

Electrocardiographic Data Quality in Thorough QT/QTc Studies

Lars Johannesen · Christine Garnett ·
Marek Malik

Published online: 20 February 2014
© Springer International Publishing Switzerland (outside the USA) 2014

Abstract

Background Most drugs with systemic bioavailability have to undergo a thorough QT (TQT) study, which includes a pharmacologic positive control. A set of QTc-quality tests was recently proposed with the possible aim of removing the need for a positive control.

Objective We evaluated the influence of QT measurement and QTc computation methodology on the proposed QTc-quality tests.

Methods The baseline ECG waveforms and fiducial points were retrieved from 34 crossover TQT studies that had full-day baseline ECGs prior to each study period. The QT measurement methodology and recorder type were retrieved from study reports. The influence of QTc computation methodology was investigated by applying the pattern-matching technique and/or using the 10-s average heart rate.

Results There were no statistical differences in any QTc-quality test values between studies using continuous or 10-s bedside recordings and, in a subset of the quality test, an increase of data quality for semi-automatically read studies compared with those manually read ($p < 0.01$). There was a significant improvement ($p < 0.01$) in all the QTc-quality test values for QTc measurements obtained by using pattern matching with or without 10-s average heart rate in comparison to the original QTc measurements.

Conclusion The findings suggest that QTc quality is mostly driven by the QTc measurement methodology rather than other study-related factors.

Key Points

The previously proposed QTc quality metrics, which quantifies intra- and intersubject consistency in QTc differences, captures factors of QTc quality.

Of the factors evaluated for their influence on QTc quality, QTc measurement methodology is the main driver of QTc quality.

L. Johannesen · C. Garnett
Division of Pharmacometrics, Office of Clinical Pharmacology,
Office of Translational Sciences, Center for Drug Evaluation and
Research, US FDA, Silver Spring, MD, USA

L. Johannesen
Department of Clinical Physiology, Karolinska Institutet and
Karolinska University Hospital, Stockholm, Sweden

L. Johannesen (✉)
US FDA, 10903 New Hampshire Avenue,
Bldg 62, RM 2127E, Silver Spring, MD 20993, USA
e-mail: lars.johannesen@fda.hhs.gov

C. Garnett
Certara, Cary, NC, USA

M. Malik
St. Paul's Cardiac Electrophysiology, University of London,
and Imperial College, London, UK

1 Introduction

Most drugs with systemic bioavailability have to undergo a thorough QT (TQT) study which aims at establishing whether the drug affects cardiac repolarization as assessed by the rate-corrected QT interval (QTc). A QTc prolongation of regulatory interest is defined as the upper one-sided 95 % confidence interval (CI) of mean baseline- and placebo-corrected QTc change exceeding 10 ms at any post-dosing timepoints [1, 2].

Overall quality of QTc data is an important factor of TQT studies. Improvement in QTc quality allows the sample size to be decreased, making the study more economical. It could also make early clinical studies sufficiently powered to satisfy the International Conference on Harmonisation (ICH) E14 requirement [1, 2]. In early clinical studies, such as the first-in-man investigations, the necessity of demonstrating assay sensitivity by introducing a separate positive control arm or period is highly impractical. Rather, different proofs of study quality need to be established. Related to this, a set of QTc-quality tests have recently been proposed with the possible aim of eventually removing the need for a positive control [3]. Among others, the proposed tests only deal with the systematicity of QTc measurements and not the bias, which is of less importance when measuring the placebo-corrected QTc change. However, before these QTc-quality tests can be considered further, an evaluation of factors that impact on their results is needed.

Having this in mind, this investigation used the data from a spectrum of existing TQT studies to evaluate the proposed QTc-quality tests. Also, to propose strategies for systematic improvement of QTc data quality, we investigated the influence of different QTc measurements and computations, which were demonstrated to improve the responses to the pharmacological positive control [4].

2 Methods

2.1 Data

TQT studies with multiple drug-free baselines submitted to the US FDA between January 2006 and December 2012 ($N = 34$) were available for the purposes of this investigation. For each study, the baseline QT and R-R interval measurements at each nominal timepoint were obtained, together with the annotated ECG waveforms from the ECG Warehouse (<http://www.ecgwarehouse.com>).

2.2 Proposed QTc-Quality Tests

A detailed description of the proposed QTc-quality tests has been previously published [3]. Briefly, the concept is to quantify known physiological relationships and to verify that the consistency of baseline-to-baseline or subject-to-subject differences are preserved, providing confidence in the measurements of the QT and RR (R-R interval).

There are three components of the QTc-quality tests, all of which are evaluated using the randomized treatment allocation to avoid period effects. Test 1 evaluates whether or not the study captures physiologically known differences in individual QTc values, i.e. that intersubject variability is

larger than intrasubject variability, with higher intrasubject variability in females compared with males. The test 1 values, or intrasubject variability, are computed by gender. Test 2 quantifies the stability of QT measurements by timepoint across treatment periods. The assumption of test 2 is that the individual baselines tightly reproduce each other. Test 2 is computed as the lower and upper CIs of the time-matched inter-baseline differences. As the CIs might narrow due to increasing sample size, we also included a test for the standard deviation of the time-matched inter-baseline differences. Lastly, test 3 evaluates the stability of intersubject differences across treatment periods; i.e. whether the same QTc difference between different subjects is maintained from baseline to baseline. For test 3, the intersubject differences can either be computed in the averaged QTc values (test 3 – average) or time-matched intersubject differences (test 3 – time-matched). In this analysis, the same software implementation of the tests was used as in their seminal proposal [3] (see the original publication of the tests for further details).

2.3 Protocol Factors

Two protocol factors were considered, potentially influencing the QT measurement quality—QT measurement methodology and recorder type. QT measurement methodology was obtained from the study protocols and defined as either semi-automatic (partial use of a computerized ECG measurement), fully manual, or unknown (if not explicitly stated in study protocol). Recorder type was defined as continuous (e.g. 10-s extractions from a 12-lead Holter) or standard 10-s bedside, if explicitly not stated in the protocol.

2.4 Influence of QTc Computation

To investigate the influence of QT computation, the QT measurements were remeasured by applying the pattern-matching technique as proposed by Hnatkova et al [5]. This technique has been shown to decrease the intrasubject variability [5] and to lower the overall data variability when evaluating the exposure-response of moxifloxacin [4].

In addition, as in the study investigating the moxifloxacin response [4], the impact of using the average heart rate from the 10-s ECG instead of the originally reported heart rate was also investigated. Four different QT and heart-rate combinations were investigated, resulting in different data variants: originally reported QT and heart rate (*data 1*); originally reported QT and 10-s average heart rate (*data 2*); QT interval corrected using pattern matching with originally reported heart rate (*data 3*); and QT interval corrected using pattern matching with 10-s average heart rate (*data 4*).

Table 1 Investigated thorough QT studies and their ECG characteristics

Study	No. of subjects (% women)	No. of periods	No. of timepoints	QT replicates	Recorder type	QT measurement	Sampling frequency (Hz)/resolution (μ V)	No. of ECGs (% measurable)
1	52 (40)	4	12	4	Continuous	Unknown	1,000/3.75	23,660 (97)
2	65 (51)	4	12	3	Continuous	Manual	1,000/3.75	20,300 (98)
3	36 (28)	4	8	3	Standard	Unknown	500/4.88	7,876 (97)
4	51 (53)	4	8	3	Continuous	Semi-automatic	1,000/3.75	10,761 (98)
5	39 (21)	4	11	3	Standard	Semi-automatic	500/4.88	10,333 (99)
6	55 (45)	3	12	3	Continuous	Manual	500/4.88	13,328 (98)
7	56 (34)	4	9	3	Continuous	Semi-automatic	500/1.00	14,728 (99)
8	53 (49)	4	5	3	Standard	Unknown	500/5.00	19,357 (99)
9	56 (16)	4	9	3	Standard	Manual	1,000/2.50	13,440 (99)
10	63 (48)	3	9	3	Continuous	Manual	1,000/2.50	17,849 (99)
11	16 (50)	4	11	3	Standard	Semi-automatic	500/1.00	4,751 (97)
12	48 (21)	4	9	3	Continuous	Manual	1,000/2.50	13,934 (98)
13	59 (53)	4	12	3	Standard	Unknown	1,000/2.50	17,598 (99)
14	59 (46)	4	10	3	Continuous	Unknown	1,000/3.75	13,780 (99)
15	47 (40)	4	15	3	Continuous	Manual	1,000/2.50	17,561 (100)
16	55 (49)	4	13	3	Continuous	Unknown	1,000/3.75	19,058 (98)
17	22 (50)	6	4	3	Standard	Manual	500/5.00	8,505 (98)
18	60 (28)	4	12	3	Continuous	Semi-automatic	1,000/2.50	19,455 (99)
19	79 (20)	4	14	3	Continuous	Semi-automatic	1,000/2.50	25,631 (98)
20	100 (44)	3	13	3	Continuous	Unknown	1,000/2.50	27,528 (98)
21	123 (0)	4	5	9	Standard	Unknown	1,000/2.50	67,007 (94)
22	53 (47)	4	12	4	Continuous	Semi-automatic	1,000/2.50	20,350 (99)
23	51 (49)	4	12	3	Continuous	Semi-automatic	500/4.88	15,972 (99)
24	29 (48)	5	28	3	Continuous	Semi-automatic	500/1.00	23,916 (90)
25	51 (41)	4	13	4	Continuous	Manual	1,000/3.75	21,390 (95)
26	41 (51)	4	9	3	Standard	Semi-automatic	500/4.88	8295 (99)
27	103 (34)	4	11	3	Continuous	Unknown	1,000/2.50	30,797 (96)
28	38 (45)	4	9	3	Continuous	Unknown	1,000/2.50	8,647 (99)
29	52 (48)	4	12	4	Continuous	Semi-automatic	1,000/2.50	19,917 (100)
30	61 (100)	4	10	3	Standard	Manual	1,000/2.50	17,270 (98)
31	57 (46)	4	9	3	Standard	Manual	500/4.88	12,548 (99)
32	38 (45)	4	10	3	Standard	Unknown	500/1.00	20,329 (97)
33	66 (42)	4	7	4	Continuous	Manual	180/6.25	23,846 (95)
34	40 (100)	4	10	12	Continuous	Manual	500/5.00	46,513 (98)

All but four studies used lead II as the primary lead. Of the remaining four, three used a global measurement lead as the primary, and one study used V3

The Fridericia-corrected QTc values [6] were used since no substantial heart rate changes in drug-free recordings were anticipated.

2.5 Statistical Analysis

The difference between the QTc-quality test results in groups based on protocol factors was assessed using the Mann–Whitney test. The differences between the numerical values of each of the QTc-quality tests derived from the four data variants were assessed using the Mann–Whitney

test. p -values <0.05 were considered statistically significant. Statistical evaluations were performed using R version 2.15.3 (R Foundation for Statistical Computing, Vienna, Austria).

3 Results

The 34 investigated studies (Table 1) included a total of 1,874 subjects (773 females, 41 %). The median number of subjects per study was 53 (range 16–123), with a median of

46 % of female subjects (range 0–100 %). The median number of nominal timepoints and treatment periods were 11 (range 4–28) and 4 (range 3–6), respectively. The median number of 10-s ECGs included in each study was 17,724 (range 4,751–67,007). The median percentage of measurable 10-s ECGs was 98 % (range 90–100 %). In 20 studies, the ECGs were sampled at 1,000 Hz, and in the remaining 13 studies at 500 Hz, except for one study in which the ECGs were sampled at 180 Hz. The median amplitude resolution was 2.5 μ V (range 1.0–6.25 μ V).

3.1 Influence of Protocol Factors

Table 2 shows the comparison of the QTc-quality test for groups of studies based on the protocol factors. There was no difference for ECG recorder type, while the QTc-quality test values were significantly lower (i.e. the QTc data quality was better) in a subset of the tests in studies with semi-automatic QT measurement compared with studies using fully-manual QT measurement.

3.2 Influence of QTc Computation

Figure 1 shows the QTc-quality test values obtained from four different data variants. For all QTc-quality tests, the test values were significantly lower (i.e. better data quality) with pattern-matched QT measurements (data 3, 4) compared with original QT measurements (data 1, 2), irrespective of method used to obtain RR intervals (original vs.

derived from 10-s signals). There was also a numerical trend of lower QTc-quality test values when combining the pattern-matched QT measurements with RR intervals from full 10-s signals compared with the pattern-matched QT combined with original RR data (data 4 vs. data 3).

4 Discussion

A statistically significant improvement of all the QTc-quality test values was found when involving QT measurement with pattern matching, with or without heart rate data derived from full 10-s signals (data 3 and 4) compared with original QT and RR measurements (data 1). The combination of improved QT measurement consistency by pattern matching with heart rate data from full 10-s ECGs (data 4) led to further lowering of the test values (compared with data 3), and thus to further improvement in data quality.

The observation of a significant difference between some of the QTc-quality tests between studies using manual and semi-automatic QT reading suggests more consistent QTc data in semi-automatically read studies. Unfortunately, we have no data available on how many automatic ECG readings were manually over-read in individual studies. There was no noticeable difference between studies with continuous recordings compared with the others. The lack of a difference might be caused by procedures of ECG extraction from continuous recordings.

Table 2 QTc-quality test values grouped by protocol factors

Test	Recorder type		QT measurement methodology		
	Continuous (<i>N</i> = 22)	Standard (<i>N</i> = 12)	Manual (<i>N</i> = 12)	Semi-automatic (<i>N</i> = 11)	Unknown (<i>N</i> = 11)
Test 1					
Female (ms)	11.14 (1.54)	10.91 (1.67)	11.55 (1.67)	10.21 (1.34), <i>p</i> = 0.03*	11.41 (1.38)
Male (ms)	10.50 (1.57)	9.84 (0.99)	10.53 (1.01)	9.50 (1.46)	10.82 (1.45)
Test 2					
Lower CI (ms)	−5.17 (1.14)	−5.44 (1.57)	−5.50 (1.16)	−5.16 (1.53)	−5.11 (1.26)
Upper CI (ms)	5.08 (1.06)	5.75 (1.64)	5.23 (1.45)	5.47 (1.32)	5.25 (1.26)
SD (ms)	11.92 (1.43)	11.86 (1.20)	12.21 (1.43)	11.16 (1.17)	12.30 (1.19)
Test 3 – average					
Proportion < −10 (%)	13.86 (2.84)	14.51 (3.46)	15.33 (2.86)	11.30 (1.45), <i>p</i> < 0.01*	15.52 (2.59)
Proportion >10 (%)	13.58 (3.38)	15.43 (3.50)	15.28 (3.83)	11.66 (2.58), <i>p</i> = 0.02*	15.66 (2.55)
Test 3 – time-matched					
Proportion < −10 (%)	22.75 (2.62)	22.12 (2.08)	23.46 (2.37)	20.52 (1.55), <i>p</i> < 0.01*	23.52 (2.08)
Proportion <10 (%)	22.50 (3.10)	22.90 (2.33)	23.19 (3.13)	20.98 (2.58)	23.69 (2.07)

The mean (SD) was computed for each QTc test (see the text for details) and comparisons between types of recording methodology (continuous vs. standard) and QT measurement (manual vs. semi-automatic) were performed. The comparisons between studies that used manual and semi-automatic QT readings are shown where they reached statistical significance (*)

CI confidence interval, SD standard deviation

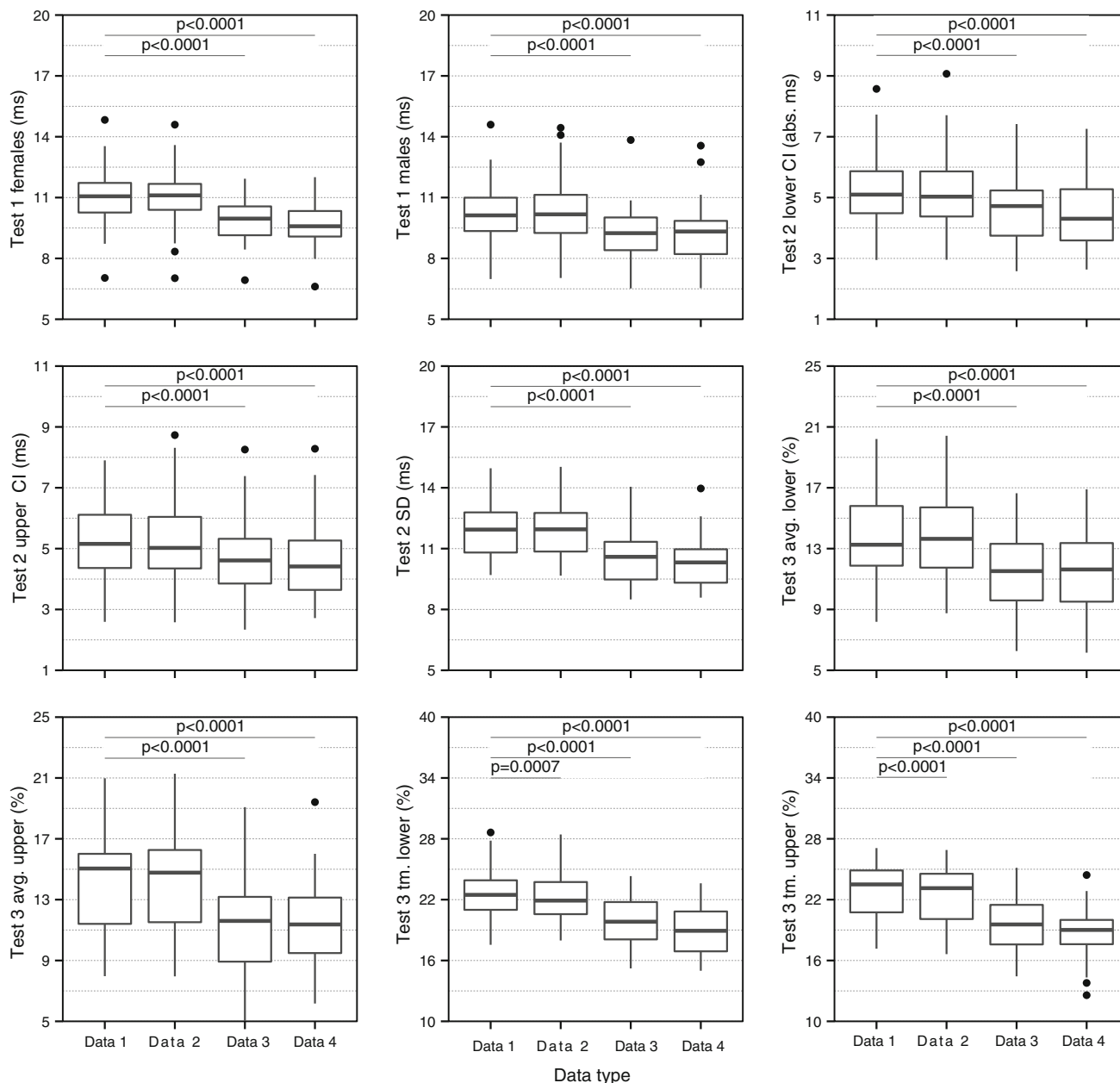


Fig. 1 Boxplot for each of the QTc-quality tests grouped by data variants of QT [original (*data 1, 2*) or pattern-matching (*data 3, 4*)] and RR (R-R interval) [original (*data 1, 3*) or 10 s average (*data 2, 4*)] interval measurements. Comparisons between each of the QTc-quality tests by data variant were done using the *t*-test with the *p*-values shown (where no *p*-value is shown, the difference was not statistically significant). *Lower* and *upper* for test 3 refer to the proportion of differences below and above -10 and 10 ms, respectively. *abs* absolute values, *avg* average, *tm* time-matched, *CI* confidence interval, *SD* standard deviation

In some studies, a fixed timepoint relative to the nominal timepoint was used for the extraction (i.e. the extractions from continuous Holter recordings paid no attention to the quality of the signal).

The observation that pattern-matching adjustment of QT measurement and the use of 10-s average RR interval values improve the data quality is important from a practical point of view. It is also consistent with the observation that these techniques increase the precision of other

characteristics of TQT measurements [4]. The implementation of these techniques can thus be universally recommended for future TQT studies, as well as other investigations that require accurate and systematic QT and RR measurements.

It has previously been reported that extracting 10-s ECG segments from continuous recordings at stable heart rates with low noise results in more robust QTc data [7]. Our findings do not confirm this observation. Possibly,

however, different studies might have used different techniques for extracting the 10-s ECGs from the continuous signals, not necessarily following the previously published assumptions [7]. It should also be noted that in some cases, the ECG recording methodology was not explicitly mentioned, and standard 10-s bedside recordings were assumed. However, it is unlikely this influenced the results because if the application of continuous recording had led to increased quality, and if some of the Holter studies had been incorrectly classified as standard, the quality metric for continuous would be trending to lower values compared with standard, which it did not.

Our observations related to the pattern-matching technique are consistent with the original report by Hnatkova et al. [5]. Using the technology, we found a similar reduction in intrasubject QTc variability, which was also found by Meyer et al. [8]. The finding is also consistent with our observation that the pattern-matching technique of QT measurement improves the response to moxifloxacin-based positive control [4]. The improvement in QTc quality by using 10-s average heart rate instead of the heart rate derived from the preceding RR interval is consistent with the reduction in QTc intrasubject variability achieved by QT/RR hysteresis correction [9]. It is likely that by applying proper hysteresis correction, i.e. using longer heart rate history, will lead to further reduction of intrasubject QTc variability. Not surprisingly, and consistent with previous reports, we also found intrasubject QTc variability larger in females compared with males.

4.1 Limitations

The relatively small number of studies available for this analysis makes subgroup comparisons difficult (e.g. continuous vs. standard recordings). There are several aspects that we were unable to analyse, including the influence of different computer algorithms used in semi-automatic measurement, ECG extraction procedures, and the differences among individual readers involved in the processing of ECGs in the same study. These details are not provided by pharmaceutical sponsors submitting a study for regulatory approval. Because of the small numbers, we were also unable to evaluate any differences between studies using single-lead and global QT measurements. It is also likely that the number of ECG replicates could influence the results [10], but as the majority of studies that were included used three replicates (27 of 34 studies), we could not explore the influence of the number of replicates on the QTc quality metrics. Finally, the large spread of data also prevented us from evaluating the influence of the duration of the period between study

baselines. Nevertheless, these limitations are unlikely to have biased our main findings.

5 Conclusion

From a practical perspective, the study offers two principal conclusions. First, the previously proposed tests of assessing QTc data quality provide valid characterization of the study data in a variety of conditions. Second, the QTc data quality can be systematically improved by applying adjustment based on the pattern-matching procedures. Similarly, the use of RR interval values derived from longer sections of continuous recordings (most likely longer than the 10-s data used in this study [9]) could substantially improve the data quality in investigations that require accurate QTc measurements.

Acknowledgments The opinions presented here are those of the authors and no official support or endorsement by the FDA is intended or should be inferred. Marek Malik is an Honorary ORISE Research Fellow of the FDA. Lars Johannesen, Christine Garnett, and Marek Malik have no conflicts of interest directly related to this article.

The authors are grateful to Dr. Norman Stockbridge of the FDA for his substantial guidance on the topic and his help with preparing this manuscript. This project was supported in part by an appointment to the Research Participation Program at the Center for Drug Evaluation and Research administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the FDA.

References

1. ICH. Guidance for Industry E14 clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs. 2005.
2. ICH. Guidance for industry E14 clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs. Questions and answers; 2008.
3. Malik M, Zhang J, Johannesen L, Hnatkova K, Garnett C. Assessing electrocardiographic data quality and possible replacement of pharmacological positive control in thorough QT/QTc studies by investigations of drug-free QTc stability. *Heart Rhythm*. 2011;8(11):1777–85.
4. Johannesen L, Garnett C, Malik M. Impact of electrocardiographic data quality on moxifloxacin response in thorough QT/QTc studies. *Drug Saf*. 2014. doi:10.1007/s40264-014-0142-2
5. Hnatkova K, Smetana P, Toman O, Bauer A, Schmidt G, Malik M. Systematic comparisons of electrocardiographic morphology increase the precision of QT interval measurement. *Pacing Clin Electrophysiol*. 2009;32(1):119–30.
6. Fridericia LS. Die Systolendauer im Elektrokardiogramm bei normalen Menschen und bei Herzkranken. *Acta Med Scand*. 1920;54(1):170.
7. Badilini F, Vaglio M, Sarapa N. Automatic extraction of ECG strips from continuous 12-lead Holter recordings for QT analysis at prescheduled versus optimized time points. *Ann Noninvasive Electrocardiol*. 2009;14(Suppl 1):S22–9.
8. Meyer O, Ferber G, Greig G, Holzgrefe HH. Pattern recognition analysis of digital ECGs: decreased QT measurement error and

- improved precision compared to semi-automated methods. *J Electrocardiol.* 2013;46(2):118–25.
9. Malik M, Hnatkova K, Novotny T, Schmidt G. Subject-specific profiles of QT/RR hysteresis. *Am J Physiol Heart Circ Physiol.* 2008;295(6):H2356–63.
 10. Natekar M, Hingorani P, Gupta P, Karnad DR, Kothari S, de Vries M, et al. Effect of number of replicate electrocardiograms recorded at each time point in a thorough QT study on sample size and study cost. *J Clin Pharmacol.* 2011;51(6):908–14.