E14 Implementation Working Group

ICH E14 Guideline: The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs

Questions & Answers (R1)

Current version
dated 5 April 2012

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use

ICH Secretariat, Chemin Louis-Dunant 15, P.O. Box 195, 1211 Geneva 20, Switzerland
Telephone: +41 (22) 338 32 06, Telefax: +41 (22) 338 32 30
admin@ich.org, http://www.ich.org

In order to facilitate the implementation of the E14 Guideline, the ICH Experts have developed a series of Q&As:

E14 Q&As Document History

Code	History	Date
E14 Q&As	Approval by the ICH Steering Committee under *Step 4*	4 June 2008
E14 Q&As (R1)	Approval by the ICH Steering Committee under *Step 4* of four newly added Questions on: Sex Differences; Incorporating New Technologies; Late Stage Monitoring; and Heart Rate Correction.	5 April 2012

Reference

ICH E14	The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs	May 2005

E14 Questions and Answers

Date of Approval		Questions	Answers
1	**June 2008**	The ICH E14 Guideline emphasizes the importance of assay sensitivity and recommends the use of a positive control. In order to accept a negative 'thorough QT/QTc study', assay sensitivity should be established in the study by use of a positive control with a known QT-prolonging effect. Please clarify how to assess the adequacy of the positive control in the thorough QT study.	The positive control in a study is used to test the study's ability (its "assay sensitivity") to detect the study endpoint of interest, in this case QT prolongation by about 5 ms. If the study is able to detect such QT prolongation by the control, then a finding of no QT effect of that size for the test drug will constitute evidence that the test drug does not in fact prolong the QT interval by the amount of regulatory concern. There are two conditions required for ensuring such assay sensitivity: 1. The positive control should show a significant increase in QTc; i.e., the lower bound of the one-sided 95% confidence interval (CI) must be above 0 ms. This shows that the trial is capable of detecting an increase in QTc, a conclusion that is essential to concluding that a negative finding for the test drug is meaningful; 2. The study should be able to detect an effect of about 5 ms (the QTc threshold of regulatory concern) if it is present. Therefore, the size of the effect of the positive control is of particular relevance. With this aim, there are at least two approaches: a. To use a positive control showing an effect of greater than 5 ms (i.e., lower bound of a one-sided 95% CI > 5 ms). This approach has been proven to be useful in many regulatory cases. However, if the positive control has too large an effect, the study's ability to detect a 5 ms QTc prolongation might be questioned. In this situation, the effect of the positive control could be examined at times other than the peak effect to determine whether an effect close to the threshold of regulatory concern can be detected; b. To use a positive control with an effect close to 5 ms (point estimate of

			the maximum mean difference with placebo close to 5 ms, with a one-sided 95% CI lower bound > 0). In using positive controls with smaller effects it would be very important to have a reasonably precise estimate of the drug's usual effect. Importantly, whatever approach is used, the effect of the positive control (magnitude of peak and time course) should be reasonably similar to its usual effect. Data suggesting an underestimation of QTc might question the assay sensitivity, thus jeopardizing the interpretability of the thorough QT study results.
2	**June 2008**	Please discuss who should read electrocardiograms (ECGs), including the number and training of readers and the need for readers to be blinded.	The document recommends that the reader should be skilled but does not identify specific training that is needed. A technician reading with cardiologist overread would certainly be consistent with the guidance. The attempt of the guidance to limit the number of readers represented an attempt to increase consistency. The guidance asks for assessment of intra- and inter-reader variability and suggests "a few skilled readers" (not necessarily a single reader) to analyze a whole thorough QT study, since many readers may increase variability. Training would be another way to improve consistency. It is recommended for the thorough QT Study that core ECG laboratories blind subject, time and treatment in order to reduce potential bias. The T wave analysis, which calls for all 12 leads, can be performed after the QT analyses, and requires comparison to the baseline ECG; it can, however, be blinded as to treatment.
3	**June 2008**	There are recognized differences in the baseline QTc between men and women. These were noted in early versions of the guidance. In E14, however, it is recommended that outliers be categorized as >450,>480 and >500ms, regardless of gender. Can you say why there is no gender difference in the recommendation?	The 450, 480, and 500 ms categories refer to the values the E14 document suggests sponsors might use in characterizing outliers. The numbers previously specified for males and females referred to "normal" QTc values, which may differ for men and women. This section was not included in the final document, however, and such considerations would be largely irrelevant to larger durations (e.g., 480, 500 ms). As the thorough QT/QTc study is designed to examine the propensity of a drug to prolong the QTc interval, it is appropriate to perform the study in male or female healthy volunteers.

4A	**June 2008**	What is the position of ICH regarding the role of the following reading methods in the thorough QT/QTc study and other clinical trials? • fully manual • fully automated • manual adjudication (manual over-read, computer-assisted, semi-automated)	The techniques currently in use for the measurement of ECG intervals can be classified into three broad categories: fully manual, fully automated, and manual adjudication. Within each of these general categories, many different methodologies are subsumed that differ in terms of lead selection, the conventions used for defining T wave offset, and the criteria for the inclusion and exclusion of U waves. ECG readings can be performed on the following waveform presentations: • Raw waveforms: ECG waveforms recorded from a single lead; • Representative waveforms (median beats, reference cycles): Compositional waveforms constructed by a computer-based averaging process that involves aligning and combining data from all dominant, normally conducted raw ECG waveforms from a single lead; • Global Waveforms: Composite representation of cardiac electrical activity constructed by superimposing representative waveforms from all or several simultaneously recorded leads to form a spatial-vector complex, by weighted averaging of individual representative complexes with low noise and long duration, or by other methods. **Fully Manual** When using a fully manual reading technique, a human reader is responsible for examining the ECG waveform and placing the fiducial points to mark the beginning and the end of the intervals, without the assistance of a computer algorithm. Fully manual methods of fiducial point placement can be applied to raw, representative, and global waveforms. When fully manual measurements are made from the raw ECG waveforms in a single lead, three or more cycles should be averaged where available to produce the final determination of interval duration. An advantage of this approach is that the reader will not be influenced by prior computer placement of the fiducial points, but a weakness can be inter- and intra-reader variability, especially when measurements are performed over an extended time period (e.g., several months). Laboratories using manual reading techniques should observe standard operating procedures based on

			prospectively defined criteria for determining where the fiducial points should be placed. All readers in the laboratory should be trained in the consistent application of these criteria. **Fully Automated** Fully automated reading methods rely entirely upon a computer algorithm for the placement of the fiducial points and the measurement of the ECG intervals. Automated ECG interval measurements can be performed on raw, representative, or global ECG waveforms. Most digital electrocardiographs are equipped with algorithms that perform measurements on global waveforms. Although automated methods have the advantage of being consistent and reproducible, they can yield misleading results in the presence of noise or when dealing with abnormal ECG rhythms, low amplitude P or T waves, or overlapping U waves. The techniques used for construction and measurement of representative waveforms and global waveforms vary between different computerized algorithms and between different software versions within individual equipment manufacturers. As a result, between-algorithm and within-manufacturer variability of fully automated measurements can confound serial comparisons when the equipment or algorithm is not constant. **Manual Adjudication (Manual Over-Read/Computer-Assisted/Semi-Automated)** The manual adjudication approach refers to reading methods in which a computer algorithm is responsible for the initial placement of the fiducial points on the ECG waveform. A human reader subsequently reviews the algorithmic placement of the fiducial points, performing adjustments wherever the computerized measurements are considered to be inaccurate. This approach can have the advantage of greater consistency and reproducibility than fully manual readings, while providing an opportunity to correct any mistakes made by the algorithmic readings. Laboratories using manual adjudication techniques should observe standard operating procedures based on prospectively defined criteria for determining when fiducial points should be corrected. All readers in the laboratory should be trained in the consistent application of these criteria. The adjudication

			procedure should normally be performed on all waveforms being used for interval determination. If an alternative approach is used, such as adjudication limited to outlier intervals above and below a reference range, this methodology should be validated as described in Question 4B. The ICH E14 Guideline currently recommends either fully manual or manual adjudication approaches for clinical trials in which the assessment of ECG safety is an important objective, such as the thorough QT/QTc study. When the thorough QT study is positive, fully manual or manual adjudication methods are currently recommended for an adequate sample of patients in late phase studies (see Section 2.3 in E14 document). When the thorough QT/QTc study is negative, routine ECG safety assessments in late phase clinical trials using fully automated reading methods would be adequate.
4B	**June 2008**	The ICH E14 Guideline contains the following statement: 'If well-characterized data validating the use of fully automated technologies become available, the recommendations in the guidance for the measurement of ECG intervals could be modified.' What would be expected of a sponsor that wished to validate and apply an automated reading method for regulatory submissions?	Efforts to develop more sophisticated and reliable methods for automated ECG readings for both QT interval and T wave morphology assessment are encouraged. There are at present no large scale studies to validate the use of fully automated reading methods in patients; however, there are examples of thorough QT/QTc studies in healthy volunteers in which automated methods have been used and validated for QT interval measurements against manual methods. **QT Interval measurement** There are at present no clear and widely accepted criteria for validation of new semi-automated or automated methods, but it is expected that each would be validated independently for its ability to detect the QT/QTc prolongation effects of drugs that are near the threshold of regulatory concern. Data supporting the validation of a new method should be submitted and could include descriptive statistics, Bland-Altman plots of agreement, superimposed plots of the baseline- and placebo- adjusted QTc and the RR as a function of time, together with data from any trials that have employed the method. **T wave morphology assessment** The suitability of automated ECG reading techniques for the assessment of

			morphological abnormalities has not yet been demonstrated. If a sponsor intends to develop a fully automated approach, without visual assessment for morphological changes, validation studies should include a demonstration that the automated method is capable of reading and interpreting a test set of abnormal ECGs correctly (e.g., abnormalities of T wave morphology, overlapping U waves). As with methods for QT interval determination, there are at present no clear and widely accepted criteria for validation of novel methods. Because changes in morphology can affect interval measurement, fully manual or manual adjudication (as defined in Question 4A) techniques should be performed if treatment-emergent changes in morphology are observed. If, on the other hand, no morphology changes are observed, this would support the use of automated methodologies, provided they have been validated.
5	**June 2008**	In ICH E14, the recommended metric to analyze for a cross-over study is the largest time-matched mean difference between the drug and placebo (baseline-adjusted) over the collection period. Please discuss the most appropriate metric to assess a drug's effect on QT/QTc interval when the data are collected in a placebo-controlled parallel design study (i.e., when there is no corresponding placebo value for each patient).	Regardless of the study design, "the largest time-matched mean difference between drug and placebo (baseline-adjusted)" is determined as follows: The mean QTc for the drug (i.e., averaged across the study population) is compared to the mean QTc for placebo (averaged across the study population) at each time point. The "largest time-matched mean difference between drug and placebo" is the largest of these differences at any time point. The term "baseline-adjusted" in ICH E14 implies that the baseline data are taken into account in the statistical analysis. Differences in baseline assessment between cross-over and parallel design studies are discussed in Question 6.
6	**June 2008**	Please discuss the need for baseline measurements, and when needed, how they should be collected, for cross-over and parallel design thorough QT studies.	Adjustment for baseline measurements is potentially useful for several purposes, including detection of carry-over effects, reducing the influence of inter-subject differences and accounting for diurnal effects such as those due to food. There is no single best approach for baseline adjustment, but all planned baseline computations should be prospectively defined in the clinical trial protocol. Two kinds of baseline are commonly used: "time-

			matched" baseline (taken at exactly the same time-points on the day prior to the beginning of treatment as on the treatment day) and "pre-dose" baseline (taken shortly prior to dosing). The "pre-dose" baseline is used for adjustment for inter-subject differences but not for diurnal effects. The choice of baseline is influenced by whether the study is parallel or cross-over. For a parallel-group study a time-matched baseline allows the detection of differences in diurnal patterns between subjects that would not be detected by a predose baseline. In a parallel study a 'time-matched' baseline day, if performed, would ideally occur on the day before the start of the study. In contrast, in a cross-over study a time-matched baseline is usually not necessary because adjustments for subject- and study-specific diurnal variation are implicit by design in the assessment of time-matched drug-placebo differences in QT/QTc effect. The "pre-dose" baseline is therefore usually adequate for cross-over studies. Obtaining replicate ECG measurements (for example, the average of the parameters from about 3 ECGs) within several minutes of each nominal time point at baseline and at subsequent times will increase the precision of the estimated changes in QT/QTc effect.
7	**June 2008**	Please clarify the need for blinding the positive control in the thorough QT study.	The use of a double-blinded positive control does not appear to be essential, provided that the reading of ECGs is performed in a blinded manner as described in Question 4A and the study is carefully designed to ensure that specified study procedures are followed uniformly. This means that the same protocol for administering the test drug and placebo, taking blood samples and collecting the ECG data should also be used when giving the positive control. This does not mean that other aspects of the study, such as the duration of treatment with the positive control and the other treatment groups, would be identical. If blinding of the positive control is performed, common methods include the use of double-dummy techniques and over-encapsulation.
8	**April 2012**	Should we enroll both sexes in a thorough QT study, and does the study need to be	Post-pubertal males have lower heart-rate corrected QT intervals than do pre-pubertal males or females generally. Women are generally smaller than

		powered for independent conclusions about each sex?	men, so their exposure to a given fixed dose of a drug will generally be higher, and, if a drug prolongs QT, it can be expected to prolong it more in women because of the higher exposure. It is not settled whether and how often there are sex differences in response to QT-prolonging drugs not explained by exposure alone. The thorough QT study is primarily intended to act as a clinical pharmacology study in a healthy population using a conservative primary objective defining the drug's effect on QT. It is unlikely that any of a variety of baseline demographic parameters would introduce a large difference in QT response to a drug in subpopulations defined by factors such as age, co-morbidity, and gender that is not explained by exposure. It is encouraged, but not mandatory, to include both men and women in the thorough QT study. Analyses of concentration response relationship by sex can be helpful for studying the effect of the drug on QT/QTc interval in cases where there is evidence or mechanistic theory for a gender difference. However, the primary analysis of a thorough QT study should be powered and conducted on the pooled population. If the primary analysis is negative and if there is no other evidence suggesting gender differences, subgroup analysis by sex is not expected.
9	**April 2012**	How does a sponsor incorporate new technology or validate new methodology into the measurement and/or analysis of the QT interval?	The ICH process is better suited to the determination of regulatory policy once the science in a particular area has become more or less clear. In general, it is not well-suited to the qualification or validation of new technology. Sections 2.5.1 and 2.5.2 of the ICH E14 Guideline are rather discouraging about methodology outside conventional carts and human-determined measurements. Since ICH E14 was issued, 12-lead continuous recording devices have largely supplanted cart recorders in thorough QT studies without a formal validation process because of their performance in the context of a positive control. The impact of other innovative technologies can be assessed in studies incorporating a positive control. While some technologies could be assessed using other techniques in the absence of a positive control, this is more complex and beyond the scope of this Q&A.

			12-lead continuous recording devices and other new technologies can be used in late phase clinical trials. Even though a positive control is not used in late stage studies, the new technology could be validated in other studies (such as the thorough QT study). In cases where a thorough QT study is not done, a sponsor can provide alternative methods for validating the technology.
10	**April 2012**	The ICH E14 Guideline describes in Section 2.3 (Clinical Trial Evaluation After the 'Thorough QT/QTc Study') that "adequate ECG assessment to accomplish this [monitoring] is not fully established." Is there now a reasonable approach to evaluating QTc in late stage clinical development in the case of a finding of QT prolongation prior to late phase studies?	**Clarification of Approach to Evaluating QTc in Late Stage Clinical Development** The purpose of a thorough QT study is to characterize the effect of the drug on ventricular repolarization (QT interval). It is not the purpose of the thorough QT study to assess the risk of torsade de pointes (TdP) in the target population, but rather to determine whether further data are warranted to assess risk. A finding of QT prolongation above the regulatory threshold of interest (a positive thorough QT study) might call for further electrocardiographic follow up in late phase studies. The extent of the follow up would be affected by the magnitude of the estimated prolongation at doses and concentrations at which this occurs. If prolongation is substantial at concentrations expected to occur in clinical studies, it is important to protect patients in later trials and to obtain further information on the frequency of marked QT prolongation. In some cases in which there is a large margin of safety between therapeutic exposures and the exposures that result in significant ECG interval changes, an intensive ECG follow-up strategy might not be warranted. The recommended intensity of the monitoring and assessment in late-stage trials will depend on: A. The magnitude of QTc prolongation seen in the thorough QT study or early clinical studies. B. The circumstances in which substantial QT prolongation might occur (i.e., in ordinary use or only when drug concentrations are markedly increased [e.g., by renal or hepatic impairment, concomitant medications]). C. Pharmacokinetic properties of the drug (e.g., high inter-individual

			variability in plasma concentrations, metabolites). D. Characteristics of the target patient population that would increase the proarrhythmic risk (e.g., structural heart disease), E. The presence of adverse effects that can increase proarrhythmic risk (e.g., hypokalemia, bradycardia, heart block). F. Other characteristics of the drug (e.g., pharmacodynamics, safety pharmacology, toxicology, drug class, hysteresis) The following examples delineate the scope of recommended ECG investigations based on outcome of the thorough QT study or early clinical studies. These could be modified by other factors such as A-F above. **Examples of ECG Monitoring in Late Stage:** 1. The thorough QT study results in a negative finding as defined by the E14 criteria***** at the therapeutic dose, but the supratherapeutic dose (relative to Phase 3 dose) shows mean QTc effects between 10 and 20 ms. If there is reasonable assurance that the higher dose represents drug exposures that are unlikely to be seen in the patient population, only routine ECG monitoring is recommended in late phase trials. This approach provides reassurance for safety because patients are unlikely to experience a clinically significant QTc effect. 2. The thorough QT study results in a positive finding as defined by the E14 criteria***** at the therapeutic dose, with a mean prolongation <20 ms. For drugs with this magnitude of effect on the QTc interval, intensive monitoring of phase 3 patients is called for. Intensive ECG monitoring in clinical trials has two main objectives. One objective is to provide protection to patients who might have large worrisome QT intervals > 500 ms. A second objective is identifying the frequency of marked QT increases (e.g., prolonged QT >500 ms or increases in QTc >60 ms). Given the limitations of collecting ECGs in late stage trials, the focus of the analysis is on outliers, not on central tendency. Other than descriptive statistics, detailed statistical analysis is not expected. This

			monitoring is intended to be performed locally, without the involvement of a central core laboratory. The timing of ECG collection should be based on the known properties of the drug. All patients should receive baseline, steady-state, and periodic ECGs during the trial. In addition, ECGs should be collected around Tmax at the first dose and/or around steady state in a subgroup of patients or in dedicated studies. ECG collection at around Tmax is not important for drugs with low fluctuations between peak and trough concentrations. If the drug shows a delayed effect in QT prolongation, then the timing of ECG collection should reflect this delay. 3. The thorough QT study results in a negative finding as defined by the E14 criteria***** at the therapeutic dose, but the supratherapeutic dose shows a mean effect between 10 and 20 ms. If supratherapeutic exposure is anticipated at the clinical dose only in a well-characterized subgroup, intensive monitoring as described in Example 2 above could be carried out in this subset of the phase 3 population. In this case, there should be reasonable assurance that the higher exposure is unlikely to be seen in the general patient population. In contrast, if people in the general patient population (who can not be readily identified in advance) will in some cases achieve this higher exposure, intensive ECG monitoring in the phase 3 population is expected, as in Example 2. 4. The therapeutic dose results in a mean QTc prolongation of > 20 ms. For drugs with large QTc prolongation effects, intensive ECG assessment would be appropriate in all patients in phase 2/3. Because of the risk of TdP, another important use of ECG monitoring in late phase trials would be to assess any risk mitigation strategies (e.g., electrolyte monitoring, dose reduction strategies). Additional ECG assessment over and above what is recommended earlier in the Q&A might also be called for (e.g., 24 hour ECG recording, telemetry, multiple trough ECGs through steady state). The sponsor is encouraged to discuss these approaches with the relevant

			regulatory agency(ies) prior to initiation of the Phase 3 program. *A negative study as defined by the ICH E14 criteria is an upper one-sided 95% CI of QTc prolongation effect <10 ms.
11	**April 2012**	The ICH E14 Guideline states that QT interval corrected by Fridericia's and Bazett's correction should be submitted in all applications; is this still necessary? Is there a recommended approach to QT correction that is different than that specified in ICH E14?	Changes in heart rate could variably influence a drug's effect on repolarisation (i.e., QT interval) and correction methods with different characteristics are often applied. The principles set below would be applicable in all clinical studies (thorough QT or other studies). In adults, Bazett's correction has been clearly shown to be an inferior method of correcting for differences in heart rate among and within subjects. Therefore, QT interval data corrected using Bazett's corrections is no longer warranted in all applications unless there is a compelling reason for a comparison to historical Bazett's corrected QT data. Presentation of data with a Fridericia's correction is likely to be appropriate in most situations but other methods could be more appropriate. There is no single recommended alternative (see Question & Answer 9 on Incorporating New Technologies), but the following are some considerations. 1. Analyses of the same data using different models for correcting QT can generate discordant results. Therefore, it is important that the method(s) of correction, criteria for the selection of the method of correction, and rationale for the components of the method of correction be specified prior to analysis to limit bias. Model selection ought to be based on objective criteria and consider the uncertainty in parameter estimates. Alternative methods of correction should be used only if the primary method fails the pre-specified criteria for selection of the method of correction. 2. Corrections that are individualized to a subject's unique heart rate QT dynamic are not likely to work well when the data are sparse or when the baseline data upon which the correction is based do not cover at least the heart rate range observed on study drug.