Citation: Clin Transl Sci (2019) 12, 257–266; doi:10.1111/cts.12596

ARTICLE

Comparison of Electrocardiographic Biomarkers for Differentiating Drug-Induced Single vs. Multiple Cardiac Ion Channel Block

Marina Brockway[1,*], Jay W. Mason[2,3] and Brian P. Brockway[1]

Since introduction of the International Conference on Harmonization proarrhythmia guidelines in 2005, no new marketed drugs have been associated with unacceptable risk of Torsade de Pointes. Although cardiac safety improved, these guidelines had the unintended consequence of eliminating potentially beneficial drugs from pipelines early in development. More recently, it has been shown that a corrected QT (QTc) prolonging drug may be safe if it impacts multiple ion channels vs. only human ether-a-go-go related gene (hERG) and that this effect can be discriminated using QT subintervals. We compared the predictive power of four electrocardiogram (ECG) repolarization metrics to discriminate single vs. multichannel block: (i) traditional 10-second signal averaged triplicates, and (ii) three metrics that used increasing density of automatically measured beat-to-beat (btb) intervals. Predictive power was evaluated using logistic regression and quantified with receiver operating characteristic (ROC) area under the curve (AUC). Compared with the traditional 10-second signal averaged triplicates, the reduction in classification error ranged from 2–6 with increasing density of btb measurements.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ Evaluation of biomarkers that can improve arrhythmia risk assessment by differentiating pure hERG block and multichannel block are under regulatory consideration for ECG studies and may impact cardiac safety assessment of future medications.

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ To date, the biomarker assessment has only been reported on the 10-second ECG recorded during rest, whereas arrhythmia liability is often associated with events during changes in autonomic states.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ This study demonstrated that examination of continuous ECG measurements of QTcF, JTpc, and TpTe intervals over 24 hours is possible with highly automated software. The traditional 10-second ECG sampling methods may not be sufficiently representative of the diversity of autonomic states, limiting their predictive power.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?
☑ It may be possible to improve predictive power of repolarization biomarkers if ECG sampling is expanded beyond beats with low and stable heart rates.

A guidance issued in 2005 by the International Committee for Harmonization[1] has prevented drugs that increase the risk of the fatal ventricular arrhythmia, Torsade de Pointes (TdP), from reaching the marketplace. However, this has come with a cost. It is now recognized that many pharmaceutical sponsors abandon drugs with preclinical and clinical repolarization signals, although many such drugs are potentially safe and beneficial.[2] Consequently, a consortium of multiple global drug regulators, industry, and academia is developing a new Comprehensive *In Vitro* Proarrhythmia Assay paradigm that is examining additional electrocardiogram (ECG) end points for assessing cardiac safety.

In 2014, the US Food and Drug Administration (FDA) conducted its own investigation of alternative ECG biomarkers using drugs known to cause QT prolongation and associated with varying incidences of TdP risk.[3] The focus of their biomarkers was on the phases of repolarization reflected as changes in the JTpc interval (measured from the J point to the peak of the T-wave, corrected for heart rate) and TpTe interval (measured from the peak of the T-wave to the end of the T-wave, not corrected for heart rate). Using four drugs and a placebo in a prospective, randomized, crossover design, they examined the changes of the JTpc and TpTe intervals as well as the traditional corrected QT Fridericia's formula (QTcF) end point.

[1]VivaQuant, St. Paul, Minnesota, USA; [2]Spaulding Clinical Research, West Bend, Wisconsin, USA; [3]Division of Cardiology, University of Utah, Salt Lake City, Utah, USA.
*Correspondence: Marina Brockway (mbrockway@vivaquant.com)
Received: 16 July 2018; accepted: 8 October 2018. doi:10.1111/cts.12596

Subjects were given two drugs associated with a high risk of arrhythmia, one being dofetilide, a pure inhibitor of the human ether-a-go-go related gene (hERG) potassium channel and the other, quinidine, a compound that, in addition to inhibiting hERG, also inhibits the peak sodium current and several other potassium currents. The remaining two drugs, ranolazine and verapamil, are associated with a lower risk of arrhythmia.[4,5] Both of them inhibit hERG in combination with either the late sodium current (ranolazine) or the L-type calcium current (verapamil). They found that the drugs associated with high risk of arrhythmia (quinidine and dofetilide) caused lengthening of the JTpc interval in contrast to ranolazine, which shortened or did not lengthen JTpc from baseline, despite corrected QT (QTc) prolongation,[3,6] and verapamil, which did not alter the QT interval or the T-wave segments.

To date, the FDA assessments of the JTpc and TpTe end points have only been reported on 10-second ECGs recorded during supine rest. To further advance the effort to find improved ECG biomarkers, we retrospectively examined beat-to-beat (btb) measurements of ECG intervals from the entire continuous 24-hour data set and compared the biomarker derived from btb to those obtained from the traditional triplicate approach. The btb intervals were processed under three protocols (i) averaged over three 10-second windows at 16 prespecified time points (BTBT), averaged over (ii) 5-minute and (iii) 30-minute windows every 30 minutes over 24 hours (referred to as BTB5 and BTB30) and compared with the traditional interval measurements at 16 prespecified time points (TSAT) posted on Physionet.[7] Biomarker performance was determined by accuracy of discriminating selective potassium block with a high arrhythmia risk from multichannel block with lower risk.

METHODS

All ECGs were obtained from the E-OTH-12-0109-020 (FDA-1) and EOTH-12-0109-021 (FDA-2) databases archived in the Telemetric and Holter ECG Warehouse at the University of Rochester Medical Center, Rochester, NY. The downloaded signals were processed using a previously validated, fully automated software system[8,9] Rhythm Express (VivaQuant, St. Paul, MN) to generate the vector magnitude ECG of the vectorcardiogram[10] and automatically measure cardiac intervals, as previously reported.[11] JTp was corrected for heart rate using the published formula[3] ($JTpc = JTp/RR^{0.58}$ with RR in seconds), and QT was corrected with Fridericia's correction ($QTcF = QT/RR^{1/3}$ with RR in seconds). R-R interval is the interval between peaks of sequential QRS complexes on ECG.

Study design

The archived ECG data were from randomized controlled five-way single-dose crossover clinical trials in 22 healthy volunteers (11 women) conducted at a phase I clinical research unit (Spaulding Clinical Research, West Bend, WI). The study details and inclusion criteria were previously reported,[3] and the trial was approved by the FDA Research Involving Human Subjects Committee and the local institutional review board. All subjects gave written informed consent.

As per previous description,[3] the morning of each treatment period the subjects received one of four drugs or placebo under fasting conditions. There was a 7-day washout period between each 24-hour treatment period, so subjects received treatment on days 1, 9, 17, 25, and 33. Prior to dosing, a continuous 12-lead ECG recorder (Surveyor, Mortara Instrument, Milwaukee, WI) using the Mason-Likar electrode configuration[12] was connected to each subject. The continuous ECG recordings were acquired at 500 Hz with an amplitude resolution of 2.5 μV. Subjects were required to rest in a supine position only for 5 minutes during each predefined time point. At prespecified time points, following the supine position, the subjects were also requested to perform various activities, such as sitting, standing, and leg raising. The remainder of the time between each time point, subjects were permitted to walk freely and conduct normal, but not stressful, activities (i.e., walk, sit up, eat, and sleep). After each of the 16 predefined extraction time point periods, a blood sample was drawn for pharmacokinetic analysis. Plasma drug concentration was determined using a validated liquid chromatography with tandem mass spectroscopy method by Frontage Laboratories (Exton, Philadelphia, PA).

Statistical methods

Four protocols of data postprocessing were evaluated:

1. TSAT: Intervals measured on signal averaged ECG complexes derived from each of three 10-second ECG strips by the FDA software at 16 prespecified time points. The interval measurements were extracted from a public database posted on Physionet[7] following the FDA study.
2. BTBT: Beat-to-beat interval measurements averaged over three 10-second ECG strips at each of 16 prespecified time points defined in the original protocol.[3]
3. BTB5: Beat-to-beat interval measurements averaged over 5-minute extraction periods, at 30-minute intervals.
4. BTB30: Beat-to-beat interval measurements averaged over consecutive 30-minute extraction periods.

In the third and fourth protocols, the resulting trends were smoothed with a cubic spline with lambda = 0.8. The resulting metrics were used to develop predictive models to discriminate pure hERG blockers vs. multichannel blockers.

The estimated means for placebo-corrected change from baseline were computed using lme4 and lsmeans packages in R 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria). The change from baseline for each ECG biomarker (ΔQTcF, ΔTpTe, and ΔJTpc) by time point was the dependent variable, for which baseline was defined as predose measurements averaged over the respective time windows. Period, time, treatment, and an interaction between treatment and time were included as fixed effects, and subject was included as a random effect.

Individual placebo-corrected changes from baseline were computed for each metric using Matlab (Mathworks). All data points were limited to an active treatment window,

estimated as time when drug concentration is at 40% of maximum individual drug concentration (C_{max}) or higher. The data in the active treatment window were used to predict drug class (risk or no risk). Specifically, the selective hERG potassium channel blockers (dofetilide and quinidine) were identified as having cardiac liability and multichannel blockers ranolazine and verapamil were assumed to have no cardiac liability. The values were pooled for all treatments and time points for analysis of their predictive power. In all model development, the data from two thirds of subjects were randomly selected into the training set and the remaining one third into the validation set. We developed predictive models of each individual metric and their combinations using single and multiple logistic regression. Prior to computing logistic regression models, LOWESS smoothing was performed to verify that the underlying model was appropriate. The logistic regression model predictive power was assessed using receiver operating characteristic (ROC) area under the curve (AUC), similar to Vicente *et al*.[6] Additionally, decision tree models were used to assess the relationship between the ECG biomarkers and cardiac risk. Their predictive performance was measured with classification loss, reported as percentage of error measured in cross-validated classification model.

RESULTS

Demographics and vital signs for the study were previously reported.[3] The average baseline ECG values (± SD) determined in this study by the four data processing protocols were:

1. TSAT: Heart rate 56.4 ± 7.0 bpm; QTcF 395.9 ± 16.7 ms; JTpc 225.5 ± 20.3 ms; and TpTe 73.2 ± 5.5 ms
2. BTBT: Heart rate 56.5 ± 7.1 bpm; QTcF 395.0 ± 16.7 ms; JTpc 228.3 ± 20.3 ms; and TpTe 76.6 ± 5.5 ms
3. BTB5: Heart rate 68.0 ± 10.6 bpm; QTcF 389.2 ± 15.0 ms; JTpc 238.7 ± 22.0 ms; and TpTe 75.5 ± 6.7 ms
4. BTB30: Heart rate 69.3 ± 9.6 bpm; QTcF 384.7 ± 13.9 ms; JTpc 234.5 ± 21.7 ms; and TpTe 74.9 ± 6.2 ms

Baseline values between days of treatment showed no statistically significant differences in any parameter between BTB30 and BTB5. The 12–13 bpm difference in heart rate between 10-second triplicate protocols (BTBT and TSAT) and protocols with longer btb averages (BTB5 and BTB30) was also reflected in a trend toward lower QTcF values for BTB5 and BTB30. During the 5-minute and 30-minute sampling windows, the subjects may have been ambulatory. Unlike the traditional triplicate extraction time points, there was no selection of beats based upon heart rate stability. Autonomic effects, limitation of heart rate correction methodology, or both might account for the shorter QTcF observed for BTB5 and BTB30 relative to TSAT and BTBT.

Time-response curves over 24 hours

The placebo-corrected change from baseline (ΔΔ) in QTcF, JTp, and TpTe, measured under BTB30, BTB5, BTBT, and TSAT are shown in **Figures S1–S4**, respectively, for dofetilide, quinidine, ranolazine, and verapamil. **Table S1** shows the mean maximum placebo-corrected change from baseline ΔΔQTcF, ΔΔJTpc, and ΔΔTpTe and the time of occurrence for each sampling regimen. There are statistically significant differences in ΔΔQTcF maximum between QT measurements derived from the triplicate 10-second ECGs at the 16 prespecified time points (TSAT and BTBT) and protocols with sampling at 30-minute intervals (BTB5 and BTB30) for the dofetilide ($P < 0.01$) with mean ± CI TSAT: 78.2 ± 6.6, BTBT: 62.9 ± 5.3, BTB5: 49.5 ± 4.5, BTB30: 50.3 ± 3.8 and quinidine arms ($P < 0.01$) with mean ± CI TSAT: 78.8 ± 6.7, BTBT: 60.5 ± 5.2, BTB5: 44.7 ± 4.6, and BTB30: 46.0 ± 3.9. There are also statistically significant differences in ΔΔTpTe between TSAT and the other protocols ($P < 0.01$; **Table S1**). The differences between mean maximum drug effects for each method were compared using a two-sample *t*-test. These differences may be attributed to the choice of T-peak between the two measurement algorithms, intermittent inclusion of the U-waves in the FDA algorithm for dofetilide and quinidine,[11] and inclusion of cardiac cycles with a wider range of heart rates in protocols with longer averaging periods. Further, a progressive reduction in average 95% confidence intervals was observed when measurements were derived from btb intervals vs. TSAT: 25% lower for BTBT, 35% lower for BTB5, and 43% lower for BTB30. The increase in the number of analyzed beats was probably an important factor in reduction in confidence intervals.

With all three extraction periods of 30-minute, 5-minute, and 10-second triplicates, the mean placebo-corrected maximum effects ΔΔQTcF, ΔΔJTpc, and ΔΔTpTe with dofetilide (C_{max} = 2.5 hours) and quinidine (C_{max} = 2.0 hours) treatments closely coincided with the previously reported peak plasma concentrations of those two drugs.[3] However, for ranolazine and verapamil, differences were observed with extraction periods and the sampling protocols. The peak plasma concentrations of ranolazine was achieved at 4.0 hours after dosing,[3] yet the maximal ECG changes in ΔΔQTcF and ΔΔJTpc occurred at nocturnal time, 16.5 hours after dosing (**Table S1**). For BTB30 ΔΔQTcF maximum = 7.1 ± 3.9 ms, and for BTB5 ΔΔQTcF maximum = 7.5 ± 4.6 ms. For BTB30 ΔΔJTpc maximum = 5.2 ± 4.6 ms, and for BTB5 ΔΔJTpc maximum = 5.4 ± 5.3 ms. QTcF effects during the first 6 hours after dosing, while drug concentration was highest, produced mean placebo-corrected increases no greater than 4.27 ms with 30-minute sampling and 5.77 ms with 5-minute sampling.

Single logistic regression

Individual metric performances on the test data set are summarized in **Table 1** for the four protocols. The best performing metric was QTcF in all protocols with AUC = 0.99 for BTB30 and BTB5 compared with an AUC = 0.96 for BTBT and TSAT. JTpc generally performed better than TpTe in BTB30, BTB5, and BTBT protocols with AUCs between 0.86 and 0.94 vs. 0.73 and 0.79, but TpTe performed better in TSAT protocol (AUC = 0.91).

Table 1 Predictive performance of ΔΔQTcF, ΔΔTpTe, and ΔΔJTpc measured under BTB30, BTB5, BTBT, and TSAT sampling protocols, per single logistic regression

Sampling period	Sensitivity	Specificity	AUC	R^2	Coefficient	SE	*P* value
BTB30							
JTpc	90.1	80.1	0.939	0.40	0.14	0.01	7.469E-48
QTcF	96.9	98.7	0.994	0.80	0.35	0.027	1.126E-37
TpTe	73.8	91.7	0.746	0.40	0.30	0.022	1.478E-42
BTB5							
JTpc	84.5	82.4	0.898	0.50	0.13	0.009	6.37E-49
QTcF	92.6	97.8	0.987	0.80	0.29	0.021	5.92E-44
TpTe	70.6	83.4	0.727	0.40	0.29	0.021	6.50E-44
BTBT							
JTpc	77.1	82.8	0.857	0.60	0.18	0.016	4.17E-31
QTcF	91.0	90.4	0.965	0.80	0.20	0.02	1.87E-25
TpTe	69.2	85.3	0.787	0.50	0.27	0.024	2.06E-28
TSAT							
JTpc	70.5	71.3	0.808	0.60	0.16	0.01	6.82E-32
QTcF	94.5	89.8	0.960	0.80	0.17	0.02	3.96E-26
TpTe	81.4	85.3	0.914	0.60	0.21	0.02	2.33E-29

ΔΔ, placebo-corrected change from baseline; AUC, area under the curve; BTB, beat-to-beat; BTBT, beat-to-beat time; JTpc, QTcF, and TpTe are intervals measured on ECG; JTpc, measured from the J point to the peak of the T-wave,corrected for heart rate; QTcF, corrected QT Fridericia's formula; TpTe, measured from the peak of the T-wave to the end of the T-wave, not corrected for heart rate; TSAT, traditional interval measurements at prespecified time points.

Multivariate logistic regression

Multivariate logistic regression was applied to all metrics and their combinations from BTB30, BTB5, BTBT, and TSAT to find the most predictive combination of metrics and their relative contribution (**Table 2**). Limiting the sampling period from 30 minutes (AUC range = 0.987–0.994) to 5 minutes (AUC range = 0.974–0.985) had little downside on model predictivity. However, the model using 10-second triplicate BTBT showed lower predictivity with an AUC range of 0.959–0.967. With traditional triplicate TSATs, the model predictive power dropped further to AUC range of 0.927–0.94. The relative contribution from QTcF decreases with shorter averaging period, from maximum of 0.62 with BTB30 to maximum of 0.23 for TSAT and 0.22 for BTBT. However, only in BTBT protocol a combination of other metrics (JTp and TpTe) outperforms QTcF alone. In TSAT protocol, combination of two or more metrics has lower predictivity than QTcF alone. The scatter plots of metric-based classification into low- and high-risk clusters on training and testing sets in the model logit ~1 + JTpc + QTcF for all protocols are shown in **Figures 1–4**. The scatter plots show the QTcF on x-axis and JTpc on y-axis. The classification errors marked as FP (cyan) and FN (magenta) are progressively decreasing with longer averaging windows.

Risk thresholds

Decision trees selected the ECG biomarkers in the same order as the ROC-AUC analysis of logistic regression (QTcF and JTpc). This method identified the range of uncertainty around QTcF threshold and its dependency on averaging method (**Table S2**). For the BTB30 the range was between 14 and 20 ms in ΔΔQTcF, for BTB5 the range was between 8 and 16 ms, for BTBT the range was between 21 and 27 ms, and for TSAT the range was between 18 and 28 ms. The higher thresholds with 10-second triplicates reflect the selection bias of the beats with stable, lower heart rate (HR) that may not be fully compensated for by HR correction formulas. In btb averages BTBT, BTB5, and BTB30, the first two decision splits are based on ΔΔQTcF thresholds and those resolve most of the decisions with 90–94% certainty. For the TSAT, the second decision split is based on ΔΔJTpc = 12 threshold. The decision splits in TSAT protocol are similar to the ones proposed by Vicente *et al*.[6] (JTpc = 9 ms and QTcF = 29 ms).

Experiments and observations on differences in predictive power

The reasons behind the differences in predictive power are not apparent, but understanding the phenomena associated with these differences may provide insight into the applicability and possibly the limitations of this methodology. Three phenomena were examined that might explain the increased predictive power of BTBT5 and BTB30: (i) the increased number of available time points, (ii) ambulatory vs. supine state, and (iii) beat sampling within a time segment.

Increased number of available time points. To test whether loss in power with triplicates can be explained by the fact that fewer time points are used with triplicates vs. 5-minute averages acquired every 30 minutes (BTB5) within the active drug windows, the 5-minute averages that did not occur around prespecified time points were excluded. In single and multiple logistic regression models, limiting the number of data points resulted in $<1\%$ reduction in predictive power for all metrics and their combinations. Therefore, this seems to have a minor influence on predictive performance.

Influence of ambulatory vs. supine state. To test whether loss in power with triplicates can be explained by the fact that subjects are always supine during triplicate

Table 2 Predictive performance of combinations of ΔΔQTcF, ΔΔTpTe and ΔΔJTpc measured under BTB30, BTB5, BTBT and TSAT sampling protocols, per multiple logistic regression models.

	Performance				Coefficients				SE				*P*-value			
Model	**Sensitivity**	**Specificity**	**AUC**	R^2	**Intr.**	**JTpc**	**QTcF**	**TpTe**	**Intr.**	**JTpc**	**QTcF**	**TpTe**	**Intr.**	**JTpc**	**QTcF**	**TpTe**
BTB30																
QTcF+JTpc+TpTe	92.6	93.6	0.987	0.9	−4.63	−0.25	0.62	−0.27	0.38	0.06	0.08	0.09	1.E-33	1.E-04	2.E-13	4.E-03
QTcF+JTpc	94.7	98.3	0.993	0.8	−4.74	−0.08	0.42	–	0.38	0.03	0.04	–	6.E-35	3.E-03	1.E-26	–
JTpc+TpTe	93.0	99.1	0.992	0.8	−4.06	0.25	–	0.41	0.31	0.02	–	0.04	1.E-38	3.E-36	–	1.E-29
QTcF+TpTe	95.8	98.7	0.994	0.8	−4.55	–	0.34	0.05	0.37	–	0.03	0.04	1.E-34	–	2.E-33	1.E-01
BTB5																
QTcF+JTpc+TpTe	92.7	99.1	0.981	0.8	−4.23	0.12	0.14	0.22	0.32	0.04	0.05	0.06	5.E-40	7.E-03	3.E-03	5.E-04
QTcF+JTpc	92.6	97.8	0.985	0.8	−4.10	−0.02	0.3	–	0.31	0.02	0.02	–	9.E-41	3.E-01	3.E-34	–
JTpc+TpTe	91.9	98.2	0.974	0.8	−4.13	0.24	–	0.38	0.31	0.02	–	0.03	4.E-40	9.E-36	–	5.E-32
QTcF+TpTe	92.3	98.6	0.982	0.8	−4.20	–	0.27	0.07	0.31	–	0.02	0.03	1.E-40	–	2.E-35	2.E-02
BTBT																
QTcF+JTpc+TpTe	90.4	90.9	0.968	0.9	−5.02	0.2	0.07	0.14	0.57	0.03	0.03	0.05	1.E-18	3.E-09	1.E-02	2.E-03
QTcF+JTpc	91.1	91.0	0.960	0.9	−4.53	0.13	0.15	–	0.5	0.03	0.02	–	6.E-20	3.E-07	1.E-15	–
JTpc+TpTe	89.9	91.4	0.967	0.9	−5.11	0.25	–	0.24	0.58	0.03	–	0.04	7.E-19	1.E-17	–	5.E-11
QTcF+TpTe	91.0	90.4	0.959	0.8	−4.04	–	0.22	−0.05	0.42	–	0.02	0.03	6.E-22	–	2.E-19	1.E-01
TSAT																
QTcF+JTpc+TpTe	91.0	90.0	0.938	0.8	−4.62	0.19	0.07	0.08	0.52	0.04	0.04	0.05	1.E-18	1.E-05	8.E-02	1.E-01
QTcF+JTpc	90.8	88.4	0.927	0.8	−4.47	0.14	0.13	–	0.5	0.03	0.02	–	4.E-19	1.E-07	4.E-17	–
JTpc+TpTe	91.7	90.1	0.940	0.8	−4.67	0.25	–	0.17	0.53	0.03	–	0.02	9.E-19	4.E-16	–	1.E-15
QTcF+TpTe	92.2	89.1	0.939	0.8	−4.09	–	0.23	−0.11	0.44	–	0.03	0.03	2.E-20	–	1.E-18	3.E-04

ΔΔ, placebo-corrected change from baseline; AUC, area under the curve; BTB, beat-to-beat; BTBT, beat-to-beat time; Intr Intercept; Tpc, QTcF, and TpTe are intervals measured on ECG; JTpc, measured from the J point to the peak of the T-wave,corrected for heart rate; QTcF, corrected QT Fridericia's formula; TpTe, measured from the peak of the T-wave to the end of the T-wave, not corrected for heart rate; TSAT, traditional interval measurements at prespecified time points.

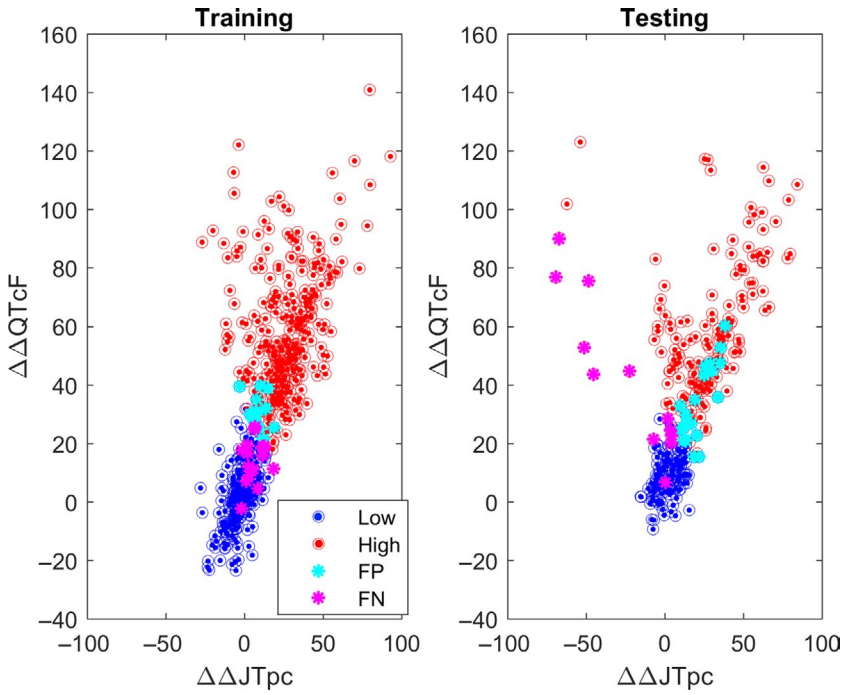

Figure 1 TSAT protocol. Scatter plot of multiple logistic regression results. Model used logit ~1 + JTpc (measured from the J point to the peak of the T-wave, corrected for heart rate) + corrected QT Fridericia's formula (QTcF). FP, false positive; FN, false negative. All values in milliseconds.

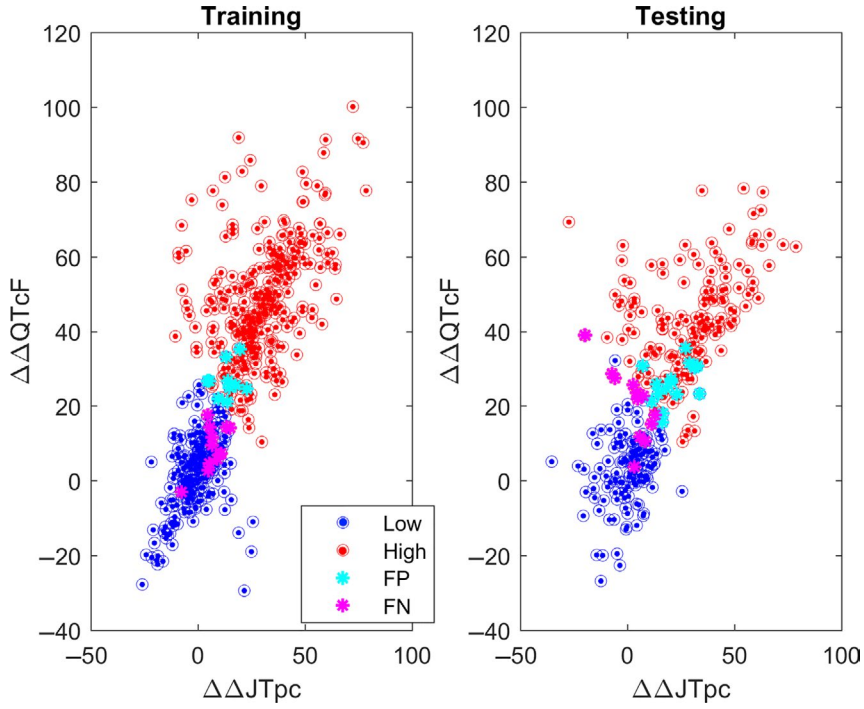

Figure 2 BTBT protocol. Scatter plot of multiple logistic regression results. Model used logit ~1 + JTpc (measured from the J point to the peak of the T-wave, corrected for heart rate) + corrected QT Fridericia's formula (QTcF). FP, false positive; FN, false negative. All values in milliseconds.

recordings, BTB5 measurements were separated to (i) supine data synchronized to the supine period of the prespecified time points and (ii) ambulatory data. Separate logistic regression models were developed for each group. Prespecified time points were scheduled at 30-minute intervals during the period of the highest drug

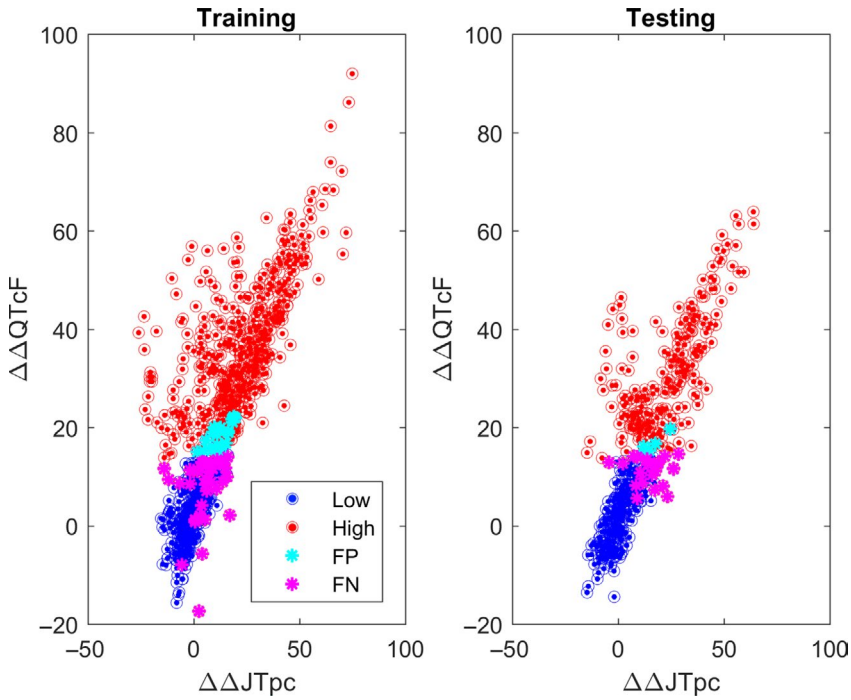

Figure 3 BTB5 protocol. Scatter plot of multiple logistic regression results. Model used logit ~1 + JTpc (measured from the J point to the peak of the T-wave, corrected for heart rate) + corrected QT Fridericia's formula (QTcF). O, low-risk measure; FN, false negative. All values in milliseconds.

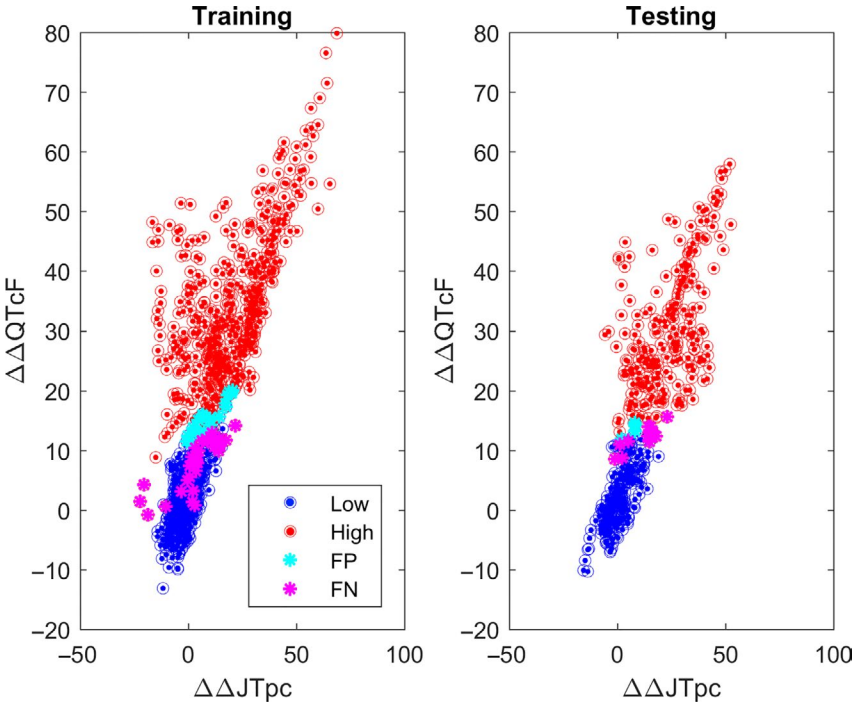

Figure 4 BTB30 protocol. Scatter plot of multiple logistic regression results. Model used logit ~1 + JTpc (measured from the J point to the peak of the T-wave, corrected for heart rate) + corrected QT Fridericia's formula (QTcF). FP, false positive; FN, false negative. All values in milliseconds.

concentration (0.5–4 hours). These data were classified as supine and excluded from the BTB5 ambulatory data set. Despite excluding data with highest drug concentration, the ambulatory data set had higher predictive power (1% increase) with increased sensitivity and specificity over the supine data set. The predictivity of BTB5 supine data

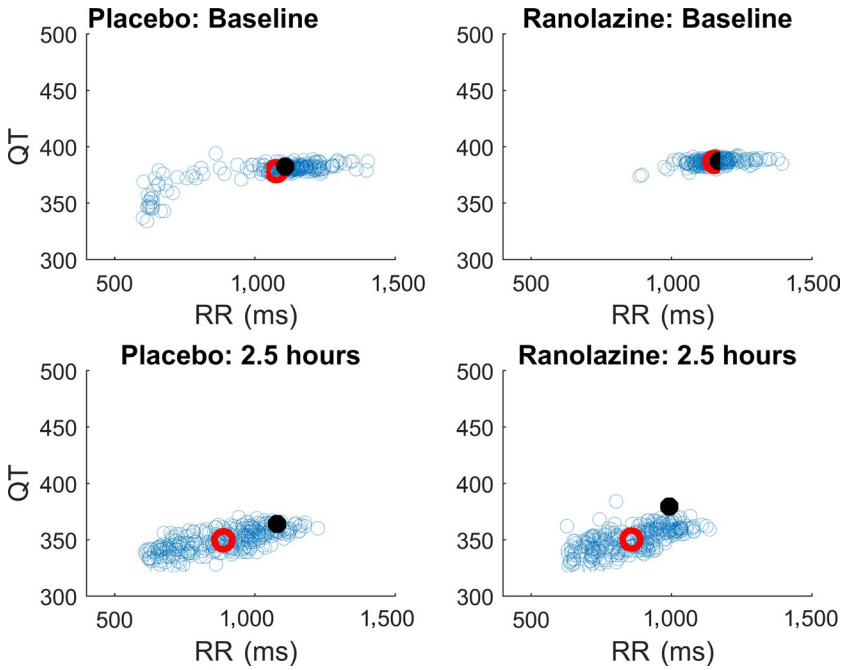

Figure 5 Scatter plots of QT vs. RR during 5-minute time points (blue), traditional interval measurements at prespecified time points (black), beat-to-beat 5 (red) for the following time points: baseline, placebo, and ranolazine at 2.5 hours postdose.

was 1% and 3% higher than the predictivity of BTBT and TSAT protocols, respectively. Therefore, it would seem that ambulatory data are advantageous, but it is even more important to use btb averages instead of selecting beats with low and stable heart rate only.

Influence of beat sampling within a time segment. To test whether degradation in power with triplicates can be explained by the methodology used to obtain TSAT measurements, we plotted the btb QT/RR clouds for 5-minute time points synchronized with triplicate measurements at the prespecified time points. QT/RR distributions are shown for baseline as well as placebo and ranolazine at 2.5-hour time points in **Figure 5**. TSAT (black circle) and BTB5 measurements (red ring) were superimposed on the QT/RR cloud. Numerous cloud plots generated from this study indicate that the relationship between TSAT and BTB5 shown in **Figure 5** is common for a high percentage of QT/RR measurements for multiple subjects and time points. These plots indicate that TSAT is biased toward selection of beats with lower HR. We believe this is an artifact of the TSAT methodology that prescribes selection of beats with stable HR prior to computation of a signal averaged beat representing the 10-second ECG strip. Often, the HR distribution within each 10-second segment tends to cluster around lower HR with occasional beats of higher HR that are typically excluded in order to satisfy TSAT beat-selection protocol.

The differences observed in QT measurements shown in **Figure 5** may be explained by different methodologies[11] as well as inclusion of a larger set of beats with BTB5 protocol. Further, the Fridericia correction has limited ability to compensate for HR differences that exceed 10 bpm.[13]

DISCUSSION

This study retrospectively analyzed archived data from a randomized controlled clinical trial conducted by the FDA to examine multiple biomarkers that can differentiate multichannel block and improve assessment of arrhythmia risk compared with use of QTcF alone. Both 10-second ECGs and continuous Holter recordings were archived by the FDA, but to date only the measurements of 10-second ECGs from resting, supine time points have been reported.[3] Because arrhythmia liability is often associated with events during changes in autonomic state,[14,15] examination of data throughout a 24-hour period to evaluate the performance of biomarkers is warranted. The purpose of this study was to compare the traditional placebo-adjusted end points derived from triplicate signal averaged 10-second ECG measures during defined supine rest periods to end points derived from btb measurements obtained from longer time segments at regular intervals.

Predictive models using traditional triplicate 10-second measurements have AUC = 0.94. When predictive models were computed using btb intervals averaged over consecutive 30-minute segments (BTB30), predictivity improved to AUC = 0.99, resulting in a sixfold reduction in classification error compared with traditional triplicate 10-second measurements. When predictive models were computed using btb intervals averaged over 5-minute segments every 30 minutes (BTB5), AUC = 0.98, a threefold reduction in error. Further, when the predictive models were limited to 5-minute

btb interval averages at 16 prespecified time points, AUC = 0.97, a twofold reduction in error resulted.

In general, the placebo-corrected averages of repolarization intervals are longest with 10-second triplicate data and become shorter with the 5- and 30-minute sampling periods. This is most likely due to the fact that during the longer collection periods, averages included beats obtained during ambulation or nonsupine positions under greater sympathetic tone, less vagal tone, or both. Direct autonomic effects, failure of the heart rate correction methods to fully correct, or both might account for repolarization shortening. The triplicate measures are generally acquired from signal averaged ECG complexes selected for stability of heart rate while subjects are resting in a supine position. This essentially results in clamping of autonomic tone to a high vagal state. The QT/RR clouds in **Figure 5** demonstrate how favoring low heart rate in TSAT measurements can bias QT selection toward outliers. It is likely that this effect will be more pronounced for drugs that lower heart rate and increase heart rate variability. Given the importance of heart rate heterogeneity in arrhythmia generation mechanisms, this bias seems to have the potential to impair predictive power of triplicates.

The mean ΔΔQTcF was prolonged with dofetilide and quinidine as expected. This provided good positive control data for modeling assessment of risk predictions. However, to assess the risk prediction of false-positive results, ranolazine was specifically included in the protocol design of the study because it has been considered to be a QTc prolonging agent with mixed ion channel activity resulting in minimal if any arrhythmia risk.[3,16] JTpc and TpTe measures were examined to determine if they could help further discriminate the safety in a potential false-positive scenario, assuming the study showed QTcF prolongation for ranolazine. Johannesen *et al*.[3] reported a mean increase of 12.6 ms in ΔΔQTcF with ranolazine at ~4 hours. However, in BTB30 and BTB5 protocols, it was found that the maximum mean ΔΔQTcF increase with ranolazine was 7.1 and 7.5 ms, respectively, and that these increases occurred at 16.5 hours (nocturnal time points) that had been excluded from the analyses because they were below 40% maximum drug concentrations. The lack of a QTcF prolongation in these protocols removed the potential for a false-positive scenario, which may have contributed to the high predictive value of QTcF (AUC = 0.99).

LIMITATIONS

In the TSAT protocol, the FDA reported values were used. However, the predictive power of the resulting models was higher than reported by Vicente *et al*.[6] Possible explanations are: (i) training and test set for the FDA models were derived from different studies, specifically study 2 data[17] (FDA study 2: NCT02308748) were used for training set and study 1 data[3] for validation set (FDA study 1: NCT01873950). In this investigation, only data from study 1 were used; (ii) data were limited to active drug window estimated as time when drug concentration is at 40% of C_{max} or higher. The application of an active drug window improved our ability to detect the drug's effects. Those effects, if present, are expected to be greater at higher plasma concentrations. Thus, our technique is expected to magnify the visibility of drug effects. In Vicente *et al*.,[6] placebo-corrected changes from baseline in each ECG biomarker were pooled from all time points.

This study reported on predictive power in discriminating pro-arrhythmic risk of drugs with strong QTcF effect, thus favoring QTcF as a predictor. It would be important to validate the observations from this study on drugs with more moderate QT effects, and drugs that have a mixed QT signature due to their modulation of other ion channels and other factors, such as autonomic tone.

This article reported retrospective analysis results with the drugs that are well studied. This design was important in accessing the predictive power of the biomarkers. The results are intended to inform the study design of future prospective studies with prospectively identified btb time points with 30-minute intervals measured from dose intake.

CONCLUSION

This study demonstrated that fully automated continuous btb measurements of QTcF, JTpc, and TpTe intervals over 24 hours can be obtained using the Rhythm Express software and used to discriminate multichannel ion block from pure hERG block. A comparison of predictive power gauged by AUC for four ECG repolarization metrics showed that, for this data set, the use of btb measurements was superior to the traditional method of deriving interval measurements from 10-second triplicates, with btb measurements exhibiting a twofold to sixfold reduction in classification error while providing a modest improvement in sensitivity. The authors speculate that the improvement observed in predictivity of btb measurements may be due to the increased contribution to the predictive model of a diversity of autonomic states vs. the traditional signal averaged triplicates in which autonomic tone is clamped as a result of the need to derive the signal averaged complex from beats of homogenous heart rate. Traditional 10-second signal averaged ECGs not only bias selection of beats toward those with the least variation between cardiac cycles, they also bias the selection of included beats toward slower heart rates, conditions that may be less prone to detection of ventricular arrhythmia liability. Further, 10-second signal averaged ECGs are not representative of the natural nonstationary QT/RR relationship. Based upon the results on this data set, the use of btb interval measurements shows promise for significantly reducing classification error and improving the specificity of assays targeted at discriminating multichannel ion block from pure hERG block.

Supporting Information. Supplementary information accompanies this paper on the *Clinical and Translational Science* website (www.cts-journal.com).

Table S1. Maximum placebo-corrected changes in QTcF, JTpc, and TpTe (mean ± CI) measured under BTB30, BTB5, BTBT, and TSAT sampling protocols.

Table S2. Predictive performance of decision tree models and the decision splits for BTB30, BTB5, BTBT, and TSAT sampling protocols.

Figure S1. Placebo-corrected changes from baseline (mean ± 95% confidence interval) (ΔΔ) QTcF, JTpc, and TpTe intervals averaged over 30-minute sampling periods over 24 hours (BTB30) after dosing with dofetilide, quinidine, ranolazine, or verapamil.

Figure S2. Placebo-corrected change from baseline (mean ± 95% confidence interval) (ΔΔ) QTcF, JTpc, and TpTe intervals averaged over 5-minute sampling periods every 30 minutes over 24 hours (BTB5) after dosing with dofetilide, quinidine, ranolazine, or verapamil.

Figure S3. Placebo-corrected change from baseline (mean ± 95% confidence interval) (ΔΔ) QTcF, JTpc, and TpTe intervals averaged over 10-second triplicates measured with RE (BTBT) after dosing with dofetilide, quinidine, ranolazine, or verapamil.

Figure S4. Placebo-corrected change from baseline (mean ± 95% confidence interval) (ΔΔ) QTcF, JTpc, and TpTe intervals using 10-second triplicates measured by the US Food and Drug Administration (TSAT) after dosing with dofetilide, quinidine, ranolazine, or verapamil.

Acknowledgments. The authors thank Heather Simenson of VivaQuant for her technical assistance in the analyses of ECG, and Douglas Hawkins, PhD, for his assistance in data interpretation and validation of statistical models.

Funding. Research reported in this publication was supported in part by the National Institute on Drug Abuse of the National Institutes of Health under Award Number R44DA041815 and by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number R43HL110739.

Conflict of Interest. M.B. is a shareholder of VivaQuant. J.W.M. is a consultant to VivaQuant. B.P.B. is a shareholder of VivaQuant.

Author Contributions. M.B., J.W.M., and B.P.B. wrote the article. J.W.M. designed the research. M.B. performed the research. M.B. analyzed the data. M.B. contributed new reagents/analytical tools.

1. International Conference on Harmonization. Guidance for Industry. E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs.<https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM073153.pdf>.
2. Stockbridge, N., Morganroth, J., Shah, R.R. & Garnett, C. Dealing with global safety issues: was the response to QT-liability of non-cardiac drugs well coordinated? *Drug Saf.* **36**, 167–182 (2013).
3. Johannesen, J. *et al.* Differentiating drug-induced multichannel block on the electrocardiogram: randomized study of dofetilide, quinidine, ranolazine and verapamil. *Clin. Pharmacol. Ther.* **96**, 549–558 (2014).
4. Aiba, T. *et al.* Cellular and ionic mechanism for drug-induced long QT syndrome and effectiveness of verapamil. *J. Am. Coll. Cardiol.* **45**, 300–307 (2005).
5. Antzelevitch, C. *et al.* Electrophysiological effects of ranolazine, a novel antianginal agent with antiarrhythmic properties. *Circulation* **110**, 904–910 (2004).
6. Vicente, J. *et al.* Electrocardiographic biomarkers for detection of drug-induced late sodium current block. *PLoS One* **11**, e0163619 (2016).
7. Goldberger, A.L. *et al.* PhysioBank, PhysioToolkit, and PhysioNet; components of a new research resource for complex physiologic signals. *Circulation* **101**, e215–e220 (2000).
8. Brockway, M. & Hamlin, R.L. Evaluation of an algorithm for highly automated measurements of QT interval. *J. Pharmacol. Toxicol. Methods* **64**, 16–24 (2011).
9. Brockway, M. & Brockway, B. Extraction of cardiac signal data. US Patent 9,050,007 (2013).
10. Guldenring, D. *et al.* Transformation of the Mason-Likar 12-lead electrocardiogram to the Frank vectorcardiogram. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2012**, 677–680 (2012) <https://doi.org/10.1109/embc.2012.6346022>.
11. Brockway, M., Fossa, A. & Mason, J.W. Comparison of two highly automated ECG algorithms for detection of drug-induced cardiac ion channel block. *Clin. Pharmacol. Ther.* **104**, 356–363 (2018).
12. Mason, R.E. & Likar, I. A new system of multiple-lead exercise electrocardiography. *Am. Heart J.* **71**, 196–205 (1966).
13. Malik, M., Hnatkova, K. & Batchvarov, V. Differences between study-specific and subject-specific heart rate corrections of the QT interval in investigations of drug induced QTc prolongation. *Pacing Clin. Electrophysiol.* **27**, 791–800 (2004).
14. Shen, M.J. & Zipes, D.P. Role of the autonomic nervous system in modulating cardiac arrhythmias. *Circ. Res.* **114**, 1004–1021 (2014).
15. Schwartz, P.J. *et al.* Genotype-phenotype correlation in the long-QT syndrome gene-specific triggers for life-threatening arrhythmias. *Circulation* **103**, 89–95 (2001).
16. Scirica, B.M. *et al.* Effect of ranolazine, antianginal agent with novel electrophysiological properties, on the incidence of arrhythmias in patients with non ST-segment elevation acute coronary syndrome: results from the metabolic efficiency with ranolazine for less ischemia in non ST-elevation acute coronary syndrome thrombolysis in myocardial infarction 36 (MERLIN-TIMI 36) randomized controlled trial. *Circulation* **116**, 1647–1652 (2007).
17. Johannesen, L. *et al.* Late sodium current block for drug-induced long QT syndrome: results from a prospective clinical trial. *Clin. Pharmacol. Ther.* **99**, 214–223 (2016).