Comparison of automated measurements of electrocardiographic intervals and durations by computer-based algorithms of digital electrocardiographs

Paul Kligfield, MD,[a] Fabio Badilini, PhD,[b] Ian Rowlandson, MS,[c] Joel Xue, PhD,[c] Elaine Clark, MA,[d] Brian Devine, MSc,[d] Peter Macfarlane, DSc,[d] Johan de Bie, PhD,[e] David Mortara, PhD,[e] Saeed Babaeizadeh, PhD,[f] Richard Gregg, MS,[f] Eric D. Helfenbein, MS,[f] and Cynthia L. Green, PhD[g] *New York, NY; Milwaukee, WI; Glasgow, United Kingdom; Andover, MA; and Durham, NC*

Background and Purpose Automated measurements of electrocardiographic (ECG) intervals are widely used by clinicians for individual patient diagnosis and by investigators in population studies. We examined whether clinically significant systematic differences exist in ECG intervals measured by current generation digital electrocardiographs from different manufacturers and whether differences, if present, are dependent on the degree of abnormality of the selected ECGs.

Methods Measurements of RR interval, PR interval, QRS duration, and QT interval were made blindly by 4 major manufacturers of digital electrocardiographs used in the United States from 600 XML files of ECG tracings stored in the US FDA ECG warehouse and released for the purpose of this study by the Cardiac Safety Research Consortium. Included were 3 groups based on expected QT interval and degree of repolarization abnormality, comprising 200 ECGs each from (1) placebo or baseline study period in normal subjects during thorough QT studies, (2) peak moxifloxacin effect in otherwise normal subjects during thorough QT studies, and (3) patients with genotyped variants of congenital long QT syndrome (LQTS).

Results Differences of means between manufacturers were generally small in the normal and moxifloxacin subjects, but in the LQTS patients, differences of means ranged from 2.0 to 14.0 ms for QRS duration and from 0.8 to 18.1 ms for the QT interval. Mean absolute differences between algorithms were similar for QRS duration and QT intervals in the normal and in the moxifloxacin subjects (mean ≤6 ms) but were significantly larger in patients with LQTS.

Conclusions Small but statistically significant group differences in mean interval and duration measurements and means of individual absolute differences exist among automated algorithms of widely used, current generation digital electrocardiographs. Measurement differences, including QRS duration and the QT interval, are greatest for the most abnormal ECGs. (Am Heart J 2014;167:150-159.e1.)

Most electrocardiograms (ECGs) in the United States are performed with digital electrocardiographs that are capable of simultaneous 12-lead signal acquisition and provide computer-based analysis of ECG waveforms, including measurement of the RR interval, the PR interval, the QRS duration, and the QT interval. Particular interest has focused on the QT interval as a marker for potential heterogeneity of repolarization[1-3] because prolongation of the QT has prognostic implications in clinical practice and in epidemiological studies as well as regulatory implications for drug development.[4-8] Advances in accuracy and widespread availability of computerized ECG interpretation have led to increasing reliance on automated measurement of global ECG intervals, including the QT interval, as a routine alternative to manual measurement of intervals from single ECG leads.[9-12] However, there is no universally accepted medical definition of the QT interval, and there are numerous methods for determination of the end of the T wave.[13-15] As a result, measurement of the QT interval (and other diagnostic ECG intervals) has become a proprietary

From the [a]Division of Cardiology, Weill Cornell Medical College, New York, NY, [b]AMPS-LLC, New York, NY, [c]GE Healthcare, Milwaukee, WI, [d]Glasgow Program, University of Glasgow, Glasgow, United Kingdom, [e]Mortara Instrument, Milwaukee, WI, [f]Philips Healthcare, Andover, MA, and [g]Duke Clinical Research Institute, Duke University Medical Center, Durham, NC.

Submitted October 10, 2013; accepted October 10, 2013.

Reprint requests: Paul Kligfield, MD, Department of Medicine, Division of Cardiology, Weill Cornell Medical College, 1300 York Ave, New York, NY 10065.
E-mail: pkligfi@med.cornell.edu

0002-8703/$ - see front matter

http://dx.doi.org/10.1016/j.ahj.2013.10.004

engineering solution of individual manufacturers of electrocardiographs.[9] These algorithms evolve with hardware and software innovations within and between manufacturers, often with dramatic differences in resulting measurements, so the direct comparability of measurements is not assured when clinicians and investigators use different generations of electrocardiographs within studies or within individual patients.[16] Differences in automatic interval measurements based on electrocardiograph selection would have important consequences in practice and in research. This study was designed to test whether clinically significant systematic differences exist between different automated computer-based algorithms for the measurement of ECG intervals in widely used, current generation digital electrocardiographs and whether differences between measurements by different electrocardiographs increase with increasing abnormality of the underlying ECGs.

Methods

Four major manufacturers of digital interpretive electrocardiographs that are widely used in the United States were invited to participate in an analysis of automated, computer-based measurements of ECG intervals and durations. Engineers from GE Healthcare (Milwaukee, WI), the Glasgow Program (Glasgow, UK, used in Burdick and other electrocardiographs), Mortara Instrument (Milwaukee, WI), and Philips Healthcare (Andover, MA) agreed to the conditions of the study and to publication of the findings. It was proposed that 600 XML waveforms would be assembled from ECGs stored in the US FDA ECG warehouse under auspices of the Cardiac Safety Research Consortium (CSRC), which approved the study design and released the waveforms for this purpose.[17]

Electrocardiograms were randomly selected from tracings collected from clinically normal volunteers participating in thorough QT (TQT) studies submitted to the US FDA during the course of drug development and from patients with genotyped long QT syndrome (LQTS).[18] Three distinct ECG groups were constructed based on the expected QT durations, including (*a*) 200 ECGs from subjects at baseline or during the placebo period of TQT trials (group, normal), comprising the most normal QT expected; (*b*) 200 ECGs from subjects during peak moxifloxacin effect of TQT trials (group, moxi), not matched to the subjects used for the normal QT group (moxifloxacin is used in TQT studies as an active control drug that is known to have modest QT prolonging effects on the ECG); and (*c*) 200 patients with genotype positive LQTS from within the CSRC database (group, LQTS) and expected to have the most abnormal QT measures. Equal numbers of men and women were sought within each QT group, with all tracings required to be simultaneous 12-lead recordings digitized at 500 samples per second. Because a number of the ECGs in the congenital long QT data set were originally digitized at lower sampling rates, unequal numbers of men and women were included in the present group to maintain the higher 500 sample per second standard throughout the study population.

This study was designed only to establish whether important systematic differences exist between measurements obtained with automated electrocardiographs from different manufacturers that are widely used in clinical practice and for clinical investigation. Investigators and participants agreed in advance that outcomes would not be presented in terms of better or worse or as more or less accurate. Accordingly, no gold measurement standard for ECG intervals was used in this evaluation, which focuses only on relative and systematic differences between methods. It was agreed by all participants that blinded automated ECG analysis would be performed to assure that all reported measurements received no manual adjudication. To accomplish this, the 600 randomly ordered and de-identified ECGs were processed simultaneously by automated algorithms on laptop computers of the participants at a single group meeting during the April 2012 annual sessions of the International Society for Computerized Electrocardiology, under the direct supervision of the study authors. To prepare for this session, each manufacturer had previously been provided with 2 sets of sample XML ECGs similar to but not identical with the final study blinded tracings to assure that the study waveforms could be analyzed by all participants. In addition, a study output file for storage of the blinded measurements was developed in cooperation with the participants, and its usability by each manufacturer and its ability to subsequently be analyzed by the nonindustry study investigators were confirmed. A brief description of the methods used by each participant for measurement of global ECG intervals is contained in the online Appendix.

Study participants were aware of the nature of the population groups, but none of the tracings used for the primary blinded analysis had been previously examined by the manufacturers. One of us (C.L.G.) assembled the data set of 600 anonymized ECGs in random order with unique identifiers, which was given to the participants only at the time of blinded analysis for measurements that were incorporated into the standardized output files and immediately submitted for central analysis (C.L.G. at the Duke Clinical Research Institute [DCRI]) for the purpose of the study. At DCRI, measurements were identified by sex and by QT group for each of the participating algorithms. Accordingly, no modifications of algorithm-based intervals or durations were possible by blinded study design, and all data represent intrinsic ECG measurements used routinely by the participating manufacturers with no human adjudication.

For each standard digitized 12-lead ECG, each manufacturer analyzed and provided measurements of average 10-second cardiac cycle length (RR interval), atrioventricular conduction time (PR interval), intraventricular conduction time (QRS duration), and the total duration of depolarization and repolarization from the onset of the QRS complex to the end of the T wave (QT interval). QT intervals were not corrected for heart rate because the same tracings were used by all participants. Global measurements rather than single-lead measurements of intervals and durations are used by each of the automated algorithms of the 4 manufacturers, but the individual algorithms may differ in technical implementation, as further defined and discussed below.[9] Findings were re-identified and assembled at the DCRI for analysis according to manufacturer, QT group, sex, and individual interval measurements. The PR interval for 3 ECG tracings could not be analyzed by all manufacturers; the PR interval for each of these tracings was excluded from all analyses.

The total population was separated by sex and also by normal, moxifloxacin, and LQTS groups for analysis. Differences between groups according to measurement algorithm were examined as differences between means, presented in

Table I. Mean intervals, by sex and algorithm

Interval	Sex	n	Algorithm	Mean ± SD (ms)	Lower 95% CI (ms)	Upper 95% CI (ms)	Minimum (ms)	Maximum (ms)
RR*	Men	280	GE	1015 ± 165	995	1034	495	1621
		280	Glasgow	1014 ± 165	995	1034	494	1614
		280	Mortara	1009 ± 164	990	1029	493	1614
		280	Philips	1013 ± 165	993	1032	492	1620
	Women	320	GE	933 ± 169	914	951	431	1764
		320	Glasgow	933 ± 169	915	952	473	1767
		320	Mortara	928 ± 167	909	946	487	1767
		320	Philips	931 ± 167	913	950	488	1764
PR†	Men	279	GE	157 ± 24	154	160	112	392
		279	Glasgow	154 ± 21	152	157	104	242
		279	Mortara	160 ± 41	155	165	104	465
		279	Philips	156 ± 21	153	158	104	240
	Women	318	GE	153 ± 19	151	155	94	214
		318	Glasgow	151 ± 20	148	153	82	210
		318	Mortara	153 ± 26	151	156	106	455
		318	Philips	152 ± 20	150	154	88	212
QRS‡	Men	280	GE	91 ± 11	90	93	56	120
		280	Glasgow	94 ± 11	92	95	62	154
		280	Mortara	97 ± 11	96	98	66	141
		280	Philips	96 ± 10	95	97	59	151
	Women	320	GE	82 ± 9	81	83	56	108
		320	Glasgow	86 ± 8	85	87	62	122
		320	Mortara	89 ± 9	88	90	59	118
		320	Philips	91 ± 12	90	93	65	166
QT§	Men	280	GE	421 ± 43	416	426	284	560
		280	Glasgow	425 ± 44	420	430	286	578
		280	Mortara	417 ± 42	412	422	269	545
		280	Philips	426 ± 44	421	431	286	568
	Women	320	GE	429 ± 48	424	434	310	628
		320	Glasgow	432 ± 48	426	437	304	650
		320	Mortara	420 ± 43	415	424	176	564
		320	Philips	430 ± 50	425	436	310	639

* $P < .05$ by Bonferroni-corrected repeated-measures analysis of variance for all comparisons of RR within sex except GE versus Glasgow and Glasgow versus Philips for men and GE versus Glasgow and GE versus Philips for women.
† $P < .05$ for all comparisons of PR within sex except GE versus Mortara, GE versus Phillips, and Mortara versus Philips for men and GE versus Mortara and Mortara versus Philips for women.
‡ $P < .05$ for all paired comparisons of QRS duration within sex.
§ $P < .05$ for all paired comparisons of QT within sex except Glasgow versus Philips for men.

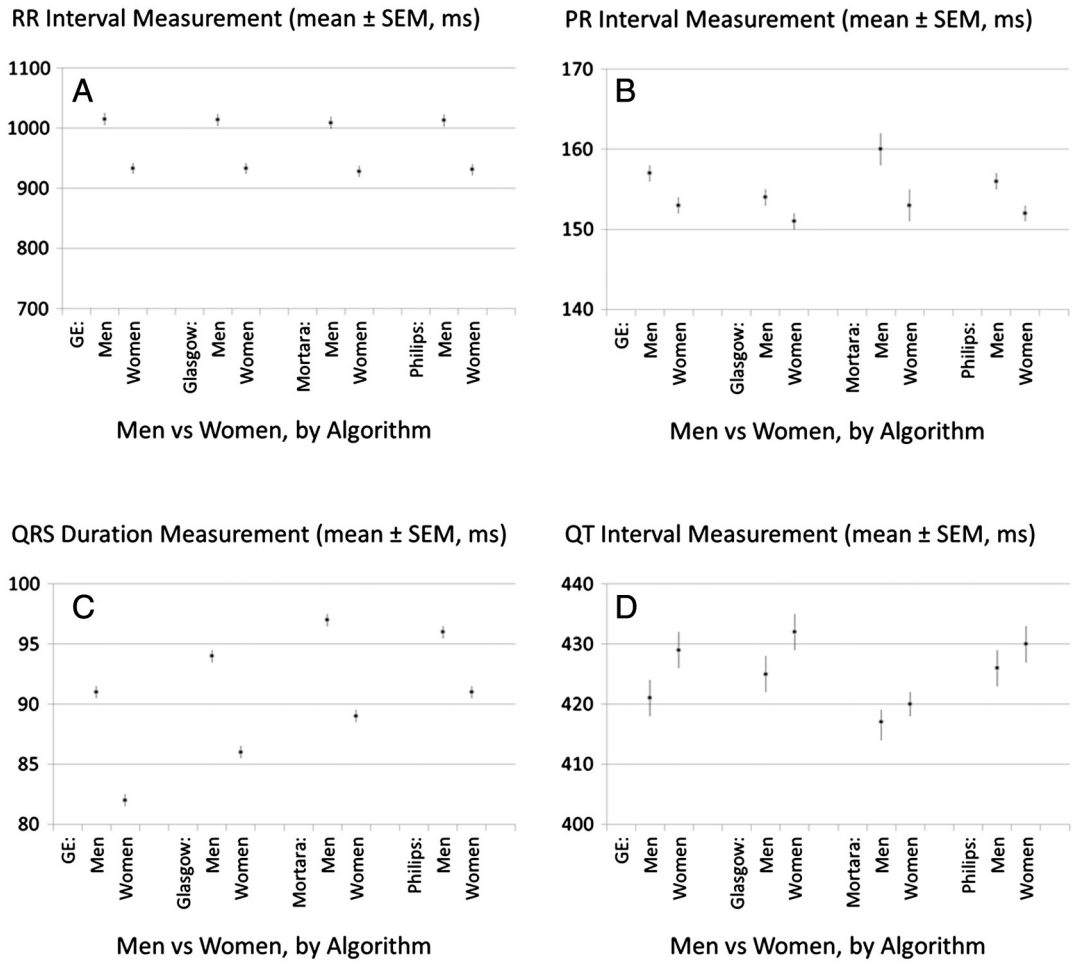
the text and tables with the SD as the index of dispersion; in the figures, SEM is used for clarity. Group differences between measurements according to sex were assessed using the unpaired *t* test. Differences in mean measurements between manufacturers and differences within each manufacturer by study group were examined using repeated-measures analysis of variance, with the Bonferroni adjustment for multiple comparisons. In addition to differences of means, we also examined the means of absolute differences among measurements by algorithm for individual ECGs. These were calculated by averaging all 6 of the absolute differences present between pairs of algorithm measurements for each single ECG as a pooled single value. Analysis of variance with Bonferroni adjustment was used to compare the resulting pooled absolute differences of intervals and durations. The magnitudes of pooled absolute differences in measurements for all comparisons are illustrated by boxplots as well as by means.

Results

Mean intervals by algorithm, according to sex, are listed in Table I and shown in Figure 1. Of note, each algorithm detected highly significant differences between men and women for RR intervals and QRS interval durations ($P < .001$ for each comparison within algorithm method). It is apparent from Figure 1 that mean RR intervals are systematically shorter (ie, rates are faster) in women than in men, as are QRS durations within each of the automated methods. Strong trends or significant differences only at the $P < .05$ level were found within algorithm methods for shorter PR intervals in women than in men and for longer QT intervals in women than in men in the total study population. The higher variability for QT prevented statistically significant differences from being detected.

Statistically significant differences between algorithms were present for many of the mean values for RR, QT, PR, and QRS interval durations, separately in men and women in the total population, as shown in Table I. However, mean differences between algorithms for automated interval measurements in the total population of men and women, not separated by study group, were small: 0.4 to 5.6 ms range for RR interval, 0.3 to 5.7 ms range for

Figure 1



Automated interval measurements (mean ± SEM) by sex, according to manufacturer, for RR interval (**A**), PR interval (**B**), QRS duration (**C**), and uncorrected QT interval (**D**). All algorithms detect significant differences between measurements in men and women, whereas small systematic differences in mean values are apparent for some comparisons (see Table I).

PR interval, 1.1 to 9.4 ms range for QRS duration, and 1.0 to 12.2 ms range for uncorrected QT interval.

Differences between algorithms for measurements according to normal baseline, moxifloxacin administration, and congenital long QT groups of combined men and women in the population are shown in Table II and Figure 2. For each algorithm, there were no within-methods significant differences among groups for automated RR interval measurements or for PR interval measurements or between normal and moxifloxacin groups for QRS durations. In contrast is the progressive, highly statistically significant increase in QT interval measurements from the normal to the moxifloxacin group and from the moxifloxacin to the congenital long QT group (and also from the normal to the long QT group), comparably detected by each of the measurement algorithms (Figure 2D). For several of the algorithms, QRS duration was significantly reduced in the congenital long QT group (Figure 2C).

Differences between mean values for study groups according to algorithm and statistical significance of differences by Bonferroni adjustment of repeated-measures analysis of variance are shown in Table III. Even where differences reach statistical significance, the absolute magnitudes of differences within groups by the tested algorithms were in general clinically small. Mean differences for RR interval were ≤6.5 ms for any comparison, whereas mean differences for PR and for QRS duration were ≤5.2 ms in the normal and moxifloxacin groups. Mean differences between algorithms were somewhat larger for measurements of QRS duration and QT interval in the congenital long QT group, ranging from 1.8 to 14.0 ms for QRS duration and 0.8 to 18.1 ms for QT interval comparisons (Figure 2).

Table II. Mean intervals, by clinical QT group and algorithm (statistical significance of individual comparisons is shown in Table III)

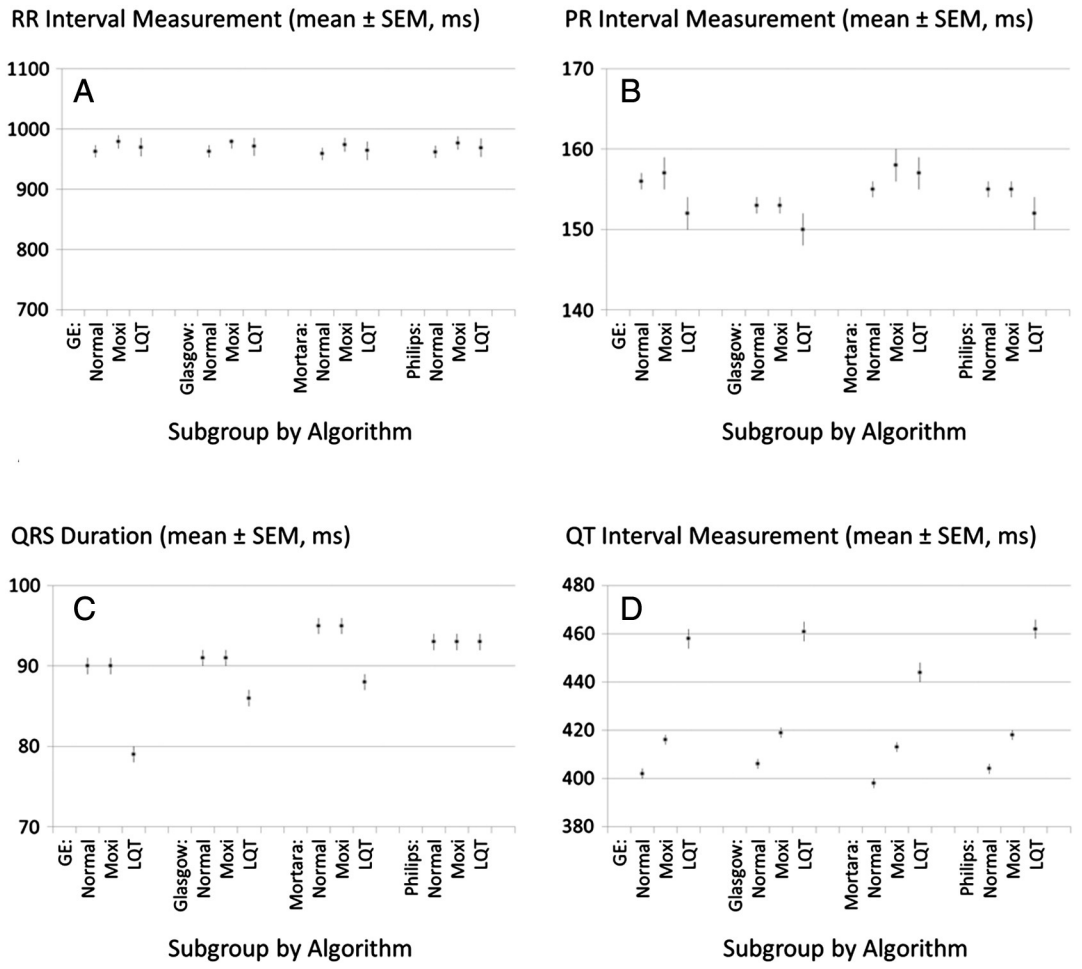
Interval	Group	n	Algorithm	Mean ± SD (ms)	Lower 95% CI (ms)	Upper 95% CI (ms)	Minimum (ms)	Maximum (ms)
RR	Normal	200	GE	963 ± 145	943	984	606	1463
		200	Glasgow	963 ± 146	943	984	587	1451
		200	Mortara	959 ± 144	939	979	600	1451
		200	Philips	962 ± 146	941	982	588	1464
	Moxi	200	GE	979 ± 151	958	1000	652	1463
		200	Glasgow	979 ± 151	958	1000	651	1452
		200	Mortara	974 ± 151	953	995	653	1380
		200	Philips	977 ± 150	956	998	652	1464
	LQTS	200	GE	970 ± 212	941	1000	431	1764
		200	Glasgow	971 ± 211	941	1000	473	1767
		200	Mortara	964 ± 209	935	993	487	1767
		200	Philips	969 ± 210	940	998	488	1764
PR	Normal	199	GE	156 ± 17	153	158	112	226
		199	Glasgow	153 ± 19	151	156	104	226
		199	Mortara	155 ± 18	152	157	109	224
		199	Philips	155 ± 18	152	157	107	224
	Moxi	200	GE	157 ± 25	154	160	114	392
		200	Glasgow	153 ± 19	150	156	108	212
		200	Mortara	158 ± 35	153	163	110	455
		200	Philips	155 ± 18	152	158	108	212
	LQTS	198	GE	152 ± 22	149	158	94	204
		198	Glasgow	150 ± 24	147	155	82	242
		198	Mortara	157 ± 44	150	163	104	465
		198	Philips	152 ± 24	149	155	88	240
QRS	Normal	200	GE	90 ± 10	88	91	70	118
		200	Glasgow	91 ± 10	90	92	68	116
		200	Mortara	95 ± 9	93	96	74	121
		200	Philips	93 ± 9	92	95	71	123
	Moxi	200	GE	90 ± 10	89	91	68	120
		200	Glasgow	91 ± 10	90	92	68	122
		200	Mortara	95 ± 10	94	96	74	141
		200	Philips	93 ± 10	92	95	67	133
	LQTS	200	GE	79 ± 10	78	81	56	106
		200	Glasgow	86 ± 11	85	88	62	154
		200	Mortara	88 ± 11	87	90	59	130
		200	Philips	93 ± 15	91	95	59	166
QT	Normal	200	GE	402 ± 26	398	405	326	482
		200	Glasgow	406 ± 27	402	409	322	490
		200	Mortara	398 ± 26	394	402	319	489
		200	Philips	404 ± 26	401	400	335	490
	Moxi	200	GE	416 ± 27	412	419	364	496
		200	Glasgow	419 ± 28	415	423	360	502
		200	Mortara	413 ± 28	409	417	357	504
		200	Philips	418 ± 28	415	422	360	407
	LQTS	200	GE	458 ± 56	451	466	284	628
		200	Glasgow	461 ± 56	454	469	286	650
		200	Mortara	444 ± 54	437	452	176	564
		200	Philips	462 ± 58	454	470	286	639

Mean values of differences, as contrasted with differences of means, are shown for pooled absolute measurement differences in single recordings for each interval in Figure 3, which also includes boxplot illustrations of medians, 25% to 75% percentile ranges, and outliers. Expressed as pooled absolute mean difference, these were similar for QRS duration and QT intervals in the normal and in the moxifloxacin subjects (≤6 ms), but variation among algorithms in patients with LQTS was larger for QRS duration (8.9 ms, $P < .001$) and for measured QT interval (14.1 ms, $P < .001$).

Discussion

Automated measurements of intervals and durations by widely used digital electrocardiographs demonstrate statistically significant small differences of mean values by method and generally small mean absolute differences for individual comparisons. Differences were least in normal subjects and greatest in patients with abnormal ECGs, as represented by our LQTS group. All fully automated algorithms clearly detected differences in measurements between men and women and also

Figure 2



Automated interval measurements (mean ± SEM) in normal, moxifloxacin, and congenital long QT groups, according to manufacturer, for RR interval (**A**), PR interval (**B**), QRS duration (**C**), and uncorrected QT interval (**D**). All algorithms detect significant differences in QT by group. Small systematic differences in mean values are apparent for some of the other comparisons (see Table III).

detected progressively increasing QT measurements in normal subjects, in subjects taking moxifloxacin and in patients with LQTS.[5,18-20]

Our study population was selected primarily to compare potential differences in QT interval measurements among different automated ECG algorithms without manual adjudication by human overreaders. Related comparisons of other ECG intervals and durations are also of interest. Our normal subjects are generally young volunteers from baseline or placebo arms of TQT studies in the ECG warehouse administered by CSRC[2,17,18,21]; they are often selected for absence of ECG abnormalities, including high normal QT intervals, and, therefore, should not be considered representative of normal populations of older adults, in whom ECG intervals may differ. Moxifloxacin is known to produce small increases in the QT interval and is used commonly as an active control substance to document sensitivity of ECG analysis in TQT studies submitted to the US FDA during drug development[18,21]; these volunteer subjects are also not representative of the general population. A population of genotyped patients with documented congenital LQTS were also selected from the CSRC database to represent clinically important abnormalities of repolarization for comparison of measurement performance between algorithms.[18]

In current digital electrocardiographs, interval measurements are routinely made as "global" rather than single-lead findings because simultaneous 12-lead recording allows total durations that are independent of isoelectric intervals in single leads.[14] As one implementation, global QT interval can be defined as the earliest onset of a QRS complex in any lead to the latest end of the T wave in any lead. By definition, global measurements

Table III. Differences between algorithms for each interval measurement using Bonferroni-adjusted repeated-measures analysis of variance for comparisons in Table II

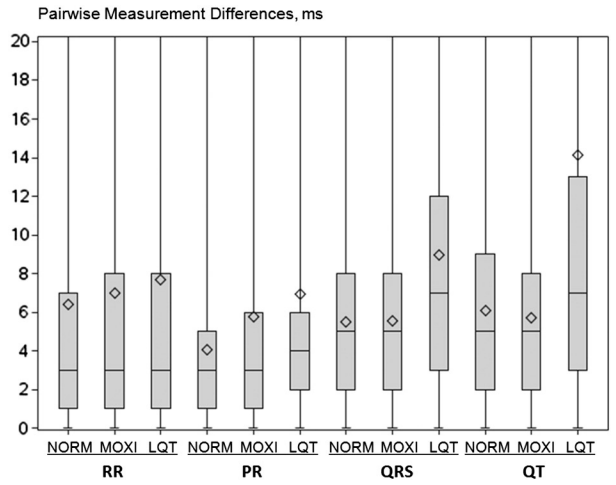
Interval	Group	Algorithm	Versus algorithm	Difference of means (ms)	P
RR	Normal	GE	Glasgow	0.1	NS
		GE	Mortara	4.1	.001
		GE	Philips	1.5	NS
		Glasgow	Mortara	4.2	<.001
		Glasgow	Philips	1.7	.043
		Mortara	Philips	−2.6	NS
	Moxi	GE	Glasgow	0.3	NS
		GE	Mortara	5.3	<.001
		GE	Philips	1.9	<.035
		Glasgow	Mortara	5.0	<.001
		Glasgow	Philips	1.6	NS
		Mortara	Philips	−3.4	<.011
	LQTS	GE	Glasgow	−0.5	NS
		GE	Mortara	6.0	<.001
		GE	Philips	1.4	NS
		Glasgow	Mortara	6.5	<.001
		Glasgow	Philips	1.9	.046
		Mortara	Philips	−4.6	.031
PR	Normal	GE	Glasgow	2.3	<.001
		GE	Mortara	0.9	NS
		GE	Philips	1.0	.010
		Glasgow	Mortara	−1.4	.018
		Glasgow	Philips	−1.3	.018
		Mortara	Philips	0.1	NS
	Moxi	GE	Glasgow	4.0	.022
		GE	Mortara	−0.8	NS
		GE	Philips	2.1	NS
		Glasgow	Mortara	−4.8	NS
		Glasgow	Philips	−1.9	<.001
		Mortara	Philips	2.9	NS
	LQTS	GE	Glasgow	2.1	<.001
		GE	Mortara	−4.4	NS
		GE	Philips	0.2	NS
		Glasgow	Mortara	−6.6	NS
		Glasgow	Philips	−1.9	<.001
		Mortara	Philips	4.7	NS
QRS	Normal	GE	Glasgow	−1.1	.015
		GE	Mortara	−4.8	<.001
		GE	Philips	−3.7	<.001
		Glasgow	Mortara	−3.7	<.001
		Glasgow	Philips	−2.6	<.001
		Mortara	Philips	1.1	.044
	Moxi	GE	Glasgow	−1.1	.027
		GE	Mortara	−5.2	<.001
		GE	Philips	−3.4	<.001
		Glasgow	Mortara	−4.0	<.001
		Glasgow	Philips	−2.3	<.001
		Mortara	Philips	1.8	<.001
	LQTS	GE	Glasgow	−7.0	<.001
		GE	Mortara	−8.8	<.001
		GE	Philips	−14.0	<.001
		Glasgow	Mortara	−1.8	.002
		Glasgow	Philips	−7.1	<.001
		Mortara	Philips	−5.2	<.001
QT	Normal	GE	Glasgow	−3.9	<.001
		GE	Mortara	3.6	<.001
		GE	Philips	−2.5	<.001
		Glasgow	Mortara	7.5	<.001
		Glasgow	Philips	1.1	.004
		Mortara	Philips	−6.1	<.001
	Moxi	GE	Glasgow	−3.3	<.001
		GE	Mortara	2.8	<.001
		GE	Philips	−2.8	<.001
		Glasgow	Mortara	6.1	<.001
		Glasgow	Philips	0.5	NS
		Mortara	Philips	−5.6	<.001
	LQTS	GE	Glasgow	−2.8	<.001
		GE	Mortara	14.4	<.001
		GE	Philips	−3.7	<.001
		Glasgow	Mortara	17.2	<.001
		Glasgow	Philips	−0.8	NS
		Mortara	Philips	−18.1	<.001

Abbreviation: *NS*, nonsignificant.

tend to be systematically longer than their single-lead counterparts because initial and terminal isoelectric waveforms are present in many single-lead measurements.[14] Recent American Heart Association/American College of Cardiology recommendations endorse the use of global rather than single-lead interventions for routine electrocardiography.[9] Changes in measurement algorithms over the past 10 years have significantly increased automated values found for QT intervals in normal subjects and in patients with disease,[16] an observation that has in part prompted the present study. Similarly, measured global QRS durations from current digital electrocardiographs are generally longer than measurements previously obtained from single leads because these include the earliest onset of the QRS and latest offset of QRS in any of the leads. This implies that normal values may have some dependence on algorithm methodology of different electrocardiographs.

Each of the automated algorithms detected established differences between men and women and progressive prolongation of QT across the normal, moxifloxacin, and congenital long QT groups.[17,22,23] Because of the relatively large numbers in our groups, many statistically significant differences of means exist within groups for the ECG algorithms examined in this study. These within-group differences are generally of small to modest clinical significance. Even so, our findings suggest value in standardizing or adjusting for measurement methodology within or between studies, particularly for epidemiological investigations that seek to establish normal limits for ECG intervals within populations and also for studies that seek to reproduce or to compare ECG interval limits between populations. Differences in measurements between the different automated computer-based algorithms are most apparent in our patients with congenital LQTS, who in turn have the most markedly abnormal

Figure 3



Pairwise absolute measurement differences (milliseconds), calculated from pooled values of all individual differences within a single ECG, by group, shown as mean (diamond) and as box and whiskers plots for medians and 25% to 75% percentile range (horizontal lines within bars). Absolute differences in QRS duration and QT interval are largest in the congenital long QT group. Abbreviations: *NORM*, normal; *MOXI*, moxifloxacin; *LQT*, LQTS group.

ECGs. The QT interval has important diagnostic and prognostic value,[4,5,7,24,25] but even here, differences among algorithms are relatively modest and less than had been hypothesized: the largest difference of means of 18 ms between methods of QT interval measurement is only approximately 4% of the mean QT value. On the other hand, technical variability should not confound biologic variability in pharmacology; a mean QT prolongation >10 ms is generally considered to indicate potential arrhythmogenicity of a drug during development or during post–market surveillance, so consistency of measurement methodology can be important in the serial evaluation of ECG intervals.

QRS duration also has diagnostic and prognostic value in a wide range of clinical populations.[26-34] Although differences between algorithms are generally small in our normal subjects, these might have some consequences for ranges of normal limits and for serial comparison studies. Differences of group mean values of up to 14 ms for QRS measurement are noted for the different algorithms in our patients with congenital LQTS, a difference of approximately 16% of the underlying mean value. The QRS measurements in patients with congenital LQTS are lower in comparison with QRS durations in the other groups for several of the algorithms examined in this study. Our data provide no direct explanation of this finding, but it should be noted that our long QT subjects had a higher prevalence of women and also an admixture of children and adolescents, both factors that might tend to lower QRS duration.[22,23,35] Definition of the end of the QRS complex as it merges with the ST segment can be difficult, so the effect of individual algorithms on QRS findings requires further study that is beyond the scope of the present report.

With respect to single individual ECGs in a given subject or patient, the mean absolute difference in measurements for all methods is <7 ms for subjects in the normal and moxifloxacin groups. Mean absolute differences among algorithms are larger for QT interval and QRS duration in the LQTS patients, as seen in Figure 3. This again indicates that the automated algorithms tested are most concordant in normal subjects and least consistent in the presence of important abnormalities of the underlying waveforms. Whether waveform measurement differences among algorithms are also less consistent in the presence of other types of depolarization and repolarization abnormality, such as those associated with ischemia and infarction, remains to be examined.

It was a considered decision of all investigators not to attempt to establish criterion standard reference values for the ECG intervals that are the bases of these comparisons, for several reasons. The purpose of this study was to establish whether important systematic differences exist between measurements obtained with automated electrocardiographs from different manufacturers that are widely used in clinical practice and for clinical investigation and not to examine the relation of measurements obtained with these algorithms to a separately defined criterion standard or to claim superior accuracy for one algorithm or another. The small differences reported here serve to help investigators and clinicians put a variety of individual and population ECG findings in perspective, with an understanding that specific ECG methodology has the potential to systematically affect test outcomes, regardless of the relative accuracy of individual algorithms.

Two other arguments were involved in the decision not to use a human criterion standard in this study. It has not been usual for major medical equipment manufacturers to cooperate in blinded comparisons of proprietary equipment when there is a perception that a possible competitive disadvantage might result from ranking of outcome differences according to purported accuracy. Absence of relative performance by reference to criterion standards, therefore, enhanced the likelihood of cooperation among the present participants and perhaps was required to implement the study. However, separately and in addition, human adjudication of ECG interval measurement is itself variable. Measured or adjudicated values by humans can differ according to the experience and habits of the electrocardiographer,[12,15,36-38] making it difficult to decide how criterion standard intervals should be defined and measured to establish absolute reference standards that would be satisfactory for this kind of comparison study. The absence of clear medical

definition of some of the intervals examined in this study has left their computer-based measurement to proprietary engineering solutions of individual algorithms, and it would be difficult to prove precisely where one solution might or might not be better than another.

Disclosures

Paul Kligfield, MD; None, Fabio Badilini, PhD; Employee of AMPS, Ian Rowlandson, MS; Employee of GE Healthcare, Joel Xue, PhD; Employee of GE Healthcare, Elaine Clark, MA; None, Brian Devine, MSc; None, Peter Macfarlane, DSc; None, Johan de Bie, PhD; Employee of Mortara Instrument, David Mortara, PhD; Employee of Mortara Instrument, Saeed Babaeizadeh, PhD; Employee of Philips Healthcare, Richard E. Gregg, MS; Employee of Philips Healthcare, Eric D. Helfenbein, MS; Employee of Philips Healthcare, Cynthia L. Green, PhD; None.

References

1. Roden DM. Drug-induced prolongation of the QT interval. N Engl J Med 2004;350:1013-22.
2. Shah RR. Drugs, QT interval prolongation and ICH e14: the need to get it right. Drug Saf 2005;28:115-25.
3. Elming H, Brendorp B, Kober L, et al. QTc interval in the assessment of cardiac risk. Card Electrophysiol Rev 2002;6:289-94.
4. Robbins J, Nelson JC, Rautaharju PM, et al. The association between the length of the QT interval and mortality in the Cardiovascular Health Study. Am J Med 2003;115:689-94.
5. Dekker JM, Crow RS, Hannan PJ, et al. Heart rate-corrected QT interval prolongation predicts risk of coronary heart disease in black and white middle-aged men and women: the ARIC study. J Am Coll Cardiol 2004;43:565-71.
6. Zhang Y, Post WS, Blasco-Colmenares E, et al. Electrocardiographic QT interval and mortality: a meta-analysis. Epidemiology 2011;22: 660-70.
7. Zhang Y, Post WS, Dalal D, et al. QT-interval duration and mortality rate: results from the Third National Health and Nutrition Examination Survey. Arch Intern Med 2011;171:1727-33.
8. de Bruyne MC, Hoes AW, Kors JA, et al. Prolonged QT interval predicts cardiac and all-cause mortality in the elderly. The Rotterdam Study. Eur Heart J 1999;20:278-84.
9. Kligfield P, Gettes LS, Bailey JJ, et al. Recommendations for the standardization and interpretation of the electrocardiogram: part I: the electrocardiogram and its technology: 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 2007;115:1306-24.
10. Garg A, Lehmann MH. Prolonged QT interval diagnosis suppression by a widely used computerized ECG analysis system. Circ Arrhythm Electrophysiol 2013;6:76-83.
11. Tyl B, Azzam S, Blanco N, et al. Improvement and limitation of the reliability of automated QT measurement by recent algorithms. J Electrocardiol 2011;44:320-5.
12. Hnatkova K, Gang Y, Batchvarov VN, et al. Precision of QT interval measurement by advanced electrocardiographic equipment. Pacing and Clinical Electrophysiology: PACE 2006;29:1277-84.
13. Kautzner J. QT interval measurements. Card Electrophysiol Rev 2002; 6:273-7.
14. Kligfield P, Tyl B, Maarek M, et al. Magnitude, mechanism, and reproducibility of QT interval differences between superimposed global and individual lead ecg complexes. Ann Noninvasive Electrocardiol 2007;12:145-52.
15. Azie NE, Adams G, Darpo B, et al. Comparing methods of measurement for detecting drug-induced changes in the QT interval: implications for thoroughly conducted ECG studies. Ann Noninvasive Electrocardiol 2004;9:166-74.
16. Kligfield P, Hancock EW, Helfenbein ED, et al. Relation of QT interval measurements to evolving automated algorithms from different manufacturers of electrocardiographs. Am J Cardiol 2006;98:88-92.
17. Kligfield P, Green CL, Mortara J, et al. The Cardiac Safety Research Consortium electrocardiogram warehouse: thorough QT database specifications and principles of use for algorithm development and testing. Am Heart J 2010;160:1023-8.
18. Kligfield P, Green CL. The cardiac safety research consortium ECG database. J Electrocardiol 2012;45:690-2.
19. Williams ES, Thomas KL, Broderick S, et al. Race and gender variation in the QT interval and its association with mortality in patients with coronary artery disease: results from the Duke Databank for Cardiovascular Disease (DDCD). Am Heart J 2012;164:434-41.
20. Couderc JP, Garnett C, Li M, et al. Highly automated QT measurement techniques in 7 thorough QT studies implemented under ICH E14 guidelines. Annals of noninvasive electrocardiology: the official journal of the International Society for Holter and Noninvasive Electrocardiology, Inc. 2011;16:13-24.
21. Green CL, Kligfield P, George S, et al. Detection of QT prolongation using a novel electrocardiographic analysis algorithm applying intelligent automation: prospective blinded evaluation using the Cardiac Safety Research Consortium electrocardiographic database. Am Heart J 2012;163:365-71.
22. Macfarlane PW, McLaughlin SC, Devine B, et al. Effects of age, sex, and race on ECG interval measurements. J Electrocardiol 1994; 27(Suppl):14-9.
23. Mandic S, Fonda H, Dewey F, et al. Effect of gender on computerized electrocardiogram measurements in college athletes. Phys Sportsmed 2010;38:156-64.
24. Shah RR. The significance of QT interval in drug development. Br J Clin Pharmacol 2002;54:188-202.
25. Drew BJ, Califf RM, Funk M, et al. Practice standards for electrocardiographic monitoring in hospital settings: an American Heart Association scientific statement from the Councils on Cardiovascular Nursing, Clinical Cardiology, and Cardiovascular Disease in the Young: endorsed by the International Society of Computerized Electrocardiology and the American Association of Critical-Care Nurses. Circulation 2004;110:2721-46.
26. Oikarinen L, Nieminen MS, Viitasalo M, et al. QRS duration and QT interval predict mortality in hypertensive patients with left ventricular hypertrophy: the Losartan Intervention for Endpoint Reduction in Hypertension Study. Hypertension 2004;43:1029-34.
27. Brenyo A, Zareba W. Prognostic significance of QRS duration and morphology. Cardiol J 2011;18:8-17.
28. Schinkel AF, Elhendy A, van Domburg RT, et al. Prognostic significance of QRS duration in patients with suspected coronary artery disease referred for noninvasive evaluation of myocardial ischemia. Am J Cardiol 2009;104:1490-3.
29. Petrina M, Goodman SG, Eagle KA. The 12-lead electrocardiogram as a predictive tool of mortality after acute myocardial infarction: current status in an era of revascularization and reperfusion. Am Heart J 2006;152:11-8.

30. Jimenez-Candil J, Cruz Gonzalez I, Martin F, et al. Relationship between QRS duration and prognosis in non–ST-segment elevation acute coronary syndrome. Int J Cardiol 2008;126:196-203.
31. Kashani A, Barold SS. Significance of QRS complex duration in patients with heart failure. J Am Coll Cardiol 2005;46: 2183-92.
32. Gold MR, Thebault C, Linde C, et al. Effect of QRS duration and morphology on cardiac resynchronization therapy outcomes in mild heart failure: results from the Resynchronization Reverses Remodeling in Systolic Left Ventricular Dysfunction (REVERSE) study. Circulation 2012;126:822-9.
33. Greve AM, Gerdts E, Boman K, et al. Impact of QRS duration and morphology on the risk of sudden cardiac death in asymptomatic patients with aortic stenosis: the SEAS (Simvastatin and Ezetimibe in Aortic Stenosis) Study. J Am Coll Cardiol 2012;59:1142-9.
34. Morin DP, Oikarinen L, Viitasalo M, et al. QRS duration predicts sudden cardiac death in hypertensive patients undergoing intensive medical therapy: the LIFE study. Eur Heart J 2009;30:2908-14.
35. Dickinson DF. The normal ECG in childhood and adolescence. Heart 2005;91:1626-30.
36. Darpo B, Agin M, Kazierad DJ, et al. Man versus machine: is there an optimal method for QT measurements in thorough QT studies? J Clin Pharmacol 2006;46:598-612.
37. Hnatkova K, Malik M. Automatic adjustment of manually measured QT intervals in digital electrocardiograms improves precision of electrocardiographic drug studies. IEEE Computers in Cardiology 2002;2002:697-700.
38. Malik M. Errors and misconceptions in ECG measurement used for the detection of drug induced QT interval prolongation. J Electrocardiol 2004;37(Suppl):25-33.

Appendix. Summary of global measurement methods for determination of ECG intervals according to manufacturer and algorithm

GE: GE Healthcare ECG waveform onset/offset determination

In the GE Healthcare 12SL ECG Analysis Program, all intervals and measurements are made from the median complex. The median complex is the representative 12-lead complex formed by time aligning all beats of the dominant morphology and using a proprietary nonlinear type of signal averaging. After the median complex is formed, the onsets and offsets are determined in the following order: QRS onset, QRS offset, T offset, P onset, and P offset. Immediately after the T offset is determined, the median complex is searched for a synchronous P wave. The P onset and offset are determined only if a P wave is found. The exact method for identifying each onset and offset is tuned for each of the markers, but all use variations of the same approach. The fundamental detection function for each marker search is a "superlead," which is the sum of the absolute value of all independent leads (I, II, V_1, ...V_6). In some cases, the first or second derivatives of the superlead are calculated, and in other cases, the derivatives are calculated first and then summed to form the superlead. Such detection functions accentuate the slope changes that accompany a wave onset or offset. After the onset and offset points are found, the intervals are calculated from the time differences between the appropriate markers.

Glasgow: Fiducial point recognition in the Glasgow program

Based on the availability of an average beat, different approaches to finding fiducial points have been tried, including a simple form of threshold crossing to a more complex template matching technique. Ultimately, a combination of these approaches has been adopted where, for example, QRS onset was found to perform best with respect to a noisy test set using a threshold technique. On the other hand, T end performed best using a template matching method. All QRST amplitudes are referred to QRS onset, as are P wave measurements. Individual QRS and T wave fiducial points are derived for all leads, and a method of selecting the earliest QRS onset for example is used to determine a global QRS onset. A similar approach is adopted for QRS termination, and the difference between the 2 global measurements is taken as the overall QRS duration. It was found optimum to use a common P onset and P termination in view of the unreliability of P-wave detection in many ECGs.

Mortara: Landmark detection in Mortara ECG 12-lead ECG analysis

All ECG landmarks, P onset/offset, QRS onset/offset, and T offset, are global, with a single index spanning all leads for each landmark. The detection of these landmarks is generally done using a spatial velocity magnitude, defined as the absolute differences of neighboring samples, summed over the available leads. The first step in landmark detection is the formation of a representative cardiac cycle from the cycles labeled as part of the dominant rhythm. Premature beats, even with QRS morphologies similar to the dominant rhythm, are excluded to avoid influencing P wave and repolarization details. The representative cycle is referred to as a median, although the actual process is a median of 3 averages, with the 3 averages found from the modulo 3 normal beat cycles. The representative cycle is recursively low pass filtered until the high-frequency noise is brought below a threshold, with the aim of robust landmark detection in the presence of noise. P-wave landmark detection first requires locating the peak spatial magnitude of a high pass filter applied to the T-P segment. Onset and offset are determined by fitting straight lines to 16-ms linear segments and locating the boundaries where the straight line fit improves (decreases) below a threshold. This straight line model allows P onset/offset to be properly located even when the P is superimposed on the terminal part of a T wave. QRS landmarks use a similar straight line fit to refine the details of onset/offset. Initially, spatial velocities are used to crudely locate estimates of the onset and offset. The straight line tests again work well in cases of steeply sloped PR/ST segments. T wave offset detection poses special problems because there is no precise end of repolarization. To avoid too early/late offset marking in cases of low/high amplitude T waves, the offset slope threshold is scaled to the amplitude of the largest T wave in any lead.

Philips: Global QT measurement method by DXL algorithm

The Philips DXL algorithm makes measurements on an averaged representative beat. QRS onset and T-wave end are measured on each lead separately. QRS onset is determined from peaks and zero crossings of the first and second derivatives. To find T end for each lead, a line is drawn from the last significant T-wave peak to the next QRS onset. If the line is too steep or too flat, the slope is adjusted to the maximum or minimum allowed slope. T end is taken to be the waveform sample farthest from the line in the vertical dimension. Global QRS onset is determined from the 25th percentile earliest QRS onset across all 12 leads. Global T end is equal to the 50th percentile latest T-wave end across all 12 leads. Only leads with T-wave amplitude >100 μV are used for the global T end determination. The global QT interval is calculated from the difference of global QRS onset and global T-wave end.