

ABBOTT LABORATORIES

Global Pharmaceutical Regulatory Affairs

Douglas L. Sporn
Divisional Vice President
Regulatory Intelligence and FDA Liaison Office
Global Pharmaceutical Regulatory Affairs
D-RA78, AP30-1
Abbott Laboratories
Abbott Park, IL

Abbott FDA Liaison Office
1700 Rockville Pike, Suite 400
Rockville MD 20852

Tel: 301-998-6144 Fax: 301-984-9543
E-mail : doug.sporn@abbott.com

December 13, 2004

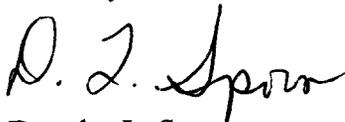
Division of Dockets Management (HFA-305)
The Food and Drug Administration
5630 Fishers Lane, room 1061,
Rockville, MD 20852

Re: **Docket No. 2004D-0377, CDER 200495. International Conference on Harmonization; Draft Guidance on E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs.**

Abbott Laboratories (Abbott) is very pleased to have the opportunity to comment on the above mentioned draft guidance, published in the Federal Register on September 13, 2004.

We thank the Agency for their consideration of our attached comments. Should you have any question, please contact Ivone Takenaka, Ph.D. at (301) 998-6144 or by FAX at 301-984-9543.

Sincerely,



Douglas L. Sporn

2004D-0377

C8

**Comments on Draft – ICH-E14
The Clinical Evaluation of QT/QTc Interval Prolongation and
Proarrhythmic Potential for Non-Arrhythmic Drugs**

Docket No. 2004D-0377

The following comments on the above draft are provided on behalf of Abbott Laboratories (Abbott).

GENERAL COMMENTS

There is a high level of uncertainty as to the magnitude of risk associated with small QT or QTc interval prolongations, how this varies across subpopulations, and what level of risk is acceptable when considering the utility of a therapeutic intervention. Both preclinically and clinically, we do not have validated means to discriminate between proarrhythmogenicity of 2 agents with comparable QT prolongation. A number of risk factors have been proposed, but the relative importance and degree of interaction of these factors with QT prolongation have not been quantified. As a result, we are considering guidance as to what threshold of effect might be associated with incremental risk, and have adopted a very conservative approach, that if not updated with emergent findings, will generate prodigious activities, with little net result on patient safety. With guidance-imposed generation of voluminous data, there should be attendant responsibilities. A mechanism must be established for continuous consolidation of knowledge, with open sharing of extant and emergent information about experimental techniques, study designs, response variance for positive controls, normal intrasubject variability, and more quantitative modeling of risk. Nonetheless, the current draft of the guidance is evolving as a valuable document for understanding and characterizing that risk. We recommend consideration of the following points for further refinement.

SPECIFIC COMMENTS

1.3 Scope

Overall, we appreciate the need for a thorough QT evaluation of new agents, but there are concerns about the mechanisms available for scientific integration and discussion of all available data generated for a given agent once the guidances become finalized. We understand that elements of the S7B battery of tests may produce false negatives. Once these are compiled in terms of sensitivity and specificity, they should be presented in the public domain so that the reasons for predictive failure are discussed and adapted to. We believe that false positives also need special attention and perspective, since a large number of agents are expected to have at least low affinity binding in the hERG channel given the ability to explore sufficiently high concentrations. The approach that is taken for drug-drug interactions, in which a C_p/K_i ratio is considered, should be a feature of the evaluation for the need of a definitive human study.

Docket No. 2004D-0377

Lines 137-138. *The recommendations contained in this document are generally applicable to new drugs having systemic bioavailability.*

The issue of biologics, specifically antibodies, requires attention in the guidance. Granted, scorpion toxin is suspected as having hERG effects, but therapeutically targeted biologics will probably not. For example, at a molecular weight of 150 kDa, an antibody is too large to reside in the hERG channel and is unlikely to interact with cell surface regions of the channel (which can be tested preclinically). Is it possible for the guidance to allow dispensation for cases like this in which some preclinical experiments would be adequate for exemption from thorough clinical study?

Lines 143-144. *Additional ECG data might also be considered appropriate if a new indication or patient population were being pursued.*

We support the modeling approaches being investigated by the FDA pharmacometrics (PM) group, and would hope that new populations or conditions could be addressed by modeling coupled with assessment of population differences in the pharmacokinetics. The use of healthy volunteers to determine the magnitude of a signal is predicated on the premise that highly controlled experiments in subjects without confounding effects is best suited for detection of small signals in the presence of significant “noise” from normal within-day and inter-day variability. If regulators are likely to request new thorough QT studies for previously “negative” agents in each new indication or patient population, then sponsors are potentially susceptible to ever escalating development costs. We would thus propose that the statement read: **“Additional analyses, including modeling, might also be considered appropriate if a new indication or patient population were being pursued.”**

2.1.2 The ‘Thorough QT/QTc Study’: Dose-Effect and Time Course Relationships

Lines 212-214. *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.*

“Substantial multiples” requires careful thought and some level of definition. We agree that concentrations well above normal therapeutic exposures are quite valuable in regression-based approaches taken by PM scientists at the FDA to define the magnitude of the signal under normal conditions. We would propose that the “no effect” calculations and labeling are based on normal exposures, and that special case circumstances be handled in labeling based on the regression data plus pharmacokinetic effects derived from special population and drug interaction data. The classic case of terfenadine is an example of an agent with extensive intestinal and hepatic first pass elimination by a single CYP isoform, resulting in $F \sim 0.02-0.04$, and the potential of $>50 \times$ increase in concentrations in the presence of a drug interaction or significant impairment

Docket No. 2004D-0377

of hepatic function. Although there are rare exceptions, most new candidates selected by sponsors do not have the amplification characteristics of terfenadine.

Lines 214-216. *Alternatively, if the concentrations of a drug can be increased by drug-drug or drug-food interactions involving metabolizing enzymes (e.g., CYP3A4, CYP2D6) or transporters (e.g., P-glycoprotein), these studies can be performed under conditions of maximum inhibition.*

We agree that the thorough study should be conducted with coverage over the concentration range expected in the presence of inhibition of major clearance pathways. However, we believe that conducting thorough QT studies in the presence of clearance inhibitors is generally a troublesome alternative. Some standard inhibitors (KTZ for CYP3A; quinidine, paroxetine or fluoxetine for CYP2D6) either have intrinsic QT effects or possible effects on autonomic tone that can greatly complicate the generalizability of the results. For certain other clearance pathways, inhibitors are either not known (e.g., UDP-GTs) or are not specific (PgP, OATPs).

There is also concern about what level of assurance might be required from regulators in the case of agents with multiple fractional clearance pathways. For a hypothetical agent with three clearance pathways, e.g., from CYP3A (60%), renal (20%) and CYP2D6 (20%), one might pose a question about risk from KTZ amplification in CYP2D6 PMs with renal failure. If KTZ coadministration and renal failure each have a 1% likelihood in the treated population, and the frequency of 2D6 PMs in the US population is 4%, then the combined likelihood of all occurring, assuming independence, is $4/10^6$. If the agent has been studied for QT effects at 5-fold therapeutic, with estimated effects of 3 and 15 ms at 1X and 5X, would the triple risk study be required, and if conducted, how would it be labeled? Have regulatory agencies decided on a format for reporting QT and risk data for normal usage and under DDI conditions?

For a typical 50 year old, the actuarial risk for all cause mortality (1996 data) is around 0.45%/y. The public, congress, regulators and consumers demand no risk in their pharmaceuticals, but in reality all drugs taken have some order of risk, as do other life activities. It has been estimated that the frequency of TdP is around $1/10^6$ for moxifloxacin. The warning in its labeling reflects incremental risk of $<10^{-6}/0.0045$, or about 0.02% for the average all cause-mortality risk for a 50-year-old. This issue will not be solved with the E14 guidance, but as a matter of policy, has a threshold been defined by regulators for marginal risk?

The example above assumed a frequency of TdP with moxifloxacin use of $1/10^6$. However, this does not speak to causality. As a reference for the E14 document and for perspective, there is a need for quantification of the prevalence of TdP and other polymorphic ventricular arrhythmias (PVA) for various populations of interest. Would the cardiologist experts on the panel or a broader group of experts attempt to quantify the % of the US population that have TdP/PVA as a result of hERG mutations, as a result of

viral infections, as a result of NYHA Class 3 or 4 congestive heart failure, as a result of a previous MI or as a result of other risk factors? In addition to that perspective, one also needs to appreciate the relative risk of QTc prolongation in the absence of confounding risk factors.

Lines 229-231. The ‘thorough QT/QTc study’ would typically be conducted early in clinical development to provide maximum guidance for later trials, although the precise timing will depend on the specifics of the drug under development.

We agree conceptually that signal detection is important “early in clinical development;” however, proceeding to a thorough study is sometimes difficult to achieve until there has been an assessment of metabolic amplification risk, and until the likely therapeutic range is defined. Although translational efforts from preclinical models may provide some sense of target exposures for efficacy, adequate characterization of exposure/response for both efficacy and adverse effects is usually not available until some time in Phase 2; thus, the reference exposure for marketing and general use is not known.

Lines 239-241. For drugs with short half-lives and no metabolites, a single dose study might be sufficient. Studies should characterize the effect of a drug on the QT/QTc throughout the dosing interval.

A large fraction of small molecules currently coming from pharma are metabolized, and have half-lives sufficient to support QD dosing; thus, many agents would default to a multiple dosing study, perhaps without adequate cause. Under the catenary model of parent drug pharmacokinetics, even high clearance metabolites will have a terminal half-life equal to that of parent. We suggest that the wording be changed to read: **“Single dose studies may be sufficient for the cases in which exposures of parent drug and metabolites with plausible electrophysiologic effects exceed those observed at steady state at therapeutic doses. Multiple dose studies, when required, should characterize the effect of a drug on the QT/QTc throughout the dosing interval.”**

We propose that the wording be sufficiently flexible to allow valid scientific arguments to prevail: for example, if drug X produces a circulating glucuronide (unlikely to enter cells) and an oxidative metabolite with hERG activity 10% of parent and an AUC of 50% of that of parent, one should be able to discount the contributions of the metabolites.

Lines 247-249. The ‘thorough QT/QTc study’ should be adequate and well-controlled, with mechanisms to deal with potential bias, including use of randomization, appropriate blinding, and concurrent placebo control group.

“Appropriate blinding” may need some clarification. First, is it agreed that moxifloxacin or other positive controls need not be blinded for the investigator and subjects? For blinding of placebo vs. active comparisons, particularly with examination of higher than normal exposures, it must be recognized that the agent’s pharmacology may unblind the subjects and investigators. Does this necessarily invalidate the ECG findings of the

study? How important is this, compared to the blinding of the over-readers of the ECGs with respect to treatment? For the over-readers of the ECGs, a case can be made that more consistent interval measurements might be possible if the cardiologist is unblinded with respect to subject identity (but still blinded to treatment). This may be important in cases in which the pretreatment ECGs have unusual characteristics (e.g., deviations in QRS or T axes or waveform morphology), or cases in which there are emergent changes in the ECG (e.g., U waves becoming more pronounced)

Lines 253-254. Absence of a positive control should be justified and alternative methods to establish assay sensitivity provided.

We agree that in the near-term that positive controls are valuable, but at some stage, we have to retain flexibility in guidance to get past this hurdle. If a given reference positive control, when studied under certain defined experimental conditions, has 20, 30, or 40 studies across industry, with a 0% false negative rate, and some defined range for central estimate, will it ever be acceptable that future experiments need not employ the reference, if experimental conditions meet some standard? Alternatively, if a large centralized database for the positive control is constructed, is it possible that sponsors might be able to use a reduced cohort size for the control, and employ Bayesian techniques to ensure assay sensitivity?

Granted, there are a number of studies evaluating moxifloxacin 400 mg, with divergent results in terms of maximum and average effect. Aside from showing assay sensitivity, we wonder if the positive control has additional interpretation. If 5 agents from different sponsor's studies have the same estimated effect on QTc, and the range for moxifloxacin 400 mg spans from 6 to 15 ms, will this have an impact on their relative risk, and labeling of such? At what stage will we be able to rely on the statistics of active vs. placebo comparisons, without having to replicate positive control? Anonymized, routinely updated, public posting of moxifloxacin or other positive control results, with some description of study design would be highly valued.

Lines 257-260. On that basis, the positive control (whether pharmacological or non-pharmacological) should be well-characterized and consistently produce an effect corresponding to the largest change in the QT/QTc interval that is currently viewed as clinically not important to detect (a mean change of around 5 ms or less).

It would be valuable to know what is meant by a "non-pharmacological" control. It should be made clear whether the mean change of around 5 ms or less referred to here is the mean over several hours, perhaps a dosing interval, or whether it is the mean peak effect, corresponding to the 'largest time-matched mean difference between the drug and placebo' referred to in the next paragraph of the concept paper.

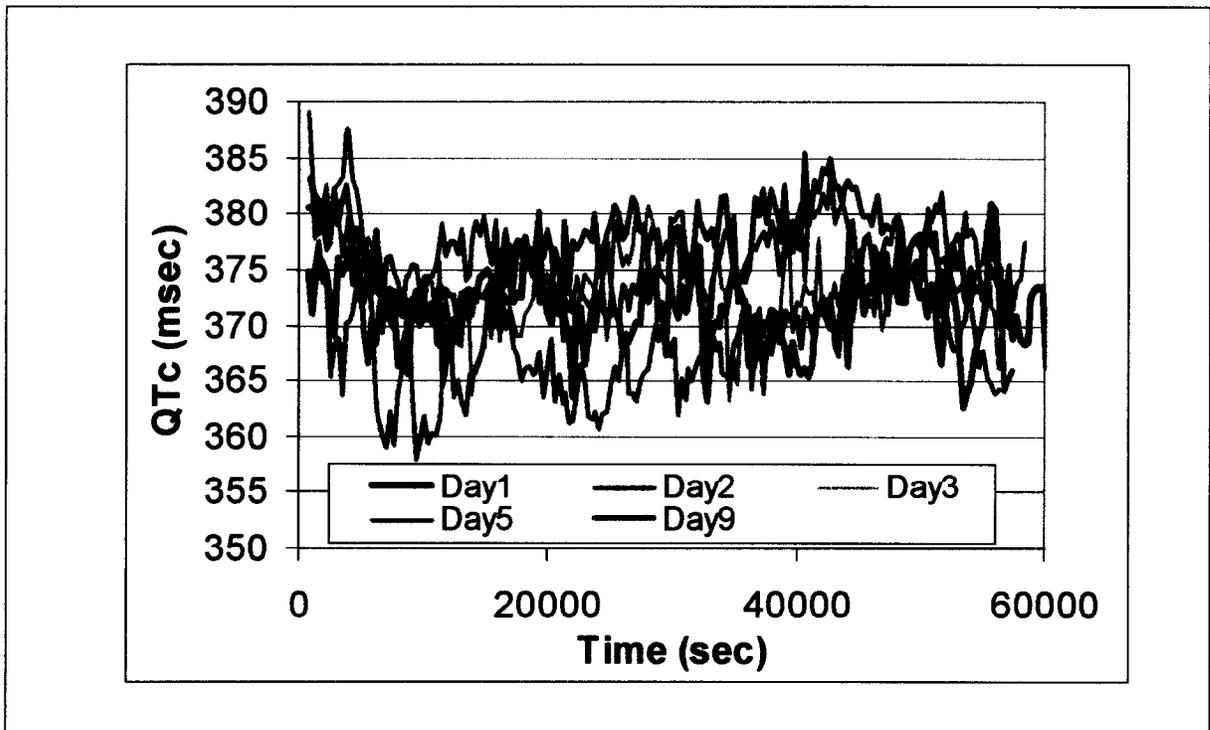
Apparently, many sponsors are using moxifloxacin as a positive control, and would prefer to continue to do so because of the reproducibility of its kinetics. As more data

Docket No. 2004D-0377

emerge, we expect it to become apparent that the peak effect of a single 400 mg moxifloxacin dose exceeds 10 ms.

With respect to the issue of whether a 5 ms mean peak effect is clinically significant, it is highly likely that it is not important in the general population (without pre-existing risk factors), and the threshold could perhaps be at least 10 ms. The FDA and its consultants have access to far more data than we do, and it would be instructive to review documentation of the rationale for 5 vs. 10 ms.

We are particularly interested in the expert consensus about the normal range of QTc within an individual, within day, and across days. From our experience, we believe that normal variation in QT/QTc is under autonomic control and that QTc varies by approximately 20 ms throughout the day for a given individual with the best available correction factor. Shown below is one typical individual under placebo conditions with 5 days of monitoring (~17 hours per day) by ambulatory device; each data point represents a 5-minute, Fridericia-corrected value based on beat-to-beat correction.



On each of the days, the range of values is around 20 ms, and the daily means have a range of 368 to 377 ms. When considering a regulatory threshold of 5 to 10 ms, it is clear that this is easily within the normal range of intra-day and interday variation for an individual. In the individual above, there is a tendency for values to drop somewhat after waking (due in part to loss of the vagal dominance occurring during sleep, but also due to

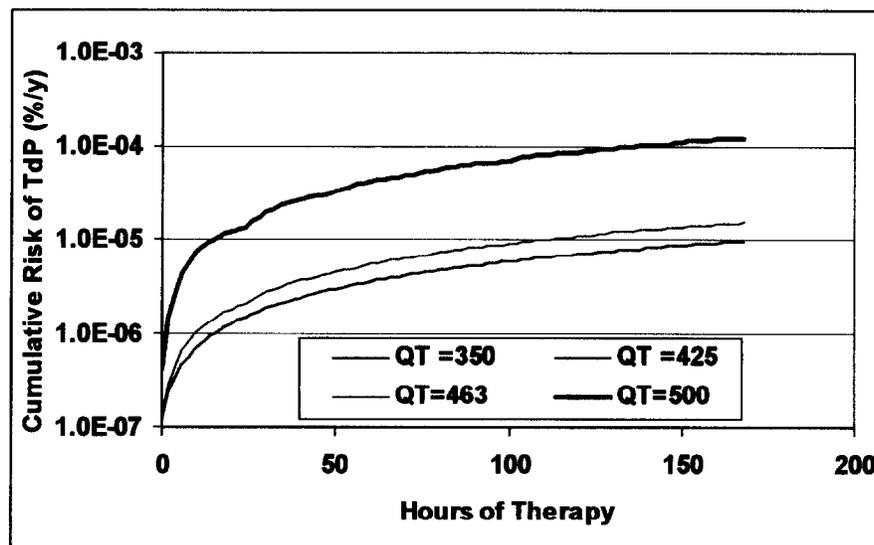
inadequacies in correction equations). Other than that, there is a low order of correlation for time-matched pairs ($R^2 \sim 3\%$ for Day 2 vs Day 1). As more data are analyzed by sponsors and regulatory agencies with refinements in measurement methodologies, the estimates of variability will improve, but it is doubtful that smaller intra-individual fluctuation will be found. We accept that mean diurnal variation for a cohort of volunteers might have reduced fluctuation compared to individuals, but from the perspectives of individual risk, and attempts to define maximal change within a day for an individual, the within-day fluctuation is troublesome. When considering this variability in background “noise,” if one assumes that peak drug concentrations occur at 10,000 sec (and that peak real QT effect occurs at the same time), the measured peak difference in QTc may not occur at C_{max} for a given individual, although the cohort mean might be coincident with C_{max} . For illustration purposes, if one makes time-matched comparisons of Day 1 values as reference, and Day 9 values plus a constant 10 ms signal, the daily mean effect would be 14.5 ms (the difference reflecting day to day variability), with a maximum time-matched effect of 29 ms (at 11 h) and a minimum time matched effect of 0.1 ms. We mention this just to highlight the difficulties in ascertaining maximum effect in the presence of diurnal variation that may not be highly correlated on a day-to-day basis. It is for this reason that we would prefer to focus effect characterization on average effect over a dosing interval or over a prescribed time interval around C_{max} (e.g., interval where concentrations are $> C_{max}/2$). From the pharmacometricians at the regulatory agencies who have access to all the positive control data (agents with reversible effects and no active metabolites), it would be interesting to know the degree of correlation between time of C_{max} and time of maximal effect, both within and across studies. If this is low, and appears to be due largely to underlying random day-to-day variation in the diurnal variation, then more robust and efficient estimation of drug effects may be found for relatively wide time windows around C_{max} .

While a consensus among opinion leaders might exist that QT prolongation less than 10 ms provides negligible marginal risk, the logic and data supporting this have not been translated to a unifying model, although the foundations have been set by Shah (Br. J Clin. Pharm, 2002, 54:188). Selection of the threshold value in the 0-10 ms range for prolongation is somewhat arbitrary, with a value of 0 being the least risk tolerant, but also having a specificity of 0. Considering the normal intraindividual range of QTc of perhaps 20 – 30 ms, selection of a mean effect of 5 ms can be defended as being very intolerant of risk, but our simulations show that it will be virtually impossible to be discriminated from the choice of 10 ms. While we accept the selection of 5 ms for a time-averaged mean effect, we must comment on the specification that **peak** effect needs to be 5 ms and that the UCL needs to be 8 ms or less. If we assume that effect is linearly related to drug concentration, and that typical QD drugs administered chronically have a fluctuation of 2- to 4-fold, then the guidance proposal thus requires that the 24-h average effect needs to be substantially less than 5 ms, because $C_{max} > C_{avg}$.

In addition to lowering the overall threshold for no effect, the specification of time-matched maximum effect as the primary analysis unduly creates a greater statistical

Docket No. 2004D-0377

hurdle since greater precision is afforded by averaging over a larger time interval. An argument might be advanced that some agents might have a very “steep” exposure-effect relationship and that risk near peak is disproportionately higher than at other times in the dosing interval. While there are examples of rate of rise being important (nifedipine CV effects), examples supporting this argument need to be carefully analyzed and openly discussed, particularly with orally administered agents. With IV administration, even greater care needs to be exercised since cardiovascular pharmacologic effects associated with high peak concentrations may have significant transient effects on autonomic tone and thus undermine the use of correction factors based on daily average autonomic tone. Since we will never have data for time-course PK/PD modeling of TdP, should we not adopt the most parsimonious model, and assume that inhibition is reversible, with little lag from the plasma kinetics of the active moieties, and thus instantaneous risk is proportional to exposure, and cumulative risk is dependent on the exposure*duration? An illustration of this is shown below for an agent causing average effects of 17 ms (peak =25 ms) administered for 1 week to individuals with baseline QT values of 350, 425, 463 and 500 ms. The risk equation has E_{max} characteristics, with a baseline risk, and risk enhancement as QT interval gets progressively larger, particularly as QT exceeds the safe range of <470 ms. The PK/PD model assumes that QT prolongation is directly dependent on circulating drug concentrations. From this simulation, one would infer that cumulative risk is mostly dependent on average daily risk and the number of days exposed, and that for oral agents with typical fluctuation indices within a dosing interval, focus on maximal effect may not be the proper metric, particularly if other factors beyond QT prolongation dominate the within-day risk profile.



Lines 262-267. Based on similar considerations, a negative ‘thorough QT/QTc study’ is one where the largest time-matched mean difference between the drug and placebo (baseline-subtracted) for the QTc interval is around 5 ms or less, with a one-sided 95% confidence interval that excludes an effect >8.0 ms. This upper bound was chosen to reflect the uncertainty related to the variability of repeated measurements. As with other data, the presence of outliers (see section 3.2.2) should also be explored.

Simulation studies based upon our data from QT/QTc studies show that with 8 ms as the non-inferiority bound for the peak effect (the maximum difference between investigational drug mean and placebo mean over the several times of measurement), it is quite difficult to get a negative thorough QT/QTc study if the investigational drug has a peak effect of 4 ms. The data that we have access to indicates that the peak effect of a moxifloxacin 400 mg oral dose (the maximum difference between moxifloxacin mean and placebo mean over several times of measurement) exceeds 10 ms. For these reasons, we think that it would be more appropriate for the non-inferiority bound for the maximum drug effect to be 10 ms instead of 8 ms. Information on the simulation studies is in Appendix A. As noted in the previous sections, selection of a mean maximum effect of 5 ms appears to be somewhat arbitrary (less than ½ the peak effect of moxifloxacin in some studies) and can not be distinguished clinically from a **mean average effect** of 5 ms. When considering the choice of an upper confidence bound of 8 ms vs. 10 ms, we are not aware of any expert opinion that the two are somehow clinically different with regard to patient risk. Nonetheless, this 2 ms difference in the UCB has a significant impact on the cost of a thorough study. The authors of the guidance need to clarify whether the objective of the boundary is to ensure that risk is below that for moxifloxacin.

The parenthetical expression ‘(baseline-subtracted)’(line 263) should be deleted or qualified. See the comment on Section 3.2.1, where this expression also appears.

Lines 276-278. They usually call for smaller numbers of subjects than parallel group studies, as the subjects serve as their own controls and hence reduce variability of differences related to diurnal variations and inter-subject variability;

A crossover study has no advantage over a parallel group study for reducing variability due to diurnal variations except as may already be covered by the term ‘inter-subject variability’ that immediately follows the mention of diurnal variations. The effect of consistent diurnal variations is reduced by controlling the times at which ECGs are collected, and the schedule of study activities (e.g., meals) and/or utilizing in the statistical analysis baseline measurements that are obtained at the same time of day as the measurements during treatment. Therefore, the bullet point should be simplified to: **“They usually call for smaller numbers of subjects than parallel group studies, as the subjects serve as their own controls and hence reduce variability of differences related to inter-subject variability;”**

2.2.2 Assessment of Standard 12-Lead ECGs

Lines 343-344. Several methods ..., and for a given trial, the sponsor should describe the accuracy and precision of QT/QTc interval measurements using the selected system.

There are several issues with accuracy and precision. For accuracy, there is no gold standard about the proper procedure for determination of the end of the T wave. There are numerous algorithms for this. If expert A uses a tangent method for 95% return to isoelectric line and expert B uses first derivative of the T wave and another algorithm, and the 2 experts have readings that are extremely highly correlated, but consistently 5 ms different, would the difference matter if both were to conduct the analysis of a thorough QT study? For precision, it is relatively common for different readers, using the same technique, or the same reader for replicate ECGs on different occasions to vary by 20 ms between readings, especially when not operating in an environment in which a computer algorithm sets the initial estimate. On the other hand, computer-based algorithms place the fiduciary marks exactly in the same place, providing the ECGs are electronically captured and software “settings” are identical.

Lines 346-350. At present, this would usually involve the measurement by a few skilled readers operating from a centralized ECG laboratory, although other methods (e.g., semi-automated ECG reading) can be acceptable when appropriately supported. Readers of ECGs should be blinded to time, treatment and subject identifier, and one reader should read all the ECG recordings from a given subject.

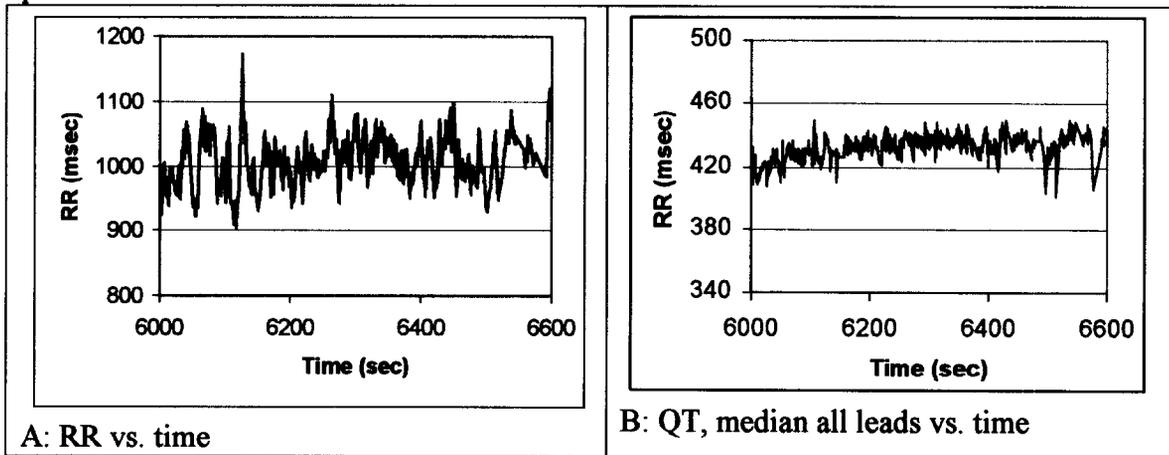
We generally agree with the suggestions in this section. However, it has been our experience (with 3 centralized labs) that in some cases over-readers are capable of very significant errors, due to fatigue, attempting measurements on low amplitude or noisy signals and general subjectivity in marking the end of the interval. Moreover, it is becoming more routine practice that over-readers are already using semi-automated (software) techniques for initial placement of the fiduciary marks. If semi-automated, or fully automated software (with professional review) is used in conjunction with a positive control, and the study results for the positive control are within historical ranges, then all of the experimental techniques, including the interval measurements, should be considered “appropriately supported.”

If readers of ECGs should be blinded to subject identifier, and one reader should read all ECGs for a given subject, then in the cases of multiple readers at a centralized lab, the sponsor will have to sort the ECGs so that both conditions hold. Considering the cases we have observed of unusual waveforms in certain subjects (or certain leads within subject), we are not convinced that knowledge of subject ID is detrimental or produces bias, as long as treatment is blinded. For automated procedures, we are investigating techniques in which the software “learns” about an individual subject’s waveform characteristics under control conditions and thus may be more discriminating under test

conditions. Overall, we would suggest this paragraph be restructured to: **At present, this would usually involve the measurement by a few skilled readers operating from a centralized ECG laboratory, although other methods (e.g., automated or semi-automated ECG reading) can be acceptable when appropriately supported by the behavior for the positive control. Readers of ECGs should be blinded to time and treatment, and one reader should read all the ECG recordings from a given subject.**

Lines 364-365. While the most appropriate lead(s) and methodology to measure the QT interval have not been established, lead II is often used. A consistent approach should be used for a given trial.

We agree that comparisons of methodologies are evolving and that recommendation of a given approach in the guidance should be avoided until fully discussed. Lead II, of course, is often used because the lead II axis closely approximates the typical repolarization axis in a population sense; however, there will be cases of unusual axes deviations in which lead II will not be a high amplitude signal. We and others find that precision is enhanced using the median of multiple leads, as is illustrated in the following 1000 Hz ambulatory data taken over a 10 minute period of relative inactivity, in a sitting position:



Over the interval shown, the RR interval had a CV of 4% (41 ms), whereas the QT interval at lead II had a CV of 1.9% (8.2 ms). When the median of 12 leads with adequate signal quality is analyzed, the CV is reduced to 1.6% (SD=6.8 ms). With individual correction for this subject, QTc(median) on a beat to beat basis has a SD=8.3 ms; however, if a 10-beat average is used for QTc, the variability is reduced to SD=4.8 ms. The tracings above illustrate, with the goal of detecting an average 5 ms signal, that considerable attention be paid to reduction of noise that is inherent in surface ECGs, resulting from measurement errors, and transient shifts in autonomic tone at the SA node and ventricular myocardium (that are not perfectly correlated), and that the best lead(s), or best methodology is far from being defined, although it is clear that noise and signal averaging represents a significant opportunity for enhancing signal to noise ratio.

2.2.3 Ambulatory ECG Monitoring

Lines 379-382. *However, as QT/QTc intervals measured by this methodology might not correspond quantitatively to those from standard surface ECGs, data obtained from the two methodologies might not be suitable for direct comparison, pooling, or interpretation using the same thresholds of concern.*

We understand that resting supine ECGs with bedside equipment might show a somewhat different drug effect than that obtained from ambulatory monitoring, due to a larger range of postural effects and autonomic tone for the latter. However, for modern ambulatory devices and electronic capture, we are not sure the assertion above is otherwise correct (Sarapa et al. A.N.E. 9:48-57, 2004). We will be testing the correspondence of a 12 lead ambulatory device to a standard bedside ECG, all connected to the same leads. We expect to find no difference for the raw signals. If the devices differ at all with respect to signal, it will be the result of internal software smoothing algorithms aimed at noise reduction. The devices might differ in terms of interval estimation, for the algorithms are likely to be different, and assignment of one being “better” than the other is difficult.

Ambulatory monitoring data are probably more clinically relevant than the data from supine ECGs using bedside equipment because the former are closer to real life. It must be recognized that if different hardware, over-reader CRO selection or even different readers at a given CRO is used, one can still raise hypothetical issues about pooling. As noted in subsequent sections, one would not want to pool data from a thorough study with Phase III multicenter data collected 2 years later and probably over-read by different staff, although if all of this is modeled simultaneously, the between-study differences could be interrogated in the model. Ultimately, it may turn out that as long as the data are collected electronically with modern equipment, with close attention to experimental details, the best data to pool are those that are all processed by computer-based automatic or semi-automatic interval measurement (with human oversight for gross errors)—which is the best source of reproducibility across conditions.

3.1.2 Correction Formulae Derived from Within-Subject Data

Lines 445-451. *Corrections for heart rate using individual subject data have been developed, applying regression analysis techniques to individual pre-therapy QT and RR interval data over a range of heart rates, then applying this correction to on-treatment QT values. As adaptation of the QT/QTc interval to changes in heart rate is not instantaneous, care should be taken to exclude ECG recordings collected during times of rapid heart rate changes due to this QT/RR hysteresis effect.*

We support the use of the individualized correction, and would prefer that Bazett’s correction not even be required for regulatory submissions, unless individual or population corrections show that it is the best approach for a given dataset or individuals

within that set. We have examined the distribution of exponents for individually corrected data when the individualized correction is selected from the family of corrections defined by the formula $QT_c = QT/RR^\alpha$. We find the Bazett exponent to be appropriate only in rare cases with healthy volunteers, although the distribution of exponents may be different in some other populations. Considering how we have advanced in our techniques for determination of QT_c , and considering the volume of QT data in the modern era will far exceed the historical data linked to Bazett's correction, we need to abandon this correction when others are superior.

While the advice given about rapid heart rate change is generally appropriate, it is usually not an issue with the collection of supine ECGs after an appropriate rest period and accommodation to the postural change. The issue of hysteresis is complicated and controversial. The key issues are: when are changes in autonomic tone at SA node and ventricular myocardium coordinated, when is there delay, when are differences independent of the length of the RR cycle, and when is the delay differentially dependent on other factors, such as postural changes, exercise and stress level and a host of other factors? In the evaluation of postural changes, hysteresis is clearly evident, but the question arises whether similar delays are observed when there are transient RR spikes and posture is constant. Prof. Malik (J Electrocard 2003; 35:187) has also explored weighted regression approaches to compensate for hysteresis, finding distinct nonlinearities over time that vary considerably across individuals. The root cause of the effect is not at all understood, which raises the issue about the possible pharmacologic effect of drugs invalidating a hysteresis correction factor derived under placebo conditions. A very informative study described by Kowallik et al (J Cardio Electrophysiol, 2000, 11:1063), showed that during sleep, changes in sinus node automaticity (PP) are not necessarily indicative of the autonomic control of ventricular myocardium (QT). Moreover, the data "cloud" of QT vs. RR values for individual beats over a day is composed of families of QT/RR lines that may have different slopes and vertical displacement within the same RR region (bin), depending on current autonomic tone and other factors. Within the same individual on the same day, in some cases hysteresis is apparent, in other cases RR and QT change coordinately without appreciable delay, and in still other cases (for example at slow heart rates), RR may change over the long term, with only minor change in QT. A number of approaches have been investigated to censor or correct non-equilibrated QT and RR intervals, but we are unaware of an emergent consensus about the best approach.

3.2.1 Analyses of Central Tendency

The parenthetical expression '*(baseline-subtracted)*' (line 473) should be deleted or qualified. For a multiple dose crossover study in which the baseline measurements are more than two days before the during-treatment measurements, we have found that analyzing changes from baseline reduces the power. In some cases this may also be true for a single dose crossover study. Note that in a crossover study an analysis of the during-treatment measurements themselves is based upon within subject variability. For

a parallel group study, it is likely that the baseline measurement would be included in the model as a covariate, in which case the test for comparison of the treatments would be the same whether or not the during-treatment values themselves or the changes from baseline are analyzed. If the expression is not deleted, it should be replaced by an expression such as **'(with the respective baseline measurements subtracted from the during-treatment values if desired)'**. As noted before, this expression appears in Section 2.1.2 (line 263), where it also should be deleted or qualified.

More generally, proposals for completely replacing the current text of Section 3.2.1 are provided below. These are motivated by the definition of a negative 'thorough QT/QTc study' that is given in Section 2.1.2 (lines 262-266). The proposals address the fundamental characteristics that must be present whether the study has a crossover design or a parallel group design, but allow for the optimization on more detailed matters for each particular situation and also allow for more powerful statistical methods that may be developed in the future. The proposals are written using 8 ms as the boundary for non-inferiority in accord with Section 2.1.2 of the current draft, but this could be changed if a different boundary point is subsequently chosen.

Proposal 1:

The first proposal is written assuming that the essence of the definition of a negative 'thorough QT/QTc study' that is given in Section 2.1.2 (lines 262-266) will remain. That is, it is assumed that the investigational drug will be assessed on the basis of the maximum time-matched mean difference between the drug and placebo. The proposal follows:

Active drug treatments will be compared to placebo in terms of the differences of means at pre-specified times of measurement relative to the time of dosing. The effect of a drug at a given time refers to the difference in means between the drug and placebo. Here the drug mean and the placebo mean may be a mean change from baseline.

For a study with a positive control to be deemed adequate, the positive control should show an effect consistent with previous clinical trials. Evidence that the active control has an effect on QT/QTc interval must be shown in one of two ways, with the option specified in advance. One option is that the hypothesis of no difference between the active control and placebo be rejected by a one-sided test at significance level 0.05 for one or more of appropriate pre-specified times of measurement relative to dosing (no adjustment for multiplicity). The second option is that for the average over appropriate pre-specified times of measurement, the hypothesis of no difference between the active control and placebo be rejected by a one-sided test at significance level 0.05.

For the results of a 'thorough QT/QTc study' to be declared negative with respect to an investigational drug (i.e., no clinically significant effect associated with

QT/QTc interval prolongation expected), an assessment on non-inferiority will be conducted at a significance level of 0.05. The hypothesis that the maximum effect of the drug over pre-specified times of measurement is ≥ 8 ms must be rejected in favor of the alternative hypothesis that the effect is < 8 ms at all of the pre-specified times. The pre-specified times for the investigational drug need not be the same as the pre-specified times for the assessment of a positive control that may also be one of the treatments in the study

To further characterize the effect of the investigational drug, other analyses may be performed. Time-averaged QT/QTc interval over an appropriate interval of time may be analyzed. The relationship between effect on QTc interval and drug concentration may provide valuable information.

Proposal 2

In comments on Section 2.1.2 of the draft concept paper we suggested that for many or most drugs, the more appropriate criterion for a negative thorough QT/QTc study would be defined in terms of an average over a dosing interval or over a prescribed time interval around C_{max} . The criterion for a given drug must be specified in the protocol for the thorough QT/QTc study. That is, it must be specified in the protocol whether the investigational drug is to be assessed on the basis of the maximum time-matched mean difference from placebo or on the basis of the difference of means for time averaged QTc interval over an appropriate interval of time. If the latter option is chosen, the interval of time must be defined. If the definition of a negative thorough QT/QTc study is modified to allow for the assessment of the investigational drug on the basis of time averaged QTc interval over an appropriate interval of time, we propose the following as a replacement for Section 3.2.1.

Active drug treatments will be compared to placebo in terms of the differences of means at pre-specified times of measurement relative to the time of dosing or the difference of means for time-averaged QT/QTc interval over an appropriate interval of time, with the choice between these two approaches specified in the protocol. The choice may be different for an active control and the investigational drug. The effect of a drug refers to the difference in means between the drug and placebo. Here the drug mean and the placebo mean may be a mean change from baseline.

For a study with a positive control to be deemed adequate, the positive control should show an effect consistent with previous clinical trials. Evidence that the active control has an effect on QT/QTc interval must be shown in one of two ways. One option is that the hypothesis of no difference between the active control and placebo be rejected by a one-sided test at significance level 0.05 for one or more of appropriate pre-specified times of measurement relative to dosing (no adjustment for multiplicity). The second option is that for the average over appropriate pre-specified times of measurement, the hypothesis of no difference

between the active control and placebo be rejected by a one-sided test at significance level 0.05.

For the results of a 'thorough QT/QTc study' to be declared negative with respect to an investigational drug (i.e., no clinically significant effect associated with QT/QTc interval prolongation expected), an assessment on non-inferiority will be conducted at a significance level of 0.05. If the criterion for non-inferiority is defined in terms of the maximum mean difference between the investigational drug and placebo over pre-specified times of measurement, the hypothesis that the maximum effect of the drug over pre-specified times of measurement is ≥ 8 ms must be rejected in favor of the alternative hypothesis that the effect is < 8 ms at all of the pre-specified times. The pre-specified times for the investigational drug need not be the same as the pre-specified times for the assessment of a positive control that may also be one of the treatments in the study. If the criterion for non-inferiority is in terms of time averaged QTc interval over a specified interval of time, the hypothesis that the mean of the investigational drug is ≥ 8 ms larger than the placebo mean must be rejected in favor of the alternative hypothesis that the difference of means is < 8 ms.

To further characterize the effect of the investigational drug, other analyses may be performed. In particular, the relationship between effect on QTc interval and drug concentration may provide valuable information.

3.2.2 Categorical Analyses

The categorical outliers may depend on the selection of the correction factor. It would be valuable if this section would state that categorical analyses should be conducted using the correction factor shown to have the least correlation between QTc and RR. For the statement about 500 ms as representing concern, one should recognize that if this is based on Bazett's corrections, the real threshold for concern might be different. As a matter of policy, if Bazett's QTc correction in a study is shown to be highly correlated with RR, why must we identify the extreme values in the study report, particularly in cases in which the reference and test treatments have different heart rates?

3.3 Morphological Analyses of ECG Waveforms

Granted, a change in U-wave amplitude, or in T-U shape might be meaningful. Would the cardiologists on the panel suggest what amplitude, relative to the preceding T-wave is considered as significant (e.g., $>50\%$ of T height)?

5.1 Relevance of QT/QTc Interval Prolonging Effects to the Approval Process

Lines 640-643. *It is difficult to determine whether there is an effect on the mean QT/QTc interval that is so small as to be inconsequential, but the risk of arrhythmias*

appears to increase with the extent of QT/QTc prolongation. Drugs that prolong the mean QT/QTc interval by around 5 ms or less do not appear to cause TdP.

Although analyses of central tendencies are essential for robust statistical estimation, we believe that we need to move away from the concept that 5, 10 or 20 ms mean changes increase arrhythmia risk for all individuals, and focus more on the margins where risk changes more rapidly with prolongation of a given magnitude. Consider a healthy population with no known risk factors, with a mean QTc of 390 ms and a population SD of 30 ms. As a whole, this group, without taking agents affecting QT, have an extremely low likelihood of TdP. If we segregate the middle 50% of this distribution into 2 groups with QTc between 373 and 390 (25% of all) and 390.1-409 (25%), these subgroups have means of 381 and 400 ms. The means of the two groups are nearly 20 ms different. The risk difference in the groups can't be quantified because it is so low, but probably approaches 0. For members in each group, the intraday variation in QTc is in the order of 20 ms, depending on numerous factors, including autonomic tone changing as a function of wakefulness, meals, and activity, and if monitored over long periods, the range could increase to 30 ms and above, all without quantifiable incremental risk. If we modify these activities, do we believe that risk changes in the short term? If we could change the low group to have a mean QT the same as the high group, including giving a QT interval-prolonging agent, do we believe that the change will "substantially increase the risk of arrhythmic events?" We believe the answer is "no" for the selected population, because the range of QT values is in the region in which 20 ms changes are routine and inconsequential.

We do believe that risk vs. QT curves are complex functions in which there are 3 segments: 1) very short values that have higher risk (e.g., SQTS; QTc<300 ms), 2) the minima "normal" region (340-440 ms?) in which risk changes slowly, if at all, and 3) the "prolonged" region, in which risk changes progressively and then dramatically as QT increase proceeds from 500 to 650 ms. If we return to the original population, and focus on the 6% with QT values between 430 and 440 ms, and ask the question whether a mean 20 ms increase might confer greater risk, we would be more likely to answer affirmatively.

Currently, since we typically deal with small effects with new NCEs, due to signal/noise issues, we can estimate PK variability well, but have some difficulty in estimating the variability in individuals in their sensitivity to a given concentration of drug. That variability is probably relatively large, with "slope" or EC50 CVs greater than 50%. Thus, while 1 µg/mL of an active agent might cause a mean prolongation of 20 ms, the range in slopes could easily encompass 5 to 50 ms/µg/mL, with some individuals perhaps changing minimally, while "sensitive" others, at the edge of the distribution, might experience intervals approaching 500 ms, which probably confers differential risk, particularly in the presence of other underlying risk factors. If we are to succeed in advancing our understanding risk increases from drug effects on QT intervals, a

Docket No. 2004D-0377

probabilistic PK/PD model will need to be constructed, perhaps by collaboration among the stakeholders. The model will contain estimates of variabilities in PK, PK/PD for prolongation, and estimation of the QT-risk curve shapes for the various risk groups.

The last challenge is large, but not insurmountable, and is where we need to focus some of our efforts. It is our impression that $QT > 500$ is certainly a risk factor, but that other risk factors and their interactions may dominate the overall risk profile. Thus a 20 ms change, from 450 to 470 ms in a patient with no other risk factors may have a low probability of a TdP; however, if the patient has hypokalemia, a previous MI, CHF, or an underlying propensity for T-wave alternans, the probability increases substantially. It is probably mostly these cases that appear in the databases as suggestions of proarrhythmic propensities. We have been exploring a probabilistic PK/PD model coupled with hypothetical curve shapes for various risk groups. The results of initial simulations with this model reinforce the guidance statement, that incremental increase in QT will bring an incremental risk, the magnitude of which is probably nonlinearly dependent on the degree of prolongation. For an overwhelming majority of the population, the incremental risk of a 5 ms daily mean change is minute compared to competing risks, and is so small as to be undetectable in any trial. However, in the presence of one or more risk factors, including pretreatment QT value, electrolyte imbalance and underlying heart disease, the risks might be larger, but the risk basis is also higher, and the change is still small compared to the within-day fluctuation.

5.2 Labelling Issues for Drugs that Prolong the QT/QTc Interval

Although there is a well recognized significant issue with physicians reading labelling and reacting to stated risk factors, we would recommend that the expert advisors critically examine this list for important omissions (e.g., all electrolytes rather than just K^+ ; evidence of T-wave alternans; etc.) to compose a complete list and to attempt to rank them in the order of their importance. This should then be universally implemented across all labelling referring to risk factors.

Summary and Additional Recommendations

In conclusion, we agree that the QT/QTc prolongation clinically not important to detect is a **mean maximum change of 5 ms** or less, but suspect that risk does not change appreciably for a mean change of 5 ms over a dosing interval for most drugs. We would further recommend that the upper bound for defining a negative QT/QTc study be increased from 8ms to 10 ms. The difference between the two values is clinically meaningless, in the face of intraindividual diurnal variation in QTc of >20 ms, and an interday range of 10 ms; however, experimentally, the 8 ms boundary results in significant study costs, with some sponsors collecting more than 5000 ECGs from scores of subjects to protect against crossing the boundary.

Since the evaluation of these effects is becoming standard, we would finally also propose that standard labeling text be used in the reporting of results from thorough QT studies. For example: **Panacea's effect on the QTc interval was evaluated in XX subjects at the therapeutic dose and Y times that dose, the latter which provides exposure greater than that observed in any special population or drug interaction study. The steady-state mean QTc prolongation for the two doses were XX (95% UCB=X) and YY (95% UCB=Y). The positive control (XXXXX) in the study produced a QTc prolongation of ZZ ms (95% UCB=Z). Historically, mean prolongations of 5-10 ms have not been associated with clinical risks, 10-20 ms prolongations are of uncertain risk, and >20 ms indicates proarrhythmic potential (or insert Dr. Shah's designations, if the experts agree on his categories). The normal diurnal range in QTc typically is 20-30 ms. Vulnerable individuals, including those with (insert sanctioned LIST of risk factors), may have higher proarrhythmic risks when using agents with QTc effects greater than 10 ms.**

Appendix A

For a crossover study with 64 subjects on a drug with a peak effect of 4 ms, the estimated probability that the results on the investigational drug will be favorable (a negative QT/QTc study) is 54%. In actuality the estimated probability of success is lower than this because the simulations were done with no premature discontinuations, and it was also assumed that the results for the positive control would be acceptable. The simulations were done for a four-period complete crossover study in which the treatments were placebo, positive control, higher dose of investigational drug, and therapeutic dose of investigational drug. The particular simulation reported above was done for a study with 10 times of measurement in which the true differences between the mean for the higher dose of the investigational drug and the placebo mean were 1, 2, 3, 3, 4, 4, 3, 3, 2, and 1 ms, respectively. A crossover study of 96 subjects with no premature discontinuations has an estimated probability of success of 83%.

The simulation results reported above were obtained for a study in which an individualized correction for heart rate is employed. For Fridericia's correction, the estimated probability of success is a little lower.

For the statistical analysis, a linear mixed effects model was used, and the computation was done with SAS Proc MIXED. The model had fixed effects for sequence, period, treatment, the time of measurement (10 levels), and the interaction of treatment and time of measurement. The variance was assumed to be the same for the 10 times of measurement. The covariance structure provided for the correlation of measurements from the same subject, but also allowed for a higher correlation of measurements within a period than for measurements from different periods. The results reported are those for the evaluation of the higher dose of the investigational drug. The study was considered to be a 'negative thorough Qt/QTc study' if it could be concluded from a test conducted at significance level 0.05 that the mean for the higher dose of the investigational drug is less than 8 ms larger than the placebo mean for 10 times of measurement.