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