

Rank Power of Metrics Used to Assess QTc Interval Prolongation by Clinical Trial Simulation

Peter L. Bonate, PhD

Monte Carlo simulation was used to assess the type I error rate and rank order of power for six different metrics using linear mixed-effect models, including two variables recommended by the European Agency for the Evaluation of Medicinal Products (EMA) in the analysis of QTc interval data. The metrics analyzed were maximal change in QTc interval from baseline, maximal QTc interval, area under the QTc interval-time curve (AUC), average QTc interval, maximal QTc interval with baseline QTc interval as covariate, and AUC with baseline QTc interval as covariate. Two dosing regimens were studied: multiple-dose oral and multiple-dose continuous intravenous infusion. Both regimens were designed to produce similar

maximal plasma concentrations, albeit with the infusion regimen maintaining maximal plasma concentrations for a longer period of time. The ability of the metrics to detect a drug effect was examined, assuming drug effect followed either an E_{\max} or linear model. All statistics had a type I error rate near the nominal value. Regardless of pharmacokinetic or pharmacodynamic model, AUC with baseline QTc interval as a covariate had greater power than any other metric examined. The simulations also suggest that mean QTc interval data not be used.

Journal of Clinical Pharmacology, 2000;40:468-474
©2000 the American College of Clinical Pharmacology

Lately, a number of drugs have been shown to prolong cardiac repolarization possibly through their effect on the HERG-1 potassium channel.^{1,2} Regulatory agencies are particularly interested in prolonged cardiac repolarization as a safety issue because potentially life-threatening ventricular tachycardia, called torsades de pointes (TdP), may result.³ Identification of prolonged cardiac repolarization is made using an electrocardiogram (ECG), in which increased QT intervals in a patient are indicative of the condition. Because the QT interval is dependent on heart rate, QT intervals are frequently corrected using either Bazett's equation,⁴

$$QTc = \frac{QT}{\sqrt{RR}},$$

or Fridericia's correction,⁵

$$QTc = \frac{QT}{\sqrt[3]{RR}},$$

From Quintiles, Clinical Pharmacokinetics, Kansas City, Missouri. Submitted for publication June 30, 1999; revised version accepted November 16, 1999. Address for reprints: Peter L. Bonate, PhD, Quintiles, Clinical Pharmacokinetics, P.O. Box 9708 (L4-M2828), Kansas City, MO 64134.

where QT is the time interval in msec between the Q and T wave on an ECG, and RR is the time interval between consecutive R waves on an ECG in sec.

Exactly what constitutes a "prolonged QTc interval" is a matter of debate, although the European Agency for the Evaluation of Medicinal Products (EMA) has indicated that changes in QTc intervals between 30 and 60 msec are likely to represent drug effect, and changes greater than 60 msec "raise clear concerns about the potential risk [for a new drug]."⁶ Also, if the QTc interval is greater than 450 or 470 msec for males and females, respectively, then this too is evidence of prolonged QTc intervals. The Food and Drug Administration currently has no guidelines in place for determining QTc interval prolongation. Bonate and Russell⁷ have recently published clinical and statistical guidelines for the analysis of QTc intervals in clinical trials. They suggest that QTc intervals are best analyzed using mixed-effect models, either linear or nonlinear, with one of four different dependent variables: maximal change in QTc intervals from baseline, maximal QTc interval, area under the QTc interval-time curve (AUC), and average QTc interval. It was suggested that improved power can be achieved through analysis of

covariance with baseline QTc interval as the covariate. The purpose of this paper is to use Monte Carlo simulation to determine the type I error rate for each of these dependent variables and their rank order of power and to confirm that covariate inclusion results in increased statistical power.

SIMULATION

Monte Carlo simulation is a computer-intensive technique to simulate models with fixed effects and random components having specified probability distributions that can be used to determine the long-term outcome or expected value of such models. Monte Carlo simulation was done to simulate two different clinical trials for the same drug. In one trial, drug administration was by the oral route of administration, whereas in the other study, drug administration was by continuous intravenous infusion. In each trial, 40 subjects were enrolled in a randomized, placebo-controlled, four-period crossover design. Prior to drug administration within each period, each subject was given a placebo to characterize the baseline, drug-free QTc interval for that subject (day -1). The treatments for the oral administration study were placebo or 50, 100, or 200 mg drug once daily for 7 days, and the treatments for the infusion study were placebo or 5, 10, or 20 mg drug once daily for 3 days infused over a 12-hour period. In the continuous infusion study, a 12-hour infusion period is unusually long; this infusion duration was chosen to simulate a sustained-release formulation in which drug concentrations are elevated for prolonged periods of time. Both dosing regimens lead to approximately equivalent maximal drug concentrations (Figure 1). Each subject was randomly assigned to 1 of 20 unique dosing sequences. The drug's pharmacokinetics was assumed to follow a one-compartment model. The population characteristics were as follows:

- Cl ~ LN(35 L/h, CV = 30%),
- Vd ~ LN(100 L, CV = 20%),
- Ka ~ LN(0.7 per h, CV = 10%) for oral administration,
- F ~ Logit(0.3, CV = 20%) for oral administration,

where LN is log-normal distribution, Logit is the logit distribution that has range [0,1], CL is systemic clearance, Vd is volume of distribution, Ka is the absorption rate constant, F is bioavailability, and CV is the coefficient of variation in percent. A logit distribution was chosen for F since F must be constrained between 0

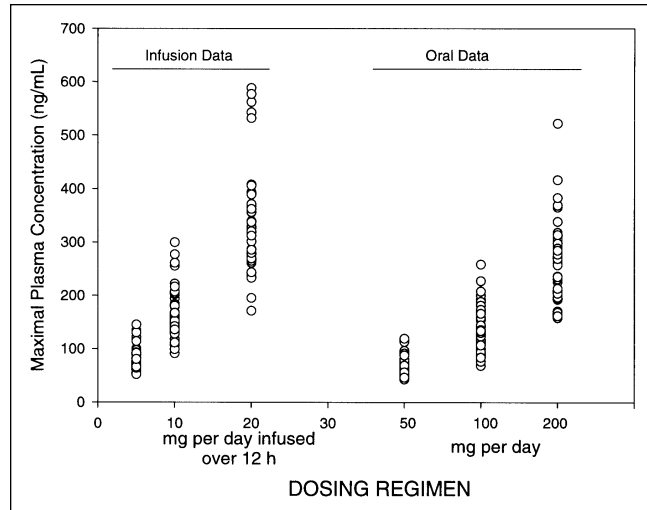


Figure 1. Representative plot of maximal plasma concentrations after oral administration and continuous intravenous infusion in 40 subjects.

and 1.⁸ Each subject's pharmacokinetic parameters on the jth day, S_{ij} , were defined as a linear function of the population mean, μ , plus a random term for deviation from the population mean (intersubject variability), δ_i , and a random term for deviation from the ith subject mean (intrasubject variability), ϵ_{ij} :

$$S_{ij} = \mu + \delta_i + \epsilon_{ij}.$$

It was assumed that inter- and intrasubject variability were independent. ϵ_{ij} was defined as normally distributed with mean zero and 10% coefficient of variation around $\mu + \delta_i$. Drug accumulation was calculated using the superposition principle.

Baseline, drug-free QTc intervals for each subject, QTc_i^b , were defined as

$$QTc_i^b = \mu + S_i + R_i \quad i = 1, 2, \dots 40 \quad (1)$$

where μ is the population mean of 400 msec, S_i is the subject-specific effect for the ith subject, R_i is the intrasubject effect for the ith subject, and the superscript b refers to baseline. When QTc intervals are actually measured using an ECG, equation (1) is modified to account for measurement error, e_i :

$$QTc_i^b = \mu + S_i + R_i + e_i \quad i = 1, 2, \dots 40. \quad (2)$$

The expected value for the population is μ , and the total variance is the sum of the intersubject, intrasubject, and measurement error variability,

$$\text{Var}(QTc^b) = \sigma_S^2 + \sigma_R^2 + \sigma_e^2, \quad (3)$$

which was reported to be 1156 msec in healthy normal volunteers in a drug-free state.⁹ The test-retest reliability, g , is given by

$$g = \frac{\sigma_R^2 + \sigma_e^2}{\sigma_S^2 + \sigma_R^2 + \sigma_e^2} \quad (4)$$

and was reported to be 0.85.¹⁰ Solving for equations (3) and (4) and assuming that measurement variability, σ_e^2 , is 64 msec ($\pm 2\%$) gives $\sigma_S^2 = 928$ msec and $\sigma_R^2 = 164$ msec.* All sources of variability were assumed to follow a normal distribution. For simplicity, it was assumed that each QTc interval was independent and that regression toward the mean did not occur on repeated measurement.

Some effect due to drug was added to each QTc interval. An E_{\max} model was chosen to model the relationship between drug concentration and QTc intervals. Table I shows each of the pharmacodynamic simulation conditions. In general, for the E_{\max} model, IC_{50} values were simulated to range from much less than average maximal plasma concentrations to much greater than maximal plasma concentrations. It should be noted that when IC_{50} equals zero or is much less than drug concentrations, the model simplifies to the case in which a constant drug effect, T , is present. Within each IC_{50} value, E_{\max} was varied from 0 to 50 msec, an interval that shows an ascending positive drug effect. QTc intervals were measured in each subject within each period 0, 2, 4, 8, 12, and 24 hours after the last drug administration on the last day of dosing and on the placebo lead-in day within each period. These data were used for statistical analysis.

Six different dependent variables were calculated for each subject: maximal change in QTc interval from baseline, maximal QTc interval, area under the QTc interval-time curve (AUC), average QTc interval, maximal QTc interval with baseline QTc interval as covariate, and area under the QTc interval-time curve with baseline QTc interval as covariate. The tests were chosen based on methods used in the literature, recommended methods by regulatory agencies, and proposed methods by Bonate and Russell.⁷ Maximal change from baseline with baseline QTc interval as the covariate was not included in the analysis because estimation of the treatment effect using this approach would produce exactly the same results as analysis of maximal QTc intervals with covariate adjustment.¹¹ The baseline QTc interval was calculated for each subject as the average of the six QTc intervals collected on

* These mean and variance estimates are based on Bazett's equation and will be different if Fridericia's equation is used.

Table I Pharmacodynamic Simulation Conditions

Simulation	E_{\max} (msec)	IC_{50} (ng/mL)	τ (msec)
1	0	0	0
2	1	1	0
3	5	1	0
4	10	1	0
5	20	1	0
6	30	1	0
7	40	1	0
8	50	1	0
9	1	200	0
10	5	200	0
11	10	200	0
12	20	200	0
13	30	200	0
14	40	200	0
15	50	200	0
16	1	2000	0
17	5	2000	0
18	10	2000	0
19	20	2000	0
20	30	2000	0
21	40	2000	0
22	50	2000	0
23	0	0	1
24	0	0	5
25	0	0	10
26	0	0	20
27	0	0	30
28	0	0	40
29	0	0	50

the placebo lead-in day within each period (day -1). AUC was calculated using the trapezoidal rule:

$$AUC = \sum_{i=1}^{n-1} 0.5(t_{i+1} - t_i)(QTc_{i+1} + QTc_i),$$

where n is the number of sampling points (in this case, $n = 6$) and t is the time point (e.g., 0, 2, . . . , etc.). As an example, consider a subject whose QTc intervals on day -1 were {412, 416, 399, 402, 432, 408 msec} and on day 7 were {414, 425, 422, 427, 444, 401 msec}. That subject's mean baseline QTc interval would be 412 msec with a maximal QTc interval of 444 msec, a maximal change from baseline of 32 msec, an AUC of 10,196 msec•h, and a mean postdose QTc interval of 422 msec.

All data points were simulated with PROC IML of SAS for Windows[®] (version 6.12), and all data were analyzed using linear mixed-effect models (PROC

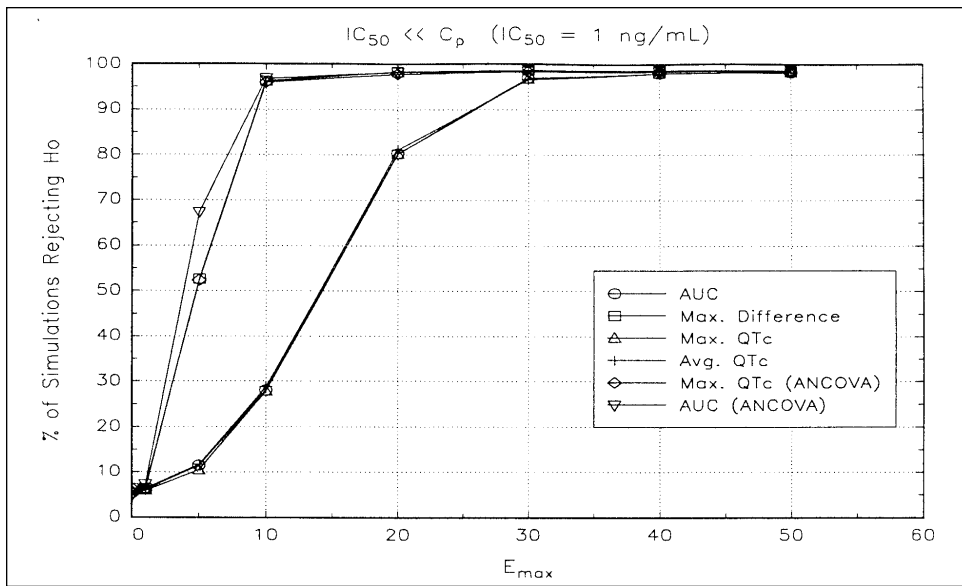


Figure 2. Percentage of simulations rejecting the null hypothesis when $IC_{50} \ll$ maximal plasma concentrations following oral administration. Data were generated using 1000 simulations. See text for details.

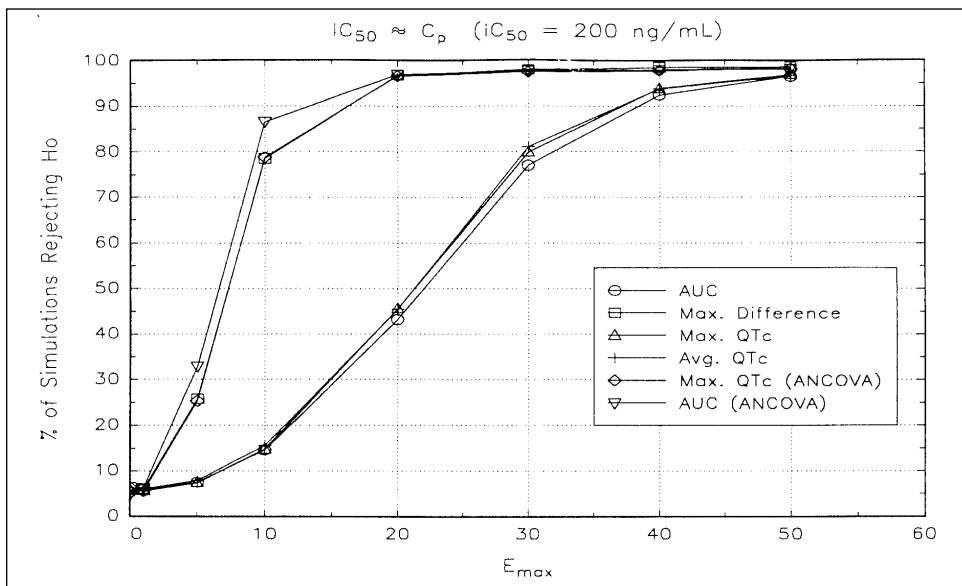


Figure 3. Percentage of simulations rejecting the null hypothesis when $IC_{50} \approx$ maximal plasma concentrations following oral administration. Data were generated using 1000 simulations. See text for details.

MIXED) with terms for sequence (19 degrees of freedom, df), subject within sequence, period (3 df), and treatment or dose (3 df). Mixed-effect models allow the analyst to specify factors as either random or fixed. In this simulation, subjects within sequence were treated as a random effect, whereas all other factors were treated as fixed. Baseline scores were used as optionally included covariates leading to an analysis of covariance. The null hypothesis for treatment effect was that the mean parameter estimate was equal among treatment groups. Treatment effects were considered statistically significant if the p -value for treatment effect

was less than 0.05. One thousand iterations were done for each pharmacodynamic simulation condition (Table I) for both the oral and intravenous formulations. The overall power for each dependent variable was calculated as the percentage of simulations that rejected the null hypothesis of no drug effect.

RESULTS AND DISCUSSION

Figures 2 through 4 show the percentage of simulations that rejected the null hypothesis for the oral formulation, and Table II shows the percentage of

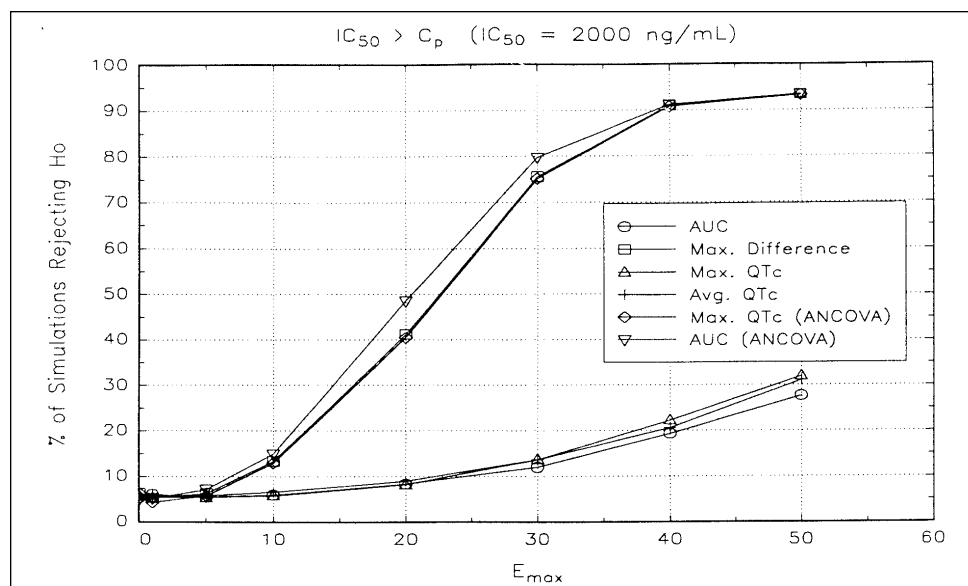


Figure 4. Percentage of simulations rejecting the null hypothesis when $IC_{50} >$ maximal plasma concentrations following oral administration. Data were generated using 1000 simulations. See text for details.

Table II Percentage of Simulations Rejecting the Null Hypothesis for the Continuous Infusion Simulation

E _{max}	AUC	Maximal Change from Baseline Scores	Maximal QTc Interval	Mean QTc Interval	Maximum QTc Interval (ANCOVA)	AUC (ANCOVA)
E _{max} model: $IC_{50} \ll$ maximal plasma concentrations ($IC_{50} = 1$ ng/ml) constant drug effect						
0	6.5	4.9	6.4	6.9	5.3	6.4
1	6.5	4.9	6.4	6.9	5.3	6.4
5	10.4	37.3	9.8	10.4	36.8	54.0
10	23.4	95.9	22.2	21.9	95.0	99.3
20	71.3	100.0	73.3	69.6	99.9	99.9
30	97.8	100.0	98.3	97.1	99.9	99.9
40	100.0	100.0	100.0	100.0	99.9	99.9
50	100.0	100.0	100.0	100.0	99.9	99.9
E _{max} model: $IC_{50} \cong$ maximal plasma concentrations ($IC_{50} = 200$ ng/ml)						
0	6.5	4.9	6.4	6.9	5.3	6.4
1	6.6	4.3	6.1	6.6	4.5	5.3
5	6.7	7.9	7.6	6.8	7.4	11.2
10	8.4	19.4	8.5	8.6	19.6	34.7
20	15.5	78.6	16.3	13.6	79.4	93.1
30	29.5	99.1	31.5	24.3	99.2	99.9
40	49.5	100.0	58.5	40.2	99.9	99.9
50	69.4	100.0	79.8	60.5	99.9	99.9
E _{max} model: $IC_{50} >$ maximal plasma concentration ($IC_{50} = 2000$ ng/ml)						
0	6.8	4.4	6.2	6.2	4.9	5.0
1	6.8	4.4	6.1	6.2	4.8	5.0
5	6.5	4.2	6.2	6.5	4.4	5.3
10	6.8	4.8	6.4	6.5	4.6	6.3
20	6.7	6.3	6.9	6.6	6.6	9.5
30	7.2	9.3	7.6	6.8	9.3	13.6
40	7.7	13.3	8.2	7.5	13.4	21.9
50	8.4	20.3	8.9	8.6	19.8	34.3

simulations that rejected the null hypothesis for the intravenous formulation. For both the oral and intravenous formulations, the results were similar and can be generalized across simulations. The type I error rates for all of the dependent variables were near their nominal value of 5%. In general, the rank of order of power was the following: area under the QTc interval-time curve with baseline QTc interval as covariate \geq maximal QTc interval with baseline QTc interval as covariate \cong maximal change in QTc interval from baseline $>$ area under the QTc interval-time curve \cong maximal QTc interval \cong mean QTc interval. As expected, as E_{\max} or τ increased, so did the percentage of simulations rejecting the null hypothesis. As the IC_{50} increased, the power of the tests decreased, as did the magnitude of the effect of E_{\max} on power. Thus, drugs with high IC_{50} s have little effect on QTc intervals. Conversely, when the IC_{50} is small, very small changes in E_{\max} lead to large increases in power.

The EMEA has indicated that maximal change in QTc intervals from baseline and maximal QTc interval be used as the primary variables in the analysis of QTc interval data. These simulations suggest that while using maximal change from baseline as the dependent variable has fairly good power, using maximal QTc intervals as the dependent variable does not. Maximal QTc intervals alone have roughly 25% of the power of maximal change from baseline, which is not surprising since maximal change from baseline corrects for individual differences at baseline while maximal QTc intervals do not. Following the EMEA guidelines and analyzing both maximal QTc intervals and maximal change from baseline may lead to the situation in which one test shows a significant drug effect and the other does not. The inclusion of baseline QTc intervals as a covariate in the linear model with maximal QTc intervals as the dependent variable may correct this possible discrepancy due to an increase in statistical power since now the model begins to account for individual differences at baseline.

Mean QTc intervals also had poor power compared to the ANCOVA models and maximal change in QTc intervals from baseline. This is not really surprising since the mean value probably dilutes the true treatment effect, a point discussed by Bonate and Russell.⁷ For example, suppose that QTc intervals are increased immediately after drug administration but rapidly return to baseline. Clearly, a drug effect is present, but as the number of ECGs increases with each subsequent ECG showing a normal QTc interval, the observed mean value will approach the baseline mean, thus masking any drug effect. Only with the collection of a

large number of ECGs showing abnormal results, relative to the total number of ECGs collected, will the mean value be an indicator of drug effect. In these simulations, even though drug concentrations were elevated in the continuous infusion study, mean QTc intervals failed to detect drug effect relative to other more sensitive measures.

These results indicate that the suggestion by Bonate and Russell⁷ to use area under the curve as a measure of pharmacodynamic effect in QTc studies is a good one, provided baseline QTc is used as a covariate. Area under the curve alone is too insensitive a metric to detect drug effect. But area under the curve with baseline QTc interval as covariate was the most powerful statistical test studied under all simulations. Under certain conditions, the increase in power using the area under the curve with baseline covariate approached 20%. When baseline QTc was not included, the power of using area under the curve as the dependent variable decreased to the dismal levels seen using maximal QTc intervals or mean QTc values as the dependent variable. These results suggest that this method should be the preferred way to analyze QTc interval data under this experimental design. However, one difficulty with its use is its interpretation. The units of area under the curve are the time scale of QTc intervals \cdot time scale of the collection interval—in this case, msec \cdot h. What does a value of 10,000 msec \cdot h mean? Some may argue that because a clinician, who ultimately makes the decision regarding whether the degree of QTc prolongation that may occur with a drug is clinically significant, cannot easily understand what a value of an area under the curve represents, it should not be used. Even if it is not easy to communicate the meaning of a test statistic, we should not discourage its use. Rather, we should find other means to present the statistic in an easily understood manner. For example, it may be more useful to report to an audience of clinicians the area under the curve after drug administration relative to the baseline drug-free area under the curve (e.g., “drug X had a 25% increase in area under the curve, relative to the drug-free state”). An alternative is to use AUC divided by the time interval used to calculate the time-averaged AUC. In this instance, the metric then becomes a time-averaged AUC with units of msec. For example, in the example data set used to illustrate the metrics used in the simulation, the AUC was calculated to be 10,196 msec \cdot h, but divided by 24 hours, the time-averaged AUC is 425 msec, a value more easily understood. Presenting the data in this manner will not change the results of an

analysis since this is a monotonic transformation using a constant value.

Two key points need to be addressed. First, the simulation was designed to simulate a clinical trial in which the goal of the study was to determine if QTc interval prolongation occurs with a particular drug. The timing of sampling was chosen to occur at or near the time of maximal plasma concentrations, at the time of trough (the time immediate prior to dosing) plasma concentrations, and during the elimination phase. Thus, it was likely that the maximal drug effect and time course of drug effect were captured. Many phase I studies, which are not specifically designed to determine if QTc interval prolongation occurs with a drug, are examined retrospectively to answer that very question. Often in these studies, few ECGs, much fewer than simulated herein, are collected. The results of this simulation cannot be generalized to those studies because it is not clear if the advantage of using area under the curve as a metric continues when the number of time points decreases to two or three. Future simulations will address that issue. Second, it must be recognized that the characterization of the baseline QTc interval within each subject is very important. It is frequent practice in phase I studies that a single ECG is collected in the drug-free state immediately prior to drug administration as a baseline on which to make future comparisons. While this may be acceptable from a safety point of view, using a singular QTc interval as the baseline measurement in an analysis of covariance is risky because (1) the baseline is actually a random variable that is indeed quite variable, and (2) the possibility of regression toward the mean may occur with future measurements.^{12,13} Collecting multiple drug-free QTc intervals and then using the mean as the baseline make for a more robust estimate of the baseline and should be encouraged in studies wishing to examine the effect of a drug on QTc intervals.¹⁴

In summary, Monte Carlo simulation was used to choose a metric with which to analyze QTc data in a clinical study. The simulations suggest that the best metric is area under the QTc interval-time curve with baseline QTc interval as a covariate. The difficulty with using this metric lies primarily with its difficult interpretation, an outcome that should not discourage its use. The results also show that the EMEA guidelines of recommending both maximal change from baseline scores and maximal QTc interval as primary

metrics in an analysis may lead to conflicting results because maximal change from baseline scores has much greater power than maximal QTc intervals. One metric may show a significant drug effect, while the other may not. Inclusion of baseline QTc intervals in the analysis of maximal QTc intervals leads to power curves that are indistinguishable from maximal change from baseline scores. Use of mean QTc intervals as the primary metric should also be discouraged.

REFERENCES

1. Rampe D, Murawsky MK, Grau J, Lewis EW: The antipsychotic agent is a high affinity antagonist of the human cardiac potassium channel HERG. *J Pharmacol Exp Ther* 1998;286:788-793.
2. Ducic I, Ko CM, Shuba Y, Morad M: Comparative effects of loratadine and terfenadine on cardiac K⁺ channels. *J Cardiovas Pharmacol* 1997;30:42-54.
3. Algra A, Tijssen JGP, Roelandt J, Pool J, Lubsen J: QTc prolongation measured by standard 12-lead ECG is an independent risk factor for sudden death due to cardiac arrest. *Circulation* 1991;83:1888-1894.
4. Bazett HC: An analysis of time-relations of electrocardiograms. *Heart* 1920;7:353.
5. Fridericia LS: Die systolendauer im elektrokardiogramm bei normalen menschen und bei herzkranken. *Acta Med Scand* 1920;53:469.
6. European Agency for the Evaluation of Medicinal Products (EMA), Committee for Proprietary Medicinal Products (CPMP): *Points to Consider: The Assessment of the Potential for QT Interval Prolongation by Non-Cardiovascular Medicinal Products*. London: EMA, 1997.
7. Bonate PL, Russell T: Assessment of QTc prolongation for non-cardiac drugs from a drug development perspective. *J Clin Pharmacol* 1999;39:349-358.
8. Hastings NAJ, Peacock JB: *Statistical Distributions*. New York: Halsted, 1975.
9. 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.
10. Myrtek M: Covariation and reliability of ECG parameters during 24 hour monitoring. *Int J Psychophysiol* 1990;10:117-123.
11. Laird N: Further comparative analyses of pretest-posttest research designs. *Am Stat* 1983;37:329-330.
12. Berntson GG, Uchino B, Cacioppo JT: Origins of baseline variance and the law of initial values. *Psychophysiol* 1994;31:204-210.
13. Nesselroade JR, Stigler SM, Baltes PB: Regression toward the mean and the study of change. *Psychol Bull* 1980;88:622-637.
14. McDonald CJ, Mazzuca SA, McCabe GP: How much of the placebo "effect" is really statistical regression? *Stat Med* 1983;2:417-427.