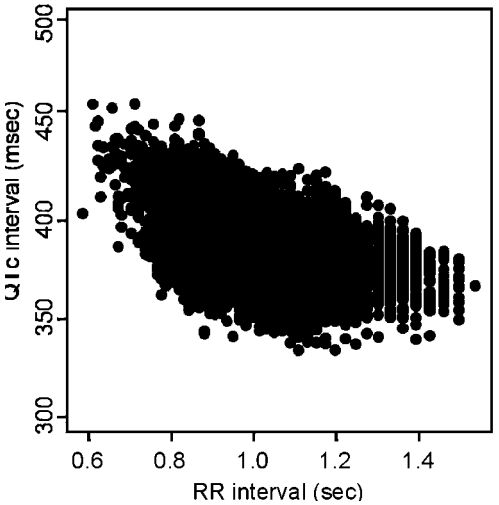

Fig. 5. Example of overcorrecting (should be no trend in QTc).

with increasing values of the RR interval, indicating that Bazett's correction over-corrected for the RR interval.

An alternative to the fixed population methods is to define the QT-RR relationship based on the observed study data. It is recommended that one use off-treatment data (baseline or baseline + on-placebo data). The first step is to model QT as a function of RR (linear or log–linear) and estimate the slope parameter, β. Then QTc is calculated using this estimate of the slope parameter.

In Table 1, the second column shows the model for the QT-RR relationship. The third column shows the formulae for calculating the corrected QT interval (QTc) based on the estimated parameters from the model for the QT-RR relationship. Note, for Bazett's and Fridericia's corrections, the slope parameter, β, equals 0.5 and 0.33, respectively, for a log-linear model.

The limitations of the population corrections are that they require the following assumptions:

1. stable and constant QT–RR relationship across subjects
2. stable and constant QT–RR relationship across time, days, and/or sessions
3. stable and constant QT–RR relationship across treatments

There are data to suggest that the QT–RR relationship varies from subject to subject and varies over time (when awakening, within a day, across days, weeks, months) *(2,4,6,7,16,17)*. Additionally, the QT–RR relationship may be altered by external factors such as autonomic balance, drugs, and other external factors, as reviewed earlier.

Individual correction methods relax the assumption about stable and constant relationships across subjects, by determining a unique relationship for each individual subject. The only assumption across subjects is that the form of the relationship (linear or log-linear) is the same. The correction is similar to the population approach, but a unique slope parameter, β, is calculated for each subject. Thus, only assumptions 2 and 3 from above are made. As with a population correction, it is recommended that individual corrections be based on off-treatment data.

The limitations of the individual correction are the need for a sufficient number of observations and a sufficient range of RR intervals. If either or both are insufficient, the QT–RR relationship may be poorly defined, adding both bias and variability to treatment effects. Figure 6 provides two such scenarios. Figure 6 illustrates two examples of insufficient data for determining an individual correction. Off-treatment RR interval (*x*-axis) vs QT intervals are plotted with the dashed line representing a "best" fit log–linear model. In the figure on the left, the data are clustered around a small range of RR values. In this case, one could imagine that any line would have provided a reasonably good fit to the data. In addition, one would not have much confidence in the modeled QT–RR relationship for an RR interval greater than 1.2 s. The figure on the right illustrates that a single observation (at RR 1.1 s) can impact the slope of the curve for sparse data. If that point was not there, the slope of the curve might be quite different, leading to a different set of corrected QT values.

The authors recommend that 20 to 50 off-treatment observations per subject are needed for use of an individual correction, with more being better. For a crossover design this should not be an issue, but it may be a limitation for parallel group designs. QTs distributed over a sufficient range of RR intervals, approx 0.7 to 1.1 s (heart rate of 55 to 86), are necessary for an adequate estimate for each individual's correction. However, a well-controlled trial by design is going to limit the range of RR intervals for a subject, by

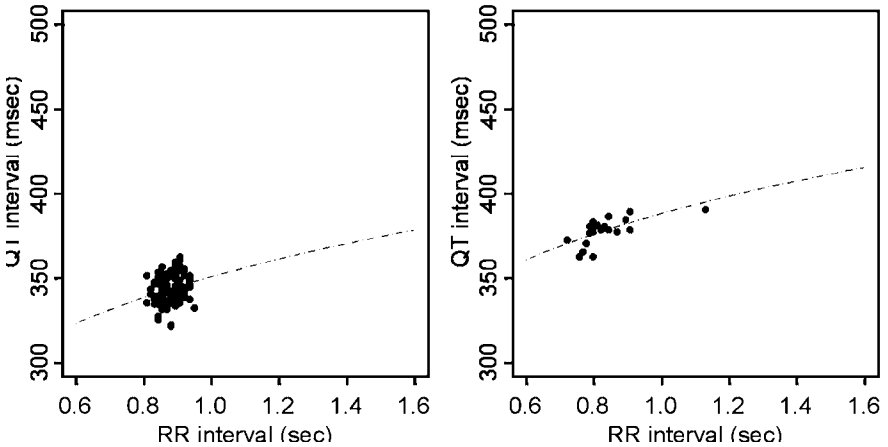

Fig. 6. Insufficient data for individual corrections.

controlling the external factors that may affect the RR interval. As such, care should be taken when using an individual approach.

As all correction methods are flawed in some manner, two statisticians have proposed using a repeated measures analysis that does not require a correction of the QT interval *(18)*. This is desirable as it avoids the potential for adding bias and variability to treatment effects caused by the correction method, and it can account for potential treatment effects on both the RR interval and the QT–RR relationship. The proposed method also allows for assumptions that the QT–RR relationship may vary over time and treatment, and takes into account the correlation between observations within a subject. As with population and individual correction methods, the QT interval is modeled as either a linear or log-linear function of the RR interval, with a unique set of parameters being determined for each sampling time (i.e., pre- and post-dose) and each treatment (i.e., placebo and each active dose). Treatment differences are then calculated for the QT interval based on the estimates of the model parameters for a given RR interval. For further description of this approach the reader is referred to the method description *(18)*. Graphically, this analysis is illustrated in Fig. 7. The placebo is represented by the circles and the experimental drug by the triangles. Open symbols are predose or baseline values and filled symbols are post-dose values. The treatment effect is the difference between experimental drug post-dose and predose values and placebo post-dose and predose values ([▲ - △] – [● - ○]).

As the QT–RR relationship is no longer restricted to be the same for each treatment and time point, treatment effects must be evaluated at a range of values for the RR interval. Specifically, if the QT–RR relationships are not parallel across treatments and time, then the treatment effect will vary with the value of the RR interval. In the left part of Fig. 7, the treatment effect increases with increasing RR interval. This may confound the study results, leading to a false negative conclusion. Because many QT-prolonging drugs change the QT–RR relationship and exhibit greater QT prolongation at slow heart rates ("reverse rate dependency"), the assumptions of this method may underestimate drug effects. When the method assumes parallel lines (i.e., assume stable and constant QT–RR relationship across time and treatment) illustrated in the right part of Fig. 7, the treatment effects are calculated by a linear combination of intercepts for the various

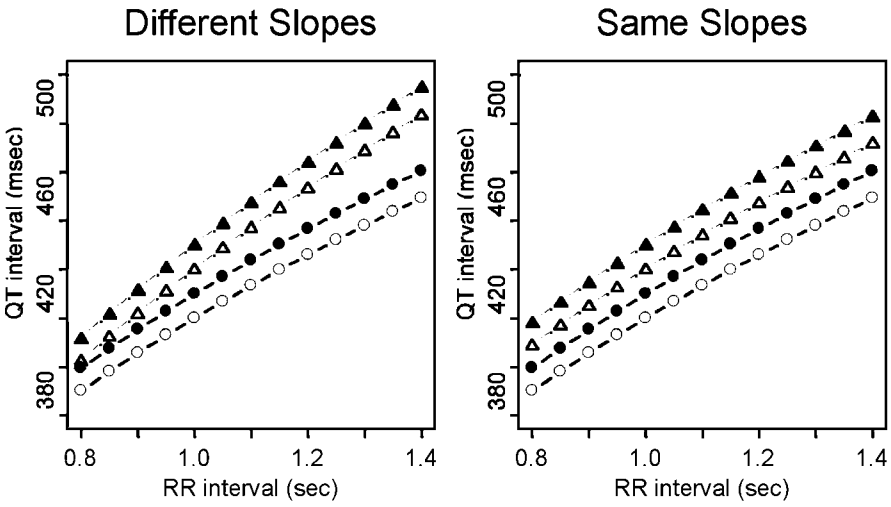

Fig. 7. QT analysis without heart rate correction: model options.

treatments and time points. By imposing parallel lines, this method provides results similar to data based on a population correction.

The draft ICH guidance document *(19)* recognizes that none of the discussed methods can be identified as "best" for all situations. As such, the guidance recommends and allows for multiple correction methods to be used, with Bazett's and Fridericia's being the standard methods. The authors recommend data-driven corrections (population or individual) or repeated measures analysis on uncorrected QT values.

One recently validated method to avoid the use of correction factors is the so-called "Holter bin" method *(20)*. This method uses continuous ECG recordings via a Holter monitor over a period of maximum pharmacodynamic effect of the drug. All of the PQRST complexes from 10 ms RR intervals ("bins") are averaged electronically and the resulting high fidelity trace is measured for the QT interval. This allows for the comparison of placebo and on-drug QT intervals at every heart rate recorded, generating the QT–RR relationship for both drug and placebo during the period of maximum drug effect. The drug and placebo QT can be compared at the same heart rate (e.g., an RR of 1000 ms = a HR of 60 bpm), across all heart rates recorded, from the RR bin where the greatest number of complexes were recorded for placebo and drug, and/or a regression of the QT–RR relationship. Because of the large number of complexes averaged within each RR bin, the within subject variability of the QT interval and its change is one-half to one-third that of replicate ECGs. This results in a large increase in the sensitivity for identifying a QT effect of a drug. The limitation of the "Holter bin" method is that recording must be performed over a period of time (2–4 h) covering the peak effect of the drug. This may dilute the maximum effect of a short half-life drug that exhibits a short-lived peak. The marked advantage of the Holter bin method is that it allows within-subject analysis via the repeated measures method using many more data points than can be obtained with ECGs. It also avoids the increases in variability caused by correction methods.

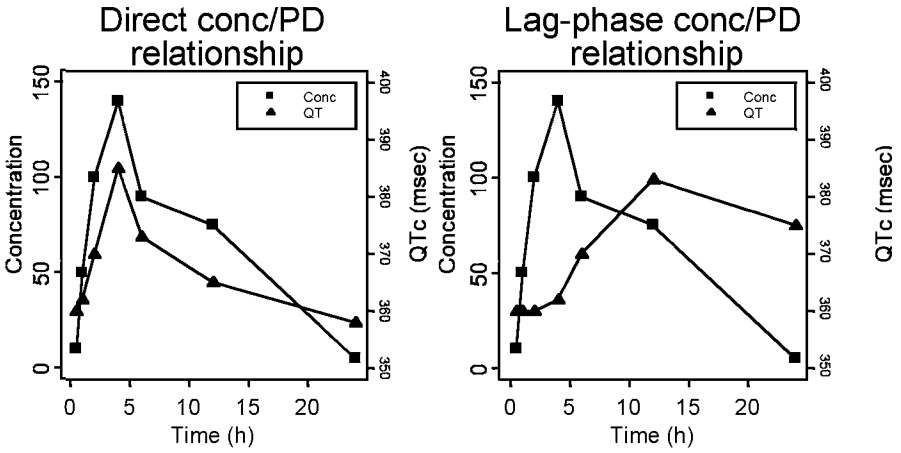

Fig. 8. Direct and lag-phase concentration QT relationships.

CENTRAL TENDENCY ANALYSIS

The draft ICH guidance recommends that the "thorough" QT trial to evaluate repolarization be designed to either detect a minimal mean effect or to "rule out" a mean effect. In either case, the mean effect can be defined by a summary measure of the time course of ECGs measured after each treatment. Possible summary measures or central tendency parameters include:

1. change at observed drug Cmax/Tmax (maximum plasma concentration and time to maximum concentration)
2. change at anticipated Tmax
3. maximum change regardless of drug concentration or time
4. average change over a specific period
5. area under the curve (or more properly, the area under the effect curve [AUEC]) of QT for a specific period.

As the QT, RR, and corrected QT (QTc) are highly correlated, all should be summarized in the same manner and analyzed at the same time points or periods chosen. The choice of the parameter depends on the experimental drug's pharmacokinetic (PK) characteristics (time and duration of Cmax, variability of Tmax, half-life) and its concentration/QTc relationship (direct or lag-phase). If the experimental drug has active metabolite(s), the PK and PK/QTc relationship should also be considered. For example, Fig. 8 illustrates a direct concentration/QT relationship (left) and lag-phase (or indirect) relationship (right). On the left, both the time profiles of the concentration data and the QT interval are similar, with peaks occurring approximately at the same time. While on the right, the time profiles are different, with the peak of the QT profile occurring several hours after the peak of the concentration profile. It is easy to see that using the change at Tmax for a compound with a lag-phase concentration–QT relationship could result in a false negative conclusion.

The change from baseline at observed Tmax or anticipated Tmax (option 1 or 2) is an appropriate parameter when the drug has the following characteristics: The concentration-QTc relationship is a direct one (i.e., no lag-phase as in the left figure); Tmax is well defined with low variability; the half-lives of the drug and its pharmacodynamic effect are fairly

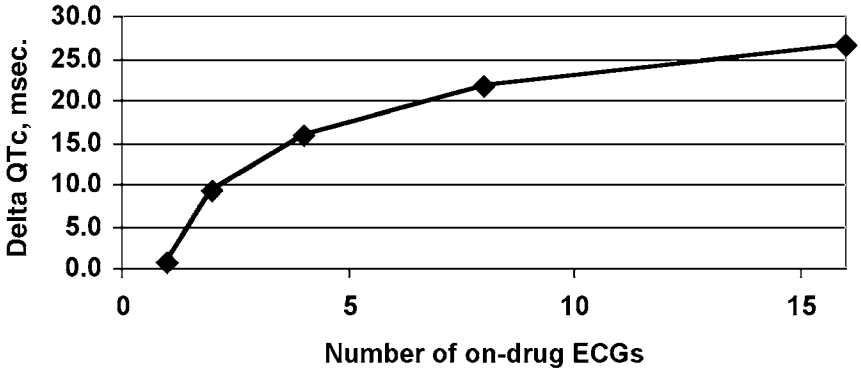

Fig. 9. False positivity magnitude with the maximum change in QTc.

short (< 24 h); and the drug either has no active metabolite or the active metabolite(s) have both a similar PK profile as the parent and direct concentration–QTc relationship.

The change from baseline at the anticipated Tmax is more easily implemented during a study than the change from baseline at the actual Tmax, especially for parallel group designs. For the change from baseline at observed Tmax, it is recommended that the placebo comparator be time-matched to reduce bias caused by diurnal variation. For a crossover trial, this would entail for each subject that the placebo comparator be the change from baseline to the same Tmax of the active treatment. Although parallel group designs cannot be analyzed in this manner, a Hodges-Lehmann-Moses non-parametric approach can be used to estimate a time-matched comparison between active and placebo treatments *(21)*. Again, the change at Tmax parameter is only valid when there is a good correspondence between the PK and the pharmacodynamic effect of the measured drug (the left part of Fig. 8).

Another concern with using the change at observed or anticipated Tmax is that the Cmax is typically the most variable PK parameter. As such, this variability will contribute to the variability of the QT parameters. This is especially important when designing parallel group designs. To help minimize variability, the study should be designed to ensure that experimental conditions are optimal at the time of maximum concentration. For example, one would not want to feed subjects within 2 h of the anticipated Tmax.

If the Tmax is highly variable, if there is a lag-phase in the concentration–QT relationship, and/or if active metabolites have a different PK profile than the parent, then the maximum change from baseline may be the appropriate measure of drug-induced QT change. For the QTc interval, the maximum change is just the maximum post-dose QTc value minus the baseline value, regardless of time and pharmacokinetics. For the QT and RR intervals, the authors recommend that the change from baseline be time-matched to the time at which the maximum QTc value occurred, so that one can assess and relate effects on QTc to those of the QT and RR interval appropriately. This parameter is dependent on both the sampling scheme for the ECG (number and timing) measurements and the experimental conditions (evaluation time, meals, sleep, etc.). Too few sampling time points may result in missing the maximal drug effect, whereas too many samples can increase the probability of spurious effect (false-positive). This is illustrated in Fig. 9, which demonstrates that the magnitude of the "maximum change in QTc" parameter is dependent on the number of ECGs obtained after drug administration, as a result of simple statistical variabil-

Table 2
Difference in Change From Baseline @ Tmax vs Maximum Change From Baseline

Drug	*Δ QTc @ Cmax (mean ± SE)*	*Tmax (h)*	*Max Δ QTc (mean ± SE)*	*Time of max QTc (h)*
X 1x	9 ± 15	2	24 ± 15	5.2
X 2x	20 ± 18	2.4	32 ± 17	4.3
Y 1x	5 ± 14	1.5	22 ± 14	5.5
Y 2x	7 ± 15	1.5	21 ± 12	5.3
Z	3 ± 14	1.3	21 ± 13	5.9
Placebo	–5 ± 15	—	16 ± 14	6.1

ity. The magnitude of this change may be reduced if replicate ECGs are obtained at each time point, but the risk of a falsely positive result is still present. The use of the maximum change in QTc parameter will consistently overestimate the magnitude of drug-induced QTc changes, and is only of value when changes can be compared to placebo data, or when hysteresis curve analysis indicates a disjunction between the pharmacodynamic effect (PD) and the PK of the measured drug. Additionally, the parameter is susceptible to diurnal variation with the sampling times chosen, another cause of false-positive results. As with the change at observed Tmax, it is recommended that the placebo comparator be time-matched to reduce bias caused by diurnal variation.

A potential pitfall of the maximum change from baseline is that, as a result of the high variability of the QT interval, often the maximum change from baseline does not occur at the same time as Tmax, even for drugs with a well defined Tmax of low variability, direct PK/PD relationship and no active metabolites. This is illustrated in the Table 2 of actual data taken from the moxifloxacin Summary Basis of Approval for three different drugs and placebo.

Table 2 demonstrates the large difference between the change in QTc at the time of Cmax compared to the maximum change in QTc at any time after drug administration. All three of these known QT-prolonging drugs (and even placebo) had their effects confounded by the parameter "maximum change in QTc." The time of Tmax and true drug QT effect occurred at about 2 h for each drug, whereas the time of the maximum change in QTc varied over the entire 12 h of data collection after drug administration. The consistent 12 to 20 ms difference between these two estimates of drug effect demonstrates the problems of false positivity with the maximum change in QTc parameter, and is completely predictable as in Fig. 9.

If the drug has a long pharmacokinetic or pharmacodynamic (QTc) half-life, then the average QTc change or the QTc area under the effect curve (AUEC) over a specific time period may be considered as the response parameters. Similar to the maximum change, the sampling scheme and experimental conditions may bias these parameters. Whereas these parameters are useful for evaluating whether a drug has a sustained effect over time, they have not been independently validated and as such are not considered sufficient for drug approval when used alone. Specifically, these parameters are not appropriate for drugs that have a direct concentration–QT relationship with a short half-life, as they may provide a false-negative result.

The statistical model for all of these central tendency parameters is similar. For a crossover design, the authors recommend a mixed effect model fitting a random term for subjects and fixed terms for sequence, period, and treatments. Additionally, baseline

Table 3
Clinical Interpretation of Mean Changes in QTc

Change from baseline	*Relative risk of TdP*
< 5 ms	So far no TdP
5–10 ms	No clear risk
10–20 ms	Uncertainty
> 20 ms	"Substantially" increased likelihood of being pro-arrhythmic

TdP, torsade des pointes.

Table 4
Null and Alternative Hypotheses for Two Statistical Approaches to Definitive QT Study

	To detect an effect	*To rule out an effect*
Null hypothesis	Ho: $\theta = 0$	Ho: $\theta \geq \delta$
Alternative hypothesis	Ha: $\theta \neq 0$	Ha: $\theta < \delta$

θ represents the difference between the experimental drug and placebo.
δ represents a clinical relevant difference, e.g., the change from baseline in Table 3.

(predose) values should be fit as a covariate to further reduce variability. For a parallel group design, an analysis of covariance (ANCOVA) is recommended, fitting a single fixed term for treatment and baseline as a covariate. As appropriate to the study design, fixed terms for other factors such as gender and age can also be included in either model.

INTERPRETATION OF THE OBSERVED CENTRAL TENDENCIES IN QTc

There are two aspects to interpreting the observed central tendencies in QTc of an experimental drug; statistical and clinical. The statistical interpretation is based on the statistical hypothesis being tested. For a "thorough" QT study, this will either be to detect a specific effect or to rule out a specific effect. The specific effects should be clinically meaningful changes in QTc.

The ranges and associated risks in Table 3 are from the draft ICH guidance and are based on clinical experience. It should be noted that there are other factors that may mitigate or enhance the risk of TdP. In addition, these ranges do not take into account the variability of the measurement, method of measurement, or correction factor. Finally, these are based on historical data using Bazett's corrected QT.

As mentioned earlier, the "thorough" QT study can be designed to test one of two hypotheses: to detect a specific difference or to rule out a specific difference. The mathematical expressions of these hypotheses are shown in Table 4.

The former is the more traditional statistical hypothesis (similar to that used to demonstrate a drug is superior to placebo in a pivotal trial). The latter is similar to that used for a bioequivalence trial or a noninferiority trial. The draft ICH document recommends that the study be designed to detect a 5 ms difference or to rule out a 5 or 7.5 ms difference. For both hypotheses, the study should be adequately powered. In general, to rule out an effect is a more stringent test and will require a larger sample size than that needed to detect a difference.

For either of these hypotheses, the comparison of interest should be between the change in QTc by the experimental drug vs the change in QTc on placebo. The draft ICH guidance does not indicate for which dose, therapeutic or supratherapeutic, of the experimental drug the hypotheses should be tested. If the hypotheses are to be tested for more than one dose, then the type I error rate, α, needs to be controlled for multiple comparisons.

For all comparisons of interest, $100(1-\alpha)\%$ confidence intervals, rather than *p* values should be employed to interpret the results, where α is the type I error rate. For the hypothesis to detect a difference, the confidence interval should not include 0. For the hypothesis to rule out an effect, the upper bound of the confidence interval should be less than the clinically relevant difference, δ. For comparisons for which no hypotheses are being tested, confidence intervals provide a range of plausible values. In addition, it is recommended that point estimates and $100(1-\alpha)\%$ confidence intervals be provided for mean effect of each treatment including placebo.

As changes in the QTc interval may be confounded with changes in the heart rate (RR interval), effects of the drug on both the QT and RR (or HR) intervals should be examined. Whereas there is no current guidance to interpret changes in the QT and RR intervals, an increase in the RR interval should correspond to an increase in the QT interval, although the magnitude of the increases should not necessarily be the same. Another signal that drug effects may be confounded is when there are differences in the results depending on the method of correcting QT for the RR interval. Bazett's method tends to overcorrect the QT interval when there are increases in heart rate. Thus, for a drug that increases the heart rate, Bazett's method may yield a larger estimate of drug effect on QTc than either a Fridericia's or population-based QT correction would.

There are two other analyses that might be done to help further assess the effect of experimental drugs. The first is to examine the dose- or concentration–response relationship, and the second is to look at individual subject changes from baseline. These are discussed in the next two sections.

DOSE–RESPONSE AND CONCENTRATION–RESPONSE MODELING

Understanding the dose–response and/or concentration–response of an experimental drug is essential to assessing the risk of QT prolongation. A shallow dose– or concentration–response may indicate a low risk of prolongation. To adequately model a dose–response of an experimental drug, a minimum of three doses should be studied, which has not been employed in "thorough" QT studies submitted to the regulatory agencies *(22)*. However, a concentration–response can be done with two doses, such as the therapeutic and supratherapeutic doses. If pharmacokinetic samples are taken over a range of times following dose, this should provide a range of concentrations from very low to supratherapeutic levels.

It is recommended that the following plots be generated to better understand the concentration–response:

- individual hysteresis plots of concentration vs QTc
- population scatter plots of drug concentration vs QTc
- mean time course plots of both concentration and QTc

Hysteresis plots account for time, and thus are useful in helping to assess whether there is lag-phase or prolonged effect on QTc. Figure 10 provides an example of what a hysteresis plot might look like for a drug with a direct concentration–response relation-

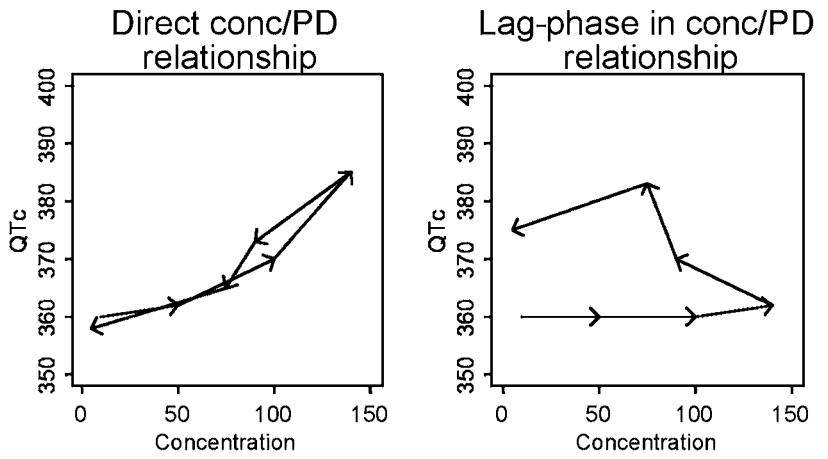

Fig. 10. Hysteresis plots for direct and lag-phase concentration/QTc relationships.

ship (left) and for a drug with a lag phase (right). These plots can also provide insight into the shape of the concentration–response relationship, will validate the adequacy of the frequency and timing of ECG collection, and may reveal additional factors (e.g., active metabolites, distribution phase, etc.) that will influence the interpretation of the results.

Figure 10 shows hysteresis plots for direct (left) and lag-phase (right) concentration–QTc relationships. Paired concentration (x-axis) and QTc interval (y-axis) values are plotted, with > indicating the time sequence of the paired observations. The figure on the left shows that concentration and QTc move together over time, whereas the right figure shows that there is delay.

Plots of the mean time courses of both concentration and QTc provide a visual assessment of the presence of a lag phase in the concentration–response relationship as illustrated in Fig. 8. These are then enhanced by the hysteresis plots of the concentration–PD effect relationship illustrated earlier.

Scatter plots of concentration vs QTc provide a visual assessment of a direct concentration–response. It is recommended that scatter plots include data from all doses of the experimental drug.

Figure 11 illustrates the concentration–response relationship for an experimental drug (open circles). A linear relationship was assumed that is represented by the solid line.

If the relationship between the drug concentration and the QTc change appears to be well defined and without a lag phase, a simple linear or nonlinear regression can be used to model the concentration–response relationship. The model should account for the study design (crossover or parallel group), for correlation between observation and within and between-subject variability. Whether three doses are tested or a concentration–response model is used, this approach rapidly increases the power of the study and minimizes the sample size.

The balance between the number of subjects and the frequency of data collection can be estimated using techniques described by Ahn and Jung *(23)*. In the presence of a lag phase, modeling of the concentration–response relationship can be performed using software designed for this purpose (e.g., Non-Mem, Win Non-Lin) and the model results tested for significance of the relationship. The model then can be

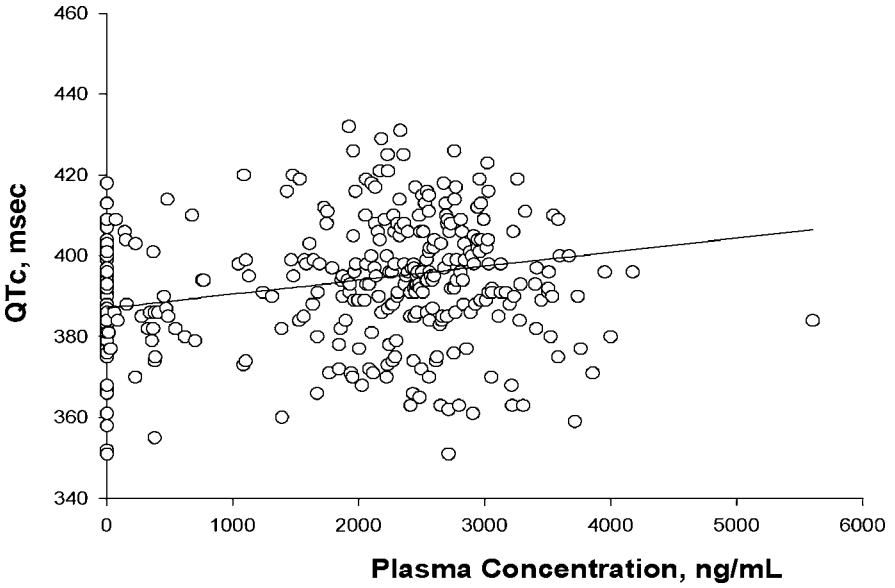

Fig. 11. Scatter plot of concentration and QTc intervals.

used to provide predictions for various concentrations or to predict at what concentration a threshold or target effect occurs. As a general rule, predictions should only be made for concentrations that fall within the range of observed concentrations, and care should be taken in interpreting results for concentrations outside this range. As a result of the large number of observations employed for concentration–response modeling, this approach may avoid some of the inherent problems with the central tendency analyses for drugs with variable Tmaxs. However, this analysis is not sufficient for drug approval and should be considered a supplemental analysis to the central tendency analyses.

CATEGORICAL AND OUTLIER ANALYSIS

One further way of assessing a drug's effect on the QT interval is to look at the individual values, as well as individual changes from baseline of QT and QTc. Both the draft ICH guidance document and a CPMP Points to Consider document *(24)* provide categories (Tables 5 and 6).

These categories are based on published clinical data and/or clinical experience, and most recently have been influenced by the terfenadine data. As with other critical ranges for QTc, these are primarily based on data using Bazett's correction and single ECG QT measurements. These may not be applicable to alternative correction factors, and specifically should not be employed for QT data averaged from two or more ECGs. Additionally, they may not account for the variability of the measurement, differences between males and females or the method of ECG measurement.

A simple way to display the data is to generate tables with a row for each category and column for each treatment arm (placebo and experimental drug). Each cell of the table should include both the number and frequency/percentage. Additionally, a total row and column is recommended (*see* Table 7).

Table 5
Categories of Risk for Absolute Values of Individual QT/QTc Intervals

Absolute QTc values	*Relative risk of TdP*
≤ 450 ms	So far no TdP
> 450 and ≤480 ms	No clear risk
> 480 and ≤ 500 ms	Uncertainty
> 500 ms	"Substantially" increased likelihood of being pro-arrhythmic

TdP, torsade des pointes.

Table 6
Categories of Risk for Changes From Baseline of Individual QT/QTc Intervals

Change from baseline	*Relative risk of TdP*
≤ 30 ms	No clear risk
> 30 and ≤ 60 ms	Uncertainty
> 60 ms	"Substantially" increased likelihood of being pro-arrhythmic

TdP, torsade des pointes.

Table 7
Sample Table of Categorical Summary of Individual Changes From Baseline

Change from Baseline	*Placebo*	*Drug X, dose 1*	*Drug X, dose 2*	*Active control*	*Total*
≤ 30 ms	25 (50%)	23 (46%)	20 (40%)	20 (40%)	88 (44%)
> 30 and ≤ 60 ms	20 (40%)	19 (38%)	20 (40%)	23 (46%)	82 (41%)
> 60 ms	5 (10%)	8 (16%)	10 (20%)	7 (14%)	30 (15%)
Total	50	50	50	50	200

As with some of the central tendency parameters, these tables may be dependent on the sampling scheme for ECGs and experimental conditions. Again, drugs that require a longer evaluation time in which meals, sleep, etc. will occur, may result in more observations in the higher categories. However, if the drug has no effect, numbers and frequencies should be similar to placebo.

SUMMARY AND RECOMMENDATIONS

A large number of factors influence the QT interval and its variability within and between subjects. Appropriate experimental design and conditions can limit those that are intrinsic to the individual (e.g., meals, sleep, physical activity, minute-to-minute QT variations, etc.). Those extrinsic factors, such as ECG data collection and measurement, correction factors, use of data from Cmax or anticipated Tmax, choice of baseline, and the time between baseline and on-drug measurements, must all be carefully controlled in order to enhance the power and efficiency of QT study design. Because the dose range employed in QT studies is often limited by subject tolerability, virtually all of the design decisions in "thorough" QT studies must be directed toward reducing the QT variability

and toward the use of the most powerful statistical analyses available. The choice of many of the study design options is often a trade-off, and the assumptions inherent in these choices must be understood and incorporated into the data analysis and interpretation.

We offer the following recommendations for the design and statistical analyses of "thorough" QT studies:

1. The experimental design and setting should minimize known sources of QT variability. The period of peak pharmacodynamic effect should avoid the post-prandial period, exercise, and sleep. Because of altered QT–RR relationships during sleep, QT data should not be compared between awake and asleep periods, nor for approx 1 h after awakening. Balancing male and female subjects and studying subjects evenly distributed throughout the age range of the intended patient population will allow these factors to be added as covariates in the statistical analysis. At least two and preferably three doses of the drug should be tested to gain the marked increase in statistical power of a dose–response or a concentration–response analysis. Crossover designs offer enhanced power by controlling for intersubject variability, but may not be practical for long half-life drugs and active metabolites, or for drugs requiring dose titration.
2. Computer-assisted manual over-read of QT intervals from at least two replicate ECGs at each time point is the most effective way to reduce QT variability. The same ECG lead should be used within subjects for estimation of QT changes.
3. The preferred baseline is a predose time point on the same day as, or as close as practicable to the on-drug ECG recordings. Data for both placebo and drug should be reported prior to any "placebo-adjusted" calculations. For long half-life drugs requiring parallel study designs, careful control for sequence effects in both the drug and placebo groups should be incorporated in the data collection.
4. The Bazett's and Fridericia's corrected QT intervals are requested by regulatory agencies, however, the QT–RR relationship of the study population should be inspected. Where necessary, a population or individualized correction factor should also be employed to avoid the increases in variability caused by standardized equations. A "Holter bin" analysis that constructs the QT–RR relationship for each subject on drug and on placebo is a highly sensitive and informative method that was accepted the FDA as supporting evidence for the alfuzosin NDA in 2003. Alternative statistical methods that avoid correction of the QT interval have been proposed but, as of the time of this manuscript, have not been the basis for regulatory approval.
5. The ICH draft document proposes that the ECG obtained at drug Cmax or anticipated Tmax is the first central tendency parameter to use. This is of value when the plasma pharmacokinetics of the drug and/or measured metabolite corresponds to the pharmacodynamic effect of the drug. The alternative parameter, the maximum change in QT, has a high false positive rate and is randomly distributed over the time of observation even in the presence of a moderate, true drug effect. The high false positive rate can be reduced by obtaining replicate ECGs at each time point, and by performing an hysteresis analysis. When hysteresis analysis indicates a correlation between drug or metabolite and the maximum change in QT, the data at this time point may be considered as valid.
6. Dose–response and concentration–response modeling analysis offers powerful statistical tools to increase the sensitivity of the "thorough" QT studies and to enhance the ability to assess risk. When careful attention is paid to the timing of the concentration–response relationships, this method is the most effective way to reduce sample size and increase statistical power.
7. The categorical analyses should be limited to comparative frequency tables between drug and placebo for individual (i.e., nonaveraged) ECG QT data.

REFERENCES

1. 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. Am J Cardiol 1991;67:774–776.
2. Molnar J, Zhang F, Weiss J, Ehlert FA, Rosenthal JE. Diurnal pattern of QTc interval: How long is prolonged? J Am Coll Cardiol 1996;27:76–83.
3. Widerlov E, Jostell KG, Claesson L, Odlind B, Keisu M, Freyschuss U. Influence of food intake on electrocardiograms of healthy male volunteers. Eur J Clin Pharmacol 1999;55:619–624.
4. Browne KF, Prystowsky E, Heger JJ, Chilson DA, Zipes DP. Prolongation of the Q-T interval in man during sleep. Am J Cardiol 1983;52:55–59.
5. Vrancianu R, Filcescu V, Ionescu V, Groza P, Persson J, Kadefors R, Petersen I. The influence of day and night work on the circadian variations of cardiovascular performance. Eur J Appl Physiol 1982; 48:11–23.
6. Lande G, Funck-Brentano C, Ghadanfar M, Escande D. Steady-state versus non-steady-state QT–RR relationships in 24-hour Holter recordings. PACE 2000;23:293–302.
7. Extramiana F, Maison-Blanche P, Badilini F, Pinoteau J, Deseo T, Coumel P. Circadian modulation of QT rate dependence in healthy volunteers. Gender and age differences. J Electrocardiol 1999;32:33–43.
8. Rautaharju PM, Zhou SH, Wong S, Calhoun HP, Berenson BS, Prineas R, Davignon A. Sex differences in the evolution of the electrocardiographic QT interval with age. Can J Cardiol 1992;8:690–695.
9. Ahnve S, Vallin H. Influence of heart rate and inhibition of autonomic tone on the QT interval. Circulation 1982;65:435–439.
10. Roden DM. Drug-induced prolongation of the QT interval. New Engl J Med 2004;350:1013–1022.
11. Camm AJ, Janse MJ, Roden DM, Rosen MR, Cinca J, Cobbe SM. Congenital and acquired long QT syndrome. Eur Heart J 2000;21:1232–1237.
12. Sun H, Chen P, Kenna L, Lee P. The chaotic QT interval variabilities on risk assessment trial designs. Clin Pharm Ther 2004;75:P55.
13. Lee SH, Sun H, Chen P, Doddapaneni S, Hunt J, Malinowski H. Sensitivity/reliability of the time-matched baseline subtraction method in assessment of QTc interval prolongation. Clin Pharm Ther 2004;75:P56.
14. Summary Basis of Approval, moxifloxacin hydrochloride, April 21, 2001. FDA document.
15. Hnatkova K, Malik M. "Optimum" formulae for heart rate correction of the QT interval. PACE 1999;1683–1687.
16. 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.
17. Batchvarov VN, Ghuran A, Smetana P, Hnatkova K, Harries M, Dilaveris P, Camm AJ, Malik M. QT–RR relationship in healthy subjects exhibits substantial intersubject variability and high intrasubject stability. Am J Physiol 2002;282:H2356–H2363.
18. Dmitrienko A, Smith B. Repeated-measures models in the analysis of QT interval. Pharmaceut Statist 2003;2:175–190.
19. The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs: Preliminary Concept Paper, November 15, 2003. http://cdernet.cder.fda.gov/QTWG/QT%20Workshop/qt4jam.pdf.
20. Badilini F, Maison-Blanche P, Childers R, Coumel P. QT interval analysis on ambulatory electrocardiographic recordings: A selective beat averaging approach. Med Biol Eng Comput 1999;37:71–79.
21. Hollander M, Wolfe DA. Non-parametric Statistical Methods, Wiley Series, 1999, pp. 125–133.
22. Kenna LA, Parekh A, Jarugula V, Chatterjee DJ, Sun H, Kim MJ, Ortiz S, Hunt JP, Malinowski H. Experience evaluating QT prolongation data. Clin Pharm Ther 2004;75:P7.
23. Ahn C, Jung S-H. Efficiency of general estimating equations estimators of slopes in repeated measurements: Adding subjects or adding measurements? Drug Info J 2003;37:309–316.
24. Points to Consider: The Assessment of the Potential for QT Interval Prolongation by Non-Cardiovascular Medicinal Products. Committee for Proprietary Medicinal Products, EMEA, December 17, 1997.