

1
2
3 **AN ANALYSIS OF THE RELATIONSHIP BETWEEN PRECLINICAL AND CLINICAL QT**
4
5 **INTERVAL-RELATED DATA**
6
7

8
9
10 Christopher E Pollard*, Matthew Skinner*, Stanley E Lazic[†], Helen M Prior*, Kelly M
11
12 Conlon*, Jean-Pierre Valentin*, Corina Dota[‡]
13
14
15
16
17
18
19

20 *Safety and ADME Translational Sciences, Drug Safety and Metabolism, [‡]Quantitative
21
22 Biology, Discovery Sciences, Innovative Medicines and Early Development, AstraZeneca
23
24 R&D, Cambridge, UK, CB4 0WG. [†]Cardiovascular Centre of Excellence, Global Medicines
25
26
27 Development, AstraZeneca R&D, Gothenburg, Sweden, 43183.
28
29
30

31 Running title: Preclinical to clinical QT concordance
32
33
34

35 Corresponding author:
36

37 Dr Christopher E Pollard
38

39 AstraZeneca
40

41 Cambridge Science Park, Milton Road
42

43 Cambridge
44

45 CB4 0WG
46

47 UK
48

49
50 Tel: +44 7824350586
51

52
53 Email: Chris.Pollard@astrazeneca.com
54
55
56
57
58
59
60

ABSTRACT

There has been significant focus on drug-induced QT interval prolongation caused by block of the hERG-encoded potassium channel. This can lead to a cardiac arrhythmia and has been a significant cause of attrition in drug discovery and development. Regulatory guidance has been implemented to assess QT interval prolongation risk: preclinical guidance requires a candidate drug's hERG potency 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 hERG and dog QTc data. Concordance was assessed by calculating sensitivity and specificity across a range of decision thresholds, thus yielding Receiver Operating Characteristic (ROC) curves of sensitivity versus (1- specificity). The area under the curve (AUC) of ROC curves (for which 0.5 and 1 indicate chance and perfect concordance, respectively) was used to summarise concordance. Three aspects of preclinical data were compared with the clinical outcome (ROC AUC values shown in brackets): absolute hERG IC₅₀ (0.78); safety margin between hERG IC₅₀ and clinical plasma exposure (0.80); safety margin between QTc effects in dogs and clinical plasma exposure (0.81). Positive and negative predictive values of absolute hERG IC₅₀ 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

INTRODUCTION

There has been considerable interest in drugs that block the hERG-encoded potassium channel as side-effect, thereby delaying ventricular action potential repolarisation and prolonging the QT interval on the surface electrocardiogram (Sanguinetti and Tristani-Firouzi, 2006; Gintant et al., 2016). 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 Harmonisation (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 (Anon. 2001) and ICH S7B (Anon. 2005a)) and another focused on clinical testing prior to marketing approval (ICH E14 (Anon. 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 (Anon. 2005a) stipulates that: "Ascending concentrations should be tested until a concentration-response curve has been 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 post-dose 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

1
2
3 assessment of QTc relative to plasma C_{max} , often expressed as the concentration not bound
4 to plasma proteins (C_{max} free).
5
6
7

8
9
10 The definitive clinical evaluation of drug-induced QT prolongation risk has been a “Thorough
11 QT/QTc Study” (TQTS) (ICH E14 (Anon. 2005b) which often involves measuring QT/QTc at
12 two dose levels of the compound of interest: the therapeutic dose and a “substantial multiple
13 of the anticipated maximum therapeutic exposure”. Since ICH E14 (Anon. 2005b)
14 recommends that QT/QTc data are collected at around C_{max} , these studies require
15 measurement of drug plasma concentrations to establish C_{max} , often expressed as C_{max} free.
16
17 A typical TQTS therefore yields an assessment of QTc versus plasma C_{max} free at two
18 doses. There are variants of this TQTS design: for example, some TQT studies use a single
19 dose that from previous studies has been established as the maximum tolerated dose, the
20 logic being that if it is “negative” at this dose (i.e. did not increase QTc), it will be negative at
21 a lower, therapeutic dose. ICH E14 defines the threshold for regulatory concern (i.e. a
22 “positive”) as a prolongation of “around 5 ms as evidenced by an upper bound of the 95%
23 confidence interval around the mean effect on QTc of 10 ms.” The statistical power required
24 to detect such a small change dictates the need for careful collection and analysis of QTc
25 data. Hence the output of TQT studies represents an opportunity to compare high resolution,
26 quantitative clinical data with that collected preclinically.
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

47 In contrast to the TQTS, which is often conducted late in development once the therapeutic
48 dose has been established, the Phase I clinical trials on a development compound are
49 usually a Single Ascending Dose (SAD) and a Multiple Ascending Dose (MAD) study. These
50 assess the pharmacokinetic and safety properties of a test compound, including QTc data,
51 but such studies have not traditionally been statistically powered to detect the very small
52 QTc prolongations used as the threshold for regulatory concern defined in ICH E14 (Anon
53 2005b). Thus only relatively large QTc increases have confidently been attributed to the
54 effect of the test compound. In other words, the apparent absence of a QTc effect in a Phase
55
56
57
58
59
60

1
2
3 I study is not deemed a definitive demonstration that the compound has no effect: only the
4
5 TQTS is powered sufficiently to define a negative. Hence additional preclinical to clinical
6
7 translational data can be obtained by studying compounds that have prolonged QTc in
8
9 Phase I studies, but not those which had no apparent effect.
10
11

12
13
14 In this assessment we have extracted clinical QTc data for 24 compounds from TQT studies,
15
16 or Phase I studies where there was a clear QTc increase, and compared them with hERG
17
18 and dog QTc data collected preclinically.
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

MATERIALS AND METHODS

Origin of compounds

Twenty-three 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 organisations.

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 ≥ 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.

1
2
3 In order to perform the relevant analyses for every compound, an IC_{50} value was required.
4
5 For inactive or low potency compounds where an IC_{50} was not defined, the following process
6
7 was used to provide an estimate. A concentration-effect curve was simulated that had a
8
9 user-defined IC_{50} value, a Hill slope of 1 and a minimum and maximum of 0% and 100%
10
11 inhibition, respectively. For compounds that were active but did not achieve 50% inhibition,
12
13 for example, 35% inhibition at the highest test concentration of X μ M, the user iteratively
14
15 inputted IC_{50} values until the 35% inhibition point occurred at X μ M and this IC_{50} was then
16
17 used as an estimate. For compounds that were inactive at the highest test concentration, the
18
19 same process was used to estimate a “worst-case” IC_{50} by assuming that the maximal
20
21 concentration tested elicited 5% inhibition of hERG current; this is the same approach as
22
23 that adopted by Gintant (2011).
24
25
26
27
28

29 **Dog safety pharmacology studies**

30
31
32 All studies were conducted according to the relevant national guidelines on animal welfare.
33
34 Six compounds were assessed in anaesthetised beagle dog studies and the remaining 18 in
35
36 conscious telemetered beagle dog studies. The key details of each study are shown in
37
38 Supplementary Table 2. Studies described the effect of test compound on a variety of
39
40 cardiovascular parameters, but for this work note was made of whether the report stated a
41
42 compound-related effect on QTc at a given dose. All studies reported the arithmetic mean of
43
44 the total plasma exposures at a time point judged to be close to C_{max} . These values were
45
46 converted to free (i.e. unbound) C_{max} values using dog plasma protein binding data. Hence,
47
48 for each dose in a study, the data extracted were the mean free C_{max} at that dose and
49
50 whether there was or wasn't an effect on QTc.
51
52
53

54 **Clinical studies**

55
56
57 The key details of each study are shown in Supplementary Table 3. All studies used digital
58
59 ECG collection. Data for sixteen of the compounds came from TQT studies. All 16 TQT
60

1
2
3 studies had a “positive” moxifloxacin arm and in 9 of these studies the C_{max} for moxifloxacin
4 was defined. In order to include moxifloxacin in the analysis, C_{max} data were extracted from
5 all 9 studies and data from the study with the median C_{max} value were used to exemplify
6 moxifloxacin clinical data.
7
8
9
10

11
12
13
14 Of the 8 remaining compounds, data for 6 came from SAD studies, 1 from day one of a MAD
15 study and 1 from a combination of SAD and MAD studies. The “positive” QTc outcome for
16 the latter was based on concentration-response modelling of QTc data.
17
18

19
20 Studies described the effect of test compound on a variety of parameters, but for this work
21 note was made of whether the report stated a compound-related effect on QTc at a given
22 dose. All studies reported the geometric mean of total plasma exposures at a time point
23 judged to be close to C_{max} . These were then converted to a free exposure using human
24 plasma protein binding data. Hence, for each dose in a study, the data extracted were the
25 mean free C_{max} at that dose and whether there was or wasn't an effect on QTc.
26
27
28
29
30
31
32
33

34 35 36 **Clinical plasma exposure reference point used to calculate exposure multiples** 37 38 **relative to preclinical data** 39

40
41 A key aspect of this work was to assess preclinical data at exposure multiples relative to
42 C_{max} free values in clinical studies. It is therefore important to clarify what clinical exposure
43 values were used. For compounds that prolonged QTc in clinical studies, the clinical
44 exposure reference point was the mean C_{max} free at the lowest dose where an increase in
45 QTc was seen. For compounds that did not prolong QTc in clinical studies, the mean C_{max}
46 free at the highest dose tested was used.
47
48
49
50
51
52
53

54 55 56 **Data classification** 57

58
59 For clinical studies, a “negative” was a compound that did not prolong QTc in a TQTS, based
60 on the definition of a “negative” in ICH E14: “a negative ‘thorough QT/QTc study’ is one in

1
2
3 which the upper bound of the 95% one-sided confidence interval for the largest time-
4
5 matched mean effect of the drug on the QTc interval excludes 10 ms.” A “positive” was a
6
7 compound that prolonged QTc in a TQTS based on the definition of a “positive” in ICH E14
8
9 (i.e. if the 95% one-sided confidence interval for the largest time-matched mean effect of the
10
11 drug on the QTc interval exceeded 10 ms). “Positive” compounds were also drawn from
12
13 those tested in other (i.e. non TQT) clinical studies where the study report concluded that
14
15 there was a compound-induced increase in QTc.
16
17
18
19
20

21 Concordance with clinical outcome was assessed for three aspects of the pre-clinical data,
22
23 using the same approach as described by Ewart et al., (2014).
24
25
26

- 27 1. The absolute hERG IC₅₀ value. This was assessed by setting thresholds for IC₅₀
28
29 below which hERG data were deemed “positive”, starting with a threshold of 0.3 μM
30
31 and increasing the value to 1, 3, 10, 30, 100, 300 and 1000 μM.
32
33
34
35
- 36 2. The exposure multiple between the hERG IC₅₀ and clinical C_{max} free reference point.
37
38 This was done by setting an exposure multiple that constituted a “positive” if the
39
40 hERG IC₅₀ fell within it, starting with an exposure multiple of 1-fold and increasing the
41
42 value to 3, 10, 30, 100, 300 and 1000-fold.
43
44
45
46
- 47 3. The exposure multiple between QTc increase in dogs and the clinical C_{max} free
48
49 reference point. This was done by setting an exposure multiple that constituted a
50
51 “positive” if the QTc increase in dogs fell within it, starting with an exposure multiple
52
53 of 1-fold and increasing the value to 3, 10, 30 and 100-fold. It is important to note that
54
55 the exposure multiples covered those both above and below the clinical C_{max} free
56
57 reference point. For example, the 10-fold exposure multiple ranged from 0.1 to 10-
58
59 fold the clinical C_{max} free reference point.
60

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 pre-clinical test. Specificity represents the proportion of compounds that did not prolong QTc in clinical studies that were correctly identified by a pre-clinical 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% confidence intervals 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-Specificity) (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.

1
2
3 PPV and NPV are influenced by outcome prevalence, which in this analysis is the
4 prevalence of compounds that prolong QTc in clinical studies as a proportion of the total
5 number of drugs tested in clinical studies. Although the prevalence (or pre-test probability) is
6 usually unknown, for this work we estimated prevalence using the data reported by Park et
7 al. (2013), who stated that 46 out of 205 TQT studies submitted to the U.S. Food and Drug
8 Administration were “positive”, thus giving a prevalence estimate of 0.22. This enabled the
9 calculation of values for PPV and NPV (Altman and Bland, 1994b), as follows:
10
11
12
13
14
15
16
17
18
19
20

$$21 \text{ PPV} = (\text{sensitivity} \times \text{prevalence}) / [\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})]$$

22
23
24

$$25 \text{ NPV} = \text{specificity} \times (1 - \text{prevalence}) / [\text{specificity} \times (1 - \text{prevalence}) + (1 - \text{sensitivity}) \times$$

26
27 prevalence)]
28
29
30

31 The 95% confidence intervals around PPV and NPV were calculated based on methods
32 described in Mercaldo et al. (2007).
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 (i.e. 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 non-hERG-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 less than 10 μ M. Conversely, 10 out of 11 of the compounds that did not prolong QTc in clinical studies had an IC_{50} greater than 10 μ M.

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

1
2
3 0.81 for that minus the two β_2 -adrenoceptor agonists) given that a value of 0.5 corresponds
4 to random chance and a value of 1 would be perfect concordance. For both datasets, the
5
6 to random chance and a value of 1 would be perfect concordance. For both datasets, the
7
8 optimal balance between sensitivity and specificity occurred at a hERG IC_{50} threshold of 10
9
10 μ M. The relative values of PPV and NPV changed with hERG IC_{50} threshold, as would be
11
12 expected, but at 10 μ M, for example, for the whole dataset, PPV = 0.63 whilst NPV = 0.87.
13
14 The PPV and NPV value for the dataset minus the two β_2 -adrenoceptor agonists was 0.66
15
16 and 0.90, respectively.
17
18

19 20 21 **Safety margin between clinical exposure and hERG activity**

22
23 Based on the well-established link between hERG block (assuming selective inhibition of
24
25 hERG) and increases in QTc (Sanguinetti and Tristani-Firouzi, 2006), it is logical to expect
26
27 that the smaller the safety margin between clinical C_{max} free and hERG IC_{50} , the greater the
28
29 chance of a QTc increase in the clinical study. In Figure 2A, we explored that expectation by
30
31 plotting the ratio of the hERG IC_{50} to clinical C_{max} free using the clinical exposure reference
32
33 points described in the Methods section. Nine out of 13 compounds that increased QTc in
34
35 clinical studies had safety margins less than 30-fold whilst 10 out of the 11 compounds that
36
37 did not prolong QTc had safety margins greater than 30-fold.
38
39
40

41
42 ROC curves both with and without the β_2 -adrenoceptor agonists were constructed for these
43
44 data (Figure 2B). A numerical summary of the data is shown in Table 2 along with the PPV
45
46 and NPV data. The AUC values for the ROC curves indicate good concordance between
47
48 hERG safety margin and clinical QTc outcome (0.80 for the total dataset and 0.92 for that
49
50 minus the two β_2 -adrenoceptor agonists). For both datasets, the optimal balance between
51
52 sensitivity and specificity occurred at a hERG IC_{50} /clinical C_{max} free ratio threshold of 30-fold.
53
54
55

56 57 **Relationship between clinical exposure and QTc effect in dog studies**

58
59 Figure 3 summarises the relationship between dog QTc data and the QTc outcome in clinical
60
studies. For each compound, the ratio of the C_{max} free in dogs at each dose tested was

1
2
3 plotted as a ratio of the clinical C_{max} free reference point for that compound, as described in
4 the Methods section. As stated in the Introduction, to comply with ICH S7A guidance (Anon,
5 2001), dog cardiovascular studies are usually designed with the aim of the lowest dose
6 achieving a plasma exposure similar to that anticipated in clinical studies and higher doses
7 giving plasma exposures that exceed those anticipated in clinical studies. However, at the
8 time the doses for dog studies are decided, the plasma exposure anticipated in clinical
9 studies is a prediction based on a variety of data types and estimates. Also, as described in
10 the Methods section, the reference clinical C_{max} free values used in Figure 3 are defined
11 differently depending on whether or not a compound prolonged QTc in clinical studies. In
12 addition, since some of the dog studies pre-date the publication of ICH S7A, its ethos was
13 not necessarily followed. For these reasons, when the exposures in dog studies are plotted
14 relative to the reference clinical C_{max} free values, there is often a lack of data around the
15 reference clinical C_{max} free value. Despite these gaps in the dataset, as shown in Figure 3A,
16 8 of the 13 compounds that prolonged QTc in clinical studies also prolonged QTc in dogs
17 within 10-fold the clinical C_{max} free. In contrast, 10 out of 11 compounds that did not increase
18 QTc in clinical studies did not prolong QTc in dogs.

19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40 Although fewer data points are available for this data set, ROC curves were generated both
41 with and without the β_2 -adrenoceptor agonists (Figure 3B). A numerical summary of the data
42 is shown in Table 3 along with the PPV and NPV data. The AUC values for the ROC curves
43 indicate good concordance between dog QTc data and clinical QTc outcome (0.81 for the
44 total dataset and 0.91 for that minus the two β_2 -adrenoceptor agonists). For both datasets,
45 the optimal balance between sensitivity and specificity occurred at a dog C_{max} free/clinical
46 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 (Wallis, 2010; Gintant, 2011; Ewart et al., 2014; Vargas et al., 2015). The concordance data improve if the two 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 pre-clinical 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 IC₅₀ 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 analysed 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 two long-acting β_2 -adrenoceptor 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 IC₅₀ to QTc outcome in clinical studies. With respect to analyses of hERG safety margins and clinical QTc data, the

1
2
3 dataset of Wallis (2010) is most similar to that used here. Its clinical dataset was drawn from
4
5 19 TQT studies or TQT-like studies from within one company (11 positives; 8 negatives).
6
7 When assessed at 2, 10 and 30-fold clinical exposure multiples, the hERG IC₁₀ had a
8
9 sensitivity of 0.82, 0.90 and 0.90, and a specificity of 0.75, 0.30 and 0.14, respectively.
10
11 Gintant (2011) conducted an analysis of 39 compounds from different sources for which the
12
13 TQTS outcome was reported (14 positives; 25 negatives). The optimal balance between
14
15 sensitivity and specificity was found to be at a 45-fold safety margin from clinical exposure to
16
17 hERG IC₅₀, at which point sensitivity was 0.64 and specificity 0.88. This compares with our
18
19 data for all 24 compounds which, at the optimal safety margin of 30-fold, had a sensitivity of
20
21 0.69 and a specificity of 0.91.
22
23
24
25
26

27 In terms of previous analyses of dog QTc to clinical QTc data, Wallis (2010) quoted
28
29 sensitivity and specificity values at 2, 10 and 30-fold clinical exposure multiples: sensitivity
30
31 was 0.83 at all multiples, whilst specificity was 0.86 and 0.33 at 2 and 10-fold, respectively
32
33 (no value was calculated at 30-fold). Ewart et al. (2014) used data from up to 95 compounds
34
35 for which there were both Phase 1 clinical QTc data and dog telemetry study QTc data. The
36
37 resulting analysis yielded sensitivity values of 0.17, 0.88 and 1.00 at 3, 10 and 30-fold
38
39 clinical exposure multiples, respectively, and specificity values of 0.92, 0.76 and 0.58 at
40
41 these multiples of clinical exposure, respectively. Finally, Vargas et al. (2015) used a text
42
43 mining approach to assess the concordance between animal (rabbit, guinea-pig, dog and
44
45 non-human primate) and human studies and reported a sensitivity of 0.91 and specificity of
46
47 0.88, although, as the authors acknowledge, this approach did not factor-in plasma exposure
48
49 values in animal or clinical studies. At a 10-fold clinical exposure multiple in dogs, the data
50
51 for 16 compounds in this report were a sensitivity of 0.73 and a specificity of 0.80.
52
53
54
55
56

57 With respect to PPV and NPV data, only Ewart et al. (2014) show data that can be
58
59 compared with that presented here. For example, using a prevalence of 0.27 to represent
60
cardiovascular safety-related attrition as a whole, at a 10-fold exposure multiple, dog

1
2
3 telemetry data are reported by Ewart et al. (2014) to have a PPV of 0.57 and a NPV of 0.94.

4
5 This compares with our values of 0.51 and 0.91 for PPV and NPV respectively.
6
7

8
9
10 Taken together, the concordance data for the 24 compounds reported here is consistent with
11 those reported by others.
12
13

14 15 16 **Limitations of the dataset**

17
18 The ideal compound set would be larger than 24 and contain data from hERG, dog and
19 clinical studies that were conducted using the same protocol, thus generating a consistent
20 set of data. However, the data forming the basis for this work come from preclinical and
21 clinical studies spanning 15 years; a period of time when QT risk assessment strategies,
22 associated technology, study design and regulatory guidance have evolved. The dataset
23 from dog studies is particularly limited; although it's comprised of 24 compounds, a lack of
24 data around the clinical free C_{max} reference point for some dog studies resulted in only 11-17
25 data points across the exposure multiples assessed. This is something also evident in the
26 work of Ewart et al. (2014), although a larger overall dataset limited the impact of this issue.
27
28
29
30
31
32
33
34
35
36
37
38
39

40 41 **Non-concordant compounds**

42 Compounds 8, 9, 12 and 13 are noteworthy false negatives: they are not potent hERG
43 blockers, had large safety margins to hERG IC_{50} relative to clinical C_{max} free exposures and
44 did not increase QTc in dog studies. Nor did they have significant activity at other key
45 cardiac ion channels (no activity within a 1500-fold margin from the clinical C_{max} free
46 reference point for: hNav1.5; hCav1.2/ $\beta_2/\alpha_2\delta$; hKv4.3-hKChIP2.2; hKv7.1-hKCNE1)
47 (unpublished data). Whilst, as mentioned earlier, a clinical QTc increase was not unexpected
48 for the long-acting β_2 -adrenoceptor agonists (compounds 9 and 12) there was nothing about
49 the primary pharmacology of compounds 8 and 13 to suggest a QTc risk. Common to all four
50 compounds is the fact that they increased heart rate (6-9 beat per minute) in the clinical
51 study from which QTc data were derived. Hence the clinical QTc effect may relate to
52
53
54
55
56
57
58
59
60

1
2
3 modulation of autonomic tone leading to imperfect heart correction of QT interval and/or a
4 real effect on repolarisation (Magnano et al. 2002). From an absolute hERG potency
5 perspective, compound 10 (moxifloxacin) and compound 11 are interesting in the sense that
6 despite increasing QTc in clinical studies, they are not potent hERG blockers. However, their
7 hERG IC₅₀ to clinical exposure safety margins are less than 30-fold and moxifloxacin
8 increased QTc in dog studies. Whilst compound 14 could be regarded as a false positive
9 based on its hERG IC₅₀, its large safety margin to clinical free C_{max} explains the lack of a
10 QTc increase in a TQT study.
11
12
13
14
15
16
17
18
19
20
21
22

23 **Use of the data in the context of drug discovery**

24
25 It is worth considering how these data could help steer early drug discovery projects. Such
26 projects are generating novel chemical entities to find a compound with an overall balance of
27 properties that is good enough to constitute a candidate drug for clinical evaluation (Pollard
28 et al., 2010). As well as exploring structure-activity relationships at the primary target,
29 projects investigate the pharmacological selectivity of their compounds via secondary (“off-
30 target”) pharmacology studies (see Bowes et al., 2012). From a QTc interval prolongation
31 perspective, potency in a hERG channel screen is an important early piece of secondary
32 pharmacology data. In terms of decision-making about which compounds should be de-
33 selected at this early stage, the relevant data are PPV and NPV, rather than sensitivity and
34 specificity. As indicated in the Methods, the PPV and NPV are ideally adjusted for
35 prevalence. This value is not known for compounds yet to be tested in clinical studies so
36 only an estimate can be applied. In this work a value of 0.22 was used based on the report
37 of Park et al. (2013). Although PPV and NPV are high at some multiples of hERG IC₅₀ to
38 clinical C_{max} free ratio, in very early discovery, predictions of the C_{max} free plasma levels in
39 clinical studies are not available. Hence the PPV and NPV of absolute hERG potency are
40 the most relevant data (see Table 1). The PPV of absolute hERG potency suggest that if a
41 compound has an IC₅₀ < 10 µM then it would have a 63% chance of causing a QTc increase
42 in clinical studies. The corresponding NPV data suggest that if a compound had an IC₅₀ > 10
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 μM it would have an 87% chance of not causing a QTc increase in clinical studies. Hence
4
5 the data support a pragmatic approach of having an absolute hERG potency cut-off in very
6
7 early discovery, when little information is available for decision-making, providing it can be
8
9 applied whilst retaining desirable compound properties. It would have to be used
10
11 intelligently, however. For example, if the eventual candidate drug was going to be given by
12
13 inhalation and was designed to be rapidly degraded on leaving the lung, compounds more
14
15 potent than 10 μM would be acceptable, given the likelihood that plasma exposure of such a
16
17 drug would be very low. Conversely, the absolute hERG potency target would need to be set
18
19 higher than 10 μM for a drug class commonly requiring high free plasma concentrations for
20
21 efficacy (e.g. antibiotics). Also, it is clear that drug-hunting teams need to be aware of known
22
23 links between the pharmacological class of their compounds and non-hERG-mediated
24
25 increases in QTc, as is the case for the long-acting β_2 -adrenoceptor agonists in this report.
26
27
28
29
30
31

32 A drug discovery project aims to produce a compound with properties likely to lead to clinical
33
34 efficacy but with minimal side-effects. In the context of QTc interval prolongation, the data
35
36 presented support the philosophy of aiming to reduce hERG potency early in the discovery
37
38 process via knowledge of structure-activity relationships whilst designing-in properties that
39
40 will lead to a good safety margin between clinical exposure and hERG potency.
41
42

43 **SUPPLEMENTARY DATA DESCRIPTION**

44
45
46 Supplementary data are available online at...

47 **FUNDING INFORMATION**

48
49
50 This work was supported by AstraZeneca.
51
52
53
54
55
56
57
58
59
60

FIGURE LEGENDS

Figure 1. Absolute hERG potency data relative to clinical QTc outcome. (A) Graphical summary of the data. Red triangles, clinical QTc increase; green circles, no clinical QTc increase. Compounds with IC_{50} s $>1000 \mu\text{M}$ are plotted at a value of $1000 \mu\text{M}$. (B) Receiver operating characteristic curves for the data in A. The complete dataset (24 compounds) is shown in black filled squares and continuous lines; the dataset minus the β_2 -adrenoceptor agonists (22 compounds) is shown in blue open circles and dashed lines.

Figure 2. Ratio of the hERG IC_{50} to clinical C_{max} free reference point, relative to clinical QTc outcome. (A) Graphical summary of the data. Red triangles, clinical QTc increase; green circles, no clinical QTc increase. Compounds with hERG IC_{50} to clinical C_{max} free ratios >1000 are plotted at a value of 1000. (B) Receiver operating characteristic curves for the data in A. The complete dataset (24 compounds) is shown in black filled squares and continuous lines; the dataset minus the β_2 -adrenoceptor agonists (22 compounds) is shown in blue open circles and dashed lines.

Figure 3. QTc effect in dog studies relative to clinical QTc outcome. (A) Graphical summary of the data plotted as the ratio of the dog C_{max} free at each dose tested, relative to the clinical C_{max} free reference point. Data are shown relative to clinical QTc outcome. Red triangles, clinical QTc increase; green circles, no clinical QTc increase. Filled symbols represent a QTc increase in dog studies. (B) Receiver operating characteristic curves for the data in A. The complete dataset is shown in black filled squares and continuous lines; the dataset minus the β_2 -adrenoceptor agonists is shown in blue open circles and dashed lines. Area Under the Curve (AUC) values are shown for each dataset. The number of compounds per data point on the ROC curve varies and is summarised in Table 3.

1
2
3 **SUPPLEMENTARY FIGURE LEGENDS**
4

5
6
7 **Supplementary Table 1:** hERG studies.
8

9
10 **Supplementary Table 2:** Dog studies.
11

12 **Supplementary Table 3:** Clinical studies.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

TABLES

Table 1. Concordance data for complete dataset (24 compounds) between clinical QTc study outcome and absolute hERG IC₅₀.

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
<i>Lower 95% CI</i>	<i>0.00</i>	<i>0.06</i>	<i>0.14</i>	<i>0.25</i>	<i>0.39</i>	<i>0.46</i>	<i>0.55</i>	<i>0.64</i>
<i>Upper 95% CI</i>	<i>0.25</i>	<i>0.54</i>	<i>0.68</i>	<i>0.81</i>	<i>0.91</i>	<i>0.95</i>	<i>0.98</i>	<i>1.00</i>
Specificity	1.00	1.00	0.91	0.91	0.73	0.64	0.36	0.36
<i>Lower 95% CI</i>	<i>0.72</i>	<i>0.72</i>	<i>0.59</i>	<i>0.59</i>	<i>0.39</i>	<i>0.31</i>	<i>0.11</i>	<i>0.11</i>
<i>Upper 95% CI</i>	<i>1.00</i>	<i>1.00</i>	<i>1.00</i>	<i>1.00</i>	<i>0.94</i>	<i>0.89</i>	<i>0.69</i>	<i>0.69</i>
PPV	-	1.00	0.54	0.63	0.42	0.37	0.27	0.29
<i>Lower 95% CI</i>	-	<i>0.29</i>	<i>0.14</i>	<i>0.19</i>	<i>0.20</i>	<i>0.21</i>	<i>0.18</i>	<i>0.20</i>
<i>Upper 95% CI</i>	-	<i>1.00</i>	<i>0.9</i>	<i>0.92</i>	<i>0.67</i>	<i>0.58</i>	<i>0.38</i>	<i>0.40</i>
NPV	0.78	0.82	0.84	0.87	0.89	0.91	0.89	0.94
<i>Lower 95% CI</i>	<i>0.78</i>	<i>0.77</i>	<i>0.77</i>	<i>0.79</i>	<i>0.77</i>	<i>0.77</i>	<i>0.65</i>	<i>0.69</i>
<i>Upper 95% CI</i>	<i>0.78</i>	<i>0.86</i>	<i>0.89</i>	<i>0.93</i>	<i>0.95</i>	<i>0.97</i>	<i>0.97</i>	<i>0.99</i>

Table 2. Concordance data for complete dataset (24 compounds) between clinical QTc outcome and safety margin to hERG IC₅₀.

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

Table 3. Concordance data for complete dataset clinical 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
<i>Lower 95% CI</i>	<i>0.03</i>	<i>0.24</i>	<i>0.39</i>	<i>0.52</i>	<i>0.52</i>
<i>Upper 95% CI</i>	<i>0.65</i>	<i>0.91</i>	<i>0.94</i>	<i>1.00</i>	<i>1.00</i>
Specificity	1.00	0.83	0.80	0.50	0.50
<i>Lower 95% CI</i>	<i>0.66</i>	<i>0.36</i>	<i>0.28</i>	<i>0.01</i>	<i>0.01</i>
<i>Upper 95% CI</i>	<i>1.00</i>	<i>1.00</i>	<i>0.99</i>	<i>0.99</i>	<i>0.99</i>
PPV	1.00	0.51	0.51	0.33	0.33
<i>Lower 95% CI</i>	<i>0.16</i>	<i>0.14</i>	<i>0.15</i>	<i>0.11</i>	<i>0.11</i>
<i>Upper 95% CI</i>	<i>1.00</i>	<i>0.87</i>	<i>0.86</i>	<i>0.67</i>	<i>0.67</i>
NPV	0.83	0.89	0.91	0.94	0.94
<i>Lower 95% CI</i>	<i>0.76</i>	<i>0.75</i>	<i>0.78</i>	<i>0.61</i>	<i>0.61</i>
<i>Upper 95% CI</i>	<i>0.88</i>	<i>0.95</i>	<i>0.97</i>	<i>0.99</i>	<i>0.99</i>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

Anonymous (2001). International Conference on Harmonisation; guidance on S7a Safety Pharmacology Studies for Human Pharmaceuticals. Notice Fed Regist 66: 36791-92

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.

Anonymous (2015). ICH E14 Guideline: The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs. Questions & Answers (R3). http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E14/E14_Q_As_R3__Step4.pdf

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

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

Altman, D.G. and Bland, J.M. (1994c). Diagnostic tests 3: receiver operating characteristic plots. *British Medical Journal*. **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, 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., McMahon, N., Mead, A., Milliken, P., Suter, W., Teisman, A., Van Ammel, K., Vargas, H. M., Wallis, R., and Valentin, J. P. (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.

1
2
3 Pollard, C. E., Abi Gerges, N., Bridgland-Taylor, M. H., Easter, A., Hammond, T. G., and Valentin,
4 J. P. (2010). An introduction to QT interval prolongation and non-clinical approaches to assessing
5 and reducing risk. *Br. J. Pharmacol.* **159**, 12-21.
6

7 Sanguinetti, M. C. and Tristani-Firouzi, M. (2006). hERG potassium channels and cardiac
8 arrhythmia. *Nature.* **440**, 463-469.
9

10 Shah, R. R. (2006). Can pharmacogenetics help rescue drugs withdrawn from the market?
11 *Pharmacogenomics.* **7**, 889-908.
12

13 Vargas, H. M., Bass, A. S., Koerner, J., Matis-Mitchell, S., Pugsley, M. K., Skinner, M., Burnham,
14 M., Bridgland-Taylor, M., Pettit, S., and Valentin, J. P. (2015). Evaluation of drug-induced QT
15 interval prolongation in animal and human studies: a literature review of concordance. *Br. J.*
16 *Pharmacol.* **172**, 4002-4011.
17

18 Wallis, R. M. (2010). Integrated risk assessment and predictive value to humans of non-clinical
19 repolarization assays. *Br. J. Pharmacol.* **159**, 115-121.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1

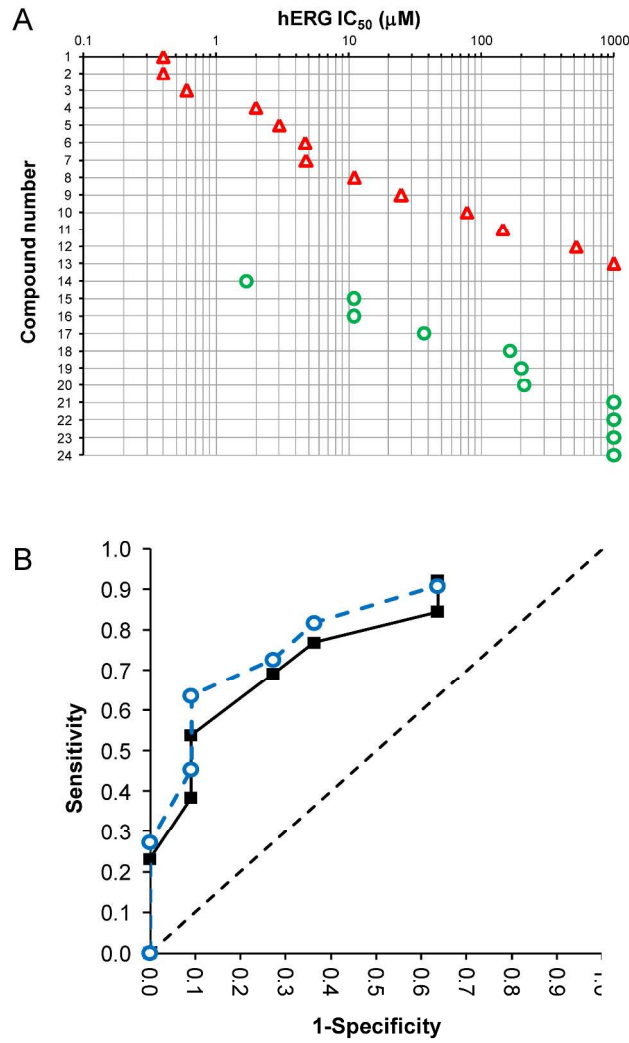


Figure 1. Absolute hERG potency data relative to clinical QTc outcome.

Figure 1

254x451mm (300 x 300 DPI)

Figure 2

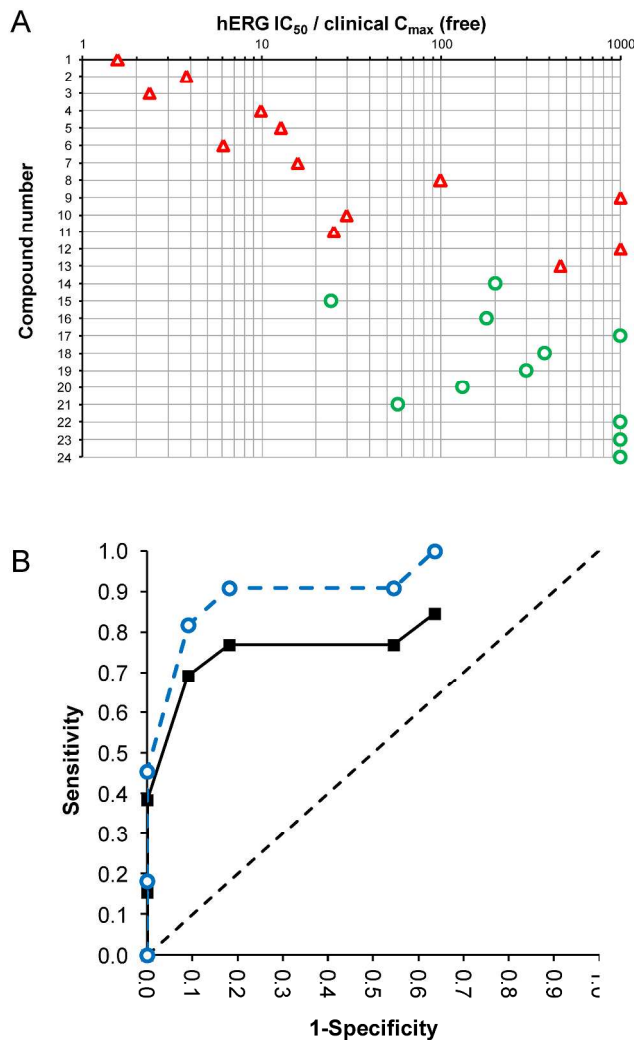


Figure 2. Ratio of the hERG IC50 to clinical Cmax free reference point, relative to clinical QTc outcome.

Figure 2

254x451mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 3

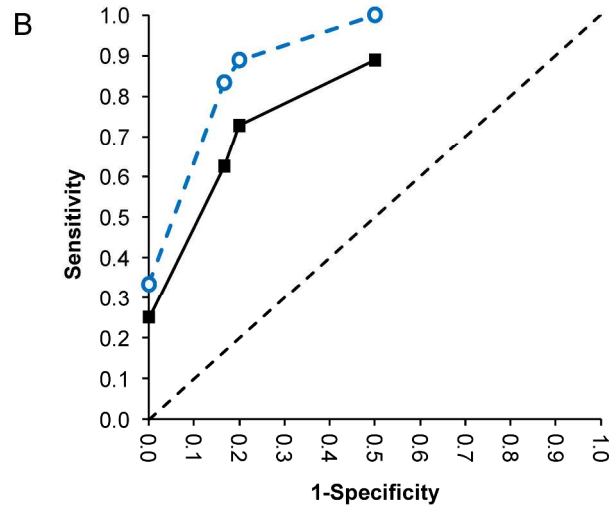
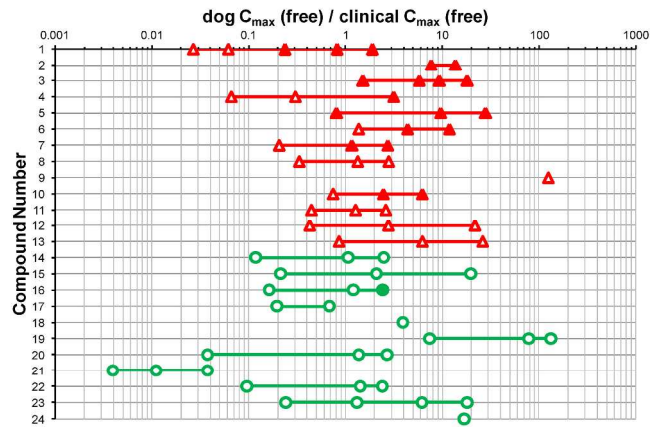


Figure 3. QTc effect in dog studies relative to clinical QTc outcome.

Figure 3

254x451mm (300 x 300 DPI)