Statistical issues regarding the analysis of ECG intervals
A. Ring, Biostatistics, Boehringer Ingelheim Pharma GmbH&Co. KG, Biberach, Germany

The newly established ICH E14 guideline leaves some flexibility regarding the statistical analysis of “thorough QT trials”. This flexibility should be used to improve the analysis, so that the sample sizes and overall costs for this type of clinical study could be reduced. Hence, most pharmaceutical companies have plans to create their own “ECG data base” to investigate both treatment-dependent as well as treatment-independent ECG data for obtaining the best strategies with regard to data sampling, data aggregation, heart rate correction or choice of the active control.

Based on the current standard for ECG recording (10 seconds – 12-lead ECG), we want to present two examples for improved statistical methodology.

1. The aggregation of the ECG data to endpoints
Thorough QT studies are data rich: Typically, triple ECGs are recorded at each point in time, and 3-5 wave forms are measured from each ECG. Due to the circadian rhythm of the ECG intervals and the undetermined relationship of the PK profile to the onset of potential QT interval prolongation, such triple ECGs are recorded about 5-10 times over each study day (ECG profile). The estimation of the placebo-adjusted treatment change from baseline requires at least 4 of these ECG profile days. Altogether, the number of observed ECG wave forms per subject is typically between 500 and 2000. The aggregation of these wave forms to the relevant subject specific endpoints is usually performed using arithmetic means. The underlying assumption of normally distributed data (especially for the RR interval and the QT interval) is typically not checked within a trial. Moreover, the arithmetic mean is not consistent with parabolic heart rate correction formulas, such as the Fridericia method.

In one thorough QT study (including 54 subjects with 500 wave forms on baseline measurements for each subject), the QT intervals appear to be log-normally rather than normally distributed. For the RR intervals, both the normal and the lognormal distributions fit similarly well to the data. To estimate the location parameter of log-normally distributed variables, the geometric mean rather than the arithmetic mean should be preferred – this aggregation is also consistent with parabolic heart rate corrections.

The observed deviation from normality was small (e.g. for heart rate corrected QT intervals less than 1 ms). As the non-inferiority margin in the E14 guideline is only 10 ms and the proposed E14 endpoint (largest mean change from baseline) is positively biased, even such a small bias could make a difference in the success of a trial. In addition, the incorporation of the statistical distribution of ECG intervals (and not only the QT/RR relation) could lead to improved heart rate correction formulas.

2. Decreasing autocorrelation of “change from baseline” over time
The intraindividually observed “change from baseline” at each point in time is the proposed endpoint in thorough QT studies. Typically in multiple dose studies, the change from baseline is observed at the first day of treatment (to investigate the acute effect) and at the last day of treatment at steady state.

We have observed in our QT studies that the between-subject variability of “change from baseline” QT(c)-endpoints is substantially smaller when the baseline profile has been observed only one or two days before the treatment profile compared to the situation when there are larger time intervals between treatment and baseline profiles. This suggests that the correlation between the ECG interval decreases over time. Hence, the single dose approach could be seen as advantageous over the multiple dosing approach, if also safety and PK allow its usage.

There are a large number of such statistical questions for further investigation. Therefore, we would propose to connect parts of the “company ECG data bases” within Germany to establish a sound foundation for statistical research.