

# Statistical Aspects of a Thorough QT Safety Assessment

Georg Ferber  
Biostatistics, Novartis Pharma AG

QT Prolongation and Safety Pharmacology  
Paris, 18 – 19 March 2008



## Overview

- What are the primary endpoints in a TQT study?
- Choosing baseline and how it affects the outcome of the TQT study.
- How many time points?
- Correction method – points to consider
- Issues with active control - do the FDA follow the ICH guideline?

## Primary endpoint

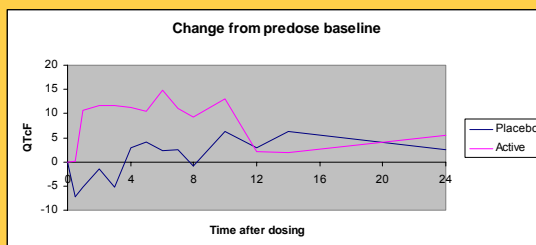
- This was the key theme during the initial statistical discussions of the guideline
- So here are some slides used at that time

3 | Statistical Aspects, Georg Ferber, 19-Mar-2008



## ICH E14 – definition of a negative TQT study

*... 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 excludes 10 ms. 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 5 ms.*



4 | Statistical Aspects/ G Ferber / 19-Mar-2008



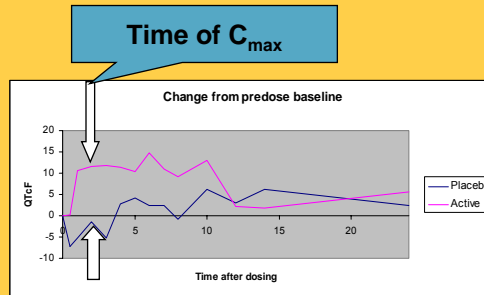
## Maximal Mean Change – Definition # 0

Mean change from baseline for study drug – mean change from baseline for placebo both taken at time of  $C_{max}$  of the drug.

Good estimator if the effect on QT is instantaneous.

Tends to **underestimate** a drug effect if there is hysteresis.

In general this is not a "maximal mean change".



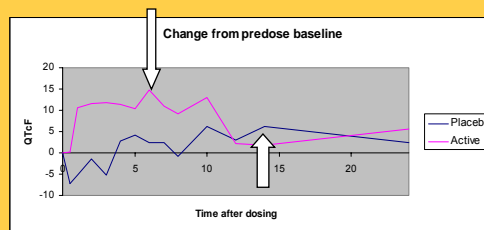
5 Statistical Aspects/ G Fer ber / 19-Mar-2008

NOVARTIS

## Maximal Mean Change – Definition # 1

Maximum over time of mean changes from baseline for study drug – maximum over time of mean changes from baseline for placebo

In the presence of circadian variability, this methods tends to **underestimate** a drug effect.



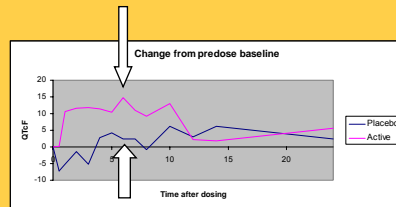
6 Statistical Aspects/ G Fer ber / 19-Mar-2008

NOVARTIS

## Maximal Mean Change – Definition # 2

[Maximum over time of mean changes from baseline for study drug] – mean change from baseline for placebo at the corresponding time point

- This can be thought of as “peak” effect on drug minus the placebo response at the same nominal time. It thus removes circadian effects, and the difference from placebo is driven by the response on drug
- But, marked circadian effect could still influence the timing of the “peak” and thus mask a drug effect. Risk of **underestimating** an effect



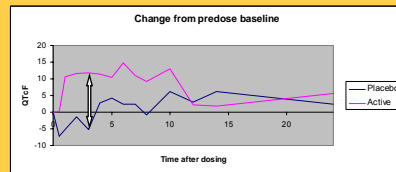
7 Statistical Aspects/ G Fer ber / 19-Mar-2008

NOVARTIS

## Maximal Mean Change – Definition # 3

Maximum over time of the means of (change from BL in study drug – change from BL in placebo) .

- This could be calculated for crossover or parallel designs.
- driven by both the largest response on drug and the smallest value on placebo.
- tends to **overestimate** a drug effect



8 Statistical Aspects/ G Fer ber / 19-Mar-2008

NOVARTIS

## Definition # 4 – Mean Maximal Change


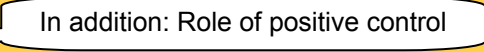
Mean (over subjects) of maximum over time of  
(change from BL in study drug – change from BL in placebo) .

- Can only be calculated in a crossover trial, since individual within subject differences are needed
- produces valid results even if PK differs interindividually
- is probably more prone to outliers
- tends to **overestimate** a drug effect.

9 Statistical Aspects/ G Fer ber / 19-Mar-2008



## Making Definition 3 Work in Practice

- Replace
  - "... confidence interval for the *largest time-matched mean effect*"
- **by**
  - "... confidence interval for the *mean effect at each time point*"
- Working with traditional confidence intervals makes handling of a TQT study simpler and statistically less demanding.
- There are still enough details to be settled, such as
  - what is "baseline" 
  - how many time points
  - how to correct for heart rate 

10 Statistical Aspects/ G Fer ber / 19-Mar-2008



## Choice of baseline – types of baseline

- "... using the largest time matched mean difference between the drug and placebo (baseline adjusted)"  
ICH E14, 3.2.1
- Common types of baseline
  - **time matched** – record a full baseline day, and match times between this baseline day and the endpoint day.
  - **predose** – take last ECG recording before start of dosing as baseline for all time points
  - **averaged** – record at several time points and take the average for all time points – taking an average reduces variability

## Choice of baseline - criteria to use

- Baseline may tell us more about the subjects we have in the trial
  - this may be of special interest for parallel group designs.
- For the statistical testing part of a TQT study, we are interested in minimising the probability of a false positive trial conditional on the sample size:
  - to achieve this, it is essential to minimize the error term in the analysis of the trial,
  - which in turn is related to the standard deviation of the change from baseline.
  - Therefore it is reasonable to choose a baseline that minimises the standard deviation of the change from baseline.
- For other scientific questions, other criteria may be important.

## Choice of baseline – Components of variability

- Components of variability in a TQT trial
  - fixed i. e. same for all subjects:
    - **period effect**
    - **circadian effect (per time point)**
    - **treatment effect per time point**
  - random: variable between subjects, but fixed within subject
    - **mean subject effect**
    - **individual period effect**
    - **individual circadian effect**
  - variable error term within each subject

13 | Statistical Aspects, Georg Ferber, 19-Mar-2008



## Choice of baseline – which components contribute

- In a crossover trial taking the double difference
  - will eliminate fixed and variable circadian effects
  - so there is no need for a time matched baseline
- In a parallel group trial taking the double difference
  - will eliminate all fixed effects except the treatment effect
  - only a time matched baseline will eliminate the **random** circadian effect
  - so a time matched baseline might reduce variability
    - if there is circadian variability that
      - **differs between** subjects but is
      - **constant within** subjects.

14 | Statistical Aspects, Georg Ferber, 19-Mar-2008



## Time matched baseline in a parallel group trial

- 4 TQT studies, looked at Moxifloxacin and Placebo data only
- First, we looked at time matched and at predose baseline
- For each time point and treatment, calculated standard deviation (SD) for change from baseline.
- Looked at the ratio of
  - SD of predose to
  - SD of time matched baseline.
- If time matched baseline is superior, this ration should be  $> 1$ .

15 | Statistical Aspects, Georg Ferber, 19-Mar-2008



## Predose vs. time matched baseline – results

- Ratio of standard deviations of change from baseline:  
predose vs. time matched baseline

Trial	Mean	Minimum	Maximum
D1	<b>0.93</b>	<b>0.63</b>	<b>1.42</b>
N1	<b>1.10</b>	<b>0.80</b>	<b>1.46</b>
N2	<b>1.12</b>	<b>0.84</b>	<b>1.48</b>
N3	<b>0.85</b>	<b>0.65</b>	<b>0.99</b>

Numbers  $> 1$  mean that time matched baseline less variable.

16 | Statistical Aspects, Georg Ferber, 19-Mar-2008



## Time matched baseline – alternatives

- If we have a full day of baseline data – how to best use it?
  - time matched or average over the baseline day?
  - averaging will reduce the variability of the baseline and therefore the variability of the difference

Trial	Mean	Minimum	Maximum
D1	0.88	0.68	1.06
N1	0.82	0.61	1.12
N2	0.84	0.64	0.96
N3	0.83	0.69	1.02

Averaging over the predose baseline is more efficient than time matched baseline

17 | Statistical Aspects, Georg Ferber, 19-Mar-2008



## How many time points?

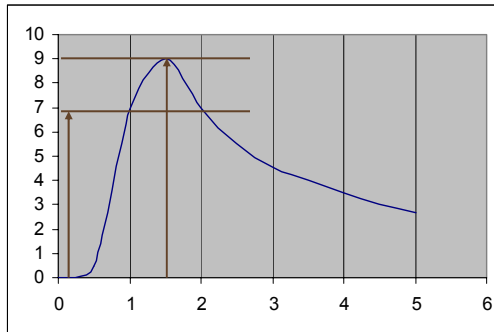
- Goal is not to miss a signal that occurs between two adjacent measurements.
- Assume a model that describes PK/PD of the drug in continuous time, then
  - the true maximum effect may be at a time point that is between two adjacent time points of our measurement scheme
  - a reasonable requirement would be that the observed maximum effect retains at least a fraction  $\phi$  of this true maximum.
  - $\phi$  could be something like 80 %
- Difficult to assess in practice

18 | Statistical Aspects, Georg Ferber, 19-Mar-2008



## How many timepoints?

- Example of missing a maximum ( $\varphi = 78\%$ )



This condition needs to be placed in a statistical context.

19 | Statistical Aspects, Georg Ferber, 19-Mar-2008



## Choice of correction method for heart rate

- In an ideal world, individual correction would be best
- However, in order to establish a stable individual estimate we need
  - a sufficient number of baseline ECGs
  - that cover the full range of heart rates to be expected under medication.
- If this cannot be guaranteed, Fridericia's correction may be preferable.
- Using a correction based on an hierarchical mixed linear model may be a compromise (Shah & Hajian, Stat in Med 2003).

20 | Statistical Aspects of a Thorough QT Study| Georg Ferber | 19-Feb-2008



## The role of active control - Guidance

- The positive control should have an effect on the mean QT/QTc interval of about 5 ms (i.e., an effect that is close to the QT/QTc effect that represents the threshold of regulatory concern, around 5 ms). Detecting the positive control's effect will establish the ability of the study to detect such an effect of the study drug.

*ICH E14 2.2*

## The role of active control – reality

- Maximum mean effect of Moxifloxacin 400 mg qd (average baseline)

Trial	Max Difference [ms]	Time [h]	Lower 95 % CI [ms]
D1	13.8	3	9.2
N1	18.5	3	11.5
N2	10.6	2	7.0
N3	9.1	4	4.3

This is definitely not what the guideline asks for.

## The role of active control – FDA requirement

- *"In order to be able to claim assay sensitivity, the one-sided 95% lower confidence bound should be greater than 5 ms at a minimum of one time point."*
- This is in contradiction to the requirement of ICH E14.
- It is understandable if Moxifloxacin 400 mg is used as active control.
- So we either need to look for alternative active controls or we may have to accept this more stringent requirement.

## Assay sensitivity – options

- We have to cope with the multiplicity problem
  - Step 1: for testing, concentrate on a few prespecified time points
  - Step 2: specify a strategy to cope with the remaining multiplicity.
- There are some conditions to keep in mind:
  - Condition 1: The procedure used needs to reflect the one used for the non-inferiority testing of our test drug.
  - Condition 2: Regardless of what we do, we should measure the ECG under active control at the same time points than those for placebo and the test drug.

## Testing for assay sensitivity – FDA proposal

- FDA recommendation:
  - We recommend you pre-specify a few time points to be examined for statistical testing to evaluate assay sensitivity based on the PK profile of moxifloxacin. (...)
  - If all or many time points are evaluated, then the overall type I error rate needs to be adjusted somehow.
  - ... recommend collecting ECG data for moxifloxacin at all the same time points as for drug and placebo ...
- FDA example:
  - 3 time points prespecified
  - use Bonferroni correction
  - chose  $\alpha = 0.02$ , i. e. a 98 % one sided confidence interval

25 | Statistical Aspects of a Thorough QT Study| Georg Ferber | 19-Feb-2008



## Assay sensitivity – some options

- In all cases, look at 3 time points between 2 and 4 h
- Options to cope with multiplicity:
  - **Bonferroni**: Use  $\alpha/3$ , i. e. 98.3 % confidence intervals, if for one of the three timepoints this excludes 0 or 5 ms, assay sensitivity is shown
  - **Hochberg** – refined procedure, see next slide
  - **Average** – compute a 95 % confidence interval for the average over the three timepoints. If it excludes 0 or 5 ms, assay sensitivity is shown.

26 | Statistical Aspects of a Thorough QT Study| Georg Ferber | 19-Feb-2008



## Testing for assay sensitivity – Hochberg procedure

- Applying Hochberg's procedure at the one-sided 5 % level:
  - Compute the one sided lower 95 % confidence interval for the 3 time points:
    - if none of them contains 0 ms (5 ms), assay sensitivity is shown.
  - If not shown, compute 97.5 % confidence intervals for the two largest values:
    - if they do not contain 0 ms (5 ms), assay sensitivity is shown
  - If not shown, compute a 98.3 % confidence interval for the largest value:
    - if it does not contain 0 ms (5 ms), assay sensitivity is shown.

27 | Statistical Aspects of a Thorough QT Study| Georg Ferber | 19-Feb-2008



## Assay sensitivity – some results

- In each of the trials, 2, 3 and 4 h time points were chosen.
  - The table gives the maximum threshold this would still result in proof of assay sensitivity.

<b>Trial</b>	<b>Hochberg</b>	<b>Sum</b>	<b>Bonferroni</b>
D1	7.9	7.6	7.9
N1	9.2	9.8	9.2
N2	6.3	7.6	5.9
N3	2.8	3.4	2.8

28 | Statistical Aspects of a Thorough QT Study| Georg Ferber | 19-Feb-2008



## Testing for assay sensitivity – discussion

- Do the procedures reflect the way we look for QT prolongation in the test drug?
  - This is questionable for the "sum" method, since the sum has a smaller variability than individual measurements
  - Bonferroni would accept assay sensitivity if there was one time point consistently outlying (positive direction)
  - Hochberg would do the same, but would "reward" more consistent results.
- Hochberg may be the best option – but it may be wise to check with regulators before applying.

## Conclusion

- In the past 5 years, we learned quite a lot
  - not only about the key question of the primary endpoint and its test
  - but also about secondary aspects that are crucial for performing a thorough QT study efficiently.
- Things to be done include
  - optimal placement of a TQT study in clinical development
  - what information to collect prior to the TQT study
  - **but above all, continue to look for a better biomarker than QT prolongation.**

Thank you!