OXFORD
SOT | Society of Toxicology
www.toxsci.oxfordjournals.org

TOXICOLOGICAL SCIENCES, 2017, 1–8

doi: 10.1093/toxsci/kfx125
Advance Access Publication Date: June 16, 2017
Research Article

An Analysis of the Relationship Between Preclinical and Clinical QT Interval-Related Data

Christopher E. Pollard,*,[1] Matthew Skinner,* Stanley E. Lazic,† Helen M. Prior,* Kelly M. Conlon,* Jean-Pierre Valentin,* and Corina Dota‡

*Safety and ADME Translational Sciences, Drug Safety and Metabolism; ‡Quantitative Biology, Discovery Sciences, Innovative Medicines and Early Development, AstraZeneca R&D, Cambridge CB4 0WG, UK; and †Cardiovascular Centre of Excellence, Global Medicines Development, AstraZeneca R&D, Gothenburg 43183, Sweden

[1]To whom correspondence should be addressed at AstraZeneca, Cambridge Science Park, Milton Road, Cambridge CB4 0WG, UK. E-mail: chris.pollard4@btinternet.com.

ABSTRACT

There has been significant focus on drug-induced QT interval prolongation caused by block of the human ether-a-go-go-related gene (hERG)-encoded potassium channel. Regulatory guidance has been implemented to assess QT interval prolongation risk: preclinical guidance requires a candidate drug's potency as a hERG channel blocker to be defined and also its effect on QT interval in a non-rodent species; clinical guidance requires a "Thorough QT Study" during development, although some QT prolonging compounds are identified earlier via a Phase I study. Clinical, heart rate-corrected QT interval (QTc) data on 24 compounds (13 positives; 11 negatives) were compared with their effect on dog QTc and the concentration of compound causing 50% inhibition (IC_{50}) of hERG current. Concordance was assessed by calculating sensitivity and specificity across a range of decision thresholds, thus yielding receiver operating characteristic curves of sensitivity versus (1-specificity). The area under the curve of ROC curves (for which 0.5 and 1 indicate chance and perfect concordance, respectively) was used to summarize concordance. Three aspects of preclinical data were compared with the clinical outcome (receiver operating characteristic area under the curve values shown in brackets): absolute hERG IC_{50} (0.78); safety margin between hERG IC_{50} and clinical peak free plasma exposure (0.80); safety margin between QTc effects in dogs and clinical peak free plasma exposure (0.81). Positive and negative predictive values of absolute hERG IC_{50} indicated that from an early drug discovery perspective, low potency compounds can be progressed on the basis of a low risk of causing a QTc increase.

Key words: hERG; QT interval prolongation; preclinical to clinical translation.

There has been considerable interest in drugs that block the human ether-a-go-go-related gene (hERG)-encoded potassium channel as side-effect, thereby delaying ventricular action potential repolarization and prolonging the QT interval on the surface electrocardiogram (Gintant *et al.*, 2016; Sanguinetti and Tristani-Firouzi, 2006). Focus on drug-induced QT interval prolongation has been intense because it is associated with a potentially fatal cardiac arrhythmia (Torsades de Pointes [TdP]) and has been a significant contributor to attrition in the pharmaceutical industry (Shah, 2006).

Under the auspices of the International Conference on Harmonization (ICH), the regulatory response to this problem was the publication of three guidance documents for those aiming to discover new medicines: two dealing with the testing requirements before starting clinical trials (ICH S7A; Anonymous, 2001) and ICH S7B (Anonymous, 2005a) and another focused on clinical testing prior to marketing approval (ICH E14; Anonymous, 2005b).

The corner-stones of the preclinical QT-related guidance are an in vitro measurement of a compound's activity as a hERG blocker and an in vivo assessment of its effect on QT interval duration in non-rodents. For the hERG test, ICH S7B (Anonymous, 2005a) stipulates that: "Ascending concentrations should be tested until a concentration-response curve has been

© The Author 2017. Published by Oxford University Press on behalf of the Society of Toxicology.
All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

characterized or physicochemical effects become concentration-limiting". The guidance around the *in vivo* preclinical QT assessment states that the dose range used "whenever feasible, should include and exceed the anticipated human exposure". This necessitates testing at more than one dose level and taking blood samples after each dose to establish plasma exposure levels. This includes, as a minimum, a sample timed to coincide approximately with the postdose time point at which the maximum plasma concentration of the drug occurs (C_{max}). ICH S7B also provides advice on appropriate heart rate correction factors to convert QT intervals into heart rate-corrected (QTc) data. Thus the preclinical data on a candidate drug usually consist of a hERG concentration-effect curve and an in vivo assessment of QTc relative to plasma C_{max}, often expressed as the concentration not bound to plasma proteins (C_{max} free).

The definitive clinical evaluation of drug-induced QT prolongation risk has been a "Thorough QT/QTc Study" (TQTS) (ICH E14 (Anonymous, 2005b) which often involves measuring QT/QTc at 2 dose levels of the compound of interest: the therapeutic dose and a "substantial multiple of the anticipated maximum therapeutic exposure". Since ICH E14 (Anonymous, 2005b) recommends that QT/QTc data are collected at around C_{max}, these studies require measurement of drug plasma concentrations to establish C_{max}, often expressed as C_{max} free. A typical TQTS therefore yields an assessment of QTc versus plasma C_{max} free at two doses. There are variants of this TQTS design: eg, some TQT studies use a single dose that from previous studies has been established as the maximum tolerated dose, the logic being that if it is "negative" at this dose (ie, did not increase QTc), it will be negative at a lower, therapeutic dose. ICH E14 defines the threshold for regulatory concern (ie, a "positive") as a prolongation of "around 5 ms as evidenced by an upper bound of the 95% confidence interval (CI) around the mean effect on QTc of 10 ms". The statistical power required to detect such a small change dictates the need for careful collection and analysis of QTc data. Hence the output of TQT studies represents an opportunity to compare high resolution, quantitative clinical data with that collected preclinically.

In contrast to the TQTS, which is often conducted late in development once the therapeutic dose has been established, the Phase I clinical trials on a development compound are usually a Single Ascending Dose (SAD) and a Multiple Ascending Dose (MAD) study. These assess the pharmacokinetic and safety properties of a test compound, including QTc data, but such studies have not traditionally been statistically powered to detect the very small QTc prolongations used as the threshold for regulatory concern defined in ICH E14 (Anonymous, 2005b). Thus only relatively large QTc increases have confidently been attributed to the effect of the test compound. In other words, the apparent absence of a QTc effect in a Phase I study is not deemed a definitive demonstration that the compound has no effect: only the TQTS is powered sufficiently to define a negative. Hence additional preclinical to clinical translational data can be obtained by studying compounds that have prolonged QTc in Phase I studies, but not those which had no apparent effect.

In this assessment we have extracted clinical QTc data for 24 compounds from TQT studies, or Phase I studies where there was a clear QTc increase, and compared them with hERG and dog QTc data collected preclinically.

MATERIALS AND METHODS

Origin of compounds. In total 23 of the 24 compounds in this investigation were from AstraZeneca's internal drug discovery and development activities or were in-licensed by AstraZeneca from another company; the exception was compound 10 (moxifloxacin). Except for the hERG data on moxifloxacin, all data were from studies conducted at AstraZeneca, at the third-party from which the compound originated or at contract research organizations.

The core pharmacodynamic and pharmacokinetic data for this work were extracted from hERG, dog safety pharmacology and clinical study reports; the conclusions of the reports were not re-interpreted. Moxifloxacin hERG potency data were taken from the paper by Gintant (2011). In order to convert from total to free drug plasma concentrations, the percentage of each compound not bound to human or dog plasma proteins was extracted from the relevant reports. The molecular weight of each compound was obtained to enable conversion from concentrations expressed in ng/ml to those expressed in μM.

hERG studies. The details of each study are shown in Supplementary Table 1. Studies on 23 of the compounds used manual whole-cell patch clamping, whilst the remaining compound was assessed using automated whole-cell patch clamping. Although voltage protocols were not the same for all compounds, they all involved an initial voltage step to $\geq$10 mV (range +10 to +40 mV) followed by a step or a ramp to ≤ -30 mV (range −30 to −80 mV) to evoke an outward tail current. Each compound was assessed for its effect on the amplitude of this tail current. Physiological intracellular and extracellular ionic concentrations were used and concentration-effect curves were based on measured concentrations of compounds where available.

In order to perform the relevant analyses for every compound, an IC_{50} value was required. For inactive or low potency compounds where an IC_{50} was not defined, the following process was used to provide an estimate. A concentration-effect curve was simulated that had a user-defined IC_{50} value, a Hill slope of 1 and a minimum and maximum of 0% and 100% inhibition, respectively. For compounds that were active but did not achieve 50% inhibition, eg, 35% inhibition at the highest test concentration of X μM, the user iteratively inputted IC_{50} values until the 35% inhibition point occurred at X μM and this IC_{50} was then used as an estimate. For compounds that were inactive at the highest test concentration, the same process was used to estimate a "worst-case" IC_{50} by assuming that the maximal concentration tested elicited 5% inhibition of hERG current; this is the same approach as that adopted by Gintant (2011).

Dog safety pharmacology studies. All studies were conducted according to the relevant national guidelines on animal welfare. Six compounds were assessed in anesthetized beagle dog studies (using 4 different anesthetic regimes) and the remaining 18 in conscious telemetered beagle dog studies. The key details of each study are shown in Supplementary Table 2. Studies described the effect of test compound on a variety of cardiovascular parameters, but for this work note was made of whether the report stated a compound-related effect on QTc at a given dose. All studies reported the arithmetic mean of the total plasma exposures at a time point judged to be close to C_{max}. These values were converted to free (ie, unbound) C_{max} values using dog plasma protein binding data. Hence, for each dose in a study, the data extracted were the mean free C_{max} at that dose and whether there was or wasn't an effect on QTc.

Clinical studies. The key details of each study are shown in Supplementary Table 3. All studies used digital ECG collection.

Data for 16 of the compounds came from TQT studies. All 16 TQT studies had a "positive" moxifloxacin arm and in 9 of these studies the C_{max} for moxifloxacin was defined. In order to include moxifloxacin in the analysis, C_{max} data were extracted from all 9 studies and data from the study with the median C_{max} value were used to exemplify moxifloxacin clinical data.

Of the 8 remaining compounds, data for 6 came from SAD studies, 1 from day one of a MAD study and 1 from a combination of SAD and MAD studies. The "positive" QTc outcome for the latter was based on concentration-response modeling of QTc data.

Studies described the effect of test compound on a variety of parameters, but for this work note was made of whether the report stated a compound-related effect on QTc at a given dose. All studies reported the geometric mean of total plasma exposures at a time point judged to be close to C_{max}. These were then converted to a free exposure using human plasma protein binding data. Hence, for each dose in a study, the data extracted were the mean free C_{max} at that dose and whether there was or wasn't an effect on QTc.

Clinical plasma exposure reference point used to calculate exposure multiples relative to preclinical data. A key aspect of this work was to assess preclinical data at exposure multiples relative to C_{max} free values in clinical studies. It is therefore important to clarify what clinical exposure values were used. For compounds that prolonged QTc in clinical studies, the clinical exposure reference point was the mean C_{max} free at the lowest dose where an increase in QTc was seen. For compounds that did not prolong QTc in clinical studies, the mean C_{max} free at the highest dose tested was used.

Data classification. For clinical studies, a "negative" was a compound that did not prolong QTc in a TQTS, based on the definition of a "negative" in ICH E14: "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". A "positive" was a compound that prolonged QTc in a TQTS based on the definition of a "positive" in ICH E14 (ie, if the 95% one-sided CI for the largest time-matched mean effect of the drug on the QTc interval exceeded 10 ms). "Positive" compounds were also drawn from those tested in other (ie, non TQT) clinical studies where the study report concluded that there was a compound-induced increase in QTc.

Concordance with clinical outcome was assessed for three aspects of the preclinical data, using the same approach as described by Ewart *et al*. (2014).

1. The absolute hERG IC_{50} value. This was assessed by setting thresholds for IC_{50} below which hERG data were deemed "positive", starting with a threshold of 0.3 μM and increasing the value to 1, 3, 10, 30, 100, 300, and 1000 μM.
2. The exposure multiple between the hERG IC_{50} and clinical C_{max} free reference point. This was done by setting an exposure multiple that constituted a "positive" if the hERG IC_{50} fell within it, starting with an exposure multiple of 1-fold and increasing the value to 3, 10, 30, 100, 300, and 1000-fold.
3. The exposure multiple between QTc increase in dogs and the clinical C_{max} free reference point. This was done by setting an exposure multiple that constituted a "positive" if the QTc increase in dogs fell within it, starting with an exposure multiple of 1-fold and increasing the value to 3, 10, 30, and 100-fold. It is important to note that the exposure multiples covered those both above and below the clinical C_{max} free reference point. For example, the 10-fold exposure multiple ranged from 0.1- to 10-fold the clinical C_{max} free reference point.

Statistical quantification of concordance. The foundation of the analysis was standard diagnostic tests for sensitivity and specificity using binary classification to define the effect of each compound (Altman and Bland, 1994a). In the context of this work, sensitivity represents the proportion of compounds that prolonged QTc in clinical studies that were correctly identified by a preclinical test. Specificity represents the proportion of compounds that did not prolong QTc in clinical studies that were correctly identified by a preclinical test. They were calculated as follows:

$$\text{Sensitivity} = \text{TP}/[\text{TP} + \text{FN}]$$

$$\text{Specificity} = \text{TN}/[\text{TN} + \text{FP}]$$

Where TP, True Positive; TN, True Negative; FP, False Positive; FN, False Negative.

The 95% CIs around sensitivity and specificity values were calculated in R (version 3.3.1) using a binomial test.

Sensitivity and specificity values for each analysis were calculated across the range of decision thresholds described above, thus enabling receiver operating characteristic (ROC) curves to be produced by plotting Sensitivity against (1-Specifcity) (Altman and Bland, 1994c). The area under the curve (AUC) of ROC curves was calculated using the trapezium rule. The optimal balance between sensitivity and specificity was defined using Youden's J statistic (Sensitivity + Specificity −1).

Positive predictive value (PPV) and negative predictive value (NPV) were also calculated, which represent the probability that a model predicts the outcome, in this case QTc effect in clinical studies. The PPV of a model is the proportion of compounds with a TP result amongst all the compounds with positive results. Conversely, NPV of a model is the proportion of compounds with a TN result amongst all the compounds with negative results. PPV and NPV are influenced by outcome prevalence, which in this analysis is the prevalence of compounds that prolong QTc in clinical studies as a proportion of the total number of drugs tested in clinical studies. Although the prevalence (or pretest probability) is usually unknown, for this work we estimated prevalence using the data reported by Park *et al*. (2013), who stated that 46 out of 205 TQT studies submitted to the U.S. Food and Drug Administration were "positive", thus giving a prevalence estimate of 0.22. This enabled the calculation of values for PPV and NPV (Altman and Bland, 1994b), as follows:

$$\text{PPV} = (\text{sensitivity x prevalence})/[\text{sensitivity x prevalence} + (1 - \text{specificity}) \text{ x } (1 - \text{prevalence})]$$

$$\text{NPV} = \text{specificity x}(1 - \text{prevalence})/[\text{specificity x}(1 - \text{prevalence}) + (1 - \text{sensitivity}) \text{ x prevalence})]$$

The 95% CIs around PPV and NPV were calculated based on methods described in Mercaldo *et al*. (2007).

RESULTS

Compound Dataset

Of the 24 compounds, 13 prolonged QTc in clinical studies and 11 did not. The compounds were all of low molecular weight (ie, not biologics) and their primary targets were G-protein-coupled receptors (10), enzymes (9), ion channels (3), transporters (1) or nuclear hormone receptors (1). They were aiming to treat indications across a range of disease areas: cardiovascular (6), central nervous system (5), gastro-intestinal (3), infection (3), oncology, (3), respiratory (3), and inflammation (1). Since it is relevant to the analysis and discussion of the data, note that compounds 9 and 12 are long-acting β_2-adrenoceptor agonists. The pharmacology of these two compounds is highlighted since irrespective of any hERG-mediated risk of a clinical QTc increase, previous data for such compounds highlight the potential for a nonhERG-mediated clinical QTc effect (Bremner *et al.*, 1993; Lecaillon *et al.*, 1999). Hence in an analysis of concordance between hERG data and clinical QTc effects, such compounds are likely to appear as false negatives, when in fact their β-adrenergic receptor pharmacology marks them out as having the potential to cause a QTc increase in man.

Relationship Between Clinical QTc Effect and hERG Data

Absolute hERG Potency

Figure 1A shows the absolute hERG potency for the 24 compounds. The data illustrate that 7 out of the 13 compounds that increased QTc in clinical studies had IC_{50}s <10 μM. Conversely, 10 out of 11 of the compounds that did not prolong QTc in clinical studies had an IC_{50} >10 μM.

ROC curves both with and without the β_2-adrenoceptor agonists (compounds 9 and 12) were constructed (Figure 1B). A numerical summary of the data is shown in Table 1 along with the PPV and NPV data. The AUC values for the ROC curves indicate a degree of concordance between absolute hERG potency and clinical QTc outcome (0.78 for the total dataset and 0.81 for that minus the two β_2-adrenoceptor agonists) given that a value of 0.5 corresponds to random chance and a value of 1 would be perfect concordance. For both datasets, the optimal balance between sensitivity and specificity occurred at a hERG IC_{50} threshold of 10 μM. The relative values of PPV and NPV changed with hERG IC_{50} threshold, as would be expected, but at 10 μM, eg, for the whole dataset, PPV = 0.63 whilst NPV = 0.87. The PPV and NPV value for the dataset minus the 2 β_2-adrenoceptor agonists was 0.66 and 0.90, respectively.

Safety Margin Between Clinical Exposure and hERG Potency

Based on the well-established link between hERG block (assuming selective inhibition of hERG) and increases in QTc (Sanguinetti and Tristani-Firouzi, 2006), it is logical to expect that the smaller the safety margin between clinical C_{max} free and hERG IC_{50}, the greater the chance of a QTc increase in the clinical study. In Figure 2A, we explored that expectation by plotting the ratio of the hERG IC_{50} to clinical C_{max} free using the clinical exposure reference point described in the "Materials and Methods" section. In total 9 out of 13 compounds that increased QTc in clinical studies had safety margins <30-fold whilst 10 out of the 11 compounds that did not prolong QTc had safety margins >30-fold.

ROC curves both with and without the β_2-adrenoceptor agonists were constructed for these data (Figure 2B). A numerical summary of the data is shown in Table 2 along with the PPV and NPV data. The AUC values for the ROC curves indicate good concordance between hERG safety margin and clinical QTc outcome (0.80 for the total dataset and 0.92 for that minus the two β_2-adrenoceptor agonists). For both datasets, the optimal balance between sensitivity and specificity occurred at a hERG IC_{50}/clinical C_{max} free ratio threshold of 30-fold.

If the same analyses were done using clinical C_{max} total (ie, plasma C_{max} values that were not corrected for plasma protein binding), the ROC curve AUC values were reduced by 0.12 (to 0.68 for the total dataset and 0.80 for that minus the two β_2-adrenoceptor agonists).

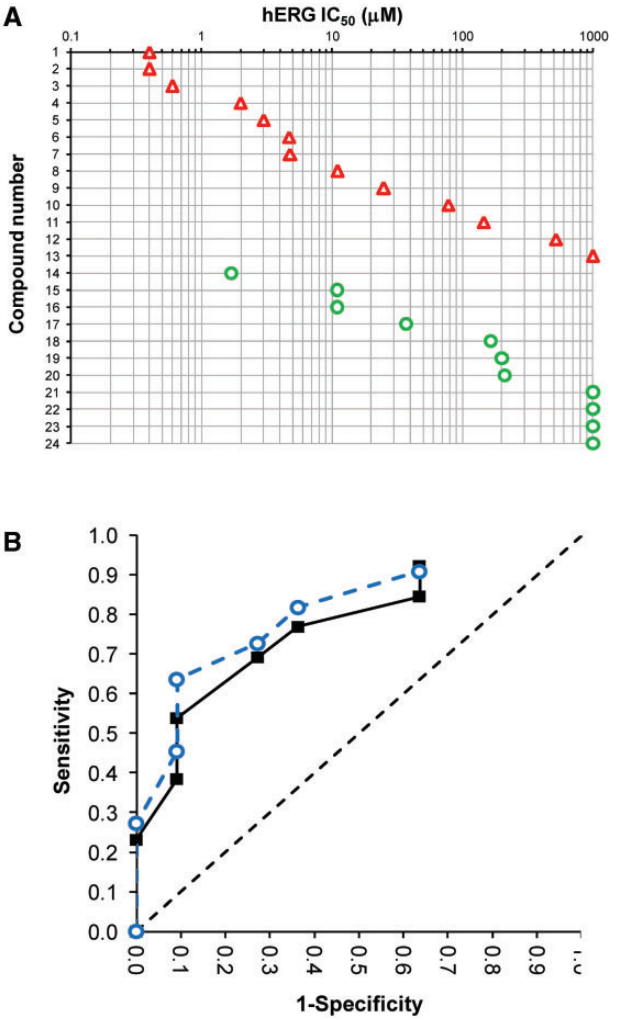

Figure 1. Relationship between effect on clinical QTc and potency as a blocker of the hERG-encoded potassium channel. A, Graphical summary of the data. Triangles, clinical QTc increase; circles, no clinical QTc increase. Potency values for each compound are plotted as hERG IC_{50}. Compounds with IC_{50}s > 1000 μM are plotted at a value of 1000 μM. B, ROC curves for the data in (A). The complete dataset (24 compounds) is shown in filled squares and continuous lines; the dataset minus the β_2-adrenoceptor agonists (22 compounds) is shown in open circles and dashed lines.

Relationship Between Clinical Exposure and QTc Effect in Dog Studies

Figure 3 summarizes the relationship between dog QTc data and the QTc outcome in clinical studies. For each compound, the ratio of the C_{max} free in dogs at each dose tested was plotted as a ratio of the clinical C_{max} free reference point for that compound, as described in the "Materials and Methods" section. As stated in the Introduction, to comply with ICH S7A guidance

Table 1. Concordance Data for Complete Dataset (24 Compounds) Between Clinical, Heart Rate-Corrected QT Interval (QTc) Study Outcome and Drug Concentration Causing 50% Inhibition of hERG-Encoded Potassium Channel

Parameter/IC_{50} Threshold (μM)	<0.3	<1	<3	<10	<30	<100	<300	<1000
Sensitivity	0.00	0.23	0.38	0.54	0.69	0.77	0.85	0.92
Lower 95% CI	0.00	0.06	0.14	0.25	0.39	0.46	0.55	0.64
Upper 95% CI	0.25	0.54	0.68	0.81	0.91	0.95	0.98	1.00
Specificity	1.00	1.00	0.91	0.91	0.73	0.64	0.36	0.36
Lower 95% CI	0.72	0.72	0.59	0.59	0.39	0.31	0.11	0.11
Upper 95% CI	1.00	1.00	1.00	1.00	0.94	0.89	0.69	0.69
PPV	—	1.00	0.54	0.63	0.42	0.37	0.27	0.29
Lower 95% CI	—	0.29	0.14	0.19	0.20	0.21	0.18	0.20
Upper 95% CI		1.00	0.9	0.92	0.67	0.58	0.38	0.40
NPV	0.78	0.82	0.84	0.87	0.89	0.91	0.89	0.94
Lower 95% CI	0.78	0.77	0.77	0.79	0.77	0.77	0.65	0.69
Upper 95% CI	0.78	0.86	0.89	0.93	0.95	0.97	0.97	0.99

(Anonymous, 2001), dog cardiovascular studies are usually designed with the aim of the lowest dose achieving a plasma exposure similar to that anticipated in clinical studies and higher doses giving plasma exposures that exceed those anticipated in clinical studies. However, at the time the doses for dog studies are decided, the plasma exposure anticipated in clinical studies is a prediction based on a variety of data types and estimates. Also, as described in the "Materials and Methods" section, the reference clinical C_{max} free values used in Figure 3 are defined differently depending on whether or not a compound prolonged QTc in clinical studies. In addition, since some of the dog studies predate the publication of ICH S7A, its ethos was not necessarily followed. For these reasons, when the exposures in dog studies are plotted relative to the reference clinical C_{max} free values, there is often a lack of data around the reference clinical C_{max} free value. Despite these gaps in the dataset, as shown in Figure 3A, 8 of the 13 compounds that prolonged QTc in clinical studies also prolonged QTc in dogs within 10-fold the clinical C_{max} free. In contrast, 10 out of 11 compounds that did not increase QTc in clinical studies did not prolong QTc in dogs.

Although fewer data points are available for this data set, ROC curves were generated both with and without the β_2-adrenoceptor agonists (Figure 3B). A numerical summary of the data is shown in Table 3 along with the PPV and NPV data. The AUC values for the ROC curves indicate good concordance between dog QTc data and clinical QTc outcome (0.81 for the total dataset and 0.91 for that minus the 2 β_2-adrenoceptor agonists). For both datasets, the optimal balance between sensitivity and specificity occurred at a dog C_{max} free/clinical C_{max} free ratio threshold of 10-fold.

DISCUSSION

This sensitivity/specificity-based analysis confirms other reports suggesting good concordance between QTc outcome in clinical studies and data from hERG and dog studies (Ewart *et al.*, 2014; Gintant, 2011; Vargas *et al.*, 2015; Wallis, 2010). The concordance data improve if the 2 long-acting β_2-adrenoceptor agonists are removed from the analysis on the basis that a clinical QTc increase is not unexpected, even in the absence of any hERG-mediated effect.

However, this sensitivity/specificity-based analysis essentially describes how the clinical outcome predicts the preclinical

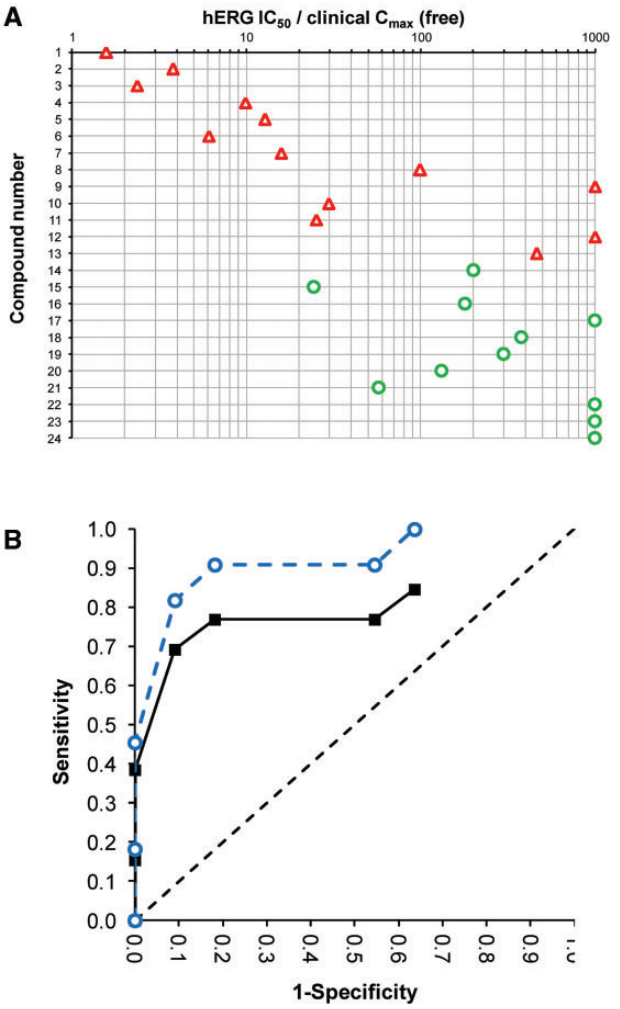

Figure 2. Relationship between effect on clinical, heart rate-corrected QT interval (QTc) and safety margin between clinical plasma exposure and potency as a blocker of the hERG-encoded potassium channel. A, Graphical summary of the data. Triangles, clinical QTc increase; circles, no clinical QTc increase. The safety margin for each compound is plotted as the ratio of: the hERG IC_{50} and the peak, free plasma concentration (C_{max} free) in the clinical study for that compound. Compounds with hERG IC_{50} to clinical C_{max} free ratios > 1000 are plotted at a value of 1000. B, ROC curves for the data in (A). The complete dataset (24 compounds) is shown in filled squares and continuous lines; the dataset minus the β_2-adrenoceptor agonists (22 compounds) is shown in open circles and dashed lines.

result. From the perspective of an early stage discovery project yet to test compounds in clinical studies, the critical question is: based on the potency of the compound as a hERG blocker, what is the likely outcome in a clinical QTc study? In this respect, the PPV and NPV data for absolute hERG IC_{50} suggest that a pragmatic selection of compounds to progress can be made on the basis of hERG potency alone, providing other aspects of the discovery project are used to put the hERG data into context.

Comparison With Other Analyses

Several pieces of work have retrospectively analyzed concordance between clinical and preclinical QT-related data, but the most relevant comparisons are with those reports for which the human information is QTc data, as opposed to an estimate of TdP incidence. Although the hERG-related concordance data in this piece of work improve if the 2 long-acting β_2-adrenoceptor

Table 2. Concordance Data for Complete Dataset (24 Compounds) Between Clinical, Heart Rate-Corrected QT Interval (QTc) Study Outcome and Safety Margin to Drug Concentration Causing 50% Block (IC_{50}) of hERG-Encoded Potassium Channel

Parameter/Exposure Multiple	1	<3	<10	<30	<100	<300	<1000
Sensitivity	0.00	0.15	0.38	0.69	0.77	0.77	0.85
Lower 95% CI	0.00	0.02	0.14	0.39	0.46	0.46	0.55
Upper 95% CI	0.25	0.45	0.68	0.91	0.95	0.95	0.98
Specificity	1.00	1.00	1.00	0.91	0.82	0.45	0.36
Lower 95% CI	0.72	0.72	0.72	0.59	0.48	0.17	0.11
Upper 95% CI	1.00	1.00	1.00	1.00	0.98	0.77	0.69
PPV	NA	1.00	1.00	0.68	0.54	0.28	0.27
Lower 95% CI	—	0.16	0.48	0.24	0.25	0.18	0.18
Upper 95% CI		1.00	1.00	0.94	0.81	0.42	0.38
NPV	0.78	0.81	0.85	0.91	0.93	0.87	0.89
Lower 95% CI	0.78	0.77	0.79	0.82	0.82	0.68	0.65
Upper 95% CI	0.78	0.84	0.90	0.96	0.97	0.96	0.97

agonists are excluded from the dataset, in order to compare with other reports, the concordance values for the total dataset are used.

There are no comparable published data relating absolute hERG IC_{50} to QTc outcome in clinical studies. With respect to analyses of hERG safety margins and clinical QTc data, the dataset of Wallis (2010) is most similar to that used here. Its clinical dataset was drawn from 19 TQT studies or TQT-like studies from within one company (11 positives; 8 negatives). When assessed at 2-, 10-, and 30-fold clinical exposure multiples, the hERG IC_{10} had a sensitivity of 0.82, 0.90, and 0.90, and a specificity of 0.75, 0.30, and 0.14, respectively. Gintant (2011) conducted an analysis of 39 compounds from different sources for which the TQTS outcome was reported (14 positives; 25 negatives). The optimal balance between sensitivity and specificity was found to be at a 45-fold safety margin from clinical exposure to hERG IC_{50}, at which point sensitivity was 0.64 and specificity 0.88. This compares with our data for all 24 compounds which, at the optimal safety margin of 30-fold, had a sensitivity of 0.69 and a specificity of 0.91.

In terms of previous analyses of dog QTc to clinical QTc data, Wallis (2010) quoted sensitivity and specificity values at 2-, 10-, and 30-fold clinical exposure multiples: sensitivity was 0.83 at all multiples, whilst specificity was 0.86 and 0.33 at 2- and 10-fold, respectively (no value was calculated at 30-fold). Ewart *et al.* (2014) used data from up to 95 compounds for which there were both Phase I clinical QTc data and dog telemetry study QTc data. The resulting analysis yielded sensitivity values of 0.17, 0.88, and 1.00 at 3-, 10-, and 30-fold clinical exposure multiples, respectively, and specificity values of 0.92, 0.76, and 0.58 at these multiples of clinical exposure, respectively. Finally, Vargas *et al.* (2015) used a text mining approach to assess the concordance between animal (rabbit, guinea-pig, dog, and non-human primate) and human studies and reported a sensitivity of 0.91 and specificity of 0.88, although, as the authors acknowledge, this approach did not factor-in plasma exposure values in animal or clinical studies. At a 10-fold clinical exposure multiple in dogs, the data for 16 compounds in this report were a sensitivity of 0.73 and a specificity of 0.80.

With respect to PPV and NPV data, only Ewart *et al.* (2014) show data that can be compared with that presented here. For example, using a prevalence of 0.27 to represent cardiovascular safety-related attrition as a whole, at a 10-fold exposure multiple, dog telemetry data are reported by Ewart *et al.* (2014) to have a PPV of 0.57 and a NPV of 0.94. This compares with our values of 0.51 and 0.91 for PPV and NPV, respectively.

Taken together, the concordance data for the 24 compounds reported here is consistent with those reported by others.

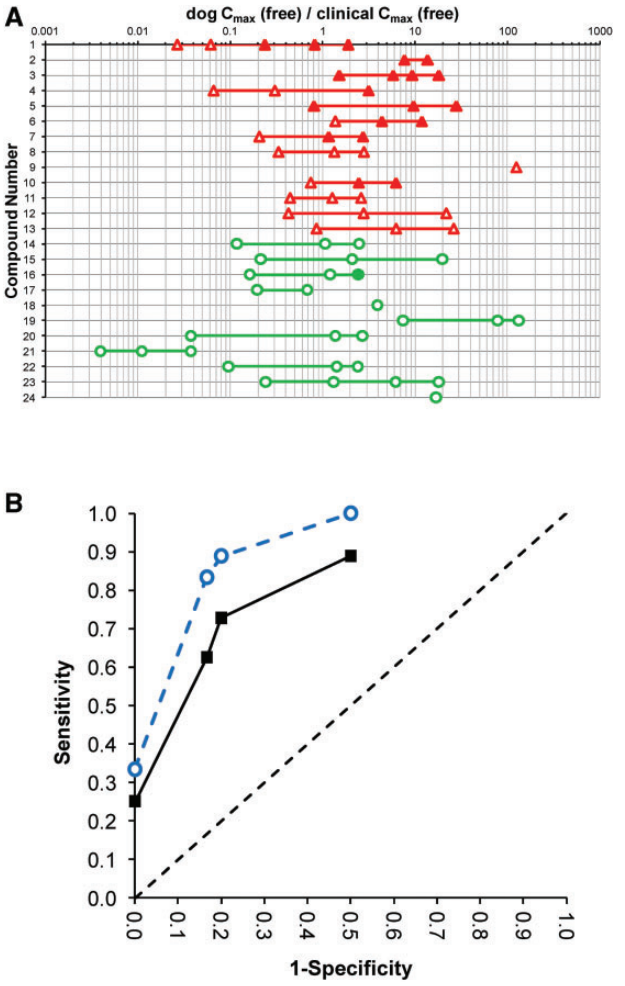

Figure 3. Relationship between effect on clinical, heart rate-corrected QT interval (QTc) and safety margin between clinical plasma exposure and QTc effects in dog studies. A, Graphical summary of the data. Triangles, clinical QTc increase; circles, no clinical QTc increase. Filled symbols indicate that there was a QTc increase in dogs. For each compound, the data are plotted as the ratio between peak, free plasma concentration (C_{max} free) in dogs at each dose used, and the C_{max} free in the clinical study for that compound. B, ROC curves for the data in (A). The complete dataset is shown in filled squares and continuous lines; the dataset minus the β_2-adrenoceptor agonists is shown in open circles and dashed lines. The number of compounds per data point on the ROC curves varies and are summarized in Table 3.

Use of Free Plasma Exposures to Calculate hERG Safety Margins

Comparable reports on hERG safety margins used free rather than total clinical C_{max} values (Gintant, 2011; Wallis, 2010), with Gintant (2011) stating that, "the relationship between QTc prolongations was more apparent with hERG margins based on free drug concentrations". The data reported here support this view, given the reduced ROC curve AUC values when using C_{max} total. Use of C_{max} free values is also consistent with usual practice in the field of secondary ("off-target") pharmacology, of which hERG potency is just one aspect (see Bowes *et al.*, 2012).

Table 3. Concordance Data for Clinical, Heart Rate-Corrected QT Interval (QTc) Study Outcome and QTc Effect in Dogs

Parameter/Exposure Multiple	1	<3	<10	<30	<100
Number of compounds	17	14	16	11	11
Sensitivity	0.25	0.63	0.73	0.89	0.89
Lower 95% CI	0.03	0.24	0.39	0.52	0.52
Upper 95% CI	0.65	0.91	0.94	1.00	1.00
Specificity	1.00	0.83	0.80	0.50	0.50
Lower 95% CI	0.66	0.36	0.28	0.01	0.01
Upper 95% CI	1.00	1.00	0.99	0.99	0.99
PPV	1.00	0.51	0.51	0.33	0.33
Lower 95% CI	0.16	0.14	0.15	0.11	0.11
Upper 95% CI	1.00	0.87	0.86	0.67	0.67
NPV	0.83	0.89	0.91	0.94	0.94
Lower 95% CI	0.76	0.75	0.78	0.61	0.61
Upper 95% CI	0.88	0.95	0.97	0.99	0.99

Limitations of the Dataset

The ideal compound set would be larger than 24 and contain data from hERG, dog and clinical studies that were conducted using the same protocol, thus generating a consistent set of data. However, the data forming the basis for this work come from preclinical and clinical studies spanning 15 years; a period of time when QT risk assessment strategies, associated technology, study design and regulatory guidance have evolved. The dataset from dog studies is particularly limited; although it's comprised of 24 compounds, a lack of data around the clinical free C_{max} reference point for some dog studies resulted in only 11–17 data points across the exposure multiples assessed. This is something also evident in the work of Ewart *et al.* (2014), although a larger overall dataset limited the impact of this issue.

Nonconcordant Compounds

Compounds 8, 9, 12, and 13 are noteworthy false negatives: they are not potent hERG blockers, had large safety margins to hERG IC_{50} relative to clinical C_{max} free exposures and did not increase QTc in dog studies. Nor did they have significant activity at other key cardiac ion channels (no activity within a 1500-fold margin from the clinical C_{max} free reference point for: hNav1.5; hCav1.2/β2/α2δ; hKv4.3-hKChIP2.2; hKv7.1-hKCNE1) (unpublished data). Although as mentioned earlier, a clinical QTc increase was not unexpected for the long-acting β_2-adrenoceptor agonists (compounds 9 and 12) there was nothing about the primary pharmacology of compounds 8 and 13 to suggest a QTc risk. Common to all four compounds is the fact that they increased heart rate (6–9 beat per minute) in the clinical study from which QTc data were derived. Hence the clinical QTc effect may relate to modulation of autonomic tone leading to imperfect heart correction of QT interval and/or a real effect on repolarization (Magnano *et al.*, 2002). From an absolute hERG potency perspective, compound 10 (moxifloxacin) and compound 11 are interesting in the sense that despite increasing QTc in clinical studies, they are not potent hERG blockers. However, their hERG IC_{50} to clinical exposure safety margins are <30-fold and moxifloxacin increased QTc in dog studies. Although compound 14 could be regarded as a false positive based on its hERG IC_{50}, its large safety margin to clinical free C_{max} explains the lack of a QTc increase in a TQT study.

Use of the Data in the Context of Drug Discovery

It is worth considering how these data could help steer early drug discovery projects. Such projects are generating novel chemical entities to find a compound with an overall balance of properties that is good enough to constitute a candidate drug for clinical evaluation (Pollard *et al.*, 2010). As well as exploring structure-activity relationships at the primary target, projects investigate the pharmacological selectivity of their compounds via secondary ("off-target") pharmacology studies (see Bowes *et al.*, 2012). From a QTc interval prolongation perspective, potency in a hERG channel screen is an important early piece of secondary pharmacology data. In terms of decision-making about which compounds should be de-selected at this early stage, the relevant data are PPV and NPV, rather than sensitivity and specificity. As indicated in the "Materials and Methods" section, the PPV and NPV are ideally adjusted for prevalence. This value is not known for compounds yet to be tested in clinical studies so only an estimate can be applied. In this work a value of 0.22 was used based on the report of Park *et al.* (2013). Although PPV and NPV are high at some multiples of hERG IC_{50} to clinical C_{max} free ratio, in very early discovery, predictions of the C_{max} free plasma levels in clinical studies are not available. Hence the PPV and NPV of absolute hERG potency are the most relevant data (see Table 1). The PPV of absolute hERG potency suggest that if a compound has an $IC_{50} < 10\,\mu M$ then it would have a 63% chance of causing a QTc increase in clinical studies. The corresponding NPV data suggest that if a compound had an $IC_{50} > 10\,\mu M$ it would have an 87% chance of not causing a QTc increase in clinical studies. Hence the data support a pragmatic approach of having an absolute hERG potency cut-off in very early discovery, when little information is available for decision-making, providing it can be applied whilst retaining desirable compound properties. It would have to be used intelligently, however. For example, if the eventual candidate drug was going to be given by inhalation and was designed to be rapidly degraded on leaving the lung, compounds more potent than 10 μM would be acceptable, given the likelihood that plasma exposure of such a drug would be very low. Conversely, the absolute hERG potency target would need to be set higher than 10 μM for a drug class commonly requiring high free plasma concentrations for efficacy (eg, antibiotics). Also, it is clear that drug-hunting teams need to be aware of known links between the pharmacological class of their compounds and nonhERG-mediated increases in QTc, as is the case for the long-acting β_2-adrenoceptor agonists in this report.

A drug discovery project aims to produce a compound with properties likely to lead to clinical efficacy but with minimal side-effects. In the context of QTc interval prolongation, the data presented support the philosophy of aiming to reduce hERG potency early in the discovery process via knowledge of structure-activity relationships whilst designing-in properties that will lead to a good safety margin between clinical exposure and hERG potency.

SUPPLEMENTARY DATA

Supplementary data are available at *Toxicological Sciences* online.

FUNDING

This work was supported by AstraZeneca.

REFERENCES

Anonymous. (2001). International Conference on Harmonisation; guidance on S7a Safety Pharmacology Studies for Human Pharmaceuticals. *Notice Fed. Regist.* **66**, 36791–36792.

Anonymous. (2005a). International Conference on Harmonisation; guidance on S7b Nonclinical evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals; availability. *Notice Fed. Regist.* **70**, 61133–61134.

Anonymous. (2005b). International Conference on Harmonisation; guidance on E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs; availability. *Notice Fed Regist* **70**, 61134–61135.

Altman, D. G., and Bland, J. M. (1994a). Diagnostic tests 1: sensitivity and specificity. *Br. Med. J.* **308**, 1552.

Altman, D. G., and Bland, J. M. (1994b). Diagnostic tests 2: predictive values. *Br. Med. J.* **309**, 102.

Altman, D. G., and Bland, J. M. (1994c). Diagnostic tests 3: receiver operating characteristic plots. *Br. Med. J.* **309**, 188.

Bowes, J., Brown, A. J., Hamon, J., Jarolimek, W., Sridhar, A., Waldron, G., and Whitebread, S. (2012). Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. *Nat. Rev. Drug Discov.* **11**, 909–922.

Bremner, P., Woodman, K., Burgess, C., Crane, J., Purdie, G., Pearce, N., and Beasley, R. (1993). A comparison of the cardiovascular and metabolic effects of formoterol, salbutamol and fenoterol. *Eur. Respir. J.* **6**, 204–210.

Ewart, L., Aylott, M., Deurinck, M., Engwall, M., Gallacher, D. J., Geys, H., Jarvis, P., Ju, H., Leishman, D., Leong, L., *et al.* (2014). The concordance between nonclinical and phase I clinical cardiovascular assessment from a cross-company data sharing initiative. *Toxicol. Sci.* **142**, 427–435.

Gintant, G. (2011). An evaluation of hERG current assay performance: Translating preclinical safety studies to clinical QT prolongation. *Pharmacol. Ther.* **129**, 109–119.

Gintant, G., Sager, P. T., and Stockbridge, N. (2016). Evolution of strategies to improve preclinical cardiac safety testing. *Nat. Rev. Drug Discov.* **15**, 457–471.

Lecaillon, J. B., Kaiser, G., Palmisano, M., Morgan, J., and Della Cioppa, G. (1999). Pharmacokinetics and tolerability of formoterol in healthy volunteers after a single high dose of Foradil dry powder inhalation via Aerolizer. *Eur. J. Clin. Pharmacol.* **55**, 131–138.

Magnano, A. R., Holleran, S., Ramakrishnan, R., Reiffel, J. A., and Bloomfield, D. M. (2002). Autonomic nervous system influences on QT interval in normal subjects. *J. Am. Coll. Cardiol.* **39**, 1820–1826.

Mercaldo, N. D., Lau, K. F., and Zhou, X. H. (2007). Confidence intervals for predictive values with an emphasis to case-control studies. *Stat. Med.* **26**, 2170–2183.

Park, E., Willard, J., Bi, D., Fiszman, M., Kozeli, D., and Koerner, J. (2013). The impact of drug-related QT prolongation on FDA regulatory decisions. *Int J Cardiol.* **168**, 4975–4976.

Pollard, C. E., Abi Gerges, N., Bridgland-Taylor, M. H., Easter, A., Hammond, T. G., and Valentin, J. P. (2010). An introduction to QT interval prolongation and non-clinical approaches to assessing and reducing risk. *Br. J. Pharmacol.* **159**, 12–21.

Sanguinetti, M. C., and Tristani-Firouzi, M. (2006). hERG potassium channels and cardiac arrhythmia. *Nature* **440**, 463–469.

Shah, R. R. (2006). Can pharmacogenetics help rescue drugs withdrawn from the market? *Pharmacogenomics* **7**, 889–908.

Vargas, H. M., Bass, A. S., Koerner, J., Matis-Mitchell, S., Pugsley, M. K., Skinner, M., Burnham, M., Bridgland-Taylor, M., Pettit, S., and Valentin, J. P. (2015). Evaluation of drug-induced QT interval prolongation in animal and human studies: a literature review of concordance. *Br. J. Pharmacol.* **172**, 4002–4011.

Wallis, R. M. (2010). Integrated risk assessment and predictive value to humans of non-clinical repolarization assays. *Br. J. Pharmacol.* **159**, 115–121.