Supporting Information:
Includes five Supplemental Tables, along with Section "Assay Descriptions".

Supplemental Table 1.

	Non-Clinical Assays			Clinical Study
	hERG Assay	APD Assay	In Vivo QTc Assay	TQT
Positive Finding	Concentration Eliciting $\geq IC_{50}$ hERG Block	Defined by Sponsor	Defined by Sponsor	Upper Bound One-Sided 95% Confidence Interval of Baseline & Placebo Corrected QTc Change Exceeds 10 msec
Negative Finding	Concentration Eliciting $< IC_{50}$ hERG Block	Defined by Sponsor	Defined by Sponsor	Upper Bound One-Sided 95% Confidence Interval of Baseline & Placebo Corrected QTc Change Less Than 10 msec

Table S1. Definitions of Outcomes Criteria, Positive and Negative Findings for Nonclinical and Clinical Assays.

Supplemental Table 2.

	Positive TQT Study (+)	Negative TQT Study (-)	
Positive Nonclinical Assay (+)	**True Positive (TP)**	**False Positive (FP)**	*Positive Predictive Value (PPV)* $\frac{\text{Sens x Prev}}{\text{Sens x Prev + (1- Spec) x (1- Prev)}}$
Negative Nonclinical Assay (-)	**False Negative (FN)**	**True Negative (TN)**	*Negative Predictive Value (NPV)* $\frac{\text{Spec x (1-Prev)}}{\text{(1-Sens) x Prev + Spec x (1- Prev)}}$
	Sensitivity TP/(TP+FN)	*Specificity* TN/(TN+FP)	*Concordance* (TP+TN)/(TP+FP+FN+TN)

LR+ = Sensitivity/(1-Specificity) ; ***LR -*** = (1-Sensitivity/Specificity)

Table S2. Dichotomous Contingency Table and Basis for Assay Characterizations. Results from each nonclinical assay (hERG, APD, *in vivo* QTc) were compared to QTc prolongation assessed in clinical thorough QT (TQT) studies. Comparisons across assays and studies were based on multiples of the free drug concentrations defined by clinical TQT study findings (CRCs, see text for definition).

Supplemental Table 3.

Study Type	Study Name	Inclusion in the Analysis		Inclusion of Positive Control		Initiation after ICH S7B / E14	
		Number	%	Number	%	Number	%
Clinical	TQT	150	100	138/150	92	116/150	77
Non-Clinical	hERG	142	95 (1)	80/89	90 (2)	29/85	34 (4)
	APD	84	56 (1)	49/59	83 (2)	6/53	11 (4)
	In vivo QTc	90	60 (1)	1/90	1	13/67	19 (4)

Table S3: Characteristics of the 150 Drug Anonymized Dataset**.** A total of 257 drugs from Investigational New Drug and New Drug Applications submitted to the FDA between March 2006 and July 2012 were evaluated for inclusion in the analysis. 107 drugs were excluded for one of the following reasons: clinical study sensitivity could not be demonstrated in non-TQT studies or negative TQT studies (n=12 and n=16, respectively), submissions describing drug combinations (n=11), absence of human plasma protein binding information (n=11), and absence of nonclinical data (n=57). The remaining 150 drugs included in the final dataset consisted of 91 New Drug Applications and 59 Investigational New Drug Applications.

[1] The earliest nonclinical data were collected from *in vivo* QTc, APD, and hERG studies conducted in 1993, 1997, and 1999, respectively. *In vivo* QTc data were excluded for 60 drugs because either plasma protein binding data were not available for the animal species studied or the maximum free drug concentration ($C_{max,free}$) was not provided. [2] Among 142 hERG data sets, the use of a positive control was not reported and unknown in 53 applications. Therefore, the percent of studies using positive controls was calculated based on the remaining 89 submissions. [3] Similarly, 59 of the 84 APD data sets reported whether or not a positive control had been used, and the percent was calculated based on 49 submissions which did use a positive control. [4] Study year was reported for only 85/142 hERG, 53/84 APD and 67/90 *in vivo* QTc assays.

Supplemental Table 4.

Youden's Index (J Statistic)	Clinical Reference Concentration Multiple						
Assay	1x	3x	10x	30x	100x	300x	1000x
hERG	-0.06	0.05	0.24	0.42	0.44	0.28	0.22
APD	-0.11	-0.08	0.13	0.10	0.01	0.05	0.09
In Vivo QTc	0.14	0.18	0.15	0.48	0.43	0.51	0.33

Table S4. Characterization of ROC curves. Youden's Index (J, also known as the Youden's J statistic), is a single value that describes the overall diagnostic performance of a dichotomous test that is calculated as J=Sensitivity + Specificity -1. A J value of 1 describes fully a fully accurate test (sensitivity and specificity =1), while a value of 0 provides no relevant information (same proportion of positive results for drugs that were or were not defined as eliciting QTc prolongation in clinical TQT studies). Assays

with comparable J values provide the same proportion of misclassified results. Results derived from dataset summarized in figure 4.

Supplemental Table 5.

FDA Review Division	Drugs in Database
Neurology	13.3%
Anti-virals	12.0%
Metabolic & Endocrine	11.3%
Cardiovascular & Renal	10.7%
Oncology	9.3%
Gastrointestinal & Inborn Errors	8.7%
Pulmonary, Allergy & Rheumatology	7.3%
Psychiatry	6.7%
Bone, Reproductive & Urology	6.0%
Anti-infectives	5.3%
Anesthesia, Analgesia & Addiction	4.7%
Hematology	4.0%
Transplant	0.7%

Table S5. Therapeutic Areas of Drugs in Dataset. Table illustrates varied therapeutic areas from which drugs were referred for review. Divisions listed in decreasing order.

Assay Descriptions.

TQT studies:

A total of 143 clinical studies were considered as TQT studies. Five of these studies did not use positive controls (such as moxifloxacin) to establish study sensitivity but were included in the analysis because the drugs tested were considered TQT positive. Data from 7 non-TQT studies were also included in the analysis because they displayed sensitivity to the test drug. The FDA classified 43 of the 150 studies (29%) as TQT positive based on ICH E14 criteria, the

remaining 107 studies classified as TQT negative. This prevalence rate was used when correcting PPV and NPV. The point estimate of the mean baseline and vehicle corrected (ΔΔ)QTc for the TQT positive drugs ranged from -2.9 ms to 38 ms. The number of drugs with (ΔΔ)QTc within the predefined stratified ranges was: 17 drugs (< 10 ms), 19 drugs (≥ 10 ms and < 20 ms) and 7 drugs (≥ 20 ms). Not surprisingly, a large proportion (77%) of clinical TQT studies were conducted after the introduction of the ICH S7B / E14 guidelines; in contrast only a small percentage (11%-34%) of nonclinical studies were conducted after the introduction of the guidelines.

hERG assay:

The study design for the hERG assay was similar across drugs. The cell type was reported in 103 studies: 77%, 21% and 2% were HEK293, CHO, and L-929 cells, respectively. Bath temperature was reported in 72 studies: 54% and 46% were conducted at room and physiological temperature (33 – 39 °C), respectively. While positive control compounds were included in most hERG studies, the information was not useful in assessing assay sensitivity because of the use of suprapharmacological concentrations resulting in near 100% inhibition of iKr current.

APD assay:

Data in the APD studies were collected over a range of stimulation frequencies (0.2 - 3Hz; 1Hz being the most common), and reported a range of relative action potential repolarization durations (APD; 10% - 90%; APD_{90} was the most common). Various species and tissues were used for the APD studies (n=84), including canine Purkinje fibers (46%), guinea pig papillary muscle tissue (24%) and rabbit tissues including Purkinje fibers, papillary muscle, and isolated heart (27%). One study used sheep Purkinje fibers.

In vivo QTc assay:

The *in vivo* QTc data (n=90) were obtained from safety pharmacology or toxicology studies in dogs or monkeys, the non-rodent species most commonly used in drug safety evaluations. The majority of studies were conducted using implanted telemetry (78%) and a cross-over study design (73%). They were performed mainly in conscious animals (79%) while fewer studies were conducted on anesthetized animals (21%), most of which were conducted prior to implementation of the ICH S7B (54/67 study year available = 81%). Approximately one third of the drugs tested in the *in vivo* QTc assays (36%; 32/90) included in-life C_{max} exposure measurements; in the other cases, C_{max} was estimated from related studies.