Available online at www.sciencedirect.com

ScienceDirect

Journal of Electrocardiology 47 (2014) 288–293

JOURNAL OF Electrocardiology

www.jecgonline.com

Assessment of reproducibility – automated and digital caliper ECG measurement in the Framingham Heart Study☆,☆☆

Gordon M. Burke, MD,[a] Na Wang, MA,[b] Sue Blease, MPH,[c] Daniel Levy, MD,[c,d] Jared W. Magnani, MD, MSc[c,e,*]

[a] Department of Medicine, Boston University Medical Center, Boston, MA, USA
[b] Data Coordinating Center, Boston University School of Public Health, Boston, MA, USA
[c] National Heart, Lung, and Blood Institute's and Boston University's Framingham Heart Study, Framingham, MA, USA
[d] Center for Population Studies and the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA
[e] Department of Medicine, Section of Cardiovascular Medicine, Boston University, Boston, MA, USA

Abstract

Background: Digitized electrocardiography permits the rapid, automated quantification of electrocardiograms (ECGs) for analysis. Community- and population-based studies have increasingly integrated such data. Assessing the reproducibility of automated ECG measures with manual measures is a critical step in preparation for using automated measures for research purposes. We recently established an ECG repository of digitally recorded ECGs for the Framingham Heart Study and we sought to assess the reproducibility of automated and manual measures.

Methods: We selected 185 digitally recorded ECGs from routine visits of Framingham Heart Study participants spanning from 1986 to 2012. We selected the following ECG measures for their relevance to clinical and epidemiologic research: P wave duration, P wave amplitude, and PR interval in lead II; QRS duration and R wave amplitude in lead V6; and QT interval in lead V5. We obtained automated values for each waveform, and used a digital caliper for manual measurements. Digital caliper measurements were repeated in a subset (n = 81) of the samples for intrarater assessment.

Results: We calculated the intraclass correlation coefficient (ICC) values for the interrater and intrarater assessments. P wave duration had the lowest interrater ICC (r = 0.46) and lowest intrarater ICC (r = 0.57). R wave amplitude had the highest interrater and intrarater ICC (r = 0.98) indicating excellent reproducibility. The remaining measures had interrater and intrarater ICCs of $r \geq 0.81$.

Conclusions: The interrater reproducibility findings for P wave amplitude, PR interval, QT interval, QRS duration, and R wave amplitude were excellent. In contrast, the reproducibility of P wave duration was more modest. These findings indicate high reproducibility of most automated and manual ECG measurements.

Keywords: Electrocardiogram; Reproducibility; P waves; QRS complex; QT interval

Introduction

Community- and population-based studies have increasingly integrated automated, computer-based analysis ECG quantification. Such analysis permits efficiently developing an ECG database comprised of a large body of data with readily accessible and reproducible measures. Establishing reproducibility between manual and automated measures is essential prior to integrating automated measures. We consequently sought to determine the intrarater and interrater reproducibility of manual and automated ECG measurements of specific waveforms in the Framingham Heart Study ECG repository.

Methods

Participants

The Framingham Heart Study is a community-based study that was initiated in 1948 to identify incident cardiovascular disease and its risk factors [1]. There has been prospective expansion of the Framingham Heart Study with

☆ Funding: Dr. Magnani is supported by American Heart Association Award 09FTF219028. This work was supported by NIH grants R21HL1060926, R01-NS17950, and NIH contract N01-HC25195.

☆☆ Disclosures: None.

* Corresponding author. 88 E. Newton Street, D-7, Division of Cardiology, Boston, MA 02118, USA.

E-mail address: jmagnani@bu.edu

0022-0736/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jelectrocard.2014.01.004

subsequent enrollment of the Offspring Cohort in 1971, the Third Generation Cohort in 2002, and the multiracial Omni cohorts in 1994 and 2002 [1,2]. Participants have ECGs as part of every Framingham Heart Study examination. In 1985, the Framingham Heart Study adopted a digital ECG recording system. Digitally recorded ECGs from 1986 to the present have been converted for contemporary analysis with the MUSE 8 ECG Management System (General Electric, Fairfield, CT), forming a repository of digitally recorded ECGs extending from 1986 to present [1].

In the current analysis, we sampled Framingham Heart Study ECGs spanning from 1986 to 2012. We randomly selected 50 ECGs from each of the following periods: 1986 to 1990, 1991 to 2000, and 2001 to 2010. We then randomly selected an additional 35 ECGs from 2011 to 2012. This approach limited overrepresentation of any single time period to account for temporal changes in ECG acquisition and recording techniques. ECGs were excluded if they had a paced rhythm, atrial fibrillation, or upon review had a technically inadequate tracing. The sample was not intended to be representative of the Framingham Heart Study or the ECG repository.

The digitally recorded ECGs were recorded at either 250 or 500 samples per second with a filter of 150Hz. They were printed on standard ECG paper at 25 mm/s and 0.1 mV/mm, followed by transformation for contemporary analysis by the MUSE 8 ECG Management System (General Electric, Fairfield, CT) [1]. P wave duration, P wave amplitude, PR interval, QT interval, QRS duration, and R wave amplitude were selected for study in specific leads because of their clinical significance. P wave duration, P wave amplitude and PR interval were measured in lead II because these waveforms in lead II can be used in evaluating for left atrial enlargement, right atrial enlargement, and sinus rhythm, respectively [3]. QT interval was measured in V5 because this is one of the recommended leads for determining QT prolongation [4]. QRS duration was measured in V6 because the QRS complex in this lead can be used to recognize certain bundle branch morphologies [5]. R wave amplitude was measured in V6 because certain methods of evaluating for left ventricular hypertrophy involve the R wave amplitude in lead V6 [5]. R wave amplitude in lead V6 has been measured manually as part of the standardized Framingham Heart Study examination. Prior studies from the Framingham Heart Study evaluating left ventricular hypertrophy have used R wave amplitude in V6 as a method of determining left ventricular hypertrophy [6,7]. The Boston University School of Medicine Institutional Review Board approved each Framingham Heart Study examination and all participants provided written, informed consent.

Measurement protocol

A single individual (GMB) used digital calipers to make manual measurements by manipulating a computer mouse. Images were maximally enlarged as allowed by the Muse 8 Management System (General Electric, Fairfield, CT) [8]. The digital caliper measurements were performed on the first complete waveform. Only waveforms in sinus rhythm were included. Incompletely recorded beats and premature ventricular beats were excluded and the next complete sinus waveform in sequential order was measured. P wave duration was measured in lead II. Measurement was conducted from the onset of the P wave, defined as the initial deflection from the isoelectric baseline of the TP segment, to the offset of the P wave, defined as the return of the P wave to the isoelectric baseline of the PR interval. P wave amplitude was measured in the same lead II waveform from the onset of the P wave to its highest amplitude. PR interval measurements were performed in the same lead II waveform. The PR interval was measured from the onset of the P wave to the onset of the QRS complex, defined as the initial deflection from the baseline of the PR interval. The QT interval was measured in lead V5, and determined as the onset of the QRS complex to the end of the T wave, defined as the return of the T wave to the isoelectric baseline of the TP segment. QRS duration was measured in lead V6, and determined as the onset of the QRS complex to the return of the complex to the isoelectric baseline of the ST segment. R wave amplitude was measured in lead V6, and quantified from the onset of the QRS complex to the highest vertical point of the R wave.

Blinded, repeated measures were obtained on different days for assessment of intrarater reproducibility. Intrarater measurements were performed on the same waveforms as those measured initially. The intrarater assessment included 20 ECGs from 1986 to 1990, 1991 to 2000, and 2001 to 2010, and 21 from 2011 to 2012 (total n = 81) from the original interrater assessment. Intrarater ECGs were selected randomly from the initial pool of tracings.

The automated measures were recorded by the MUSE 8 ECG Management System (General Electric, Fairfield, CT). The MUSE 8 ECG Management System program provides median, lead specific measures from digitally recorded ECGs.

Definitions

The definition of hypertension for this study was systolic blood pressure $\geq$140 mmHg, diastolic blood pressure $\geq$90 mmHg, or current use of antihypertensive medications for treatment of high blood pressure. Diabetes was defined as the use of oral hypoglycemic agents, insulin, or fasting blood glucose $\geq$126 mg/dL. Cardiovascular disease was defined as the presence of coronary heart disease, stroke, peripheral vascular disease, and/or congestive heart failure [9].

Statistical analysis

Means, standard deviations and descriptive statistics of continuous variables and the distributions of categorical variables were performed for the sample cohort of this study. ICCs were used to quantify both interrater and intrarater assessments of the six ECG measures. The interrater assessment calculated ICCs comparing the automated measurements with digital caliper measurements of the 185 selected ECGs. The intrarater assessment calculated ICCs comparing the digital caliper measurements performed on 81 ECGs from the interrater assessment to repeated digital