

Moving Beyond QT ...

Robb Muirhead & Kimberly Crimin

Pfizer

Wyeth

BASS XIV: November 6, 2007

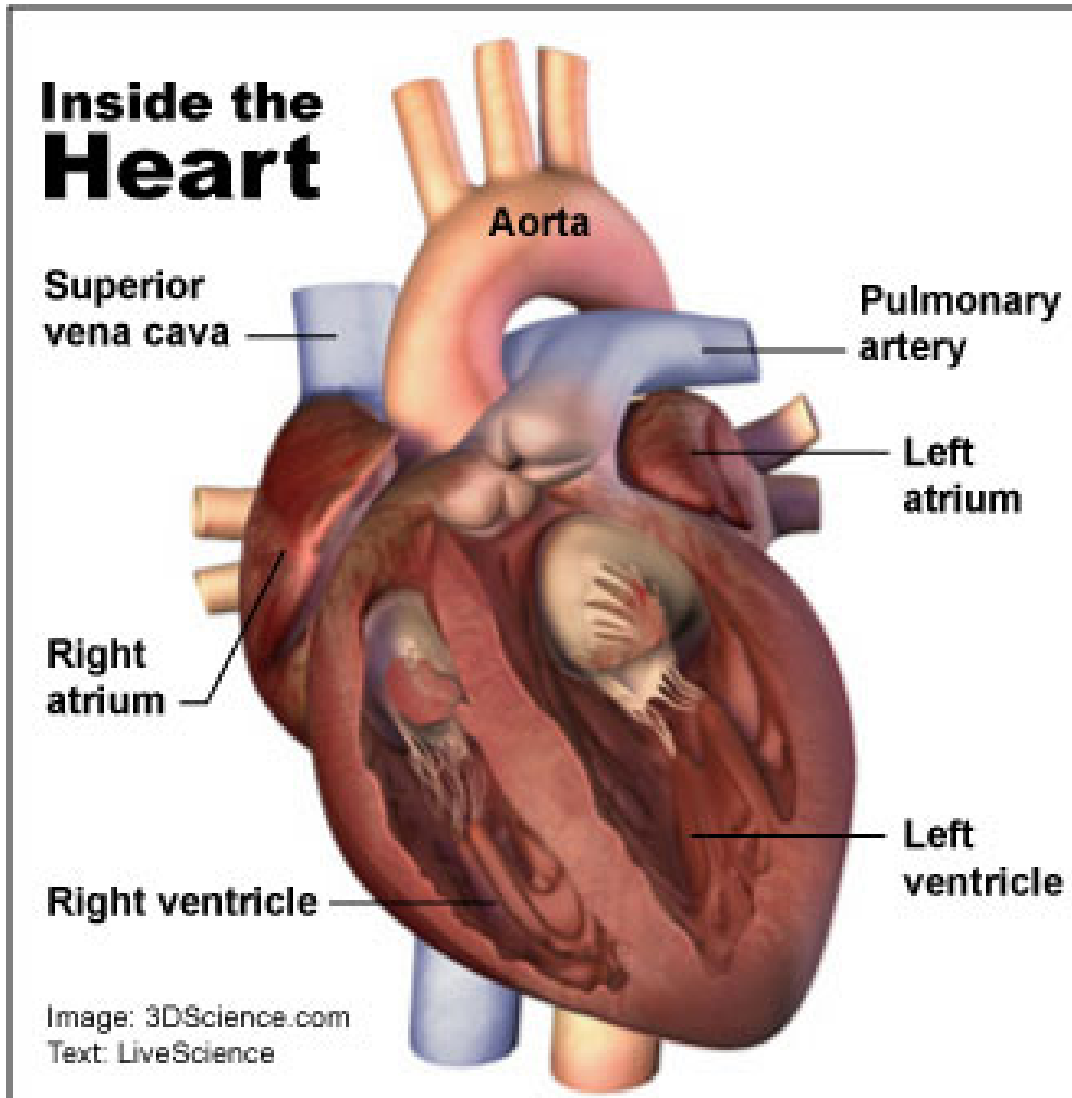
Acknowledgments

- Jay Emerson (Yale)
- Anthony Fossa (Pfizer)

Outline

- ECG and the QT interval
- The statistical approach proposed in ICH E14, and critique
- Current approaches to the assessment of arrhythmic liability
- Beat-to-beat data
- Reference regions, examples, properties
- Summary

Heart 101



De-oxygenated blood is collected in the right atrium and pumped, via the right ventricle, into the lungs to drop off CO_2 and collect O_2 (gas exchange).

Oxygenated blood from the lungs is collected in the left atrium and is pumped, via the left ventricle, out to the body through the aorta.

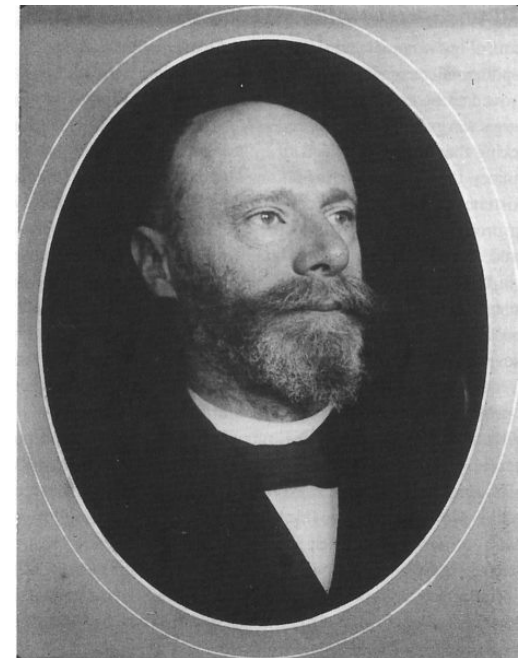
Electrocardiogram (ECG or EKG)

- An ECG is a graphic produced by an *electrocardiograph*, which records the electrical activity of the heart over time

The first practical electrocardiograph was invented by Willem Einthoven (1860-1927) in Leiden and used a string galvanometer. The original machine required 5 people to operate it and weighed about 600 lbs.

Einthoven assigned the letters P, Q, R, S, T to various features of an ECG, and described their relation to a number of cardiovascular disorders.

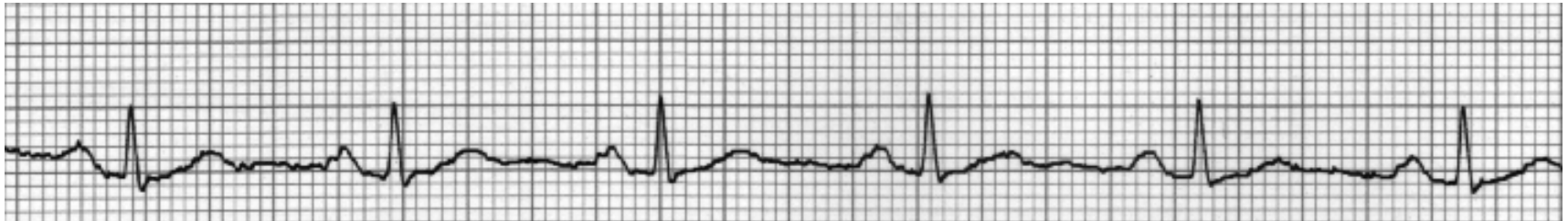
Einthoven received the Nobel Prize in Medicine in 1924 “for his discovery of the mechanism of the electrocardiogram”.



ECG (cont.)

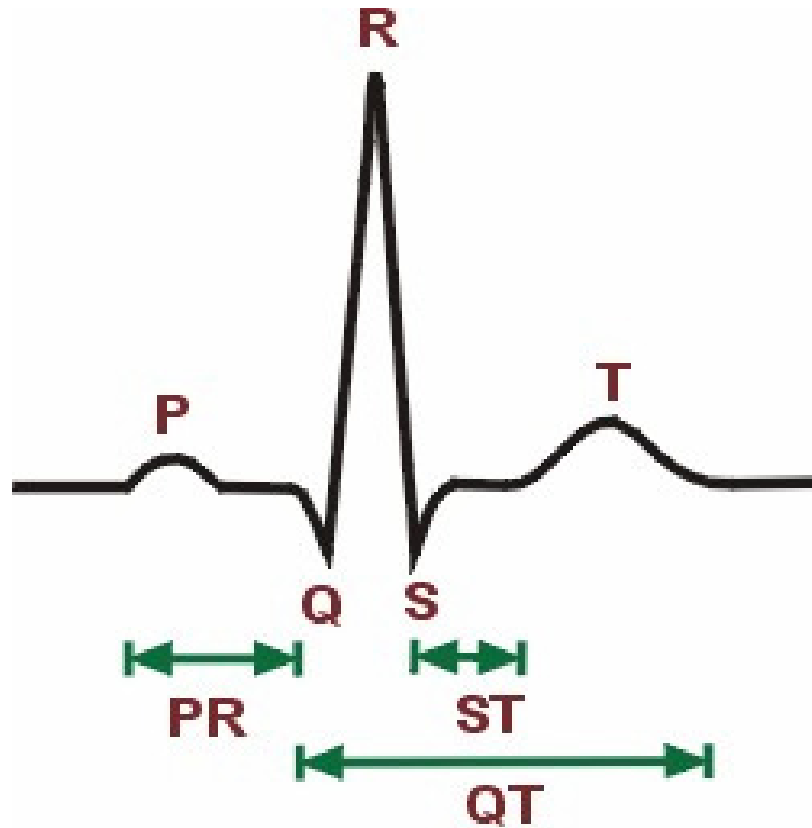
- An ECG is the gold standard for the diagnosis of cardiac arrhythmias (any of a group of conditions in which the electrical activity of the heart is irregular or is faster or slower than normal)
 - ◆ Tachycardia: Heart rate faster than 100 beats/minute
 - ◆ Bradycardia: Heart rate slower than 60 beats/minute
 - ◆ Fibrillation: Heart muscle begins a quivering motion
 - Atrial: Quivering, chaotic motion in the upper chambers of the heart (not typically a medical emergency)
 - Ventricular: Occurs in the ventricles (lower chambers) – leads to cardiac arrest

ECG of Healthy Adult



Each 1-mm square represents .04 sec (40 msec) in time and .1 mV in voltage.

“Idealized” ECG and the QT Interval

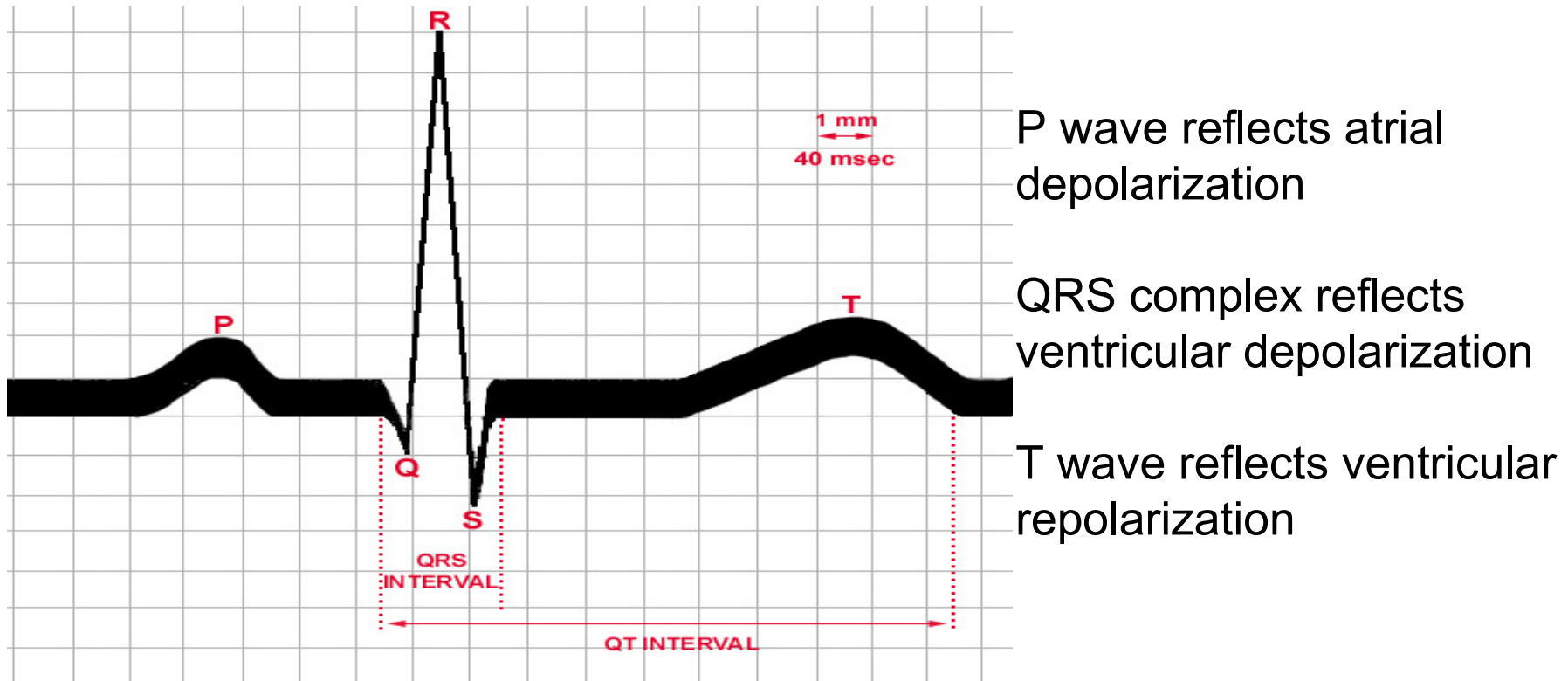


The QT interval is part of the cardiac cycle as represented by an electrocardiograph (ECG). The cardiac cycle - a single heartbeat - starts when the atria begin to fill with blood, and ends when the ventricles recover from the contraction that propels blood throughout the body.

An ECG represents the heartbeat as a series of waves that measure electrical changes in the walls of the heart—depolarization and repolarization—that accompany the blood movement.

The QT interval is the time, in milliseconds, between the depolarization (Q) at the onset of ventricular contraction and the repolarization (T) that prepares for another contraction.

QT Interval on Surface ECG



The QT interval is duration of ventricular depolarization and repolarization.

QT is the **working phase** and TQ is the **resting phase**.

Long QT Syndrome and TdP

- People with the “long QT syndrome” are susceptible to an abnormally rapid heart rhythm (arrhythmia) called “torsade des pointes” (TdP), which can lead to sudden death.
- Torsade de Pointes (TdP) is a ventricular tachycardia which may cause fatal cardiac arrest or sudden death.
- TdP is characterized by an abnormally prolonged QT interval preceding tachycardia.
- QT prolongation is used as a biomarker for TdP
- Studies that indicate significant QT prolongation are often sufficient for a company to discontinue development of a compound.

TdP

- Torsade de pointes is French for “twisting of the points” (in the QRS complex)
- It was named by a French physician (Francois Dessertenne) who described it in 1966.
- “Torsade” refers to an ornamental motif imitating twisted hairs or threads seen on classical architectural columns.
- “Pointes” refers to points or peaks



Two Examples

- The popular antihistamine **Seldane** was pulled from the market in 1998 because data linked it with sudden cardiac death and life-threatening ventricular tachyarrhythmias, such as torsade de pointes (TdP).
- **Propulsid** (heartburn, acid reflux) was withdrawn in 2000 after being associated with 341 reports of heart rhythm abnormalities including 80 deaths.
- Since 1990, at least 8 drugs have been withdrawn from major markets because of QT interval prolongation and TdP. (Some of these appear to have increased risk of TdP because drug-drug interactions.)

QT_c: Corrected QT

- Because of its inverse relationship to heart rate, the QT interval is routinely transformed (normalized) into a “corrected” index denoted by QT_c. The intent is to represent the QT interval at a standardized heart rate (the QT interval at a heart rate of 60 bpm).
- Many different formulas have been proposed. The most common have the form

$$QT_c = QT / RR^a$$

where a is a formula-specific parameter:

e.g. $a=1/2$ gives Bazett's correction

$a=1/3$ gives Fridericia's correction

- Both of these were published in 1920...
- There are now more than 50 published versions of “corrected QT”.

What's a "Normal" QT_c?

Generally accepted guidelines:

- Normal QT_c: <430 ms in males, <450 ms in females
- Borderline QT_c: 430-450 ms (males), 450-470 ms (females)
- Prolonged QT_c: >450 ms (males), >470 ms (females)

"The prolongation of the QT interval to longer than 500 msec during drug therapy should prompt a critical reevaluation of the risks and benefits of that therapy..."

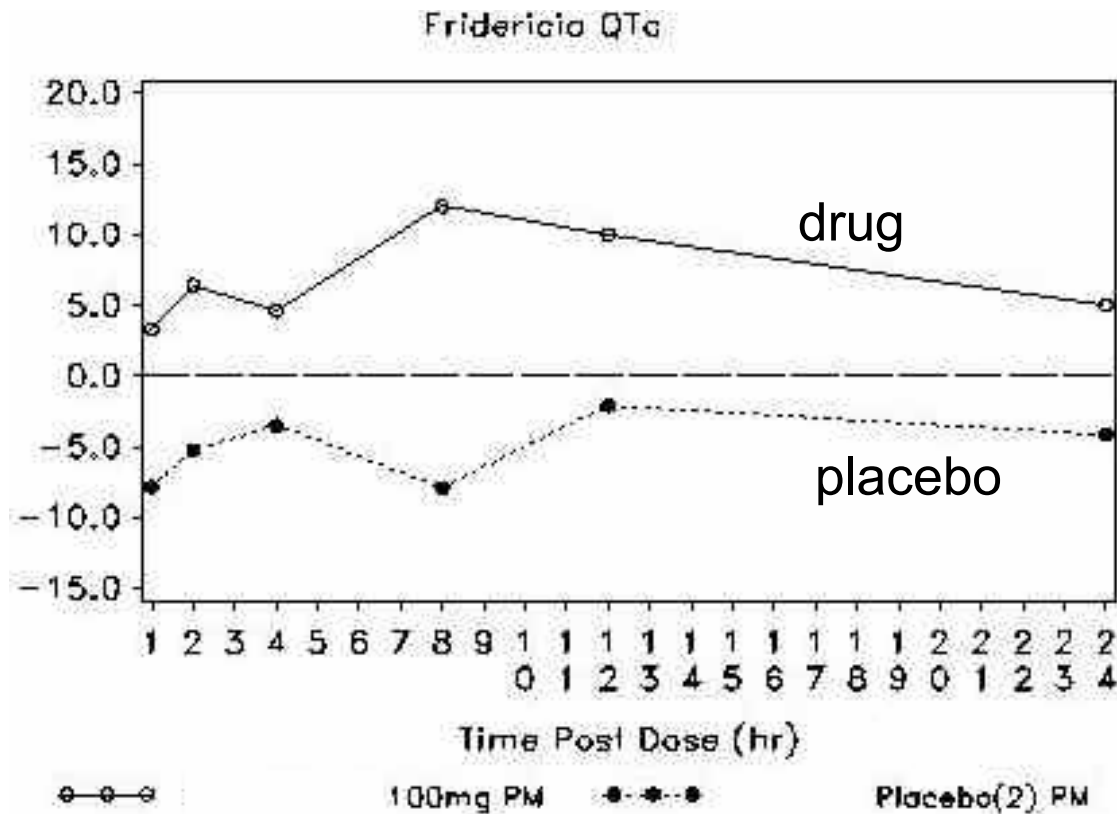
(D. M. Roden, NEJM, 2004)

Imperfect Biomarker

- *“The QT interval is used during drug development ... as a surrogate marker for the prediction of a serious adverse drug effect, syncope, or death due to TdP. However, as with many surrogate markers, its relationship to the event of interest is imperfect – the risk of TdP is not a linear function of the QT interval, nor of the extent of QT-interval prolongation during drug therapy.”*
- *“Although abnormal QT-interval morphology might predict an increased risk of TdP, analytic methods for assessing the relationship remains to be validated.”*

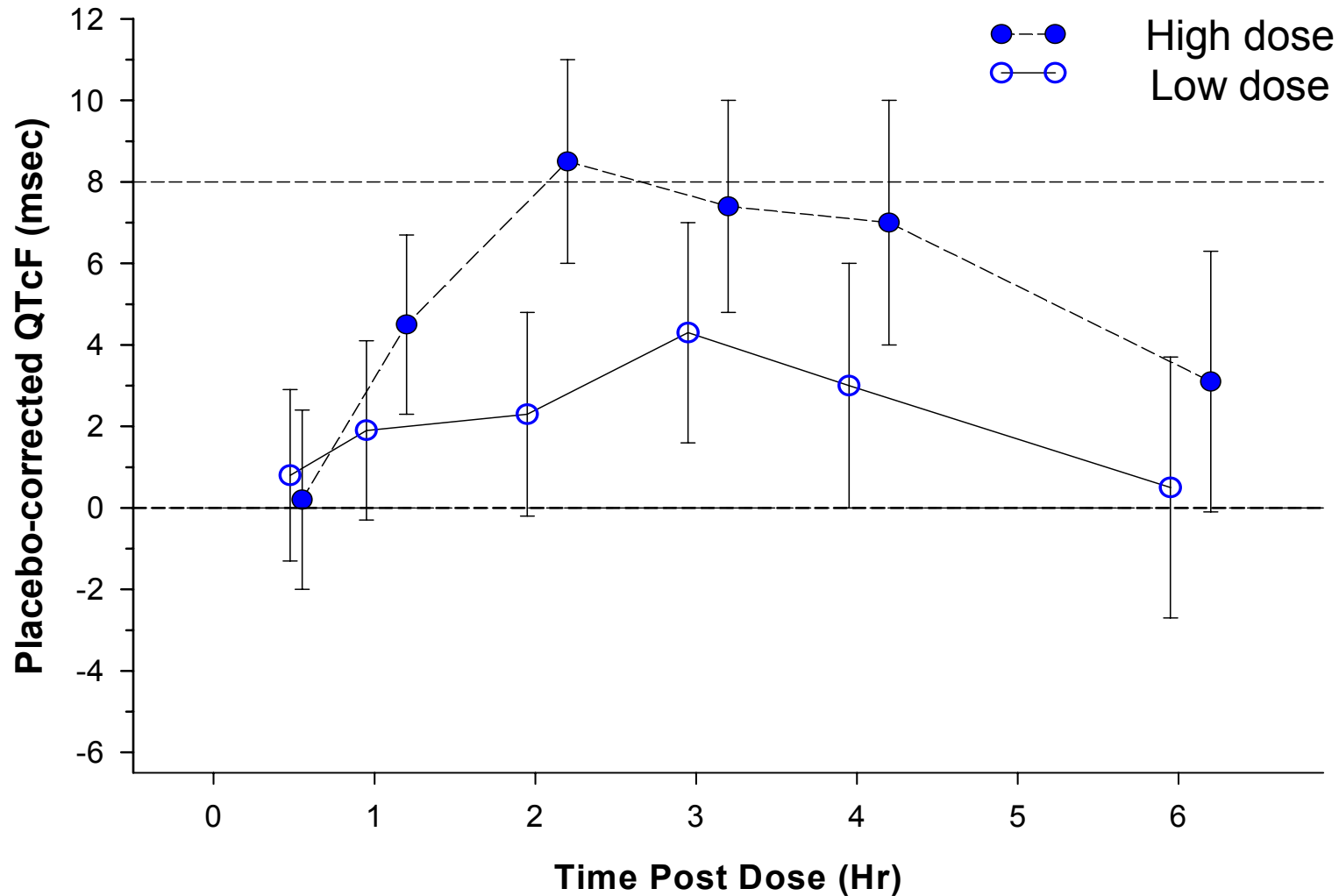
D.M. Roden, NEJM, 2004.

Example of Sample QT_c profiles



6 times post dose: 1, 2, 4, 8, 12, 24 hours

Placebo-corrected QT_c: 2 doses of a drug



The Statistical Problem in QT Study

- How do we determine if the effect of drug on QT is essentially the same (in the sense of being “non-inferior”) to that of placebo, in a “thorough QT study”?

- ICH E14 tries to address this problem.

ICH

- “ICH” is short for: *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use*
- ICH involves both regulatory agencies and industry from Europe, Japan and the US as “*equal partners in the scientific and technical discussions of the testing procedures which are required to ensure and assess the safety, quality and efficacy of medicines*”.
- The US partners are the Food & Drug Administration (FDA) and the Pharmaceutical Research and Manufacturers of America (PhRMA).

ICH Guideline (E14)

- In May, 2005, ICH released a guideline (E14): *The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs*
- Section on a “thorough QT/QTc study”:
 - “Drugs are expected to receive a clinical electrocardiographic evaluation ... typically including a single trial dedicated to evaluating their effect on cardiac repolarization...as detected by QT/QTc prolongation.”
 - “The study is typically carried out in healthy volunteers (as opposed to individuals at increased risk of arrhythmias) and is used to determine whether or not the effect of a drug on the QT/QTc interval in target patient populations should be studied intensively during later stages of drug development.”

ICH E14

- There is a single sentence devoted to statistical analysis: E14 says:
 - “... ***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.***”
- What does this mean? At best, it seems very vague.

ICH E14 (cont.)

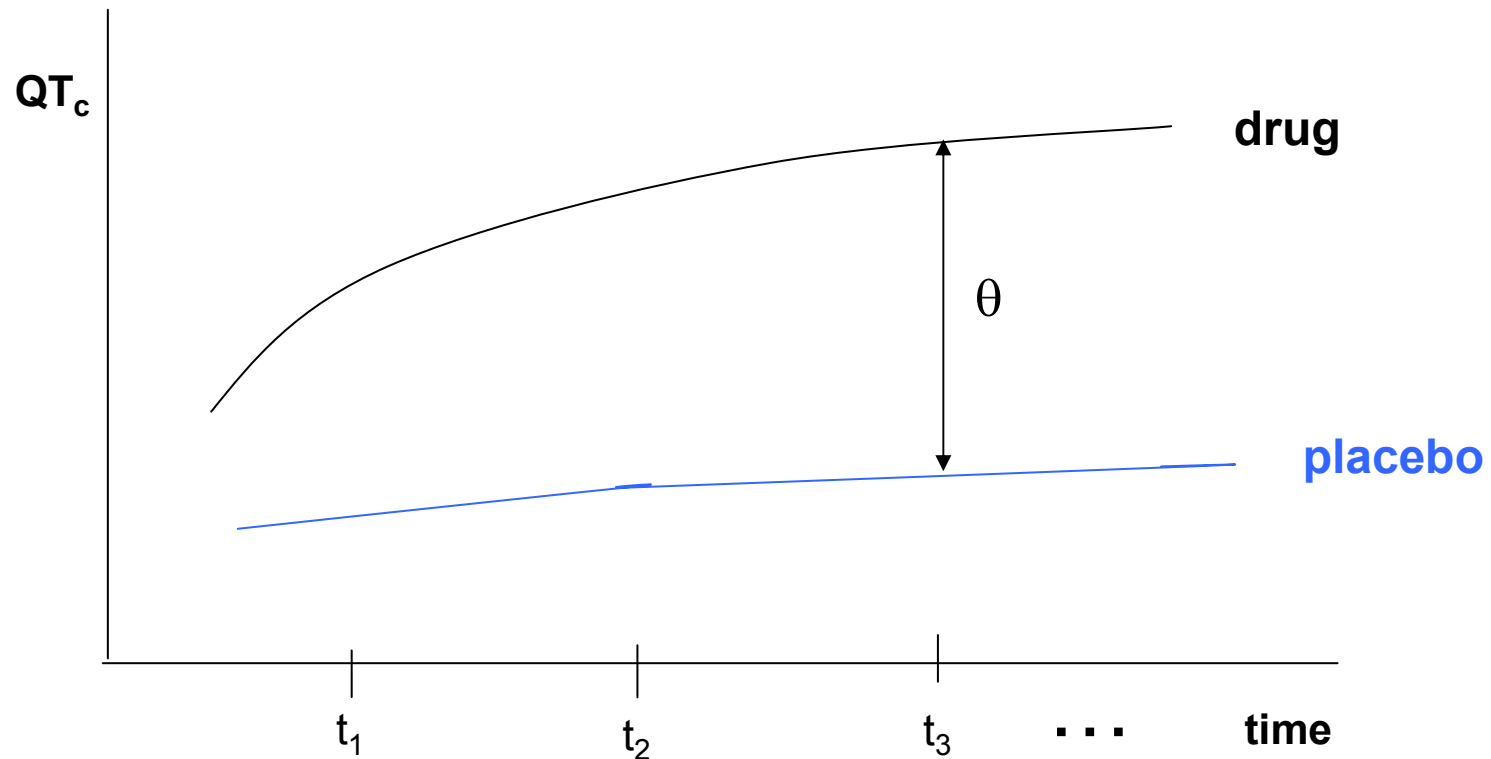
- E14 says to find:

... a one-sided confidence interval for the largest...mean effect of the drug on the QTc interval

- The only sensible interpretation of this statement is that the CI refers to a parameter that is the ***maximum coordinate of a difference of mean vectors.***

Confidence Interval for the Largest Difference in Mean QT_c

Let θ be the maximum distance between the QT population profiles (the means) of drug and placebo at the time points in the study.



A Simple Model

- Suppose we have n_D subjects on drug and n_P subjects on placebo, and QT (or QT_c) is measured at k time points.

- Model:

$$X_1^{(D)}, \dots, X_{n_D}^{(D)} \sim iid N_k(\mu^{(D)}, \Sigma)$$

$$X_1^{(P)}, \dots, X_{n_P}^{(P)} \sim iid N_k(\mu^{(P)}, \Sigma)$$

Simple Model (cont)

■ Let $\Delta = \mu^{(D)} - \mu^{(P)}$.

■ We want a confidence interval for the parameter

$$\theta = \max_{1 \leq i \leq k} \Delta_i.$$

■ There's just one minor detail (overlooked in E14).

■ ***No one knows a procedure for finding a CI for θ that always gives reliable results.***

CI for Δ (cont.)

- One method was investigated in:

Eaton, ML, Muirhead, RJ, Mancuso, JY & Kolluri, S (2006). *A confidence interval for the maximum mean QT interval change due to drug effect. Drug Information Journal, 40, 267-271*

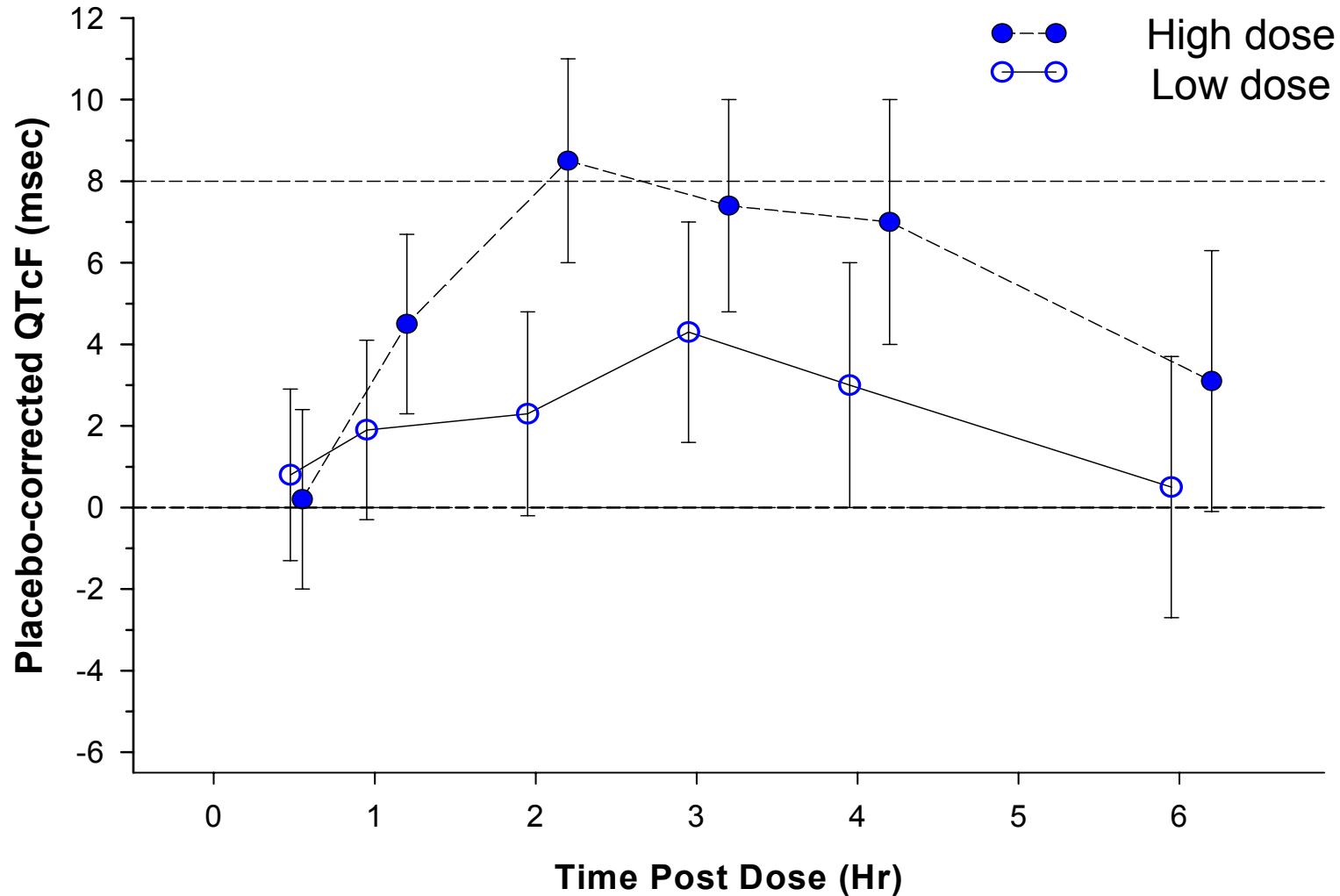
- Steps involved:

- ◆ Approximate θ by a smooth function
- ◆ Apply the “delta method” to give a CI for the smooth function
- ◆ This gives a CI for θ that asymptotically has the right coverage probability.

CI for Δ (cont.)

- This method doesn't work well if θ is close to other coordinates of Δ
- The paper was intended, not to promote the use of the method, but to sound a note of caution. Finding a CI for θ that is always reliable appears to be an unsolved problem.
- So – what are statisticians actually doing when analyzing data from a “thorough QT study”?
- Basically, they're ignoring what ICH E14 says.

Placebo-corrected QT_c: 2 doses of a drug



Traditional Hypothesis Testing Approach

- Calculate 1-sided 95% confidence intervals for each of the population mean differences and conclude that the study is negative (drug is non-inferior to placebo) if all upper bounds are less than or equal to 10 ms.
- This procedure corresponds to a well-defined hypothesis testing problem in multivariate analysis.

Hypothesis Testing (cont.)

- In testing whether drug is non-inferior to placebo, we want to be able to conclude that

$$\Delta_i < 10 \text{ for all } i = 1, \dots, k,$$

so this is the appropriate alternative hypothesis.

- Thus the null and alternative hypotheses are

$$H_0 : \Delta_i \geq 10 \text{ for at least one } i$$

$$H_A : \Delta_i < 10 \text{ for all } i = 1, \dots, k$$

Hypothesis Testing (cont.)

- The “standard” procedure involves testing each coordinate of the parameter vector Δ separately at the same significance level α using one sided t -tests, or equivalently, one-sided $100(1 - \alpha)\%$ confidence intervals.
- We reject H_0 in favor of H_A if and only if each of the individual t -tests is significant, or equivalently, if the upper limit of each one-sided confidence interval is less than 10 msec.
- This is an intersection-union test (IUT). It’s also the likelihood ratio test (LRT).
- It’s not perfect. For example, it’s conservative and biased.

Moving Beyond QT...

- Not all drugs that prolong QT are arrhythmogenic.
- An strong argument can be made that looking at QT/QTc alone is leading to poor decisions in the drug development/approval process.
- Norm Stockbridge (FDA)*

“We are left basing decisions largely on the surrogate of QT prolongation. We don’t know how to measure it. We don’t know how to correct for heart rate. We don’t know how to balance a given effect on QT against a (hopefully better studied) clinical benefit.”

* International Conference on Drug Development, Univ. of Texas. Feb, 2005

Beyond QT alone

- An ECG is a biomarker of heart activity.
- Looking at QT alone as a predictor of arrhythmia represents an attempt to reduce an inherently multivariate problem to a univariate one.
- QTc tries to correct QT for heart rate, but is still a univariate measure of a not-well-understood relationship between QT and HR.
- There is a growing awareness that there are advantages to taking the multivariate nature of ECG parameters into account.

Holter Data

Pfizer recently completed a small pilot study in which 6 subjects (healthy normal volunteers) all wore holters, and had every heart beat recorded for two days (actually, about 20 hours each day).



Example of Beat-to-Beat Data

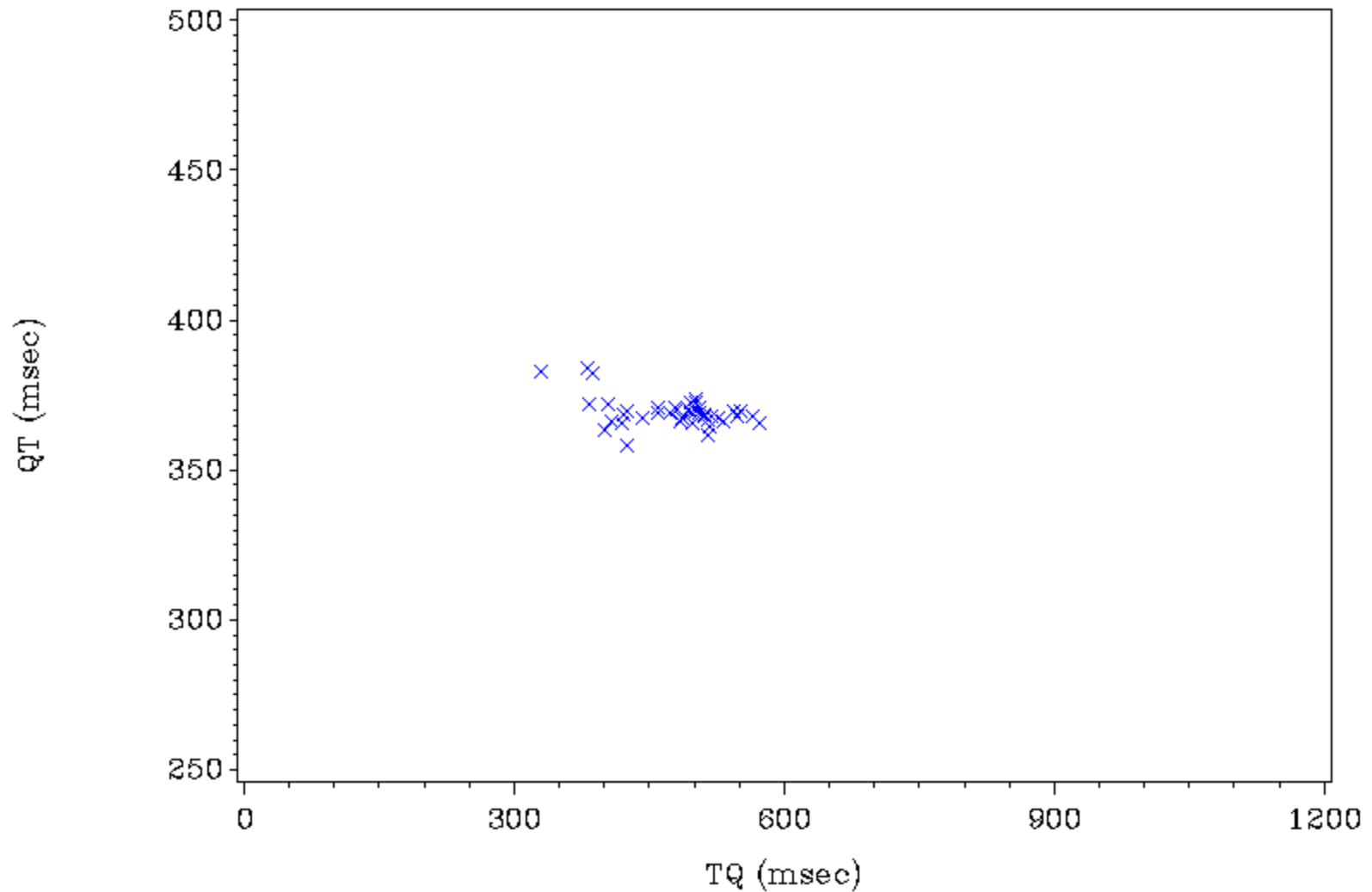
- 38 healthy volunteers from a Pharmacia study
- Study design was open-label with fixed treatment sequence administered on consecutive days. Day 1 was baseline data.
- Digital 12-lead Holter data were recorded for approximately 22.5 hours on each day.
- ECG parameters RR, QT, and TQ were extracted from digital signal. Look only at Day 1 (baseline) in the examples that follow.

Example (TQ, QT) Data

1 Subject

QT against TQ : 0 to 1375 mins

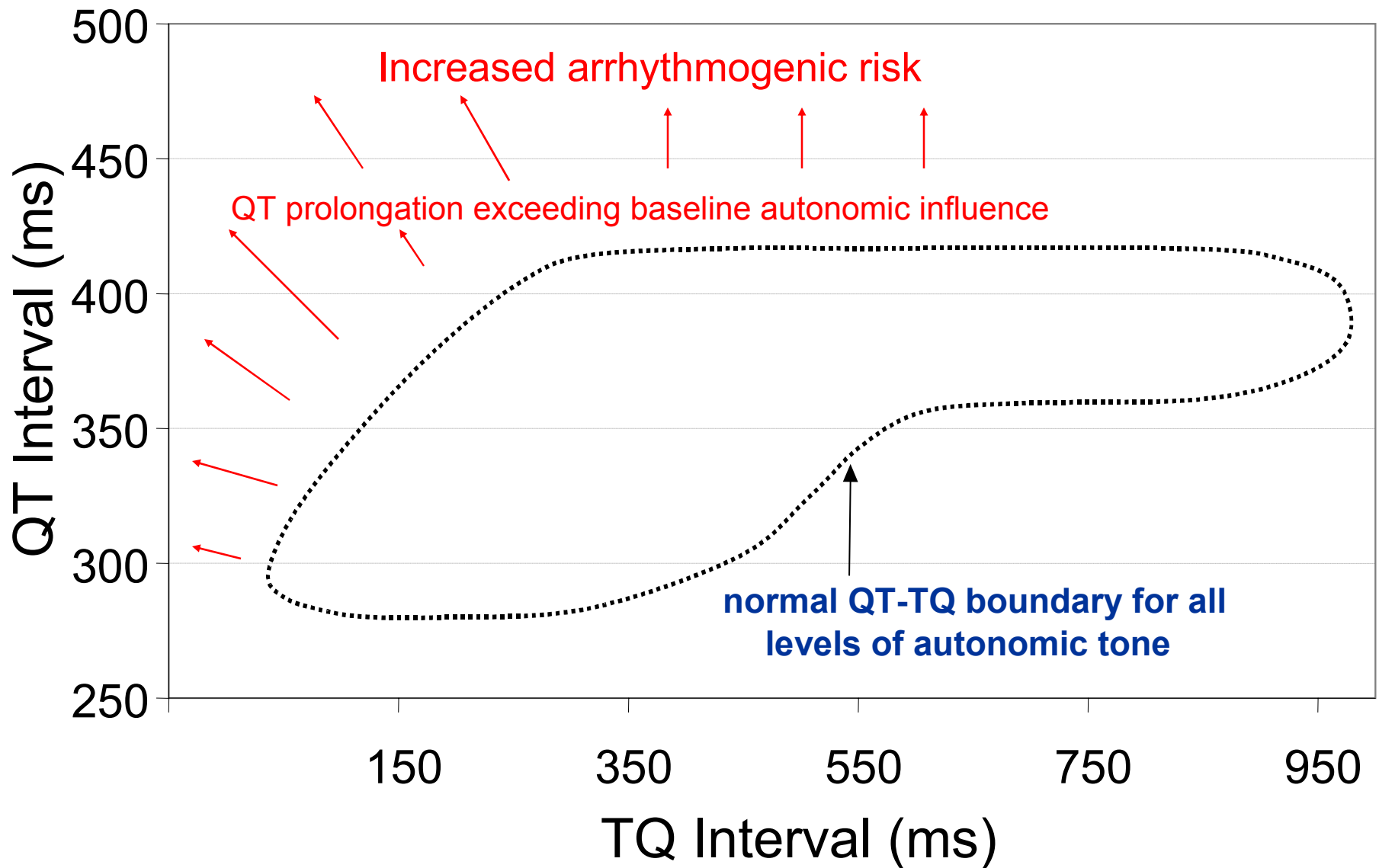
Time (in mins): 0



A Current Hypothesis

- Recent scientific work conjectures that each person has his or her own (TQ, QT) region (or (RR, QT) region).
- This region encompasses all daily activities: eating, sleeping, exercising, etc.
- Only when a compound causes (QT, TQ) points to move outside this region in a specified direction is the risk of an arrhythmia increased.

Arrhythmia Triggering Zone



“Thorough QT Study”

- All “non-arrhythmic” compounds are expected to go through a “thorough QT study”.
- Often this study is:
 - ◆ A cross-over design where each subject has QTc measured at a fixed number of time points, usually between 4 & 8
 - ◆ The times are chosen to bracket T_{max} and depend on the half-life of the compound
 - ◆ Then a test of whether the drug is “non-inferior” to placebo is carried out. ICH E14 is rather vague about the details of this test.

Thorough QT Study ...

- The (TQ,QT) relationship is a dynamic one. It isn't captured by looking at QTc alone.
- There is evidence that using QT/QTc alone is inadequate in assessing the arrhythmia liability of a drug.
- How can we improve on current practice and design studies that more effectively address the question of whether a compound increases arrhythmic liability?

Future QT Studies?

- Instead of measuring QT at (say) 6 time points, we might collect beat-to-beat (QT, TQ) data over the course of a study.
- Subjects are already in a clinic.
- Technology has advanced to allow collection of ECG data via a Holter at 1000 Hz (1 ms measurement error).
- Using Holters yields much more information.

Future Studies? (cont.)

- Placebo data collected can be used to estimate a *baseline reference region*.
- Points after administration of compound would be compared to this reference region.
 - ◆ Do any beats give (QT, TQ) values which lie outside the reference region?
 - ◆ How far outside? If we look at these points over time, do they return to the reference region quickly?

Multivariate Reference Regions...

- Given a multivariate population (for example, a bivariate population of all (TQ, QT) measurements), a **95% reference region** is a region which contains 95% of all the population.
- Regions that take into account the shape of the multivariate distribution (and thus the underlying correlation structure) are appealing and have optimal properties.
- In practice we see a sample from the population (e.g., beat-to-beat (TQ, QT) data over the course of a study), and we use this sample to estimate our desired reference region.

Multivariate Reference Region: Example

Given a sample of size n from a $N_p(\mu, \Sigma)$ distribution, where (μ, Σ) are unknown, the standard 95% reference (tolerance) region is

$$R(X) = \{z \mid (z - \bar{X})' S^{-1} (z - \bar{X}) \leq c\},$$

where

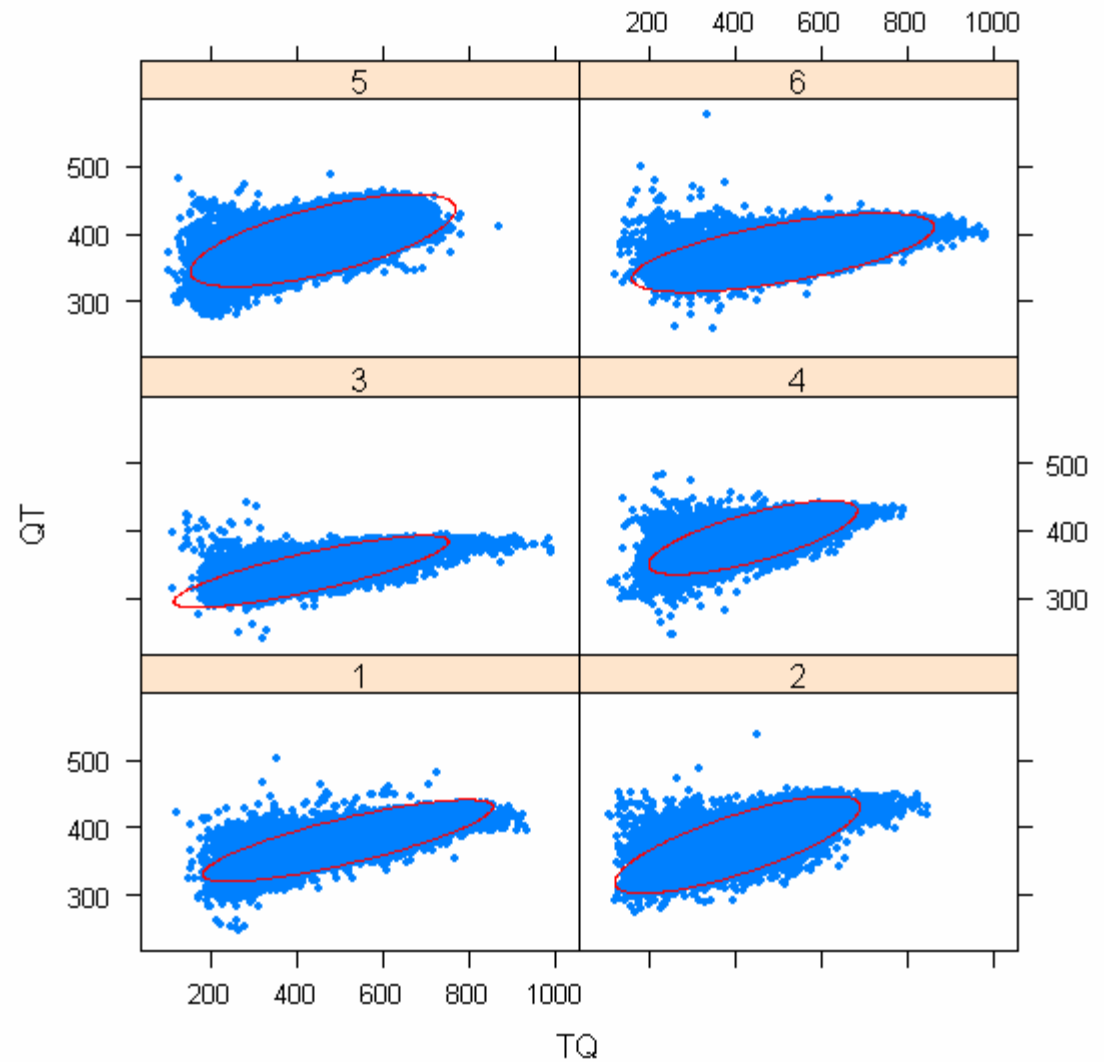
$$c = \frac{p(n+1)}{n-p} F_{p, n-p}(.95)$$

$R(X)$ is the boundary and interior of a p -dimensional ellipse centered at the sample mean and with shape and orientation determined by the sample covariance matrix S .

Reference Regions

6 subjects from example data set

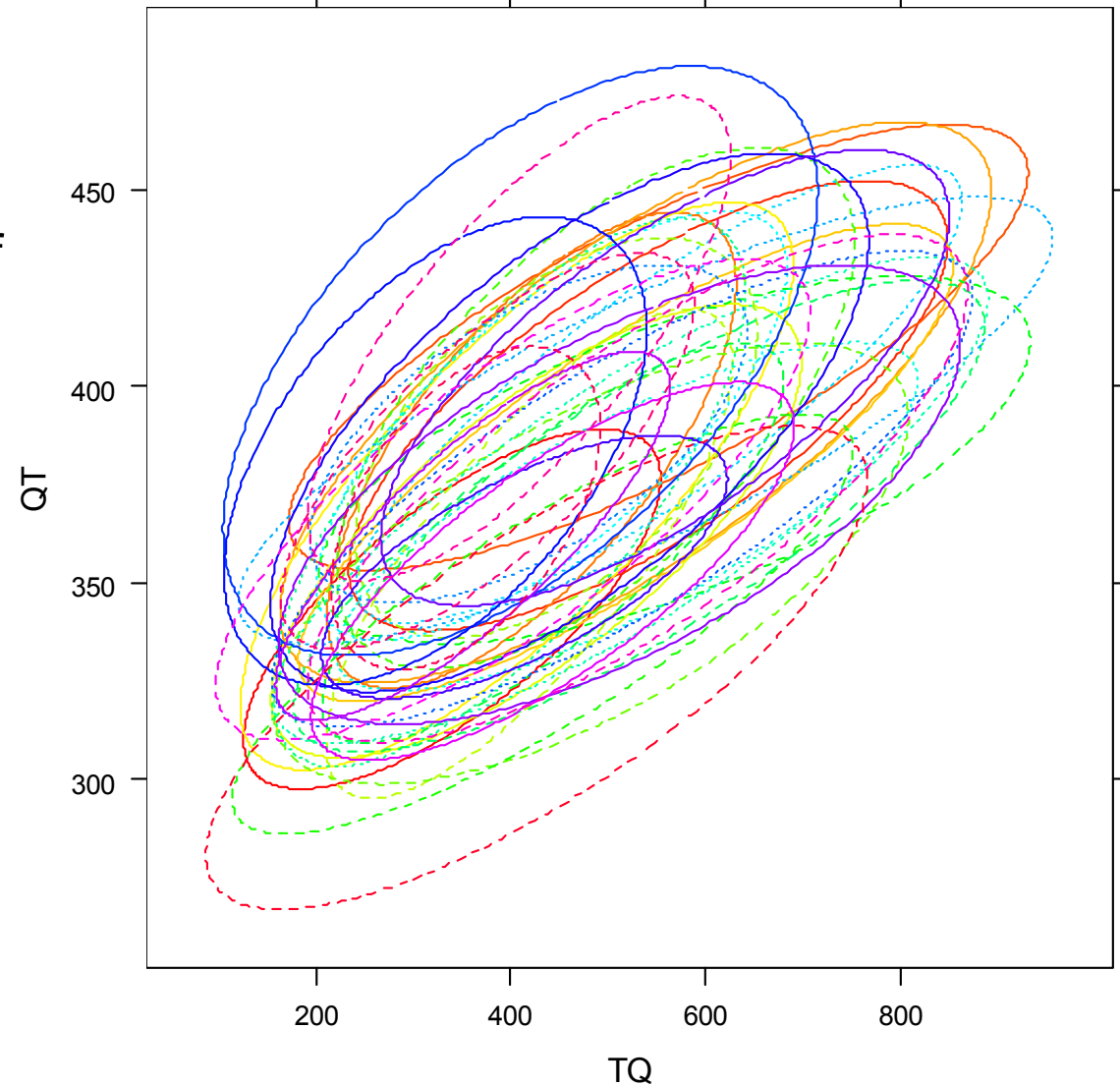
Subject	Coverage
1	95.39
2	95.66
3	96.43
4	95.22
5	94.83
6	96.93



22 Hour Reference Regions

All 38 Subjects

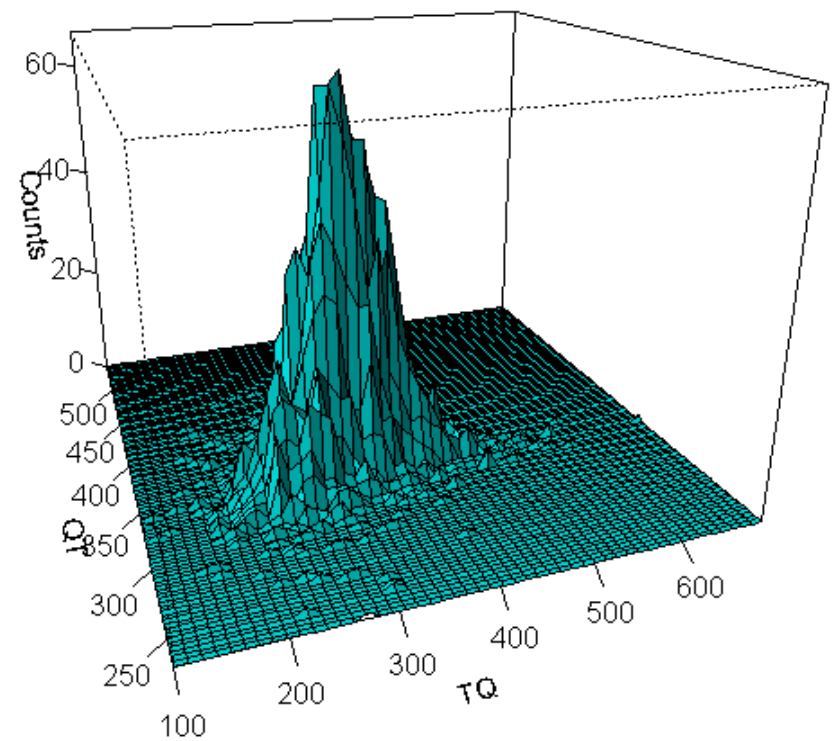
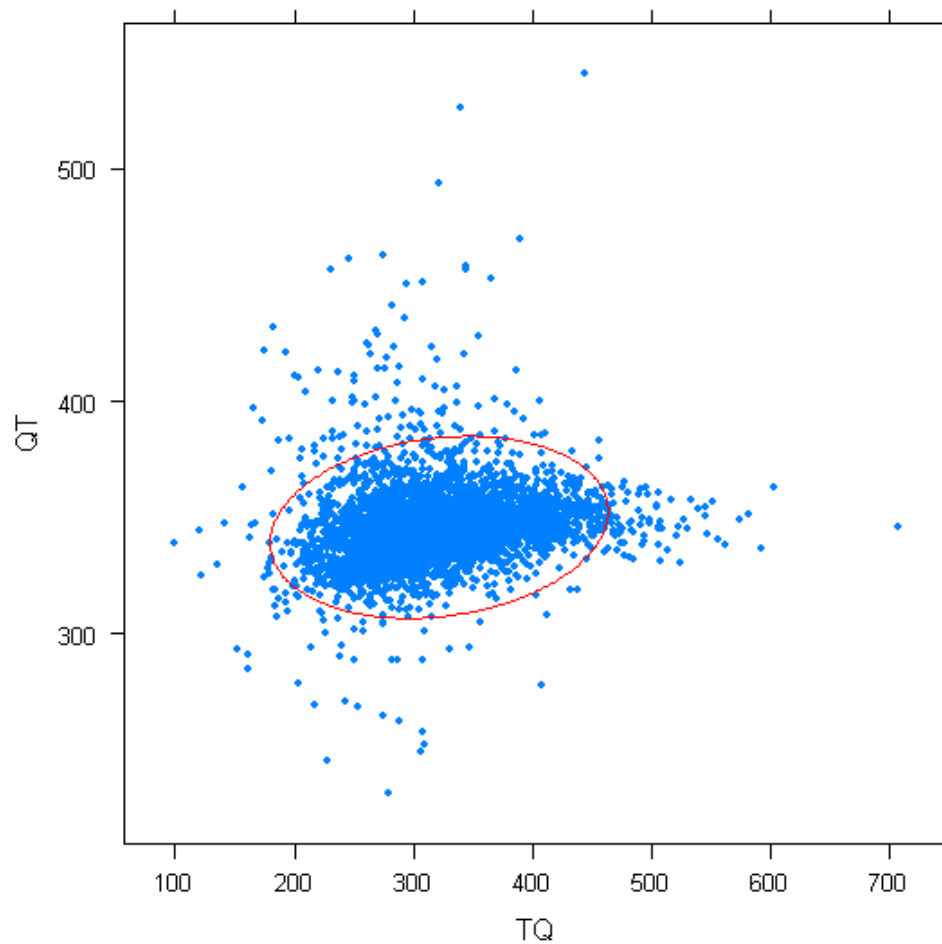
Differences in location, variability and correlation of the reference regions for all subjects.



Example

One subject data collected from 7 to 8 PM. Dinner was at 6 PM

For short time periods or periods of similar activity (e.g. sleep), the bivariate distribution looks reasonably normal.



Multivariate Normality

- The assumption of bivariate normality appears appropriate when we consider short time periods, or periods where activities are similar (e.g. sleep).
- If the elliptical 95% reference region has been constructed from $X_1, \dots, X_n \sim iid N_p(\mu, \Sigma)$, then if Z is a new observation independent of the X_i 's

$$P(Z \in R(X)) = 0.95 \text{ for all } \mu, \Sigma.$$

- Optimal properties -- see "Appendix" slides:
 - ◆ Smallest expected volume out of all invariant regions
 - ◆ Justified by Bayes predictive approach

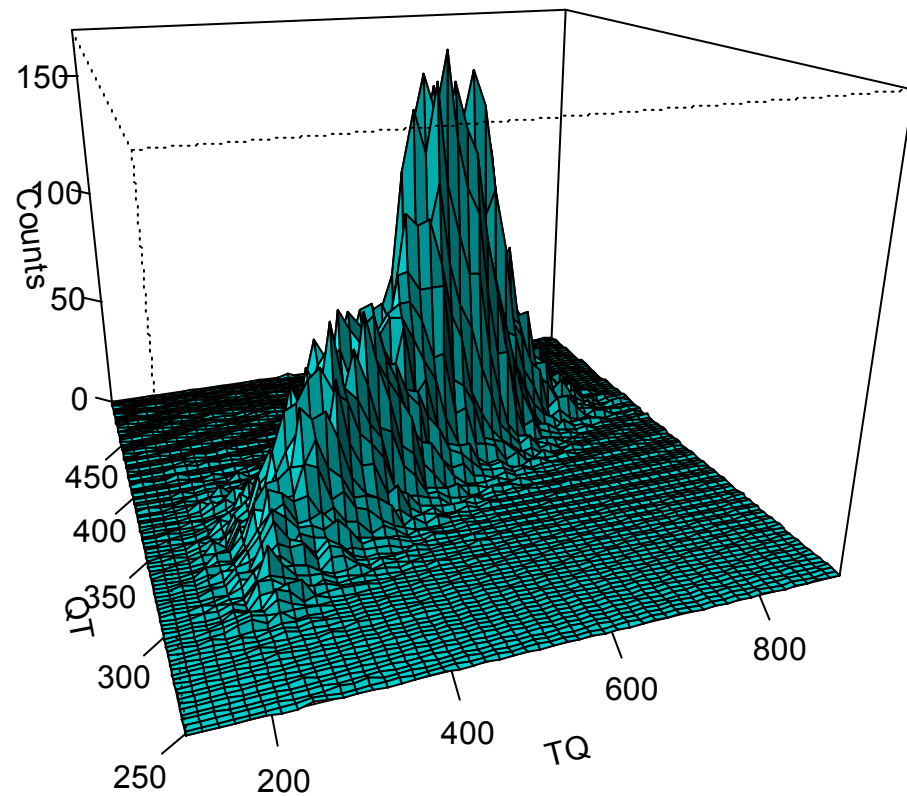
2D Histogram

1 Subject

Over the course of a full day, however, the distribution does not look bivariate normal.

Bi-modal distribution

- High heart rates (low TQ)
- Low heart rates (high TQ)



Mixture of Normal Distributions

- We know the variability in QT and TQ is different for high and low heart rates.
- A k -component mixture of bivariate normal distributions has a density function of the form

$$f(x) = \sum_{i=1}^k p_i \phi(x | \mu_i, \Sigma_i)$$

where ϕ is the bivariate normal density function;

p_1, \dots, p_k are the mixing probabilities;

μ_1, \dots, μ_k are the mean vectors;

$\Sigma_1, \dots, \Sigma_k$ are the covariance matrices

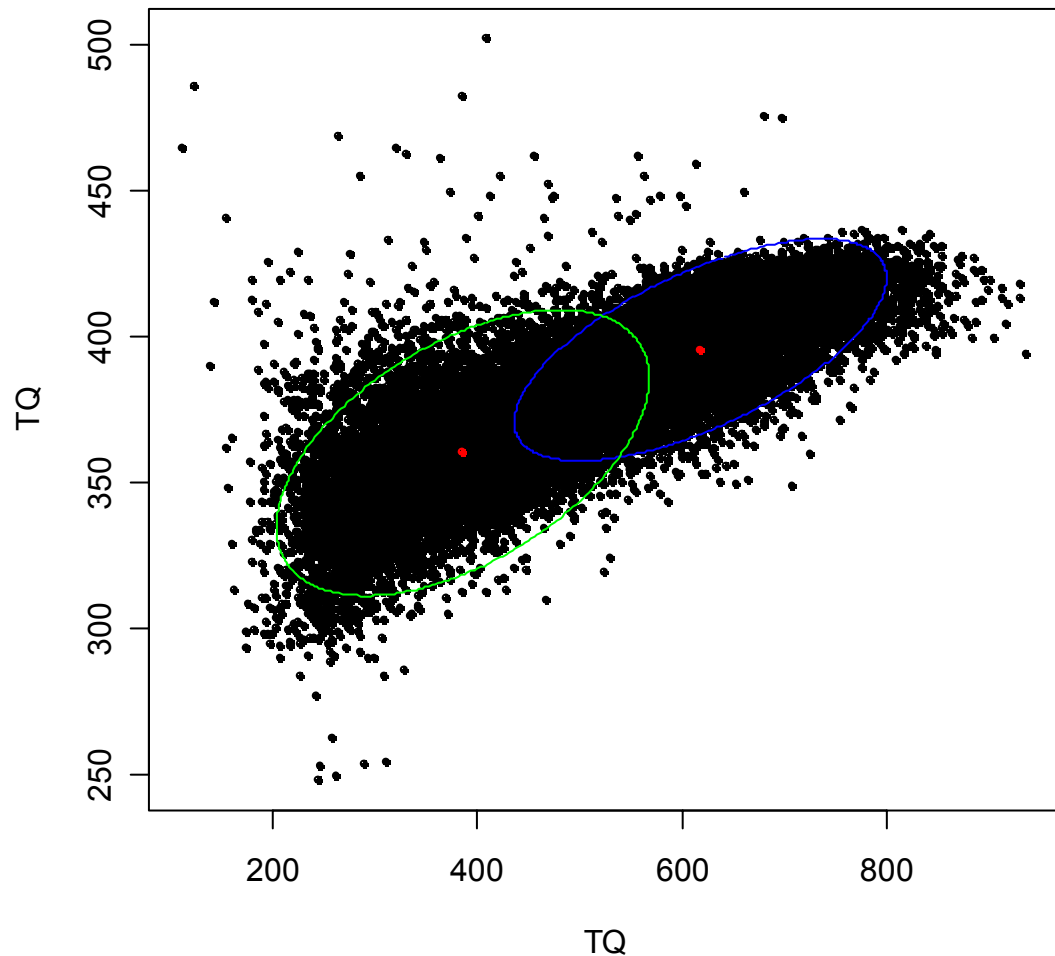
Fitting a two component mixture

In fitting a 2-component mixture, the variability is modeled separately for low and high heart rates.

Each ellipse is the 0.95 contour of the component. The reference region of the mixture is not displayed.

Two component mixture is not modeling the extreme observations.

A three component mixture model might be better



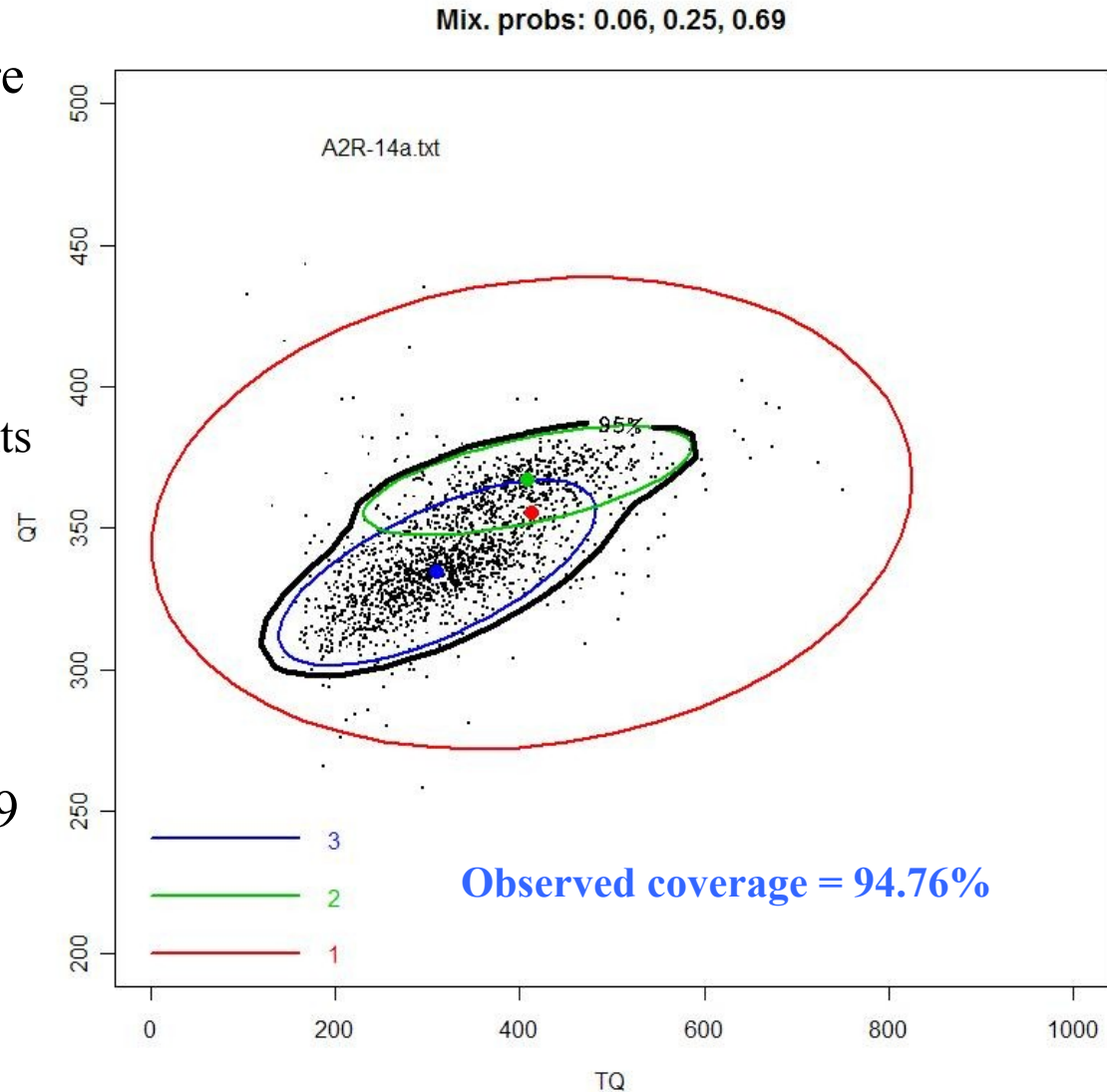
3-Component Mixture Model

1 subject

Red, blue and green lines are 95% contours for each component.

- One component models points with high variability
- The two other components model low & high HR

Black line is the reference region for 3-component mixture model with mixing probabilities 0.06, 0.25, 0.69

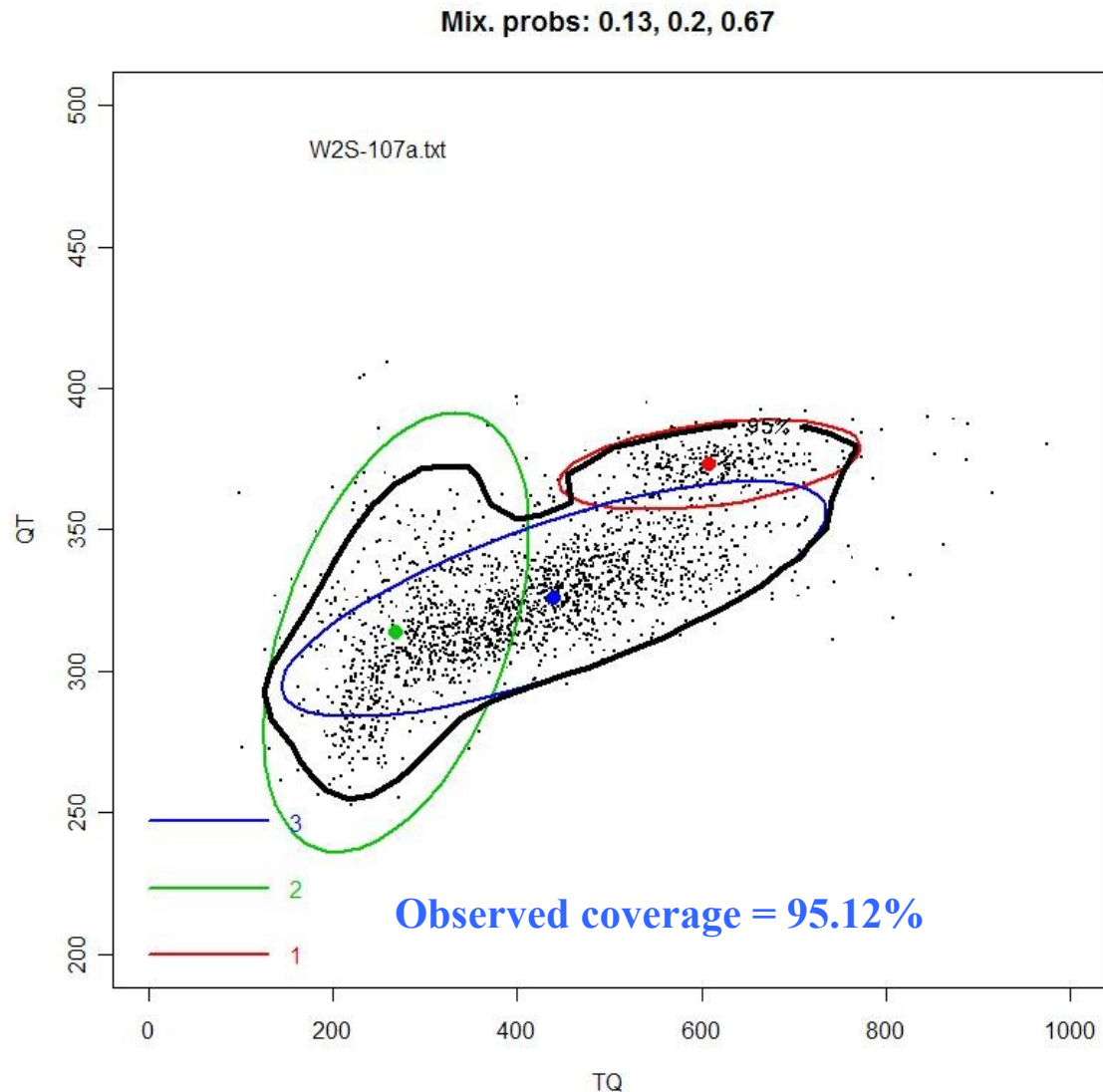


3-Component Mixture Model

Another subject

Red, blue and green lines are 95% contours for each component.

Black line is the reference region for 3-component mixture model with mixing probabilities 0.13, 0.2, 0.67

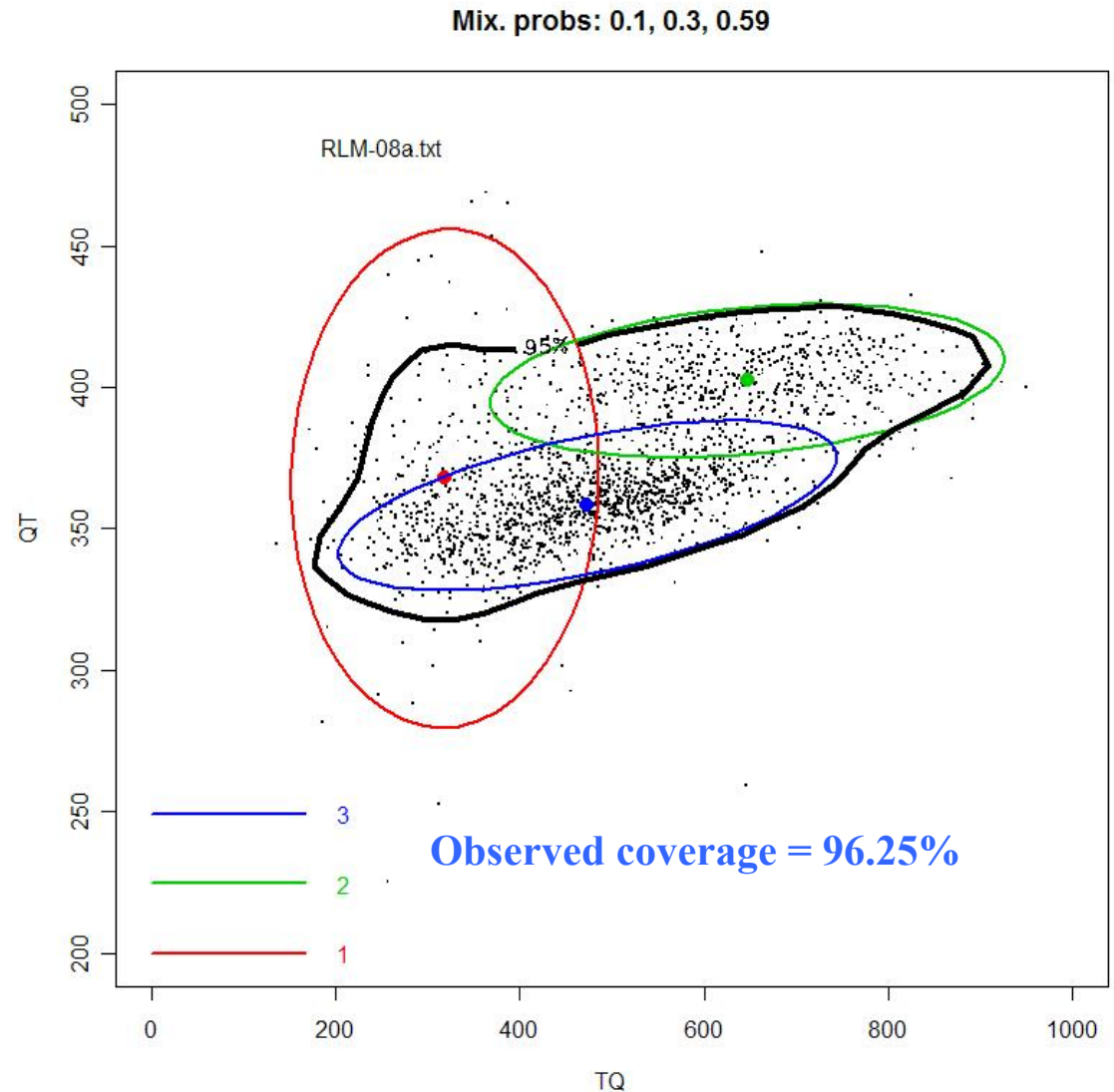


3-Component Mixture Model

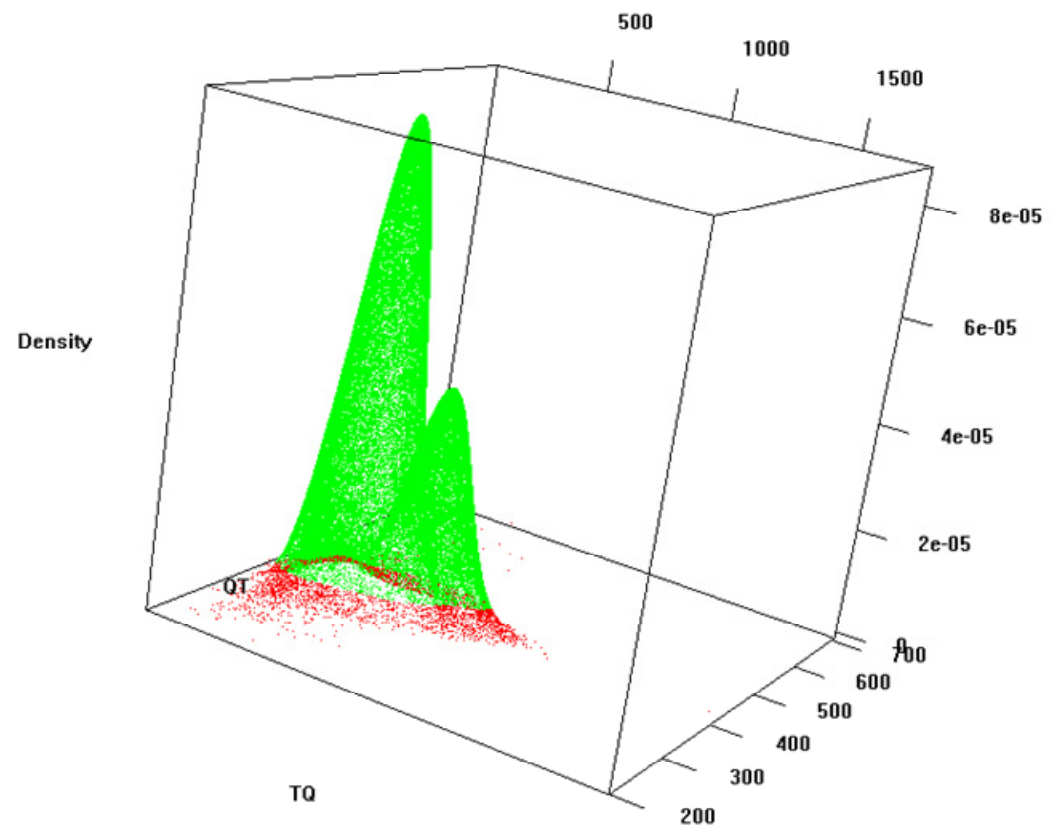
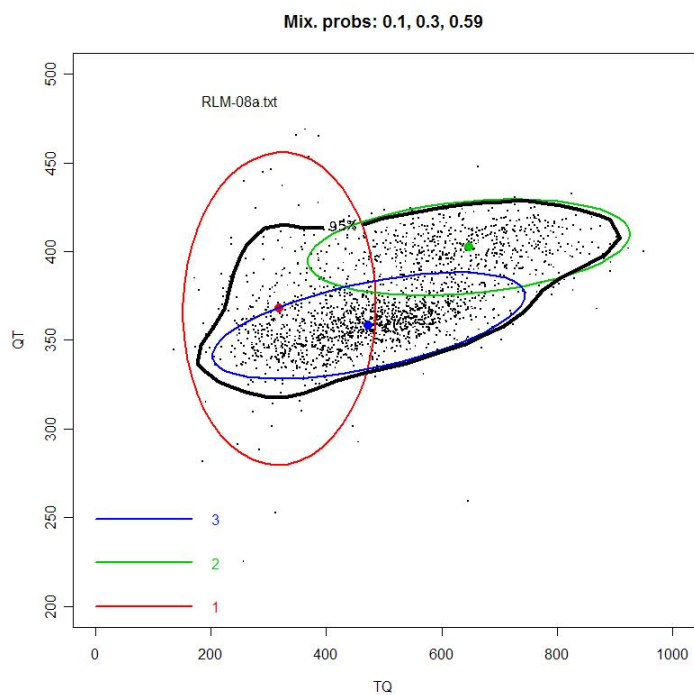
Another subject

Red, blue and green lines are 95% contours for each component.

Black line is the reference region for 3-component mixture model with mixing probabilities 0.1, 0.3, 0.59



Fitted mixture distribution and 95% reference region.



Summary

- QT is an imperfect biomarker as a predictor of arrhythmia.
- There are statistical problems with the interpretation of the recommendation in ICH E14.
- We can design more effective and more informative studies than the current “thorough QT studies”, and we should be thinking now about what these should look like.

Summary (cont.)

- Baseline reference regions could be used as potential biomarkers of arrhythmic liability by comparing (QT,TQ) points after administration of a compound.
- We have shown how beat-to-beat (QT,TQ) data may be modeled using mixtures of normal distributions.
- We used the fitted mixture models to define reference regions (or equivalently, predictive regions for new measurements).

Summary (cont.)

- In the case of a single normal model (a reasonable assumption for short time periods), the standard elliptical reference region may be justified on the grounds that:
 1. It has smallest expected volume out of all invariant reference regions
 2. It follows from the predictive distribution for a new observation generated by the standard improper prior
 3. It has desirable decision-theoretic properties.

References

- Eaton, ML., Muirhead, RJ., Pickering, EH. (2006).

Assessing a vector of clinical observations. Journal of Statistical Planning and Inference, 136, 3383-2414.

- Eaton, ML, Muirhead, RJ, Mancuso, JY & Kolluri, S (2006).

A confidence interval for the maximum mean QT interval change due to drug effect. Drug Information Journal, 40, 267-271

- Fossa, A, Wisialowski, T, Crimin, K. (2005).

Dynamic beat-to-beat modeling of the QT-RR interval relationship: analysis of QT prolongation during alterations of autonomic state vs. hERG inhibition. Journal of Pharmacology and Experimental Therapeutics, 312, 1-11.

References (cont.)

- Fossa AA, Wisialowski T, and Crimin K. (2006).

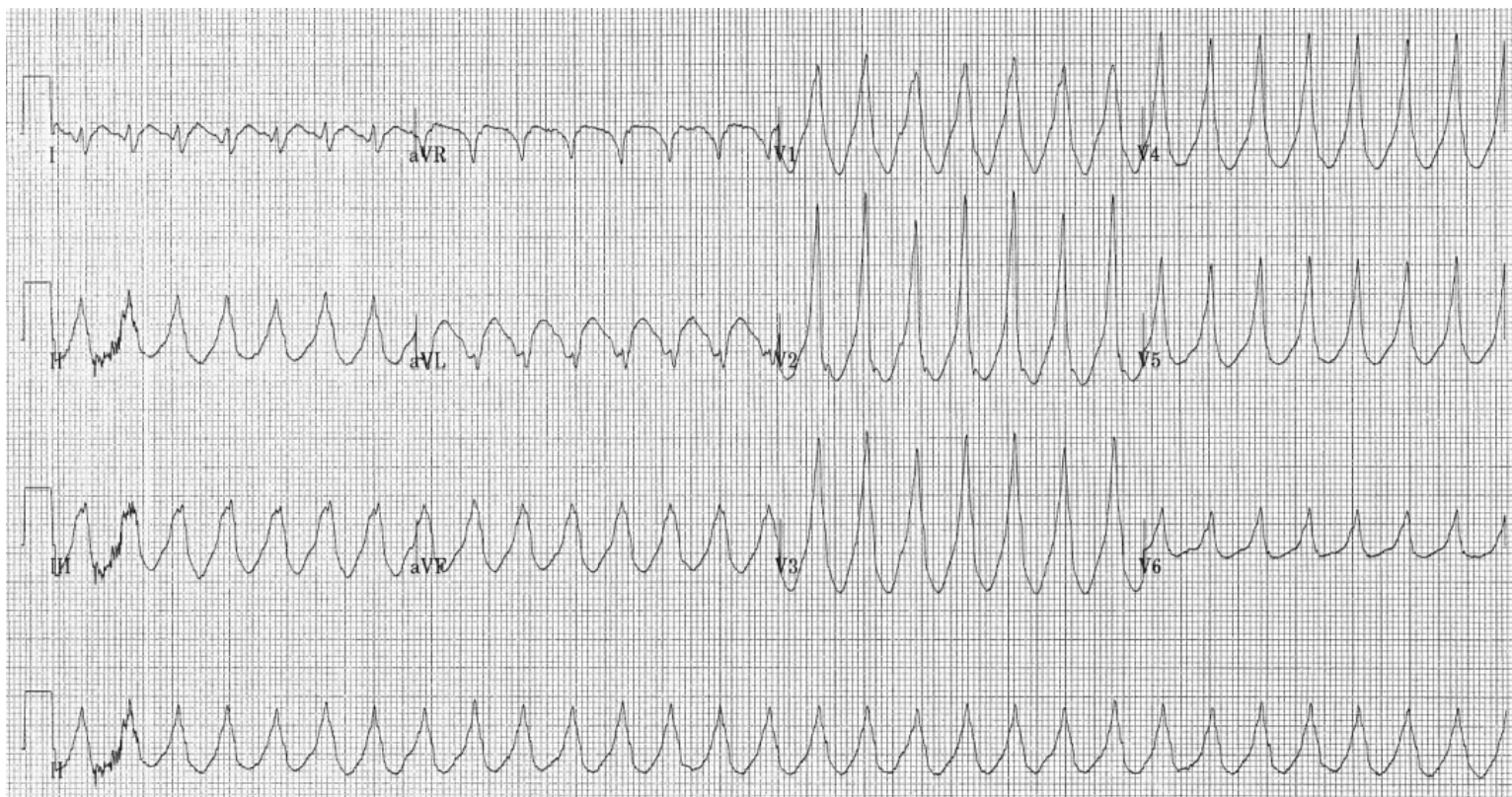
QT prolongation modifies dynamic restitution and hysteresis of the beat-to-beat QT-TQ interval relationship during normal sinus rhythm under varying states of repolarization. *Journal of Pharmacology and Experimental Therapeutics*, **316**, 498-506.

- Roden, DM (2004).

Drug-induced prolongation of the QT interval. *New England Journal of Medicine*, **350**, 1013-1022.

“Appendix” Slides

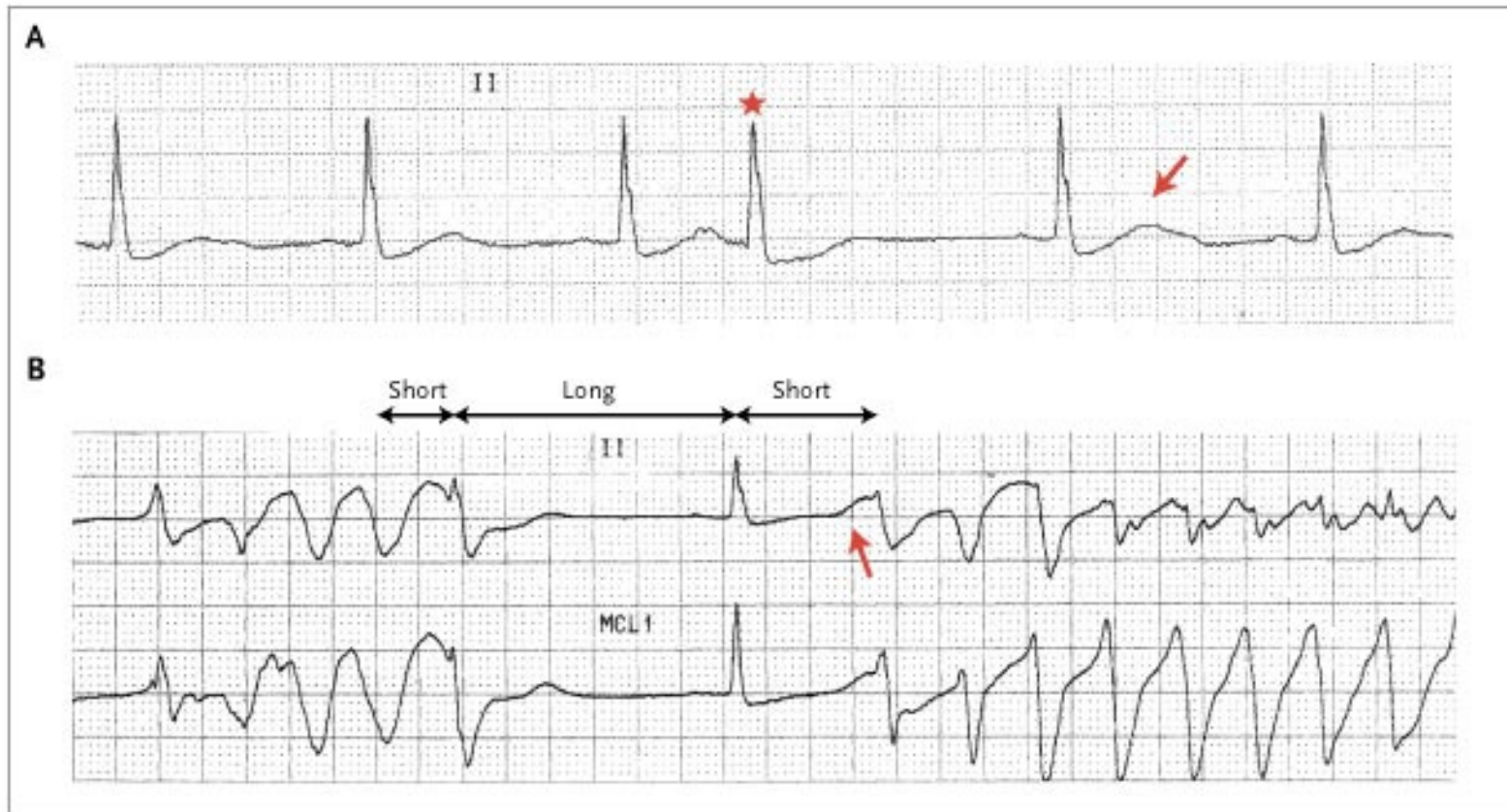
Tachycardia



Bradycardia



Drug-Associated TdP



Source: D.M Roden, NEJM, 2004.

Drugs Withdrawn

- Terfenadine
- Astemizole
- Grepafloxacin
- Terodiline
- Lidoflazine
- Sertindole
- Levomethadyl
- Cisapride

Optimality Properties of Reference Regions

- It is intuitively appealing to compare competing 95% reference regions on the basis of their average volumes (areas if $p = 2$), where average volume is

$$v(\mu, \Sigma) = E[\text{Vol}(R(X))].$$

- It is also appealing to consider reference regions that are invariant under **affine transformations**, so that properties don't change if, for example, measurement units are changed or we look at linear combinations of QT and TQ, perhaps suggested by principal components.

Optimal Volume Property

- Out of all invariant 95% reference regions, the elliptical region

$$R(X) = \left\{ z \mid (z - \bar{X})' S^{-1} (z - \bar{X}) \leq \frac{(n+1)p}{n-p} F_{p, n-p}(.95) \right\}$$

has the smallest expected volume.

Bayesian Considerations

- The reference region approach can also be regarded as the problem of constructing a predictive region for a new observation Z .
- In the usual Bayesian set-up, one way of finding a predictive region is to look at contours of equal density of a **predictive distribution** for Z .
- The first step involves specifying a joint prior distribution for μ and Σ .

Bayesian Considerations (cont.)

- The desirable invariance property we have already introduced strongly suggests using the standard (Jeffreys) improper prior distribution

$$p(\mu, \Sigma) = |\Sigma|^{-(p+1)/2},$$

which is itself invariant under affine transformations.

Predictive Distribution

- A formal Bayes calculation shows that the predictive distribution of Z corresponding to this prior (that is, the conditional distribution of Z given X) multivariate t , with density function

$$f(z|x) = c \cdot |S|^{-1/2} \left[1 + \frac{n}{n^2 - 1} (z - \bar{x})' S^{-1} (z - \bar{x}) \right]^{-n/2}.$$

- The upper sections of this predictive distribution generate p -dimensional ellipses of the exact type we have been considering.
- The standard 95% reference region is thus also justified via a Bayesian predictive approach.