



# Current State of Special Safety Analyses for Clinical Trials

Miganush Stepanians, Ph.D.

November 20, 2008

# Overview

- Introduction – Conventional Safety Analyses
- Three Examples of Safety Studies/Analyses
  - Heart Arrhythmias – QT/QTc interval prolongation
  - Hepatotoxicity- Liver function test elevations
  - Suicidality/Suicidal ideation
- Summary and Conclusions

# Introduction

- In Phase II and III clinical trials, safety is traditionally analyzed by descriptive methods:
  - AE tables (overall, by severity, by causality)
  - Laboratory/ECG/vital signs shift tables and counts of clinically significant abnormalities by timepoint
  - Descriptive statistics on continuous laboratory/ECG/vital signs parameters

## Introduction (Cont.)

- Phase II and III trials are ordinarily designed with efficacy as the primary target
- Sample sizes are designed to be sufficient for detecting a difference between active and placebo with respect to efficacy endpoints but not necessarily for detecting a safety signal

## Introduction (Cont.)

- Traditionally, the definitive investigation of the safety of a new drug is performed at the end of the development program in integrated analyses combining safety data from all Phase II and III trials in the ISS
- The same analyses as those performed on safety data in individual studies are repeated for the integrated safety database at the ISS stage

## Introduction (Cont.)

- This traditional approach is “observational” in nature rather than “experimental”:
  - The idea is to cast a wide net and identify safety signals
  - But this approach does not involve a pre-defined hypothesis about specific safety endpoints

## Introduction (Cont.)

- Increasingly, in recent years, it is becoming evident that this approach is not sufficient for most drugs
- Studies designed specifically to test hypotheses about safety are becoming a requirement for almost every new chemical entity

## Introduction (Cont.)

- This trend is no doubt a consequence of well publicized safety issues in a few widely used drugs in the past ten years:
  - Black box labels
  - Withdrawals from market of popular drugs
  - Lawsuits

## Introduction (Cont.)

- Therefore, there is a clear renewed emphasis on drug safety in today's regulatory environment with impact on study design and statistical analyses:
  - Specialized studies designed to test hypotheses about safety
  - New requirements for additional analyses involving inferential statistical methods applied to individual studies and safety databases for entire development programs

## Basis for Special Safety Concerns

- There are many such “Special Safety Concerns” resulting in additional/extra requirements for investigation of safety related hypotheses
- Usually based on
  - the drug class
  - the target patient population – age, sex, race, disease severity

# Three Examples of Special Safety Studies/Analyses

- We will address three examples of specialized safety studies/specialized safety analyses
  - Heart Arrhythmias – QT/QTc interval prolongation
  - Hepatotoxicity- Liver function test elevations
  - Suicidality/Suicidal ideation
- Examples have been selected based on their relevance to many drugs in different classes/indications/patient populations

## Other Special Safety Concerns

- Examples of more specialized population-specific and/or drug-class-specific safety studies include specialized studies to investigate the safety of corticosteroids:
  - HPA axis suppression studies
  - Pediatric growth suppression studies

# Heart Arrhythmias - Background

## Arrhythmias: QT/QTc Interval Prolongation

- In the past decade, the single most common cause of the withdrawal or restriction of the use of marketed drugs
- A significant prolongation of the QT interval is a biomarker of pro-arrhythmic risk including a potentially fatal condition: torsade de pointes

# QT Interval - Definition

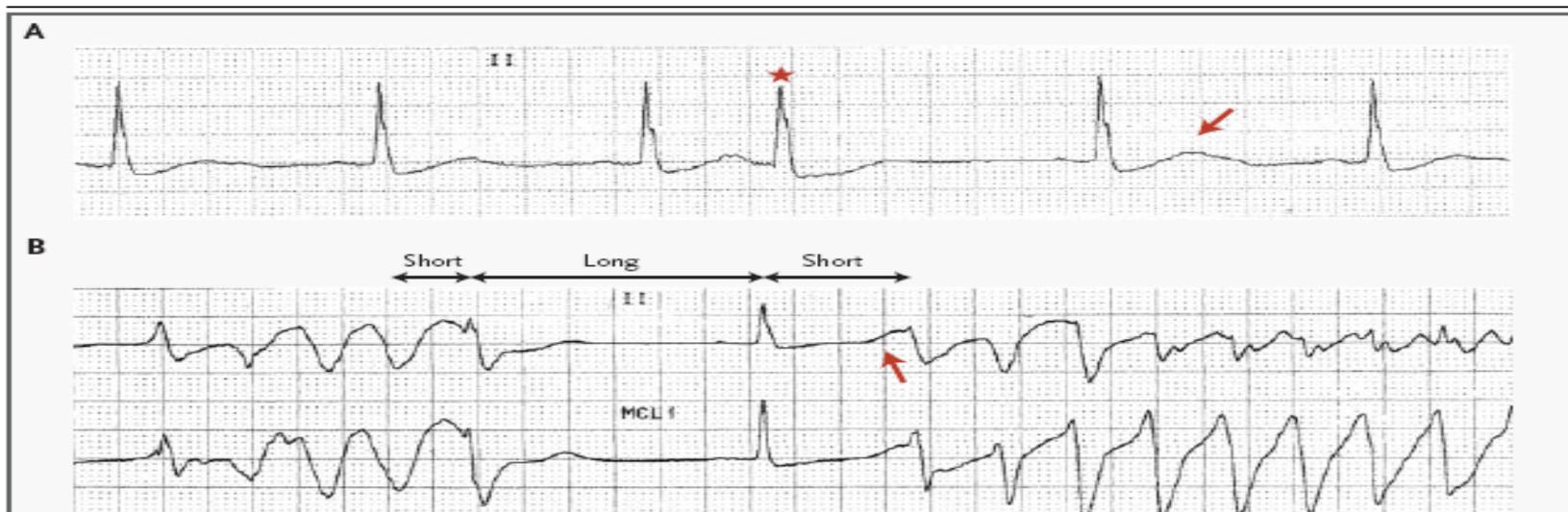
- QT Interval – ECG tracings:



# QT Interval Prolongation

- QT interval prolongation induced by the usage of a given drug is detected from ECG tracings as shown in the 2004 NEJM article by Roden

# QT Interval Prolongation (Cont.)



**Figure 1.** Rhythm Recordings from a 76-Year-Old Woman with Renal Dysfunction Who Was Treated with Sotalol for Atrial Fibrillation.

Panel A was recorded after spontaneous conversion to sinus rhythm. There is a premature atrial beat (star) followed by a pause, and the subsequent sinus beat shows marked QT prolongation and deformity (arrow). Panel B was recorded several minutes later and shows a typical episode of torsade de pointes: there is a four-beat run of polymorphic ventricular tachycardia, a pause, and a sinus beat with a long and deformed QT interval (arrow), interrupted by another episode of polymorphic ventricular tachycardia (torsade de pointes). This pattern of onset — a short cycle followed by a long one followed by a short one — is typical of drug-associated torsade de pointes. Risk factors in this case included female sex, the administration of sotalol in a patient with renal failure (causing increased drug levels), and recent conversion from atrial fibrillation.

## QT Interval Prolongation (Cont.)

- Drugs that have been withdrawn from the market because they cause torsade de pointes had a population mean increase in the QT interval estimated to be as small as 5 to 10 ms

## QTc Interval: QT Correction

- There is an inverse relationship between QT interval and heart rate. Therefore, it is necessary to correct QT for the heart rate and calculate a corrected QT or QTc.
- It is not clear whether arrhythmia development is more closely related to an increase in the absolute QT interval or QTc.

## QTc Interval: QT Correction (Cont.)

- Methods of QT correction:
  - Fridericia's correction:  $QTcF = QT/RR^{0.33}$
  - Bazett's correction:  $QTcB = QT/RR^{0.5}$
  - Pooled correction: Linear Regression
  - Individual correction: Linear Regression
  - Note: RR=60/HR

## QTc Interval: QT Correction (Cont.)

- Corrections based on linear regression techniques (pooled and individual)
- Linear regression of QT on RR
- Estimate the slope in the equation
$$QT = a + b(1-RR)$$
- Use b for adjusting the QT to a normalized heart rate of 60 and deriving the QTc

# Regulatory Guidance: ICH E14

- ICH - E14 – October 2005: Clinical Evaluation of QT/QTc Interval Prolongation and Pro-arrhythmic Potential for Non-Anti-arrhythmic Drugs
- “This guidance provides recommendations to sponsors concerning the design, conduct, analysis, and interpretation of clinical studies to assess the potential of a drug to delay cardiac repolarization. This assessment should include testing the effects of new agents on the QT/QTc interval as well as the collection of cardiovascular adverse events.”

## ICH E14 (Continued)

- Guidance: thorough investigation of potential for QT/QTc prolongation is recommended for
  - All systemically bio-available new drugs
  - Excludes topically active medications and anti-arrhythmic medications
  - Includes approved drugs investigated for new routes of administration or higher dosages, new patient populations/indications
  - Especially important for drugs within a “suspect drug class”

## ICH E14 (Continued)

- Individualized approach based on:
  - Pharmacodynamics/Pharmacokinetics
  - Safety characteristics
  - Proposed indication

## The “thorough QT/QTc study”

- Determine whether the drug has a threshold pharmacologic effect on cardiac repolarization
- “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 study is preferably conducted early in the development plan and is used to determine the intensity of ECG data collection during late stage studies, including the pivotals.

# Study Design for the “thorough QT/QTc study”

- Adequate and well-controlled
- Randomized
- Placebo-controlled
- Active-controlled for assay sensitivity
- Double-blind
- Preferable to include a suprathreshold dose
- Most often conducted in healthy volunteers (or in patients for compounds where volunteer studies are not possible; e.g., cytotoxic oncology compounds)

# Parallel or Crossover Design

- Crossover trials
  - n is usually around 50 for a crossover study (depending on the number of treatments compared)
  - An appropriate treatment, period, and carryover balanced crossover design should be used – Latin square design
  - An adequate washout between periods is necessary
  - For many drugs, crossover trials are preferred due to the small sample size and ability to correct QT on an individualized basis

## Parallel or Crossover Design (Cont.)

- Parallel group trials
  - n is usually around 60 per arm for a parallel group study
  - Parallel trials are preferred for
    - drugs with long elimination half lives and potential of carryover effects
    - situations where multiple doses of the drug are tested in the thorough QT trial

# Design and Conduct Features for Reduction of Variability

- multiple, centrally read ECGs
- multiple timepoints for measurement of ECGs with matched baseline timepoints

## Other Design Considerations

- Timing of the ECG measurements must take into consideration the PK/PD profile of the drug including most importantly the  $C_{max}$
- ICH E14: “duration of dosing or dosing regimen should be sufficient to characterize the effects of the drug and its active metabolites at relevant concentrations.”

# Analysis Methodology

- The matching baseline value for each post-dose observed ECG is the value at a timepoint on baseline day corresponding to the post-dose timepoint.
- The change from baseline is calculated for each patient at each observed post-dose timepoint.
- The maximum of these changes from baseline is calculated for each subject (or for each subject in each period for crossover studies).
- Treatment arms are compared based on this maximum change from baseline variable.

## Analysis Methodology (Cont.)

- The 95% one-sided confidence interval of the difference of the means is constructed based on appropriate models depending on the crossover or parallel study design.
- Assay sensitivity is established via the placebo-positive control comparison based on a threshold of 10 ms.
- The study drug's effect on the QT interval prolongation is assessed based on the placebo versus study drug comparison using the same threshold of 10 ms.

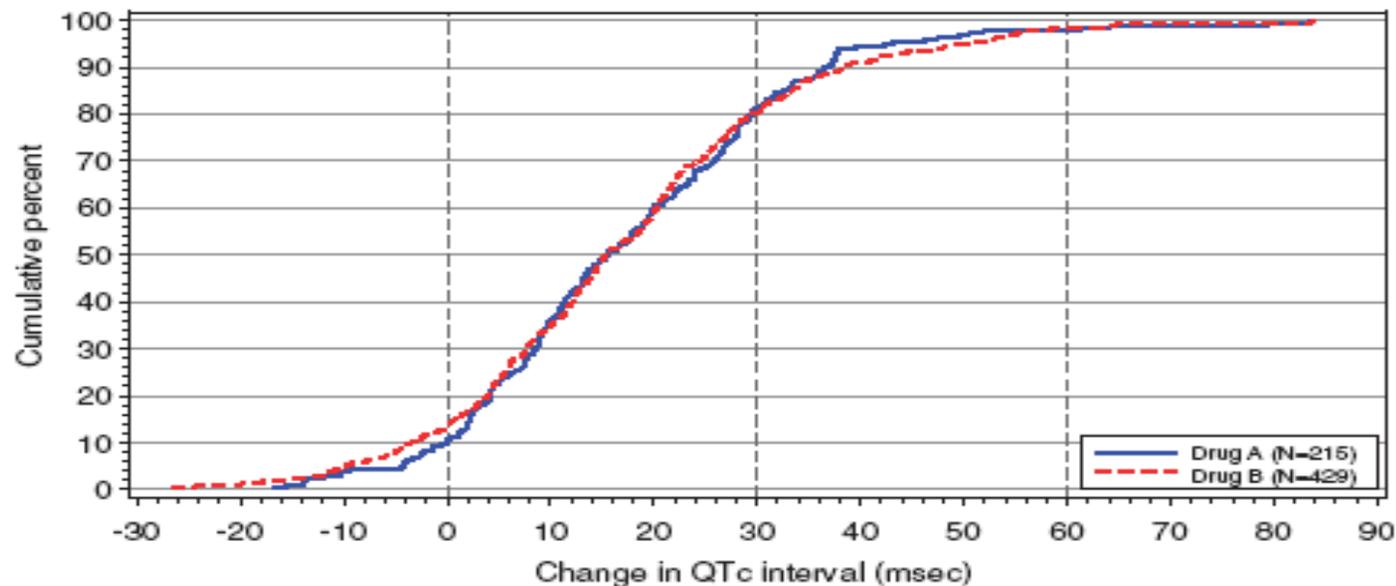
## Analysis Methodology (Cont.)

- A positive study does not necessarily mean that the drug is pro-arrhythmic, rather it is a signal for more rigorous evaluation of ECG data in later stages of the development program.

# Categorical Analyses

- In addition to the analyses of means, categorical analyses are also recommended in ICH E14 based on the number and percentage of patients meeting or exceeding some pre-defined threshold or upper limit
  - Absolute QTc interval prolongation:
    - QTc interval  $> 450$
    - QTc interval  $> 480$
    - QTc interval  $> 500$
  - Change from baseline in QTc interval:
    - QTc interval increases from baseline  $> 30$
    - QTc interval increases from baseline  $> 60$

# Graphical Techniques

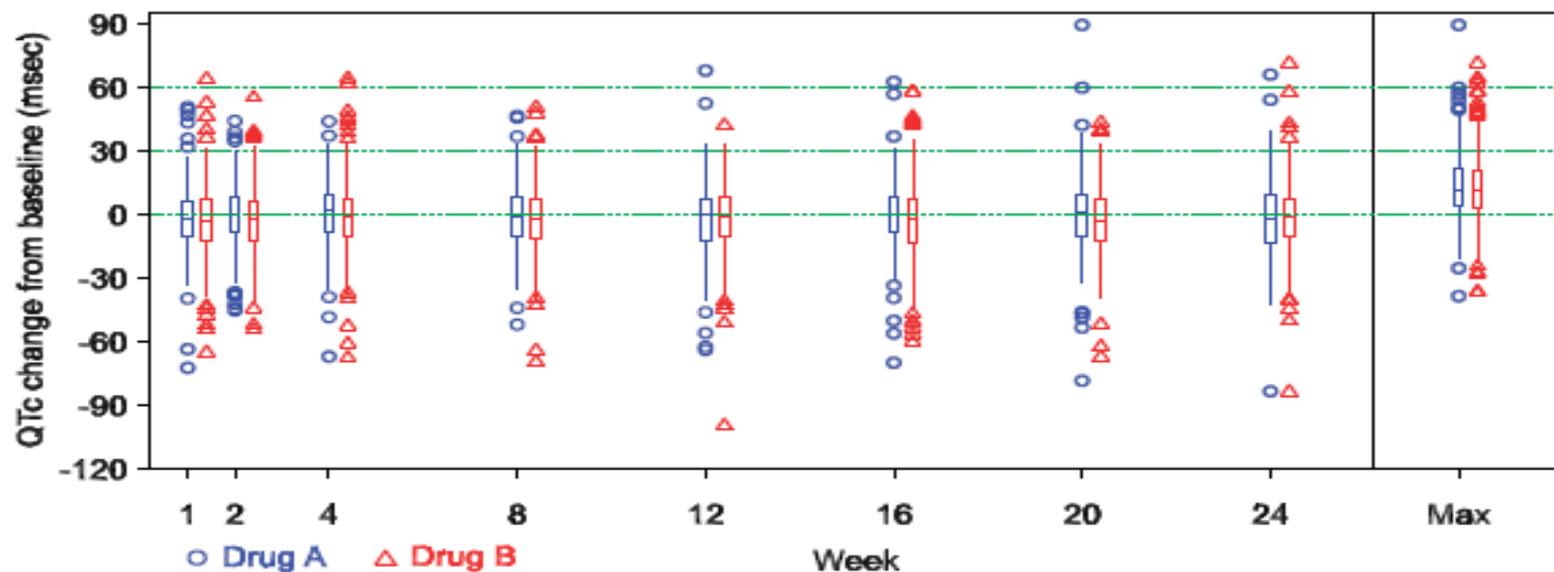


Note: Increase <30 msec 'Normal', 30-60 msec 'Concern', >60 msec 'High'

Figure 1. Empirical distribution function for maximum change in QTc.

Amit O, Heiberger RM, Lane PW. Graphical Approaches to the Analysis of Safety Data from Clinical Trials. *Pharmaceut. Statist.* 2008; 7:20-35.

# Graphical Techniques (Cont.)



	Number of subjects at visit								
Drug A	209	206	200	192	184	177	169	164	212
Drug B	405	384	362	337	316	311	299	294	413

Note: Increase <30 msec 'Normal', 30-60 msec 'Concern', >60 msec 'High'

Figure 2. Boxplot of change from baseline in QTc by time and treatment.

Amit O, Heiberger RM, Lane PW. Graphical Approaches to the Analysis of Safety Data from Clinical Trials. *Pharmaceut. Statist.* 2008; 7:20-35.

# Graphical Techniques (Cont.)

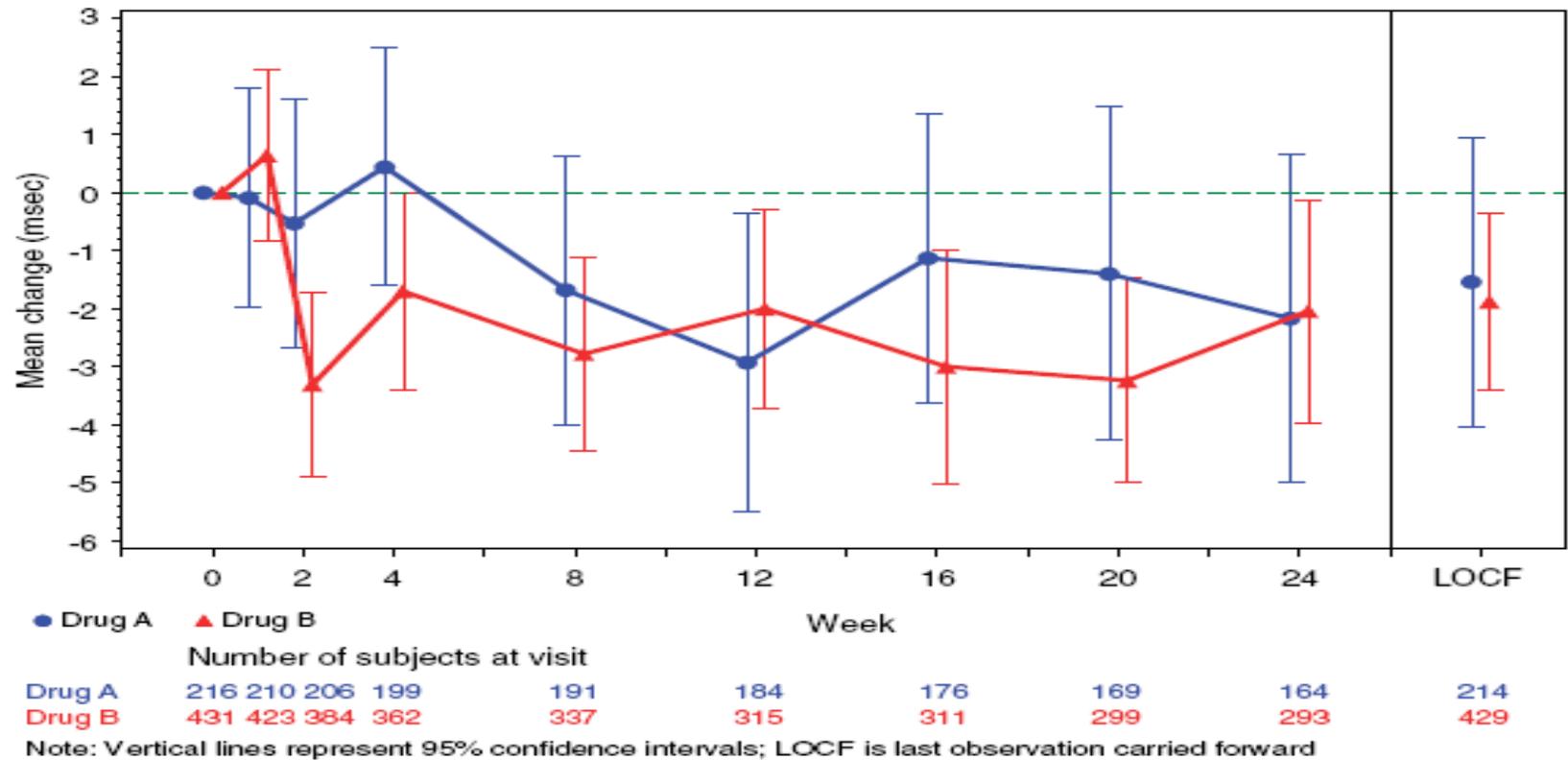


Figure 3. Mean (95% CI) change from baseline in QTc by time and treatment.

Amit O, Heiberger RM, Lane PW. Graphical Approaches to the Analysis of Safety Data from Clinical Trials. *Pharmaceut. Statist.* 2008; 7:20-35.

# Hepatotoxicity Drug Induced Liver Injury (DILI)

- DILI is the other major drug safety issue affecting a wide spectrum of drugs from various classes and for many indications
- Most common cause of marketing withdrawals in the past 50 years
- Severe DILI can lead to liver failure, need for liver transplantation, and in some cases death
- FDA draft Guidance to Industry - October 2007

# Hepatotoxicity Drug Induced Liver Injury (DILI) (Cont.)

- DILI is typically a rare side effect; hence, one may not observe any cases in a drug development program consisting of several thousand subjects, even if the drug does cause severe liver injury
- While actual severe DILI may not be observed in a program, a drug's *potential* for severe DILI can be revealed through clinical laboratory data – Liver Function Tests

# Hepatotoxicity Drug Induced Liver Injury (Cont.)

- The type of liver injury that can lead to rare cases of severe DILI is called Hepatocellular Injury
  - Hepatocellular injury results in the release of ALT and/or AST from injured liver cells into the blood
  - Many drugs cause mild hepatocellular injury as evidenced by liver enzyme (ALT or AST) elevations, but may not cause DILI unless:
    - injury severe enough to interfere with the liver's ability to clear bilirubin
- ALT is the more sensitive liver enzyme for detection of the potential for DILI

# Statistical Analyses of Liver Enzyme/Liver Function Test Data

- Elevation of liver enzymes ALT and AST beyond 3x Upper Limit of Normal ( $>3xULN$ )
  - High sensitivity but not high specificity for predicting DILI
  - Incidence of  $>3xULN$  should be analyzed in comparison to placebo or other drugs proven to have no hepatotoxicity
  - Conventional to look at the following thresholds:
    - $>3xULN$ ;  $>5xULN$ ;  $>8xULN$ ;  $>15xULN$ ; and  $>20xULN$
  - Usually performed on the integrated database of placebo-controlled studies due to rareness of these events
  - Compare drug versus placebo – Fisher's Exact test

## Statistical Analyses of Liver Enzyme/Liver Function Test Data (Cont.)

- Incidences of  $>10xULN$  and the higher thresholds above should be looked at for all studies using the entire integrated database, including uncontrolled trials, since such high elevations are a clearer signal

# Statistical Analyses of Liver Enzyme/Liver Function Test Data (Cont.)

- Elevations of other liver enzymes: GGT, LDH, ALK PHOS should be descriptively analyzed using the same cut points as used for the ALT and AST analyses
- Total Bilirubin (TBILI) is analyzed using a different set of threshold values:
  - $>1xULN$ ;  $>1.5xULN$ ;  $>2xULN$ ;  $>3xULN$

## Hy's Law

- A sensitive and specific method of identifying drugs that may cause severe DILI:
  - The drug causes hepatocellular injury: shown by a higher incidence of ALT or AST elevations ( $>3xULN$ ) in the drug versus placebo (or a non-hepatotoxic control agent)
  - Some subjects with extreme ALT or AST values also show elevation of serum TBILI  $>2xULN$ , without initial findings of cholestasis [serum alkaline phosphatase (ALK PHOS)  $>2xULN$ ]
  - No other reason can be found to explain the combination of increased ALT or AST and TBILI, such as viral hepatitis A, B, or C, pre-existing or acute liver disease, or another drug capable of causing the observed injury

## Hy's Law (Cont.)

- The incidence of severe liver injury (severe DILI) in the population is estimated to be one tenth of the rate of Hy's Law cases (DILI)
- Rule of 3 (based on binomial distribution): There will be at least a 95 percent chance of observing one or more cases of DILI in  $3n$  study subjects if its true incidence is 1 in  $n$  subjects
- Example: in a typical NDA database of 3,000 subjects, if no Hy's Law cases are observed, it can be concluded with 95 percent confidence that the true rate of such occurrences is not more than 1 per 1,000. Therefore, the expected rate of severe liver injury is  $\leq 1$  per 10,000 exposed patients
- NDA and BLA submissions should include a listing of possible Hy's Law cases and a narrative summary for each case

# Other Analyses Recommended for Drugs Suspected of Causing DILI

- Further analyses of pattern of liver injury via examination of combinations of elevated liver enzymes
- Survival analysis methods applied to the analysis of time to onset of ALT elevation  $>3xULN$
- Analyses of time to resolution of ALT elevation (both on-drug and off-drug resolution)
- Analyses of incidence of ALT and AST elevations by sex, age, and race
- Incidence of AEs associated with hepatotoxicity identified using the hepatic MedDRA SMQ

# Graphical Displays of LFT Elevation Data

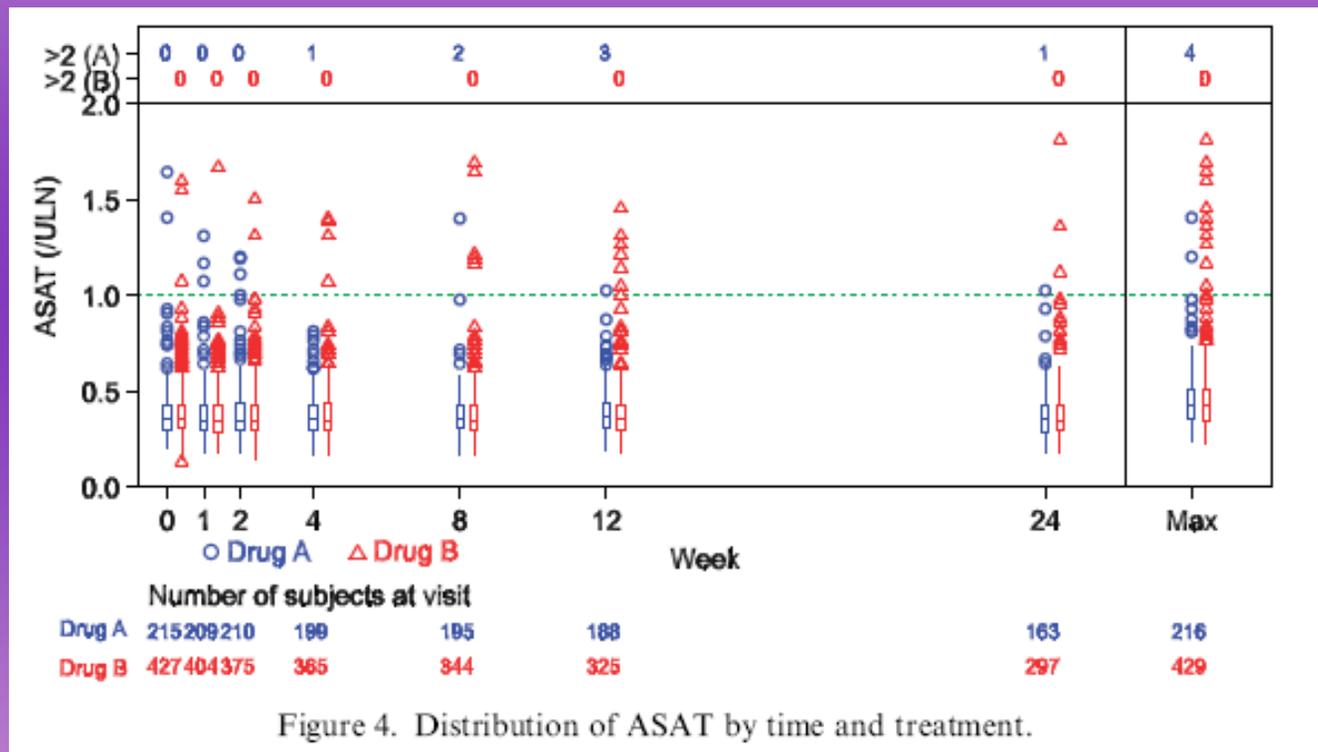
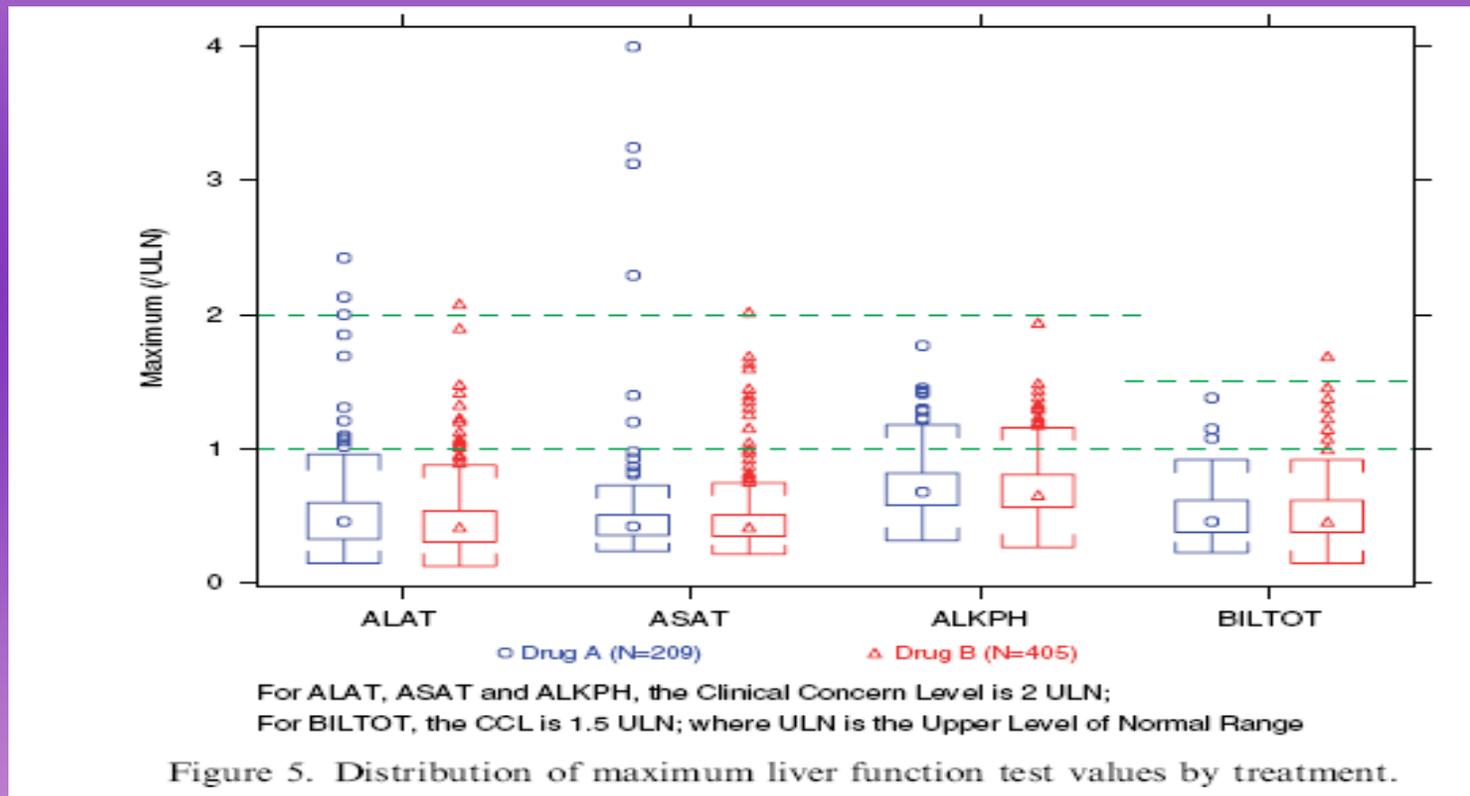


Figure 4. Distribution of ASAT by time and treatment.

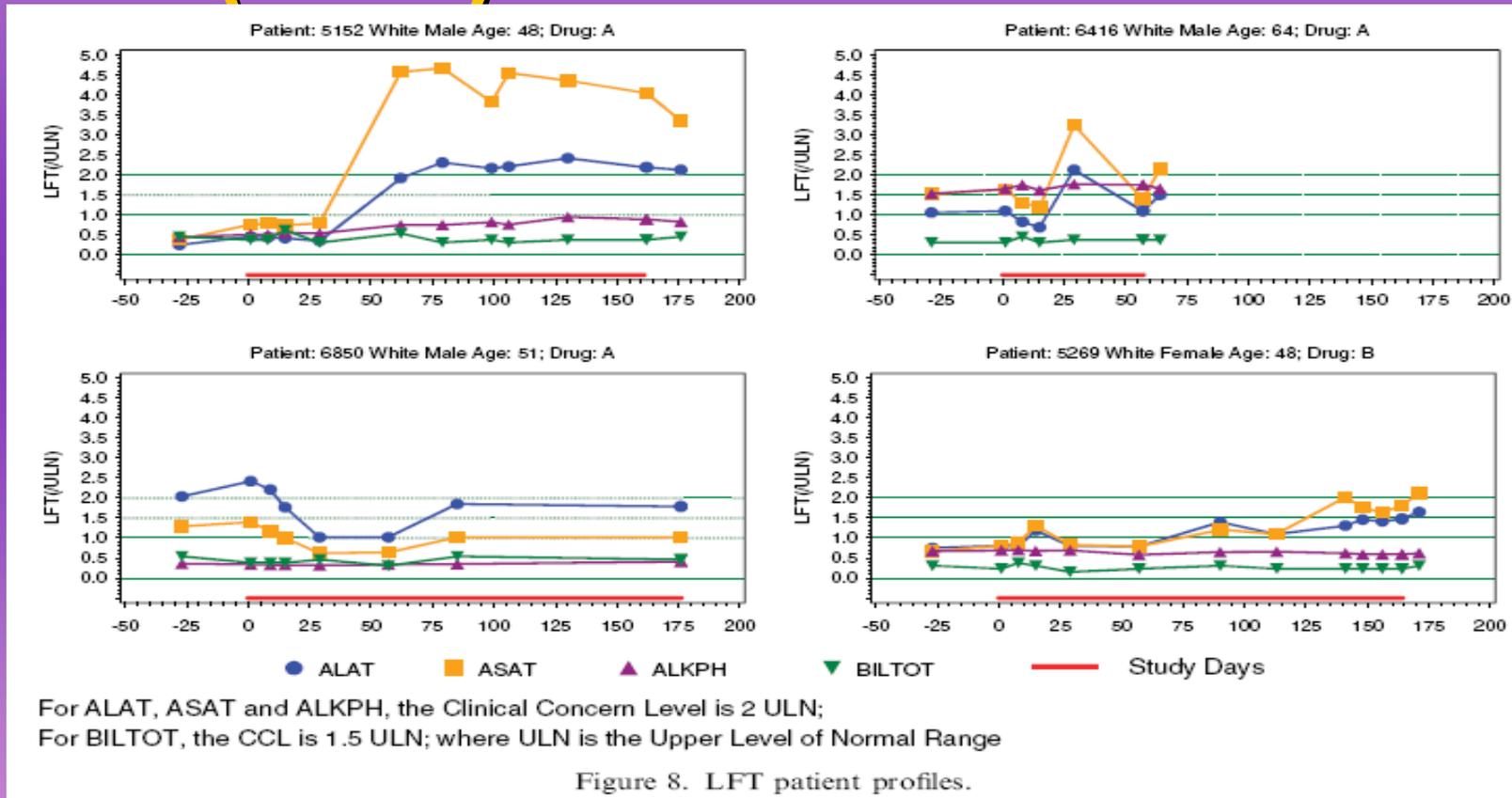
Amit O, Heiberger RM, Lane PW. Graphical Approaches to the Analysis of Safety Data from Clinical Trials. *Pharmaceut. Statist.* 2008; 7:20-35.

# Graphical Displays of LFT Elevation Data (Cont.)



Amit O, Heiberger RM, Lane PW. Graphical Approaches to the Analysis of Safety Data from Clinical Trials. *Pharmaceut. Statist.* 2008; 7:20-35.

# Graphical Displays of LFT Elevation Data (Cont.)



Amit O, Heiberger RM, Lane PW. Graphical Approaches to the Analysis of Safety Data from Clinical Trials. *Pharmaceut. Statist.* 2008; 7:20-35.

# Suicidal Behavior/Suicidal Ideation

- A relatively recent drug safety concern that started with anti-depressants and is now investigated for other classes of drugs including anti-epileptic drugs (AEDs), and others used in teenage subjects
- No guidance document
- Historically, concerns regarding the suicidality potential have arose from post-marketing reports
- Resulting in requests from regulatory bodies to sponsors for retrospective analyses of the entire clinical database of some approved drugs

# The Case of the Anti-Epileptic Drugs

- On July 10, 2008, a joint meeting was held between the Peripheral and Central Nervous System Drugs Advisory Committee (PCNS) and the Psychopharmacologic Drugs Advisory Committee (PDAC)
- The joint meeting reviewed the results of a meta analysis conducted by the FDA staff combining data across 11 different AEDs
- The search methods and statistical techniques used for these analyses appear to represent the current preferences of the Agency for investigation of suicidality potential of drugs
- These methods are detailed in a June 12, 2008 briefing by the Division of Neurology released in advance of the joint advisory meeting

# Identification of Possibly Suicide Related Adverse Events (PSRAEs)

- Include all double-blind, randomized, placebo-controlled, parallel-group, multiple-dose clinical trials conducted with the specific drug with at least 20 patients or subjects per treatment arm
- Include all multiple-dose, placebo-controlled crossover trials but use the first period data only
- Only include events that occur during the double-blind phase of treatment, or within 1 day of stopping randomized treatment in these trials

# Identification of Possibly Suicide Related Adverse Events (Cont.)

- PSRAEs are identified through a search of specified text-strings in the adverse event data casting a wide net
  - Examples of search terms: “accident”, “attempt”, “burn”, “cut”, “drown”, “self damage”
- All deaths and other serious adverse events (SAEs) are included among the PSRAEs
- All adverse events coded as “accidental injury” are included among the PSRAEs

# Identification of Possibly Suicide Related Adverse Events (Cont.)

- Any terms incorrectly identified by this search because the text string was a substring of an unrelated word is not considered as PSRAEs (false positive)
- For each PSRAE (excluding false positives), a narrative is prepared

# Identification of Possibly Suicide Related Adverse Events (Cont.)

- Based on blinded versions of the narratives, the PSRAEs are classified as
  - 1 = Completed suicide
  - 2 = Suicide attempt
  - 3 = Preparatory acts toward imminent suicidal behavior
  - 4 = Suicidal ideation
  - 5 = Self-injurious behavior, intent unknown
  - 6 = Not enough information, fatal
  - 7 = Not enough information, non-fatal

## Identification of Possibly Suicide Related Adverse Events (Cont.)

- The primary endpoint is *Suicidal Behavior or Ideation*
- A patient is considered to have suicidal behavior or ideation if the patient has any of the following suicidality events:
  - Completed suicide
  - Suicide attempt
  - Preparatory acts toward imminent suicidal behavior
  - Suicidal ideation

# Identification of Possibly Suicide Related Adverse Events (Cont.)

- Two secondary endpoints are defined as follows:
- **Suicidality:** A patient is considered as exhibiting Suicidal Behavior if the patient has any of the following suicidality events:
  - Completed suicide
  - Suicide attempt
  - Preparatory acts toward imminent suicidal behavior
- **Suicidal Ideation:** A patient is categorized as having the endpoint Suicidal Ideation if the patient has only a Suicidal Ideation event and no suicidality events

# Primary Analysis Method

- The preferred primary analysis method is the exact method for a stratified odds ratio
- The associated 95% confidence interval should be calculated
- Trial is the stratification factor
- Note: the exact method for a stratified odds ratio does not make use of trials with zero events

# Sensitivity Analyses

- The Mantel-Haenszel risk difference and the associated confidence interval which includes data from zero-event trials should be used as a sensitivity analysis
- To assess the possible effects of trial heterogeneity on the primary analysis, the trial weight of the Mantel-Haenszel odds ratio estimator can be used to quantitatively identify trials with large influence
- If one or more trials have large influence on the results of the primary analysis, a generalized linear mixed model will be used to estimate the overall odds ratio in the presence of trial heterogeneity

# Subset Analyses

- Recommended at a minimum for the primary endpoint by
  - Age Group
  - Gender
  - Race
  - Trial Setting/Location (e.g., US versus EU)
  - Indication

# Graphical Presentation of Meta Analysis of AEDs

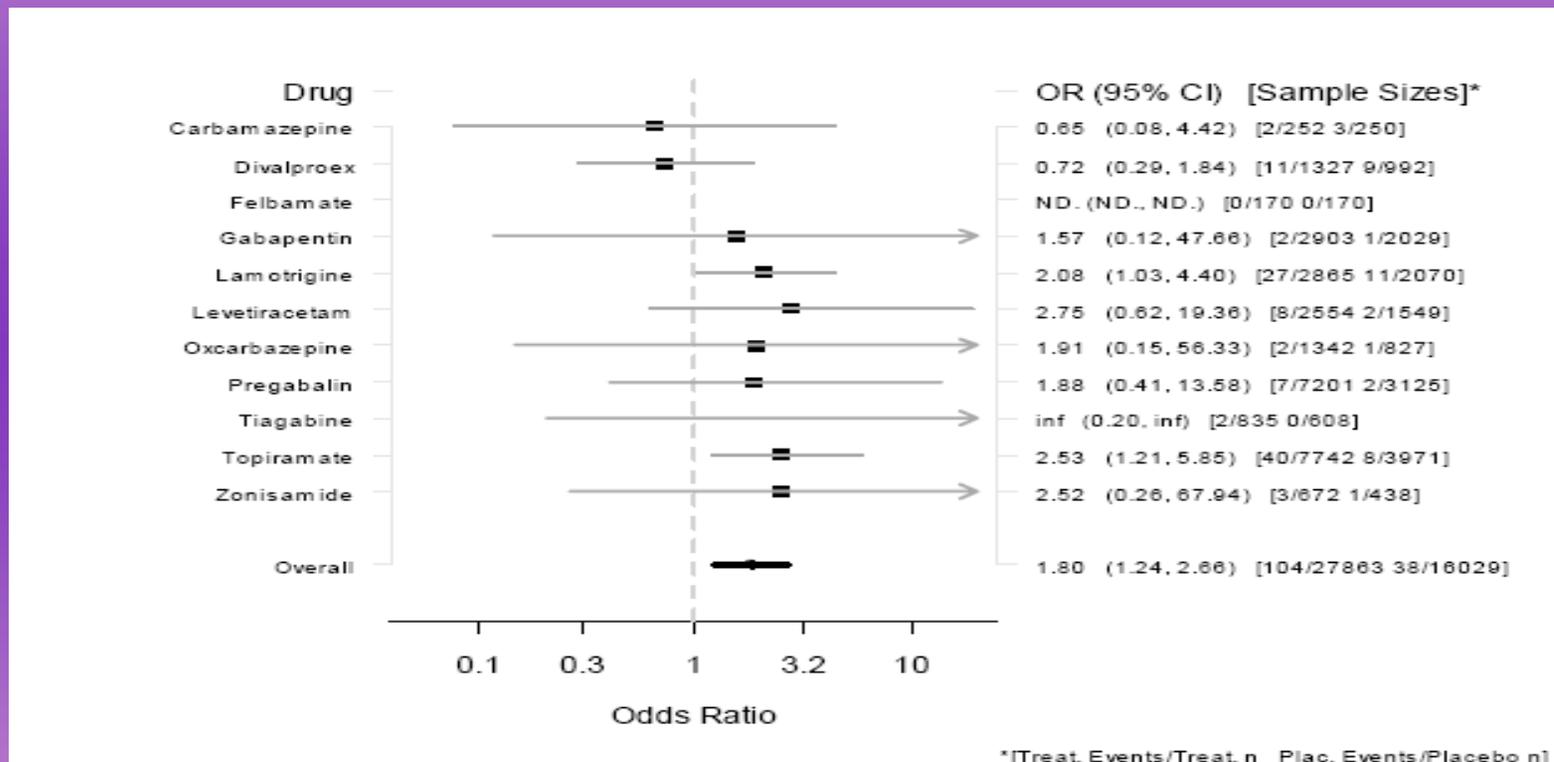


Figure 2: Suicidal Behavior or Ideation Odds Ratio Estimates, Placebo-Controlled Trials.

Katz, R. FDA Briefing Document for the July 10, 2008 Advisory Committee Meeting to Discuss Antiepileptic Drugs (AEDs) and Suicidality. June 2008.

# Summary and Conclusions

- For most new drugs submitted for marketing approval today, the traditional descriptive safety analyses are no longer sufficient
- Sponsors are expected to conduct at least one special safety study in their development programs and include complex analyses in their Clinical Study Reports and Integrated Safety Summaries
- There is clearly an expanded role for statisticians in the area of drug safety evaluation

# References

- Roden DM. Drug-Induced Prolongation of the QT Interval. *NEJM*. 2004; 350: 1013-1022.
- ICH - E14. Clinical Evaluation of QT/QTc Interval Prolongation and Pro-arrhythmic Potential for Non-Anti-arrhythmic Drugs. October 2005.
- Amit O, Heiberger RM, Lane PW. Graphical Approaches to the Analysis of Safety Data from Clinical Trials. *Pharmaceut. Statist.* 2008; 7:20-35.
- ICH Draft Guidance. Drug-Induced Liver Injury: Premarketing Clinical Evaluation. October 2007.
- Katz, R. FDA Briefing Document for the July 10, 2008 Advisory Committee Meeting to Discuss Antiepileptic Drugs (AEDs) and Suicidality. June 2008.



## CONTACT INFORMATION

**Miganush Stepanians, Ph.D.**

**725 Concord Avenue**

**Cambridge, MA 02138**

**1-617-868-2020, x211**

**1-617-844-0211**

**[mstepanians@prometrika.com](mailto:mstepanians@prometrika.com)**