Abstract. The purpose of this investigation was to define the sensitivity and specificity of the canine telemetry assay for detecting drug-induced QT interval prolongation. Data from twelve studies generated in the QT PRODACT project were used in this investigation. The study design was a $4 \times 4$ Latin square cross-over design and included the following drugs: MK-499, E-4031, terfenadine, haloperidol, cisapride, bepridil, propranolol, diphenhydramine, captopril, verapamil, amoxicillin, and ciprofloxacin. The estimated root squared error of the model, which estimated the slope of the QT-RR relationships for each animal, for all dogs during the pre-dosing period was 5.45%. Using the QT-RR model, the sensitivity and specificity in each cutoff value that judges QT prolongation were estimated based on the experiment errors and measurement errors in the 12 studies. When the cutoff value was 5%, the sensitivity in 10% prolongation was 0.978 and the specificity in 0% was 0.996. In conclusion, it was judged that a 5% cutoff value for changes in heart rate corrected QT interval using the canine telemetry assay is practical, and the sensitivity and specificity of the telemetry assay are very high when using the analytical method presented here. Based upon this information, the canine telemetry assay using the individual subject heart rate correction model is recommended as a sensitive test system for the in vivo assessment of risk for QT interval prolongation.

Keywords: dog, sensitivity, telemetry, QT, QTc
tation methods are the most frequently used in dogs as well as humans (4 – 6), they have been questioned due to the over-correction at faster heart rates and under-correction at slower heart rates (7 – 9). Lately, although the rate-correction optimized for each subject has been recommended for dogs as well as humans (8, 10), how much the individual subject correction can improve the sensitivity and specificity compared to the generic QTc formulas has not been evaluated.

The purpose of this investigation was to define the sensitivity and specificity of the canine telemetry assay to detect drug-induced QT interval prolongation. Prior to the estimation of the sensitivity and specificity, the fitness of the 4 heart rate correction/QT-RR statistical models using pre-dose QT-RR data were compared based on Akaike’s Information Criterion (AIC) (11). This is an important first step because the sensitivity and specificity of the assay are influenced by the over- or under-correction with an inappropriate QTc formula as well as the variation of the test system. The AIC was used in model selection. This is essentially the log likelihood value penalized for the number of parameters estimated. The sensitivity and specificity were estimated using the selected QT-RR statistical model from the variation (experiment errors and measurement errors).

Materials and Methods

Database of QT PRODACT

The QT Interval Prolongation: Project for Database Construction (QT PRODACT), which consists of pharmaceutical companies belonging to The Japan Pharmaceutical and Manufacturing Association (JPMA) and contract research laboratory organizations, has been proceeding to establish a database related to drug-induced QT interval prolongation. The 12 studies conducted using a 4 × 4 Latin’s square design, which is commonly used in safety pharmacology studies, were performed in 7 different laboratory facilities.

For these studies, 36 beagle dogs [Sex: male, Age: 8 – 21 months; B.W.: 7 – 14 kg] were used at least 10 days after the implantation surgery. A telemetry transmitter (TL11M2-D70-PCT; Data Science International, Inc., St. Paul, MN, USA/Primetech Corporation, Tokyo) was implanted subcutaneously into the flank of the body. Electrocardiograph (ECG) electrodes were fixed subcutaneously approximately with lead II configuration. Animal husbandry procedures during the study were in accordance with the Guideline for Animal Experimentation (Japanese Association for Laboratory Animal Science, 1987), and the experiments were performed according to the Guidelines for Animal Experimentation established in each testing facility. The test groups were divided into groups, positive drugs and negative drugs. The positive drugs have been reported to induce QT interval prolongation and associated ventricular arrhythmia such as TdP in humans and the drugs in this group were as follows: MK-499 (0.03, 0.1, and 0.3 mg/kg; Merck & Co., Inc., Rahway, NJ, USA), E-4031 (0.3, 1, and 3 mg/kg; Eisai Co., Ltd., Tokyo), bepridil hydrochloride (10, 30, and 100 mg/kg; Sankyo Co., Ltd., Tokyo), terfenadine (10, 30, and 100 mg/kg; Sigma Chemical Co., St. Louis, MO, USA), haloperidol (0.3, 1, and 3 mg/kg; Sigma-Aldrich Co., St. Louis, MO, USA), and cispapride (0.6, 2, and 6 mg/kg; Wako Pure Chemical Industries., Ltd., Osaka). The negative drugs, for which there have been few reports of QT interval prolongation or associated arrhythmias, were as follows: dl-propranolol hydrochloride (3, 10, and 30 mg/kg; Wako Pure Chemical Industries, Ltd.), verapamil hydrochloride (1, 5, and 15 mg/kg, Vasolan®; Eisai Co., Ltd.), captopril (3, 10, and 100 mg/kg, Sigma-Aldrich Co.), diphenhydramine hydrochloride (1, 3, and 10 mg/kg, Sigma Chemical Co.), amoxicillin (70, 200, and 500 mg/kg, Wako Pure Chemical Industries, Ltd.), and ciprofloxacin hydrochloride (5, 30, and 200 mg/kg; ICN Biomedicals Inc., Aurora, OH, USA). Each of the 4 dogs received orally a single low, middle, or high dose of the test drug or the vehicle alone. In each study, the 4 dogs were housed individually in metal pens or cages in a room, and ECG was recorded continuously from 24 h before to 24 h after the dosing using the DSI® telemetry system (Data Science International, Inc., St. Paul, MN, USA) and Notocord hem (NOTOCORD Systems SAS, Croissy, France). During the recording, access into the room was limited to feeding procedures, physical checks, blood collection, and dosing. The following 2 types of data sets were generated: 1) the QT and RR values of 5 beats at 6-min intervals for 24 h before dosing were measured by hand in all dogs on a personal computer and 2) the QT and RR values of more than 10 beats at each assay point after the dosing with the tested drugs were measured by hand in all experiments on personal computer in a similar manner.

Statistical model selection

To select an appropriate statistical method for the QT PRODACT telemetry data, the sequence of the QT-RR data points during the pre-dosing period were fitted with the following 4 linear models on a logarithmic scale

\[ \log(QT) = a + \beta \log(\text{RR}) \]

Model A considered the variability between the subjects of slopes and intercept; in this model, the slope \( \beta \) was estimated for each animal. Model B uses a common slope among the subjects and considered the variability between subjects of the intercept: in this model, the slope \( \beta \) was estimated for
each animal for each study. Models C and D considered the variability between subjects of the intercept; these models used a fixed slope of 0.5 (Bazett) or 0.33 (Fridericia).

Model A: $\log(QT)_i = (a_i + \alpha) + (\beta + \beta_i)\log(RR)_j + \epsilon_{ij}$.
Model B: $\log(QT)_i = (a_i + \alpha) + \beta \log(RR)_j + \epsilon_{ij}$.
Model C: $\log(QT)_i = (a_i + \alpha) + 0.5\log(RR)_j + \epsilon_{ij}$.
Model D: $\log(QT)_i = (a_i + \alpha) + 0.33\log(RR)_j + \epsilon_{ij}$.

where $\log(QT)_i$ is log-transposed QT on the $i$th assay point on the $j$th subject, $\alpha$ and $\beta$ are the population means, $\alpha_i$ and $\beta_i$ are the subject’s effects, $\epsilon_{ij}$ is random errors, $\log(RR)_j$ is log-transposed RR interval on the $j$th assay point on the $i$th subject. In all models, $\log(QT)_i$ is log-transposed QT interval on the $j$th assay point on the $i$th subject, $\epsilon_{ij}$ was assumed to be independently distributed as $N(0,\sigma^2_{ij})$.

The estimates of the parameters were derived from restricted log likelihood, and the Newton-Raphson algorithm was used for the calculations. The Newton-Raphson method is an iterative method for solving nonlinear equations, such as equations whose solution determines the point at which a function takes its maximum (12). The most appropriate model was selected with reference to AIC (smaller value is better). Moreover, residuals were visually assessed.

**Heart rate correction**

The heart rate correction was performed according to the following procedure (8, 9): First, using the QT and RR data of 5 beats at 6-min intervals for 24 h before dosing, the QT-RR relation was analyzed for each animal for each study. Models C and D considered the variability between subjects of the intercept; these models used a fixed slope of 0.5 (Bazett) or 0.33 (Fridericia).

Second, having the $\alpha$ and $\beta_i$ determined the point at which a function takes its maximum (12). The most appropriate model was selected with reference to AIC (smaller value is better). Moreover, residuals were visually assessed.

**Results**

**Statistical model selection**

The AIC values of 4 QT-RR statistical models are shown in Table 1. The models were ranked based on the AIC values: models C < D < B < A. The results indicate that Model A is the most appropriate model. Although the model B gave relatively good fit, its AIC value was higher than that of model A. In addition, the slope in model B exhibited the substantial inter-variability among studies and animals and the mean slope/intercept values were $0.22 \pm 0.05/3.95 \pm 0.32$ (Slope Max – Min: 0.31 – 0.14) in the 36 dogs used. Models C and D did not fit to the data. In addition, the estimated root squared error using model A in all dogs during the pre-dosing

<table>
<thead>
<tr>
<th>QT-RR model</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A</td>
<td>log(QT) = (a + \alpha) + (\beta + \beta_i)log(RR) + \epsilon_{ij}</td>
</tr>
<tr>
<td>Model B</td>
<td>log(QT) = (a + \alpha) + \beta log(RR) + \epsilon_{ij}</td>
</tr>
<tr>
<td>Model C</td>
<td>log(QT) = (a + \alpha) + 0.5log(RR) + \epsilon_{ij}</td>
</tr>
<tr>
<td>Model D</td>
<td>log(QT) = (a + \alpha) + 0.33log(RR) + \epsilon_{ij}</td>
</tr>
</tbody>
</table>

The sequence of the QT-RR data points during the pre-dose period were fitted to each equation, where log(QT) is log-transposed QT on the $j$th time point on the $i$th subject; $\alpha$ and $\beta_i$ are the population means; $\alpha_i$ and $\beta_i$ are the subject’s effects; $\epsilon_{ij}$ is random errors; log(QT) is log-transposed QT interval on the $j$th time point on the $i$th subject; log(RR) is log-transposed RR interval on the $j$th time point on the $i$th subject. $\epsilon_{ij}$ was assumed to be independently distributed as $N(0,\sigma^2_{ij})$. The most appropriate model was selected with reference to Aikake’s Information Criterion (AIC) (smaller AIC value is better).
The distribution of QT-RR residual difference of the individual correction (QTci) and that obtained by using Fridericia’s formula (b) are shown in Fig. 1. While the QTci was consistent at all RR intervals, the QTcf became higher at shorter RR intervals and became lower at the longer RR intervals. This indicates that the QTci worked appropriately and the QTcf over-corrected at a fast heart rate and under-corrected at a slow heart rate.

Sensitivity and specificity

The results of the ANOVA (analysis of variance) of the studies using positive drugs are summarized in Table 2. The mean experiment and measurement errors (logarithm) were calculated to be (0.0187)$^2$ and (0.0448)$^2$, respectively. The estimated error in each assay point for one test drug in this experiment was (0.1558)$^2$. When the cutoff value was 0.05, the sensitivity for a 10% change was 0.978 and the specificity for a 0% change was 0.996.

A representative (E-4031) of the time courses of the changes following the administration is shown in Fig. 2. The plasma drug concentrations of all drugs in the

<table>
<thead>
<tr>
<th>Positive drugs</th>
<th>MK-499</th>
<th>E-4031</th>
<th>Bepridil</th>
<th>Cisapride</th>
<th>Terfenadine</th>
<th>Haloperidol</th>
<th>Averaged variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental error ($\sigma^2$)</td>
<td>(0.0376)$^2$</td>
<td>(0.0457)$^2$</td>
<td>(0.0427)$^2$</td>
<td>(0.1026)$^2$</td>
<td>(0.0469)$^2$</td>
<td>(0.0858)$^2$</td>
<td>(0.0448)$^2$</td>
</tr>
<tr>
<td>Measurement error ($\sigma^2$)</td>
<td>(0.0379)$^2$</td>
<td>(0.0269)$^2$</td>
<td>(0.0405)$^2$</td>
<td>(0.0479)$^2$</td>
<td>(0.0404)$^2$</td>
<td>(0.0658)$^2$</td>
<td>(0.0448)$^2$</td>
</tr>
<tr>
<td>Assay points</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Pure exp. error ($\sigma^2$)</td>
<td>(0.0000)$^2$</td>
<td>(0.0151)$^2$</td>
<td>(0.0048)$^2$</td>
<td>(0.0371)$^2$</td>
<td>(0.0097)$^2$</td>
<td>(0.0195)$^2$</td>
<td>(0.0187)$^2$</td>
</tr>
<tr>
<td>Variation (D, T)</td>
<td>(0.0189)$^2$</td>
<td>(0.0154)$^2$</td>
<td>(0.0204)$^2$</td>
<td>(0.0303)$^2$</td>
<td>(0.0208)$^2$</td>
<td>(0.0343)$^2$</td>
<td>(0.0243)$^2$</td>
</tr>
<tr>
<td>Error (D, T)</td>
<td>(4 × 4 Latin’s square)</td>
<td>(0.1376)$^2$</td>
<td>(0.1242)$^2$</td>
<td>(0.1429)$^2$</td>
<td>(0.1740)$^2$</td>
<td>(0.1441)$^2$</td>
<td>(0.1852)$^2$</td>
</tr>
<tr>
<td>Cutoff value</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Sensitivity (10%)</td>
<td>0.995</td>
<td>0.999</td>
<td>0.992</td>
<td>0.946</td>
<td>0.991</td>
<td>0.923</td>
<td>0.978</td>
</tr>
</tbody>
</table>

Longitudinal data after each drug were analyzed. The averages of data for 6 to 8 points were used. Considering the structural formula shown below, each QT prolongation at the times was estimated using an analysis of split-plot design. The analysis was performed taking the data at each time as being independent. log(QTc)ijkl = $\mu$ + (Dog)i + (Period)j + (Drug)k + (Time)l + (Drug × Time)kl + $\varepsilon$ijkl + $\varepsilon$ijkl, where log(QTc)ijkl: log-transformed QTc; (Dog)i: effect of the ith beagle dog; (Period)j: effect of the jth period; (Drug)k: effect of the Kth drug; (Time)l: effect of the lth time; (Drug × Time)kl: the klth drug × time interaction effect; $\varepsilon$ijkl: random errors of experimentation; $\varepsilon$ijkl: random errors of measurement. It was assumed that $\varepsilon$ijkl and $\varepsilon$ijkl were independently distributed as $\mathcal{N}(0,\sigma^2_{ijkl})$ and $\mathcal{N}(0,\sigma^2_{ijkl})$, respectively.
positive drug group were determined in this database of QT PRODACT and reached the expected levels known to induce QT interval prolongation. Although the plasma drug concentrations of the negative drug group were not determined in this database of QT PRODACT, the selected dosage levels were at the clinical usage level and above and were expected not to induce QT interval prolongation.

Discussion

The present investigation demonstrated that when calculated by combining the twelve studies from variation using the model A and when the cutoff value was 5%, the sensitivity in 10% prolongation was 0.978 and the specificity in 0% was 0.996. These results suggest that the canine telemetry assay can provide sensitivity and specificity of approximately 5%.

The first investigation compared heart rate correction formulas for QT interval in order to use the optimal data to assess sensitivity and specificity of the telemetry assay. The appropriateness for the formula or model selection was based on the AIC values and the models were ranked as follows: C < D < B < A. The mean value of the slope β of each animal was 0.22; however, the slope β exhibited substantial inter-variability among studies and animals. The QT-RR relationship varies depending on each condition of measurement, for example, under the free-moving condition (7), under the restraint condition in a sling (13), and under the anesthetized condition (14). The QT-RR relation also varies depending on the method of data sampling and/or data analysis, for example, averaged QT-RR plots (15) versus beat-to-beat QT-RR plots (16) and QT-RR plots from a group of animals (9) versus QT-RR plots from individual animals (8). Therefore, the previously recommended QT-RR model would not necessarily provide the optimal fit for the telemetry data used in this investigation. Although models C (Bazett) and D (Fridericia) are the most frequently used heart rate correction formulas in dogs (6), neither provided an acceptable fit for the telemetry data in this investigation. In addition, when model D was applied to the telemetry data, the QT interval was over-corrected at shorter RR intervals and under-corrected at longer RR intervals (Fig. 1). Using estimates of QT-RR slopes for each subject and each study (models A and B), there was superior fit of the
data compared to the situation where the QT-RR slopes are fixed in the correction formulas (Bazett, model C, and Fridericia, model D). Of the two models with individual subject heart rate correction, model A was a better fit than model B. These results suggest that model A provides the highest sensitivity and specificity among the 4 heart-rate correction models based upon statistical fit and the observation that errors due to overor under-correction with model A are the smallest. Therefore, model A was selected for the estimation of the sensitivity and specificity in this investigation.

The QT interval varies with changes in the autonomic nervous tone and other factors even when the heart rate is constant (17, 18). In this investigation, the estimated root squared error using model A in all dogs during the pre-dosing period was 5.45%. This suggests that there is approximately 5% of variation in the QT interval at any heart rate (as an average). Therefore, even if an optimal heart rate correction formula was applied, the sensitivity and specificity of a test system would be limited by other influences on the QT interval and associated sources of variation. For example, the QT-RR relationship changes depending on the changes of the physical and emotional condition on the day of testing (19–21). Also, conscious dogs have a physiological respiratory sinus arrhythmia (22) so that changes in the QT interval lag behind the changes in heart rate or RR interval (QT-RR hysteresis) (23). Although the causes of QT interval variation at the same heart rate were not specifically investigated in these studies, measurement of QT and RR values are intermittent and short-term averaged data, and therefore, the variations in QT were considered due to these diurnal physical and/or emotional changes, QT-RR hysteresis, and/or arrhythmia, and so forth. Ideally, the QT interval should be assessed at the same autonomic nervous system tone. However, a QT correction method, which takes into account the autonomic tone, has not been practically established yet. This variation is a combination of the physiologically normal variation as well as the detection limit of drug-induced QT interval prolongation. The results in this investigation suggest that the cutoff value concerning the sensitivity and specificity would be approximately 5%.

On the other hand, this investigation has several limitations. First, the QT and RR data of the QT PRODACT were averaged over 5 beats at 5-min intervals for 24 h before dosing. Although models C (Bazett) and D (Fridericia) did not fit the telemetry data in this investigation, the results of models C and D may be limited to the type of telemetry data obtained in this investigation, and as mentioned above, the results of model selection are dependent on the QT-RR relationships. Second, the diurnal variation and the day-to-day variation of the QT-RR relationships were not defined in this investigation. Therefore, the error due to the difference between the QT-RR regression curve derived from the pre-dose data and those from post-dose data is unknown, although it is anticipated to be small. In addition, there was approximately 5% of variation in the QT-RR regression. Therefore, the data from our study would not be appropriate for comparing very small differences in sensitivity and specificity between regression models. As such, this investigation did give a detailed comparison among different regression models, and we employed only simple linear regression \[\log(QT) = a + b \log(RR)\] and focused on the comparison between the generic formulas (e.g., Bazett or Fridericia) and individually optimized formulas.

The optimized QT correction for each subject is beginning to be recommended as the most sensitive and appropriate method for QT/QTc assessment in humans as well as dogs (10, 21, 24). This correction method is practical since linear regression on a logarithmic scale \[\log(QT) = a + b \log(RR)\] can be easily converted to the following heart rate QTc formula: \[\log(QTci) = \log(QT) - b(\log(RR) - \log(RR_m))\], where \(RR_m\) = reference heart rate, slope \(b\) = the value for each animal (model A), for each study (model B), 0.5 (model C, Bazett), and 0.33 (model D, Fridericia). In addition, this correction method has the advantage that one can determine if the obtained QTc values are over- or under-corrected by utilizing the QT-RR baseline values and formula. Therefore, the telemetry assay using the optimized QT correction for each subject is preferable for the QT/QTc assay. Hammond et al. conducted a survey in the pharmaceutical industry and reported in 2001 that based on preliminary estimates of sample sizes and using a standard formula for power calculations, group sizes of \(n = 4\) may be inadequate to detect changes in QT interval prolongation of 10% in conscious dogs, but may detect changes of 20% (for control QT = 200 ms; control SD = 19 ms; \(P < 0.05\), 2-tailed, unpaired test; 80% power) (6). Guth et al. reported in 2004 that the analysis of the telemetric measurement of ECG using a Latin square cross-over design, with a sample size of \(N = 4\), is capable of detecting, at 80% statistical power, a QT interval prolongation of 10%, or approximately 20 ms (25). In the other activities of the QT PRODACT, Toyoshima et al. have reported the results for the canine telemetry assay (26). Although the data set of the QT PRODACT used by Toyoshima et al. was the same as that used in our present investigation, their analytical method was limited to Fridericia’s formula that is commonly used in safety pharmacology studies. Compared to these conventional test systems and/or another method, the
sensitivity and specificity of the canine telemetry assay can be largely improved by using the individual subject heart rate correction (model A) presented here, and this improvement of the sensitivity and specificity would lead to the minimization of over- or under-estimation.

In conclusion, it was judged that a 5% cutoff value for changes in heart rate corrected QT interval using the canine telemetry assay is practical, and the sensitivity and specificity of the telemetry assay are very high when using the analytical method presented here. Based upon this information, the canine telemetry assay using the individual subject heart rate correction model is recommended as a sensitive test system for the in vivo assessment of risk for QT interval prolongation.

Acknowledgments

We thank all of the QT PRODACT members for their advice and collaboration. We sincerely thank Eisai Co., Ltd. and Merck & Co., Inc. for providing us with E-4031 and MK-499, respectively. Some of the telemetry transmitters were kindly provided by Primetech Corporation (Tokyo).

References


2 ICH Harmonised Tripartite Guideline, Safety Pharmacology Studies for Human Pharmaceuticals S7A. Recommended for adoption at step 4 of the ICH process on 8 November 2000 by the ICH Steering Committee. ICH; http://www.ich.org/.


