Man versus Machine: Comparison of Automated and Manual Methodologies for Measuring the QTc Interval: A Prospective Study

Jean T. Barbey, M.D.,* Margaret Connolly, Ph.D.,† Brenda Beaty, M.S.,‡ and Mori J. Krantz, M.D.§¶

From the *Social & Scientific Systems, Inc., Silver Spring, MD; †PAREXEL International Corp., Waltham, MA; ‡University of Colorado, Aurora Colorado; §Denver Health and Hospitals, Denver, CO; and ¶CPC Clinical Research, Aurora, CO

Background: Electrocardiographic (ECG) safety evaluation is a required element of drug development. Performance characteristics of ECG measurement methodologies have rarely been studied prospectively.

Methods: We conducted a randomized, placebo-controlled, crossover study in 24 subjects to evaluate effects of moxifloxacin on the Fridericia rate-corrected QT (QTcF) interval. Five ECG replicates were obtained at 30 time points. Change from baseline QTcF (Δ QTcF) was fit by mixed-model analysis of variance to evaluate residual standard deviation. Precision was defined as intrasubject QTcF variance. Two core lab approaches were compared: QTinno, fully automated, 5 replicates and HeartSignals, computer-assisted manual, 3 replicates. Core lab values were then compared to an automated commercial algorithm (VERITAS).

Results: Twenty-three subjects provided 3450 ECGs potentially available for analysis. QTinno QTcF values were based upon 3419 ECGs, HeartSignals data on 2028 ECGs. Variance was similar between the QTinno and HeartSignals approaches (41.5 and 44 ms²). After excluding VERITAS QTcF measurements that deviated by >40 ms on visual review, variance in a set of 1907 common ECGs was lowest for HeartSignal, followed by QTinno and VERITAS (43.8, 52.6, 89.4 ms²) $P = 0.02$ HeartSignals versus QTinno, $P < 0.0001$ for both HeartSignals and QTinno versus VERITAS.

Conclusions: A fully automated core lab approach using 5 replicates and a computer-assisted manual approach using 3 replicates were equally precise. When an identical number of ECGs were compared, the computer-assisted manual method was most precise, while the commercial algorithm was relatively imprecise. Although suitable for clinical assessment the standard commercial algorithm cannot be recommended for regulated safety research.

Ann Noninvasive Electrocardiol 2016;21(1):82–90

QTc interval; electrocardiography; manual; automated; ECG core laboratories

The occurrence of torsades de pointes associated with the use of certain antiarrhythmic drugs has been known since the 1960s^{1,2} but the understanding that nonantiarrhythmic drugs may also be implicated is more recent. After several drugs were withdrawn because of torsades de pointes in the 1990s, efforts to characterize the propensity of new chemical entities to prolong the QT interval before they were administered to patients have

become standard in drug development. The European Committee for Proprietary Medicinal Products³ issued the first regulatory document in 1997. Subsequently, the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) released its E14 document, which defined the need for most new chemical entities to undergo a specialized evaluation conducted in

Address for correspondence: Mori J. Krantz, M.D., Denver Health Medical Center, 777 Bannock St, MC 0960, Denver, CO 80204. Fax: 303-602-3900; E-mail: mori.krantz@dhha.org

© 2015 Wiley Periodicals, Inc.
DOI:10.1111/anec.12277

normal subjects that has come to be known as the thorough QT study (tQT).⁴ Precise measurement of the QT interval, a potential albeit imperfect surrogate for the risk of torsades de pointes, is fundamental to the completion of a conclusive tQT study.

The techniques currently in use for the measurement of rate corrected QT (QTc) intervals in regulatory research can be classified into three broad categories: fully manual, computer-assisted with manual adjudication, and fully automated. The 2005 ICH E14 guidance recommends either fully manual or computer-assisted manual adjudication approaches for clinical trials in which the assessment of electrocardiographic (ECG) safety is an important objective, such as the tQT study.⁴ While concern remains that fully automated methods can yield misleading results in the presence of artifact, flat T waves or irregular rhythm, an increasing number of investigators have reported that the newer improved automatic algorithms can match and even surpass the precision of core lab manual or computer-assisted measurements, particularly in the context of studies conducted in normal volunteers. With the universal adoption of digital ECG technology the potential for fully automated annotation of ECG fiducial points using improved software algorithms to generate more precise measurements and create more efficient and accurate workflow is increasingly being explored.

Recently, several papers have been published comparing the precision of newer automated ECG measurement techniques with that of older methodologies.⁵⁻⁸ Typically, such studies have demonstrated greater precision of the newer methodologies but most of these comparisons have been retrospective, utilizing ECGs from previously analyzed normal volunteer studies. Given this background, the current study was designed to prospectively evaluate two proprietary core lab-based ECG management systems: a fully automated approach (QTinno, New Cardio Inc., Neshanic Station, NJ, USA) and a computer-assisted, manual approach (HeartSignals Social & Scientific Systems, Inc., Silver Spring, MD, USA) involving adjudication of all measurements by a cardiologist. The Surveyor Telemetry Central System (Mortara Instrument, Inc., Milwaukee, WI, USA) was used to collect the raw 12-lead ECG waveforms analyzed in this study; the two

core lab measurements were compared to those generated by the Mortara VERITAS algorithm. We hypothesized that the fully automated and manual core lab approaches would be similarly precise with lower intrasubject QTc variance than the "out-of-the-box" commercially available algorithm.

METHODS

Subjects and Study Design

This was in a single-center, prospective, single-blind, placebo-controlled, randomized, two-period, crossover study conducted in 24 healthy male and female volunteers to assess the effect of moxifloxacin 400 mg on the Fridericia rate-corrected QT (QTcF) interval using a fully automated system (QTinno) and a proprietary computer-assisted, manual approach (HeartSignals). Subsequently, the QTcF values generated by these two approaches were compared to those generated by the VERITAS algorithm incorporated in the Surveyor telemetry system used to collect study ECGs. Treatment periods were separated by a 1-week washout. Subjects had baseline ECGs performed five times over 5 minutes at each of three predose time points (-0.75, -0.5, and -0.25 hour) and then five times at each of 12 time points from 0.5 through 24 hours postdose. The study protocol was approved by the Institution Review Board and all subjects provided written informed consent.

ECG Interval Measurement Approaches

QTinno is a software suite that provides fully automated analysis of cardiac intervals. It uses all 12 leads and complexes in an ECG to produce a single virtual ECG lead. A confidence factor flags problematic ECGs for human review and quality control (QC). Problematic ECGs are either accepted and included for analysis without modification or rejected. In the context of this study QTinno analyzed all five available ECGs at all time points.

HeartSignals is a computer-assisted manual ECG reading platform. A proprietary automated algorithm initially preplaces ECG fiducial points. The overreading cardiologist reviews all tracings, chooses the first three consecutive good quality complexes from the preferred measuring lead V_3 ^{9,10} and adjusts the computer placed fiducial

time points to their optimal position. In the context of this study, the HeartSignals cardiologist initially read three ECGs at each time point with subsequent retrieval of the two remaining ECGs from each time point to complete the ECG triads determined to be suboptimal or incomplete after a predefined QC process.

The VERITAS algorithm used in this study is part of the commercially available Mortara Instrument set incorporated in the ECG acquisition equipment. The process is multilead, using landmarks determined by the earliest QRS onset and the latest T wave offset. The T wave offset itself is set at the point where the local slope magnitude, derived from the available leads, falls below a threshold proportional to the T amplitude. RR values used for determining cycle length specific averages, and for QT/RR modeling, are a weighted average of individual RRs preceding each beat. In the context of this study all ECGs were automatically read by the VERITAS algorithm and subsequently underwent a QC process comparable to the one applied to the HeartSignals ECGs.

Statistical Analysis

The following analyses were performed separately using HeartSignals, QTinno or VERITAS generated QTcF intervals. Change from baseline QTcF (Δ QTcF) was fit by mixed model analysis of variance to evaluate residual standard deviation (model SD or square root mean-square error) and assay sensitivity. This was adequate if the lower bound of at least one 95% one-sided confidence interval (CI) for the mean placebo-corrected Δ QTcF was greater than 5 ms.

The analysis to generate P-values for comparing the intrasubject variance estimates was based on the ANCOVA model for delta-QTcF. Fixed effects were sequence, visit, treatment, method (VERITAS/HeartSignals/QTinno), scheduled time, method by time, and the three-way interaction. There was a baseline covariate and subject sequence was random. The least mean differences on the three-way interaction estimated double-delta QTcF. In addition, intersubject variance and standard deviation were calculated for all three methodologies. For purposes of comparison, P-values < 0.05 were considered statistically significant. SAS Version 9.4 (Cary, NC, USA) was used for all statistical analyses.

RESULTS

Comparison between the Two Core Lab Approaches

A total of 23 subjects completed both periods (Moxifloxacin 400 mg and matching placebo) of the study. Five ECG replicates were acquired at 30 time points resulting in a total of 3450 ECGs (690 ECG pentads) potentially available for analysis. The QTinno QTcF values were based on measurements from 3419 ECGs it found to be suitable among the 3450 collected. The HeartSignals data were based on the analysis of 2028 ECGs (690 optimized ECG triads minus 42 tracings remaining unsuitable or unreadable after review and addition of available back up ECGs). Study sensitivity (lower bound of the 95% CI of Δ QTcF greater than 5 ms) was confirmed by both methods. The maximal mean Δ QTcF and 90% CIs were 14.3 (11.2–17.5) at 1.5 hour and 12.7 (9.5–16.0) ms at 3 hour using the automated and manual methods, respectively (Fig. 1). Assay sensitivity was demonstrated at a greater number of time points via the automated method, although these values persisted beyond the expected decay curve for moxifloxacin. As shown in Table 1, the manual and automated approaches were similarly precise with an intrasubject variance of 44.0 and 41.5ms², respectively (P = 0.60 for the comparison).

The summary of QTcF automated—manual differences suggested that intervals measured by the automated method were typically 8.3–9.8 ms shorter than intervals measured by the manual method. Mean Δ QTcF following moxifloxacin was 2.4 ms longer by the automated method, while Δ QTcF following placebo was 0.8 ms shorter by the automated method compared to interval changes measured by the manual method.

Comparison between the Three Measurement Methods

Initial review of the QTcF values generated by VERITAS from the 2028 ECGs successfully measured by both HeartSignals and QTinno were not highly correlated (Fig. 2, first panel) Measurements showed high intrasubject variance because of a number ECGs with implausibly short QTcF values. Visual inspection of outlier and control tracings as measured by VERITAS was performed with exclusion of measurements adjudicated to be incorrect by 40 ms or more

Figure 1. Change from baseline least square mean differences in QTcF for the two core laboratory approaches. The dashed line (squares) depicts QTinno and the dotted line (triangles) HeartSignals approaches. Asterisks connote demonstration of study sensitivity defined by the lower bound of the 95% CI exceeding 5 ms.

Table 1. Residual Variance of QTcF Change from Baseline (ms): Two Core Laboratory Approaches

Approach	N	Intrasubject Variance	Intrasubject Standard Deviation	Intersubject Variance	Intersubject Standard Deviation
Fully automated (QTinno)	3419	41.5	6.4	74.0	8.6
Manual (HeartSignals)	2028	44.0	6.6	66.5	8.2

P = 0.60 for difference in Intra-subject variance.

(Fig. 2, second panel). This eventually led to the deletion of 116 of 127 ECGs with a QTcF < 370 ms, 3 of 106 randomly chosen ECGs with a normal QTcF between 411 and 414 ms and 2 of 85 tracings with a QTcF > 450 ms resulting in a new "edited" data set with 1907 tracings that had been measured and quality controlled by all three methods. Figure 3 shows examples of accepted and rejected measurements.

The VERITAS QC process produced an edited machine-generated QTcF data set with 1907 values and lower intrasubject variance for the VERITAS data compared with the original 2028

ECG data set. In this edited data set of 1907 ECGs (Table 2), both QTinno and HeartSignals showed intrasubject variance values lower than VERITAS ($P < 0.0001$ for both comparisons). In addition, the intrasubject variance was significantly lower for HeartSignals than for QTinno ($P = 0.02$). Adequacy of the study sensitivity for this edited data set (lower bound of the one-sided 95% CI of Δ QTcF greater than 5 ms) was demonstrated at eight time points with VERITAS (including hours 8 and 10), five time points with QTinno (including hour 10), and three time points with HeartSignals (no point beyond hour 4). The maximal mean

Figure 2. QTcF correlation: HeartSignals vs. VERITAS, unedited and edited. The left panel depicts the relationship between HeartSignals and VERITAS QTcF values before editing and the right panel after excluding outlier measurements that deviated by >40 ms.

Table 2. Residual Variance of QTcF Change from Baseline (ms) using Common ECGs: Three Methods

Method/ Method	N	Intrasubject Variance	Intrasubject Standard Deviation	Intersubject Variance	Intersubject Standard Deviation
VERITAS*	1907	89.4	9.5	132.6	11.5
Fully automated (QTinno)**	1907	52.6	7.3	101.6	10.1
Manual (HeartSignals)	1907	43.8	6.6	65.2	8.1

*Intrasubject variance for both fully automated/QTinno and manual/HeartSignals were significantly lower than VERITAS ($P < 0.0001$ for both comparisons).
**Intrasubject variance with manual/HeartSignals was significantly lower than with fully automated/QTinno method ($P = 0.02$).

Δ QTcF for VERITAS was 17.2 (12.5–21.9) msec at 2 hours post dose, while the maximal Δ QTcF for QTinno was 12.8 (9.2–16.5) ms at 1.5 hours and the maximal Δ QTcF as measured by the HeartSignals method was 12.8 (9.4–16.2) ms, 3 hours postdose (Fig. 4). As was the case for the QTinno measurements, the VERITAS QTcF measurements were shorter than those measured by the computer-assisted manual method while mean Δ QTcF after moxifloxacin was greater.

DISCUSSION

In the last several years, numerous studies assessing different ECG measurement methodologies have been published, sometime yielding conflicting results. Fosser et al.⁵ concluded that the variability of manual and computer-assisted measurements was greater than that associated with automated methods. However, their conclusions were based on a retrospective analysis of two entirely manual

Panel A: reject: VERITAS measurement appears too short

Panel B: reject: VERITAS measurement appears too long

Panel C: accept: VERITAS measurement appears correct

Figure 3. Visual adjudication of VERITAS QTcF Measurements. (A) Reject: VERITAS measurement appears too short. (B) Reject: VERITAS measurement appears too long. (C) Accept: VERITAS measurement appears correct. Vertical gray bars show the QRS onset and T-wave offset, as measured by VERITAS.

paper studies and three on-screen computer-assisted studies compared with automated measurements generated via commercial algorithms. On the other hand, Tyl et al.⁶ using digital ECG data from a thorough QT study in normal volunteers found that measurements generated by two semiautomated methods were more precise than those produced using a fully automated method. Analyzing data from two previously completed

thorough QT studies, Darpo et al.⁷ reported that a high-precision QT measurement technique using 10 ECG replicates per time point with human validation of low confidence complexes yielded more precise measurement when compared with a standard semiautomated method using three replicate ECGs per time point. Finally, using ECGs from a phase 1 multiple ascending-dose study in normal volunteers, Sarapa et al.⁸ concluded

Figure 4. Change from baseline least square mean differences in QTcF for the three measurement methods. The dashed line (squares) depicts QTinno, the dotted line (triangles) HeartSignals and the solid line (circles) VERITAS. Asterisks connote demonstration of study sensitivity defined by the lower bound of the 95% CI exceeding 5 ms.

that the fully automated QTinno ECG analysis system produced measurements with lower QTcF variability than those generated from the same ECGs by two semiautomated techniques.

The present study demonstrates that despite technologic advances in software systems, the precision of fully automated core lab approaches to QTc interval measurement may not be greater than computer-assisted manual methods involving systematic cardiologist adjudication. Unlike many previously published studies⁵⁻⁸ evaluating ECG measurement techniques, our study was prospective and compared not only the precision of different measurement techniques applied to the same set of ECGs but also the global approaches of two core labs selecting and measuring a different number of replicates from the same predetermined time points. In addition, it should be noted that the cardiologist finalized all measurements which were all obtained from lead V₃, typically the lead displaying

the most representative and longest QT interval^{9,10} rather than from the more traditional lead II.⁴ Measurements obtained simultaneously by a commercial algorithm were included for reference and did not appear to provide similar levels of precision as assessed by intrasubject QTcF variance.

The core lab-based automated and computer-assisted manual approaches evaluated in our study used a different number of ECGs to both demonstrate assay sensitivity for a positive control with a high and comparable degree of precision. When a similar number of replicates were compared, the computer-assisted manual measurements were more precise. The second part of this study was designed to evaluate the performance of a commercially available automated ECG algorithm (VERITAS) when measuring QTcF in a set of 2028 good quality ECGs successfully read by two different core lab-based methods. Compared with the values generated by the core

lab systems, the unedited data generated by the commercial algorithm showed significantly greater QTcF variance due predominantly to erroneously short QTc measurements as determined by visual inspection. After completion of a QC process, a second, slightly smaller data set ($n = 1907$) showed improved intrasubject QTcF variance of the VERITAS measured ECGs, which however remained greater than that of either of the core lab-generated values.

Although QTcF values generated from lead V₃ by the cardiologist-adjudicated system were longer than those generated from a composite complex by the automated systems, both machine-generated moxifloxacin effect profiles were more pronounced than the profile based on manual reads, a difference that has been noted with several other automated reading systems, particularly in the descending portion of the moxifloxacin curve.¹¹

A limitation of the current analysis is that our findings apply specifically to moxifloxacin associated QTcF changes as characterized by the VERITAS automated algorithm, the HeartSignals computer-assisted manually adjudicated approach by a single cardiologist and the QTinno system. Thus, our results may not be generalizable to the performance of other ECG measurement systems nor are they applicable to populations with structural heart disease. In addition, technology is evolving such that continuous automated analysis of all complexes recorded during a tQT study may ultimately yield more detailed and reliable information when compared to sparse measurements performed at scheduled time points as is the current practice.¹² By the same token, additional computerized techniques, such as ECG morphology pattern matching, will likely continue to improve the overall quality of QTc data.¹³

CONCLUSIONS

A fully computerized core lab approach based on analyzing 5 ECG replicates per time point and a computer-assisted, fully cardiologist-adjudicated manual approach based on analyzing optimized ECGs triads generated equally precise QTcF measurements. This suggests that although we have become increasingly reliant on "machines" and technology in clinical research, superiority to "man" was not demonstrated in this prospective study. Even after deletion of clearly spurious

measured tracings, automated QTcF values generated by the standard commercially available VERITAS algorithm showed greater intrasubject variance and yielded a suboptimal moxifloxacin effect profile compared with the data generated from the same ECGs by the two core lab methods. In fact, the manual approach exhibited the lowest intrasubject QTcF variance among the methods when assessing identical tracings. Our findings suggest that both core lab-based approaches are equally suitable for drug safety evaluations but that the standard commercial automated algorithm is not. Finally, when evaluating the performance of an ECG analysis system, both the measurement technique and the approach to ECG selection and QC remain an important consideration.

Acknowledgments: We thank ICON Development Solutions for conducting the study in their San Antonio clinical pharmacology unit. We appreciate the expert technical assistance of Kenn White and John Pezzullo.

REFERENCES

1. Selzer A, Wray HW. Quinidine syncope, paroxysmal ventricular fibrillations occurring during treatment of chronic atrial arrhythmias. *Circulation* 1964;30:17-26.
2. Dessertenne F. La tachycardie ventriculaire à deux foyers opposés variables. *Arch Mal Coeur* 1966;59:263-272.
3. Committee for Proprietary Medicinal Products. Points to Consider: The Assessment of the Potential for QT Interval Prolongation by Non-Cardiovascular Medicinal Products. 17 December 1997. London, United Kingdom: European Agency for the Evaluation of Medicinal Products; 1997.
4. International Conference on Harmonization. E14: The clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs. *Fed Regist* 2005;70:61133.
5. Fossier C, Duczynski G, Agin M, et al. Comparison of manual and automated measurements of the QT interval in healthy volunteers: An analysis of five thorough QT studies. *Clin Pharm Ther* 2009;86:503-506.
6. Tyl B, Kabbaj M, Fassi B, et al. Comparison of semiautomated and fully automated methods for qt measurement during a thorough QT/QTc Study: Variability and sample size considerations. *J Clin Pharmacol* 2009;49:905-905.
7. Darpo B, Fossa AA, Couderc JP, et al. Improving the precision of QT measurements. *Cardiol J* 2011;18(4):401-410.
8. Sarapa N, Gussak I, Vajdic B, et al. Comparison of QTinno, a fully automated electrocardiographic analysis program, to semiautomated electrocardiographic analysis methods in a drug safety study in healthy subjects. *J Electrocardiol* 2009;42:358-366.
9. Rautaharju PM, Surawicz B, Gettes LS. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: Part IV: The ST segment, T and U waves and the QT Interval: A scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: Endorsed by the International

- Society for Computerized Electrocardiology. *Circulation* 2009;119:e341–e250.
10. Davey PP. Which lead for Q-T interval measurements? *Cardiology* 2000;94:159–164.
 11. Couderc JC, Garnett C, Li C, et al. Highly automated QT measurement techniques in 7 thorough QT studies implemented under ICH E14 guidelines. *Ann Noninvasive Electrocardiol* 2011;16:13–24.
 12. Meyer O, Greig G, Holzgrefe H. Comparative TQT analysis with three fully-automated platforms: Comparison to core laboratory semi-automated results. *J Electrocardiol* 2013;46:717.e1–717.e9.
 13. Hnatkova K, Smetana P, Toman O, et al. Systematic comparisons of electrocardiographic morphology increase the precision of QT interval measurement. *PACE* 2009;32:119–130.