Design and Analysis Issues with QT/QTc Studies: a Regulatory Viewpoint

Workshop on QT/QTc prolongation
German Region of the International Biometric Society

Heidelberg, Germany
November 18, 2005

Stella G. Machado
Office of Biostatistics/OPaSS/CDER/USFDA
Disclaimer

- This presentation reflects my views and not necessarily those of the U.S. Food and Drug Administration
Outline

- Introduction: Current regulatory situation
- Design considerations for Thorough QT/QTC clinical trials
- QT/QTC interval data considerations
- ICH E14 requirement and Hypothesis testing
- Example
- Other issues to consider
- Closing remarks
Introduction: Current Regulatory Situation

From the ICH E14 Step 4 Guidance (May, 2005)

A ‘Thorough QT/QTc Study (TQT)’ of a drug is a single clinical trial, conducted early in development, dedicated to evaluating the effect of the drug on cardiac repolarization, as detected by QT/QTc prolongation.

Prolonged QT may be associated with Torsades de Pointes, which may be fatal

If the TQT trial shows QT prolongation, additional evaluation in subsequent clinical studies should be performed
“A negative thorough QT/QTc study is one in which the upper bound of the 95% one-sided confidence interval for the largest time-matched mean effect of the drug on the QTc interval is $\leq$ 10ms. This definition is chosen to provide reasonable assurance that the mean effect of the study drug on the QT/QTc interval is not greater than around 5ms, which is the threshold level of regulatory concern”

* time-matched mean effect at each time point after dosing is the difference in the QT/QTc interval between the drug and placebo (baseline adjusted).
Introduction, continued

Other E14 analyses:

- Categorical analyses: tabulations and comparisons between treatment groups of absolute and baseline adjusted changes in QT/QTc

- PK/PD approach under active investigation: using modeled relationship between Drug Exposure and QT/QTc Interval changes to assess QT/QTc prolongation
Design Considerations for Thorough QT/QTc Trials

- The thorough QT/QTc study should be adequate and well controlled, randomized and blinded, with a placebo control group.

- Subjects are usually healthy, males and females.

- Use of a positive control is recommended. The positive control should have an effect on the mean QT/QTc interval of about 5 ms.

- The test drug is studied at the anticipated therapeutic dose(s). If possible, it is studied also at a supra-therapeutic dose.
Design: continued

- Designs may be crossover or parallel. The studies may be of single or repeated dosing.
  - Crossover studies have the advantages of increased efficiency (smaller numbers of subjects needed).

- Parallel studies may be preferred:
  - for drugs (and/or metabolites) with long elimination half-lives
  - when carry-over effects may be anticipated
  - if many dose levels or treatment groups are to be compared (due to economics of time versus number of subjects)
Example: frequently used cross-over design

- Treatments are DL, DH (low and high doses of drug), P (placebo), and PC (positive control, eg, Moxi 400mg)
- A Williams’ square

<table>
<thead>
<tr>
<th>Group</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>DL</td>
<td>DH</td>
<td>P</td>
<td>PC</td>
</tr>
<tr>
<td>B</td>
<td>DH</td>
<td>PC</td>
<td>DL</td>
<td>P</td>
</tr>
<tr>
<td>C</td>
<td>P</td>
<td>DL</td>
<td>PC</td>
<td>DH</td>
</tr>
<tr>
<td>D</td>
<td>PC</td>
<td>P</td>
<td>DH</td>
<td>DL</td>
</tr>
</tbody>
</table>
**QT/QTc Interval Data Considerations**

QT interval data have high intrinsic inter- and intra-subject variability. Control of sources of variability is a key consideration when choosing the design, and methods of analysis.

Examples of inter-subject variability: gender, age, activity, food intake.
Example of intra-subject variability: circadian rhythm.

1. QT interval measurements are adjusted for heart rate (to give QTc) because QT and HR are strongly inversely correlated.

Population corrections: frequently used is Fridericia’s correction: 
\[ QTc = \frac{QT}{RR^{0.33}} \]

Individual corrections: based on individual linear fits of QT with RR, or individually determined correction factors \( \gamma_i \) so that  
\[ QTc = \frac{QT}{RR^{\gamma_i}} \].
Data considerations, continued

2. Replicated QT/QTc measurements at each time point (eg, 30-60 seconds apart) are recommended; use of the average at each time point increases precision in subsequent analyses. Typically 3-6 replicates are obtained.

3. QT/QTc measurements for the drug and placebo arms are obtained at the same time points, over each dosing interval; this is time-matching.
Data considerations, continued:

4. Baseline adjustments:
   baseline QT/QTc measurements are typically either
   
   a. measurements taken at matching time-points on day prior to drug treatment
   or
   
   b. measurements taken at time 0, before dosing on same day just prior to treatment.

   In XO studies, baselines are obtained prior to each period

   Some studies have baseline measurements from many days before the study – in these cases baseline adjustment may cause a decrease in precision, not an improvement.
Baseline-subtracted drug-placebo difference in QTcF for 10 subjects
ICH E14 Requirement and Hypothesis to Be Tested

“A negative thorough QT/QTc study is one in which the upper bound of the 95% one-sided confidence interval for the largest time-matched mean effect of the drug on the QTc interval is ≤ 10ms”

⇒

Need to demonstrate that the population difference in QT/QTc response between drug and placebo is less than 10ms at all times in the dosing interval. **Response** is usually change in QT/QTc from baseline.

Let \( \Delta(D)_i \) = population QTc response to drug D at time i; \( \Delta(P)_i \) for Placebo. Then the hypothesis to be tested is:

**H0:** \( \Delta(D)_i - \Delta(P)_i \geq \delta \) for at least one i vs. the alternative

**HA:** \( \Delta(D)_i - \Delta(P)_i < \delta \) for all i

where \( \delta \) = the non-inferiority bound, i.e., 10ms.

Rejection of H0 leads to acceptance that the drug does not prolong QT/QTc.
Hypothesis Testing

Usual course of action is to estimate the population means, \( \Delta(D)i \) and \( \Delta(P)i \), and \( \Delta\Delta i = \Delta(D)i - \Delta(P)i \), by the observed means, \( \Delta\_\text{hat}(D)i \), \( \Delta\_\text{hat}(P)i \), and their difference, \( \Delta\_\text{hat}(D-P)i = \Delta\_\text{hat}(D)i - \Delta\_\text{hat}(P)i \).

Two-sided 90% confidence bounds for the true value of \( \Delta\Delta i \) are calculated by

\[
\{ \Delta\_\text{hat}(D-P)i - t_{f,1-\alpha/2} \ S, \ \Delta\_\text{hat}(D-P)i + t_{f,1-\alpha/2} \ S \}
\]

where \( t_{f,1-\alpha/2} = \) upper 100\((1-\alpha/2)\) percentile of the t-dist.(f d.f.), \( \alpha = 0.1 \), \( S = \sqrt{\sigma^2_{\text{hat}}/N} \), the estimated standard error of \( \Delta\_\text{hat}(D-P)i \).

- Test procedure: if the upper bounds are <10 ms for all times \( i \), then \( H0 \) is rejected.

- This is the \textbf{Intersection-Union test}, at the 5% level of significance.
Testing, continued

- this approach is referred to as the max-mean approach

- the particular formula for $\sigma^2_{\text{hat}}$, which measures the inherent variability of the metric being used to assess QT prolongation, depends on the study design (XO or PLL, baseline adjusted or not) but not on the number of subjects. Thus cross-study comparisons in error control may be done.
Characteristics of the Intersection-Union Test

- The Type 1 error is maintained at 5%. This is the probability of a “false negative” conclusion, i.e. concluding absence of QT effect when it exists.

- The Type 2 error (the “false positive” rate) i.e. the probability of concluding a QT effect when there is none, is not straightforward to calculate. It would be good to keep it at 20% or less (power 80% or more).

- The Type 2 error depends on:
  - shape of the true time-profile of $\Delta \Delta$ vs. time (height of peak, and extent of peakedness)
  - number of subjects, N, and design (XO or PLL)
  - sources and extent of variability: intra and inter-subject variability, and correlations between successive observations over time.
  - number of time points of observation: often more timepoints $\Rightarrow$ lower power
Intersection-Union Test: continued

Ongoing research to understand and devise approaches to quantify the Type 2 error for different study designs and drug action

- Several groups doing simulations, with various assumptions about the shape of the $\Delta \Delta$ – time profile, error structure, etc.
- Muirhead and Eaton (ref, 2005) produced a formula for the power of the IU test, under certain assumptions, and a useful lower bound, for the simplified case of independence between successive measurements.
EXAMPLE
Analysis of a placebo-controlled study of 2 drugs, $D_1$ and $D_2$ at low and high doses, and positive control, Moxifloxacin 400

**Study Basics**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>57</td>
</tr>
<tr>
<td>Design</td>
<td>XO</td>
</tr>
<tr>
<td>Washout period</td>
<td>3+ days</td>
</tr>
<tr>
<td>Number of periods/sequences</td>
<td>6/6</td>
</tr>
<tr>
<td>Treatment regimens/arms</td>
<td>Single dose oral: Placebo, $D_1$low, $D_1$high, $D_2$low, $D_2$high, Moxi400</td>
</tr>
<tr>
<td>Definition of baseline</td>
<td>Same-day: average of 18 meas in 15 to 0 mins before dosing</td>
</tr>
<tr>
<td>measurement</td>
<td></td>
</tr>
<tr>
<td>Number of replications at/close to each time point</td>
<td>6</td>
</tr>
<tr>
<td>Timepoints of measurement, post-dose</td>
<td>0.5, 1.0, 1.5, 2.5, 4.0</td>
</tr>
<tr>
<td>Dropout information</td>
<td>3 subjects dropped due to incomplete data</td>
</tr>
</tbody>
</table>
Table 1  Mean differences from placebo, baseline adjusted, by time, for QTcF, for 5 treatment comparisons versus Placebo. N=57 subjects.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Time</th>
<th>Estimate</th>
<th>Std.</th>
<th>Lower 5%</th>
<th>Upper 95%</th>
<th>( \hat{\sigma} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1low vs. Placebo</td>
<td>0.5</td>
<td>4.452</td>
<td>0.763</td>
<td>3.176</td>
<td>5.729</td>
<td>5.764</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>7.745</td>
<td>0.856</td>
<td>6.313</td>
<td>9.176</td>
<td>6.464</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>5.028</td>
<td>0.918</td>
<td>3.493</td>
<td>6.563</td>
<td>6.931</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>4.078</td>
<td>0.778</td>
<td>2.777</td>
<td>5.379</td>
<td>5.877</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.721</td>
<td>0.901</td>
<td>2.215</td>
<td>5.227</td>
<td>6.8</td>
</tr>
<tr>
<td>D1high vs. Placebo</td>
<td>0.5</td>
<td>7.248</td>
<td>0.852</td>
<td>5.823</td>
<td>8.673</td>
<td>6.433</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>9.757</td>
<td>0.822</td>
<td>8.383</td>
<td>11.131</td>
<td>6.204</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>7.891</td>
<td>0.856</td>
<td>6.46</td>
<td>9.322</td>
<td>6.462</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>6.169</td>
<td>0.879</td>
<td>4.699</td>
<td>7.639</td>
<td>6.637</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6.783</td>
<td>1.013</td>
<td>5.09</td>
<td>8.476</td>
<td>7.645</td>
</tr>
<tr>
<td>Comparison</td>
<td>Time</td>
<td>Estimate</td>
<td>Std.</td>
<td>Lower 5%</td>
<td>Upper 95%</td>
<td>( \hat{\sigma} )</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------</td>
<td>----------</td>
<td>------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------------</td>
</tr>
<tr>
<td>D2low vs. Placebo</td>
<td>0.5</td>
<td>3.542</td>
<td>0.805</td>
<td>2.195</td>
<td>4.888</td>
<td>6.080</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>6.258</td>
<td>0.852</td>
<td>4.833</td>
<td>7.683</td>
<td>6.434</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>4.7</td>
<td>0.783</td>
<td>3.391</td>
<td>6.009</td>
<td>5.911</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>2.618</td>
<td>0.845</td>
<td>1.205</td>
<td>4.031</td>
<td>6.379</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.308</td>
<td>0.933</td>
<td>1.748</td>
<td>4.868</td>
<td>7.045</td>
</tr>
<tr>
<td>D2high vs. Placebo</td>
<td>0.5</td>
<td>5.994</td>
<td>0.852</td>
<td>4.57</td>
<td>7.418</td>
<td>6.433</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>8.95</td>
<td>0.883</td>
<td>7.474</td>
<td>10.427</td>
<td>6.666</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>7.269</td>
<td>0.939</td>
<td>5.699</td>
<td>8.839</td>
<td>7.089</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>5.205</td>
<td>0.841</td>
<td>3.799</td>
<td>6.611</td>
<td>6.348</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4.792</td>
<td>0.861</td>
<td>3.353</td>
<td>6.232</td>
<td>6.502</td>
</tr>
<tr>
<td>Moxi400 vs. Placebo</td>
<td>0.5</td>
<td>2.347</td>
<td>0.873</td>
<td>0.887</td>
<td>3.807</td>
<td>6.59</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>7.645</td>
<td>0.798</td>
<td>6.312</td>
<td>8.979</td>
<td>6.022</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>7.312</td>
<td>1.072</td>
<td>5.52</td>
<td>9.104</td>
<td>8.09</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>7.765</td>
<td>1.007</td>
<td>6.081</td>
<td>9.45</td>
<td>7.606</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>9.02</td>
<td>1.004</td>
<td>7.34</td>
<td>10.699</td>
<td>7.583</td>
</tr>
</tbody>
</table>
Example, continued

Graphs show mean differences from placebo, baseline subtracted, versus time, for treatments D1low, D1high, D2low, D2high, Moxi400.

Clear evidence of QT prolongation for D1high, D2high, and for Moxi400, the positive control, since the 10msec bound is exceeded, peaking at 1 – 1.5 hours post-dose, and up to 4 hours for Moxi400.

Note the apparent dose response between D1low and D1high, and between D2low and D2high.

The \( \sigma_{\hat{}} \) values range from 5.8 to 8. These are quite small, possibly reflecting error control due to having 6 replicated measurements per time point.
Other issues to consider
Assay sensitivity

ICH E14:
“The positive control should have an effect on the mean QT/QTc interval of about 5 ms”

The positive control “should be well-characterized and consistently produce an effect on the QT/QTc interval that is around the threshold of regulatory concern (5 ms, section 2.2.).”
Statistical Procedures to assess Assay Sensitivity

- $H_0$: $\Delta_t(\text{PC}) - \Delta_t(\text{P}) < c$ ms for all $t$
- $H_A$: $\Delta_t(\text{PC}) - \Delta_t(\text{P}) \geq c$ ms for at least one $t$

- How to choose $c$? Under discussion.
- Challenge: Multiple endpoint issues (can’t apply IUT here)
PK/PD approach: using Concentration-QTc Relationship to assess QT/QTc prolongation

- Under development
- Alternative to standard max-mean approach
- Useful before full TQT
- Assumes that drug concentration values are obtained at the same time-points of QT measurement
Concentration-QT method, continued

- Assumes relationship between $\Delta\Delta QTc$ and $C$ (concentration) at each time, each person:
  - $\Delta\Delta QTc_i(t) = \beta C_i(t)$

- At average maximum concentration, of interest:
  - mean $\Delta\Delta QTc$ at $C_{max} = \beta C_{max}$

- Compare $\beta C_{max}$ with NI boundary 10ms.
Concentration-QT method, continued

- Procedure: estimate $\beta$ and 95% CI ($\beta_{0.05}$, $\beta_{0.95}$) by regressing $\Delta \Delta QTc$ versus C, adjusted for time, subject.
- If CI includes 0, the drug possibly not QT prolonger, or not enough information.
- Calculate 95% bound for mean_max $\Delta \Delta QTc$, $C_{max} \times \beta_{0.95}$, and compare with 10ms
  - Could have significant slope, but with UB $<<$ 10 for $C_{max}$ expected in clinical use
- Results guide decisions about subsequent development
Another approach to Assessing QT/QTe Prolongation

We are currently exploring the use of simultaneous confidence bounds for fitted models of the population profile of ΔΔ versus time; promising, not too wide.
Closing remarks

- Control of sources of variability via trial design is important
- More statistical thinking is recommended for choice of endpoints (eg, population max-mean vs. mean-max), metrics, appropriate hypotheses, testing approaches, and performance of testing methods
- PK/PD methods are promising as adjunct approaches – exploratory and comparative work ongoing