Published in final edited form as:
Biomed Signal Process Control. 2014 September 1; 13: 23–30. doi:10.1016/j.bspc.2014.03.009.

Statistical evaluation of reproducibility of automated ECG measurements: an example from arrhythmogenic right ventricular dysplasia/cardiomyopathy clinic

Timothy Huang, BS[1], **Cynthia A. James, PhD**[2], **Crystal Tichnell, MGC**[2], **Brittney Murray, MS**[2], **Joel Xue, PhD**[3], **Hugh Calkins, MD**[2], and **Larisa G. Tereshchenko, MD, PhD**[2]

[1]Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA

[2]The Division of Cardiology, Department of Medicine, Johns Hopkins Hospital, Baltimore, MD, USA

[3]GE Healthcare, Wauwatosa, WI, USA

Abstract

Background—Arrhythmogenic right ventricular dysplasia/cardiomyopathy (ARVD/C) is characterized by delay in depolarization of the right ventricle, detected by prolonged terminal activation duration (TAD) in V1–V3. However, manual ECG measurements have shown moderate-to-low intra- and inter-reader agreement. The goal of this study was to assess reproducibility of automated ECG measurements in the right precordial leads.

Methods—Pairs of ECGs recorded in the same day from Johns Hopkins ARVD/C Registry participants [n=247, mean age 35.2±15.6 y, 58% men, 92% whites, 11(4.5%) with definite ARVD/C] were retrospectively analyzed. QRS duration, intrinsicoid deflection, TAD, and T-wave amplitude in the right precordial leads, as well as averaged across all leads QRS duration, QRS axis, T axis, QTc interval, and heart rate was measured automatically, using 12SL TM algorithm (GE Healthcare, Wauwatosa, WI, USA). Intrinsicoid deflection was measured as the time from QRS complex onset to the alignment point of the QRS complex. TAD was calculated as the difference between QRS duration and intrinsicoid in V1, V2, V3. Reproducibility was quantified by Bland-Altman analysis (bias with 95% limits of agreement), Lin's concordance coefficient, and Bradley-Blackwood procedure.

Results—Bland-Altman analysis revealed satisfactory reproducibility of tested parameters. V1 QRS duration bias was −0.10ms [95% limits of agreement −12.77 to 12.56ms], V2 QRS duration bias −0.09ms [−11.13 to 10.96ms]; V1 TAD bias 0.14ms [−13.23 to 13.51ms], V2 TAD bias 0.008ms [−12.42 to 12.44ms].

Correspondence to Larisa G. Tereshchenko, MD, Carnegie 568, 600 N. Wolfe St., Baltimore, MD 21287. lteresh1@jhmi.edu. Phone: 410-502-2796; Fax: 410-614-8039.

Conclusion—Comprehensive statistical evaluation of reproducibility of automated ECG measurements is important for appropriate interpretation of ECG. Automated ECG measurements are reproducible to within 25%.

Keywords

electrocardiogram; automated measurement; ARVD/C; reproducibility; QRS duration; terminal activation duration

1. Introduction

Assessment of the reproducibility of any measurement technique in medicine is always needed, because only reproducible measurement techniques can provide reliable results. During recent years, remarkable advancements in biostatistics have been made, allowing for comprehensive evaluation of reproducibility. However, neither clinicians nor engineers are thoroughly familiar with available biostatistical methods for assessment of reproducibility. This fact motivated us to conduct a study with a comprehensive biostatistical evaluation of reproducibility.

Arrhythmogenic right ventricular dysplasia/cardiomyopathy (ARVD/C) is an inherited heart disorder characterized by fibrofatty replacement of the right ventricular myocardium and life-threatening ventricular arrhythmias [1,2]. Arrhythmias often precede gross structural abnormalities in the myocardium and can occur early in the natural history of ARVD/C [3,4]. Mutations in the genes encoding desmosomal proteins, responsible for cell-to-cell coupling via gap junctions, have been linked to ARVD/C [5]. Cell-to-cell uncoupling results in slow, heterogeneous electrical conduction in the right ventricular (RV) free wall and RV outflow tract, presented as the epsilon wave and QRS prolongation in the right precordial leads on a surface ECG, and as prolonged RV endocardial activation on an intracardiac electroanatomic map [6].

An International Task Force has endorsed a set of criteria for the clinical diagnosis of ARVD/C, with ECG criteria comprising an important component of the diagnostic criteria [1,2,7]. T-wave inversion in the right precordial leads (V1, V2, and V3) in individuals > 14 years of age in the absence of the complete right bundle branch block (RBBB) and the presence of the epsilon wave in the right precordial leads, were identified as 2 major criteria of ARVD/C diagnosis. Terminal activation duration (TAD) of QRS (distance from the S-wave nadir to the end of QRS) ≥ 55ms in V1, V2 or V3 in the absence of complete RBBB was identified as a minor criterion. However, a previous study has demonstrated that manual measurements of many quantitative ECG parameters relevant to ARVD/C diagnosis, particularly QRS duration, can vary greatly between readers [8].

Automated ECG analysis represents a potentially useful alternative to manual ECG measurements. Several studies have compared the reproducibility of manual and automated measurements of averaged QRS duration on 12-lead ECGs [9,10], showing the advantage of automated ECG measurements. However, reproducibility of automated ECG measurements in the right precordial leads has not been previously studied. Presence of the epsilon wave or prolonged terminal activation might result in a local QRS prolongation in the right

precordial leads, which could be measured automatically by modern ECG machines. However, only QRS duration averaged across all 12 leads is routinely reported. Local QRS duration in V1–V3, or TAD in V1–V3 are not routinely available for physicians. The goal of this study was to assess the reproducibility of automated measurements of QRS duration, TAD, and other ECG metrics on separate right precordial leads V1, V2, and V3 in ARVD/C registry participants.

2. Methods

2.1. Study Population

The study population included participants of the Johns Hopkins ARVD/C Registry (www.ARVD.com). The registry consists of prospectively enrolled consecutive subjects who were referred to the Johns Hopkins ARVD/C clinic for evaluation. All patients included in the ARVD/C Registry provided written informed consent. The study was approved by the Johns Hopkins School of Medicine Institutional Review Board.

In our retrospective study, the study population included only those registry participants who had a pair of digital 12-lead ECGs recorded at rest on the same date (mean time between recordings 3.5±2.5 hours).

2.2. ECG Recording

Serial 10-sec digital ECGs (sampling rate 500Hz, amplitude resolution 1μV) of the study participants were extracted from the JHH ECG MUSE database (GE Healthcare, Wauwatosa, WI, USA) for subsequent analysis. All 12-lead ECGs used in the study were recorded using the GE-Marquette MAC 5000 ECG system (GE Healthcare, Wauwatosa, WI, USA) on the day of the outpatient visit. All study participants had 2 consecutive ECGs recorded on the same date.

2.3. ECG Analysis

The ECGs were analyzed using 12SL™ algorithm by Magellan ECG Research Workstation Software (GE Healthcare, Wauwatosa, WI, USA). The ECG pairs were compared by examining various ECG parameters.

ECG parameters were measured on a “median beat”. Depolarization parameters measured in the right precordial leads included intrinsicoid deflection, TAD, and QRS duration. Intrinsicoid deflection in V1–V3 was measured as the time from onset of QRS complex to the alignment point of the QRS complex (Figure 1). QRS duration in V1–V3 was calculated by summing the Q, R, S, R’, and S’ durations, measured automatically in those leads by the software. As previously described, TAD was calculated as the difference between QRS duration and intrinsicoid deflection in V1, V2, and V3 (Figure 1). In addition, average QRS duration and QRS axis were calculated across all 12 leads.

Repolarization parameters measured in the right precordial leads included T-wave amplitude in V1–V4. In addition, QT interval, QTc interval (by Bazett), T axis and heart rate were averaged across 10 seconds of recording, and across all ECG leads.

2.4. Statistical Analysis

STATA 13 (StataCorp LP, College Station, TX) was used for all statistical calculations. First, the distribution of ECG parameters was evaluated. The reproducibility of the automated ECG measurements was then assessed via Bland-Altman analysis [11]. The degree of agreement was expressed as the bias (the mean difference) with 95% limits of agreement (mean±2 standard deviations), and the relative % bias, (the mean difference of two measurements divided by their mean value). Precision was defined as 100% minus relative % bias. The statistical correlation between pairs for each parameter was calculated as Pearson's correlation coefficient *r*. In addition, Lin's concordance correlation coefficient ρ_c (rho_c) was calculated to describe the strength of agreement: >0.99 indicates almost perfect agreement; 0.95–0.99, substantial agreement; 0.90–0.95, moderate agreement; <0.90, poor agreement. Furthermore, Bradley-Blackwood procedure was used to simultaneously compare the means and variances of the 2 measurements [12].

3. Results

3.1. Study Population

This study population (n=247) was heterogeneous and included ARVD/C patients, gene-negative members of gene-positive families, family members of gene-negative and untested ARVD/C patients, and individuals referred to the clinic for evaluation but who did not meet criteria for ARVD/C. The goal of this study was to evaluate reproducibility of ECG parameters, but not their diagnostic value. Therefore, we included all registry participants with an available pair of ECGs, regardless of their diagnosis. Mean age of study participants was 35.2±15.6 y. About half of the study population were men (n=144, 58.3%), and most of the participants were white (n=227, 92%). Clinical characteristics of study participants are presented in Table 1.

Definite ARVD/C according to the 2010 Task Force criteria [3] was diagnosed in 11 participants. About half of the ARVD/C cases were gene-positive (5 out of 11 [45%]). T-wave inversion in V1 was observed in all ARVD/C patients, T-wave inversion in V2 and V3 was observed in 7 out of 11 patients (64%), and T-wave inversion in V4 was seen in 5 out of 11 ARVD/C patients (45%). Complete RBBB was observed in 16 individuals, and incomplete RBBB was observed in other 3 participants. No definite ARVD/C patient had either incomplete or complete RBBB. QRS duration in V1–V3 was < 110 ms in all but the fore-mentioned participants with complete RBBB. No intermittent bundle branch block (present on one ECG, but absent on another) was seen in any study participant.

3.2 Reproducibility of depolarization metrics: automated QRS duration measurements, averaged across all leads vs. measured in right precordial leads V1–V3

Distribution of ECG parameters was normal. For averaged QRS, there was a negligible bias of −0.04 ms [0.04%] (Table 2), with 95% limits of agreement from −8.80 ms to 8.80 ms (Figure 2A). Precision of automated averaged QRS duration measurement was 99.96%. Lin's concordance coefficient confirmed substantial agreement (Figure 2B). Bradley-Blackwood F test was not significant (Table 2), which confirmed that bias did not depend on average QRS duration value. Therefore, the bias and 95% limits of agreement adequately

described the differences between two measurements. We can thus conclude that the agreement between two averaged QRS duration measurements was satisfactory; bias was negligible, and reproducibility was high. Assessment of heart rate reproducibility (Figures 2C–2D) showed that heart rate was stable for the most of patients.

Figures 3–5 illustrate agreement and concordance between 2 measurements of QRS duration, intrinsicoid, and TAD, separately for each of the right precordial leads: V1, V2, and V3. Remarkably, biases of QRS duration measurements in V1–V3 were only slightly larger than the bias of averaged QRS, and importantly, they were always less than 1 ms (0.10 ms for V1, 0.09 ms for V2, and 0.32 ms for V3). Similarly, biases of both TAD and intrinsicoid in V1–V3 were always substantially less than 1 ms, as well. However, 95% limits of agreement for intrinsicoid and TAD ranged from − 13 ms to + 13 ms, representing a substantial disagreement. It is important to note that, in some patients, differences between 2 QRS duration measurements exceeded 10 ms, which is a clinically significant dissimilarity. Figures 3–5 helped to demonstrate important differences of reproducibility between leads V1 and V2 on one hand, and V3 on the other hand. In leads V1 and V2, reproducibility of all metrics that assess depolarization (QRS duration, intrinsicoid, TAD) was very similar. However, in lead V3, only reproducibility of QRS duration fulfilled the same criteria, whereas reproducibility of intrinsicoid and TAD in V3 was lower: the concordance was poor (Lin's $\rho_c < 0.90$); precision was substantial but less than 99%, and bias of TAD in V3 was substantially greater, as compared to TAD bias in V1–V2 (Table 2).

3.3 Reproducibility of repolarization metrics: T wave amplitude measured in right precordial leads

While the relative biases of T-wave amplitude measurements in the right precordial leads exceeded the relative biases of QRS duration measurements in the same leads, the biases of T-wave amplitude were minor, and the definition of T-wave inversion never changed from one measurement to another. In other words, negative T waves remained negative, and positive T waves remained positive in both measurements. In addition, the 95% limits of agreement were relatively small, around ± 0.1 mV. The concordance within each ECG pair was particularly strong for T-wave amplitude in leads V1–V4, indicating substantial agreement (Table 2). Reproducibility of heart rate, uncorrected QT interval and T axis was high, with substantial agreement. However, reproducibility of QTc was poor. Bradley-Blackwood F test for QTc was significant (Table 2), which showed that the mean bias of QTc measurements was dependent on mean QTc value. QTc was the only parameter that demonstrated unsatisfactory reproducibility. However, this was likely due to QT correction approach [13], rather than due to imprecise measurements.

4. Discussion

Through this study, we found that the reproducibility of QRS duration measurements on separate right precordial leads V1–V3 was less than reproducibility of QRS duration averaged across 12 leads, with 95% agreement exceeding 10 ms in value. Reproducibility of QRS duration averaged across 12 leads was satisfactory. Reproducibility of automated measurements of TAD on V1–V2 was similar to the reproducibility of QRS duration on V1–

V2. However, reproducibility of TAD in V3 and QTc was poor. Bradley-Blackwood F test revealed different reasons for less than satisfactory reproducibility in case of TAD_{V3} as compared to QTc. Two TADV3 metrics were in true low agreement. However, assessment of QTc reproducibility in this study was not conclusive, because the bias of QTc metric was dependent on average QTc value. Thus, our study illustrates the importance of a comprehensive biostatistical approach to the assessment of reproducibility.

4.1. Reproducibility of Automated ECG Measurements

Our study comprehensively quantified reproducibility by measuring the following parameters: (1) bias, or the mean difference in ECG metric between 2 ECGs, with 95 % limits of agreement; (2) relative bias (mean difference in ECG metric between 2 ECGs, divided by the mean value of studied ECG metric); (3) concordance, quantified by Lin's ρ_c coefficient, and (4) Bradley-Blackwood procedure, which simultaneously compares the means and variances of the 2 ECG measurements and assesses whether or not bias depends on mean values of the examined ECG metric.

"Recommendations for Standardization of Leads and of Specifications for Instruments in Electrocardiography and Vectorcardiography" [14] and recent guidelines [15] recommended resolution of ECG metrics on the order of 1% and precision above 99%. Amongst other measures of reproducibility, relative bias estimated precision. Relative bias of most of the measured ECG parameters was less than 1%, and therefore the precision was above 99%. While, in this study, precision of QRS duration measurements in V1–V3 and TAD in V1–V2 was nearly perfect, above 99%, Bland-Altman analysis revealed wide ranges of 95% agreement values exceeding 10 ms. Differences in detecting the end of the QRS complex likely explain observed the difference in TAD reproducibility between leads V1–V3. Figure 1B clearly shows how hard could be the detection of the Epsilon wave. The low reproducibility of the TAD in this study can be largely attributed to patients displaying Epsilon waves. Our study illustrates the importance of a comprehensive assessment of reproducibility, beyond sole assessment of precision.

Pairs of ECGs in this study were recorded at the same day, but it is known that even minor changes in the physiological state between 2 recordings (e.g. food and water intake [16], smoking) can be responsible for the differences, and likely explain less-than-perfect precision of T-wave amplitude measurements in V1–V4. In addition, it is well known that variability of precordial leads placement between 2 ECG recordings could affect reproducibility of amplitudes [17,18]. Our study confirmed the previously reported finding of better reproducibility of duration measurements (QRS duration, intrinsicoid, TAD) in the right precordial leads, as compared to the reproducibility of amplitudes [19]. However, it is important to note that repeated T-wave amplitude measurements never changed direction from positive to negative or vice versa, therefore, not impacting the ARVD/C diagnostic criterion of T-wave inversion in the right precordial leads.

In this study, we employed Bland-Altman analysis [11] to quantify reproducibility. It was previously shown that a high correlation (measured by Pearson's correlation coefficient) does not necessarily mean that measured parameters on 2 ECGs agree. Correlation coefficient measures the strength of the relation between 2 variables, but not the agreement

between them. Our study illustrated that data which are in a substantial but not perfect agreement (such T-wave amplitudes in V1–V4) can produce near perfect correlations (Pearson's r = 0.98).

Concordance was another measure of reproducibility in this study. The concordance correlation coefficient is the product of the measure of precision and the measure of accuracy. Lin's coefficient increases in value as a function of the nearness of the data's reduced major axis to the line of perfect concordance (the accuracy of the data) and the tightness of the data about its reduced major axis (the precision of the data). In our study, the measure of accuracy was almost uniformly equal to 1, and the observed Pearson correlation coefficient, r, was nearly identical to the concordance correlation coefficient, rho_c.

Bradley-Blackwood test simultaneously compared the means and variances of the 2 measurements [12]. In this study Bradley-Blackwood F test was significant only for QTc measurements: the larger the mean QTc, the higher the bias that was observed. Thus, mean QTc bias did depend on average QTc value. Likely reason for such disagreement of QTc values is the dependence of Bazett equation on heart rate [13], especially pronounced at extreme heart rate values, observed in some of our study participants. Importantly, for all other ECG parameters Bradley-Blackwood F test was not significant. Therefore, the mean bias and 95% limits of agreement adequately described the differences between the two ECG measurements, with the exception of QTc.

4.2 Depolarization abnormalities in ARVD/C diagnostic criteria

Localized prolongation in the QRS complex (>110 ms) in the right precordial leads (V1–V3) and prolonged TAD (≥ 55ms) in V1–V3 are known as important features of ARVD/C [7,20]. Terminal activation duration of QRS ≥55 ms measured from the nadir of the S wave to the end of the QRS, including R', in V1, V2, or V3, in the absence of complete right bundle-branch block currently serves as a minor ARVD/C criterion [7]. Unfortunately, moderate-to-poor reproducibility of manual QRS duration measurements in the right precordial leads highlights the difficulties in accurately measuring the QRS duration, a problem that has been encountered across multiple studies when using manual ECG measurements [9,10,21]. In our study, automated measurements of ECGs parameters show a higher degree of reproducibility when compared to the inter-observer agreement reported by Jain et al using manual or digital calipers to measure ECG parameters for ARVD/C diagnosis [8]. Therefore, the use of automated measurements could likely result in more consistent evaluations of ARVD/C diagnostic criteria. Kasamaki et al [22] similarly demonstrated that automated measurements exhibited smaller inter-reader differences and improved reproducibility as compared to manual measurements. Automated measurement of ECG parameters is also considerably more time-efficient as compared to manual measurements. This potential improvement in ECG measurements via automated analysis applies not only to diagnosis of ARVD/C, but also to other conditions that rely on electrocardiographic diagnostic criteria. For example, QRS duration guides indications for cardiac resynchronization therapy.

4.3 Limitations

Our study population was rather heterogeneous, and therefore we did not compare reproducibility of ECG measurements in patients with vs. without ARVD/C. Furthermore, the ARVD/C patients in the study were not necessarily at the same stage of the disease. Nevertheless, study of the reproducibility of automated measurements of important ECG parameters provided findings that are applicable for all subjects referred to the ARVD/C clinic for evaluation.

In this study, ECG parameters were measured by only one algorithm (12SLTM by GE Healthcare). The question whether or not similar results could be obtained with other manufacturers' algorithms is required additional study [23].

Of note, our definition of TAD differed from the accepted definition for manual TAD measurements. The automated algorithm that we used robustly detected the alignment point of the QRS complex, whereas manual measurements used the nadir of the S wave instead. In addition, we measured TAD in all study participants and did not exclude ECGs of subjects with RBBB. The diagnostic value of a slightly different definition of TAD should be tested in another clinical study.

5. Conclusions

Reproducibility of automated ECG measurements was substantial and demonstrated the overall robustness of the method. A comprehensive approach is important for adequate assessment of reproducibility.

Acknowledgments

Financial support & Conflict of interest:

Study was partially supported by Dr. Francis P. Chiaramonte Private Foundation, the St. Jude Medical Foundation and Medtronic Inc. The Johns Hopkins ARVD/C Program is supported by the Bogle Foundation, the Healing Hearts Foundation, the Campanella family, and the Wilmerding Endowments. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Joel Xue is employee of the GE Healthcare. GE Healthcare did not support this study and did not influence study design, data analysis, results or conclusions.

Reference List

1. Marcus FI, Fontaine GH, Guiraudon G, Frank R, Laurenceau JL, Malergue C, Grosgogeat Y. Right ventricular dysplasia: a report of 24 adult cases. Circulation. 1982; 65:384–398. [PubMed: 7053899]
2. McKenna WJ, Thiene G, Nava A, Fontaliran F, Blomstrom-Lundqvist C, Fontaine G, Camerini F. Diagnosis of arrhythmogenic right ventricular dysplasia/cardiomyopathy. Task Force of the Working Group Myocardial and Pericardial Disease of the European Society of Cardiology and of the Scientific Council on Cardiomyopathies of the International Society and Federation of Cardiology. Br. Heart J. 1994; 71:215–218. [PubMed: 8142187]
3. Basso C, Corrado D, Marcus FI, Nava A, Thiene G. Arrhythmogenic right ventricular cardiomyopathy. Lancet. 2009; 373:1289–1300. [PubMed: 19362677]
4. Asimaki A, Saffitz JE. Gap junctions and arrhythmogenic cardiomyopathy. Heart Rhythm. 2012; 9:992–995. [PubMed: 22100709]

5. Gehmlich K, Syrris P, Peskett E, Evans A, Ehler E, Asimaki A, Anastasakis A, Tsatsopoulou A, Vouliotis AI, Stefanadis C, Saffitz JE, Protonotarios N, McKenna WJ. Mechanistic insights into arrhythmogenic right ventricular cardiomyopathy caused by desmocollin-2 mutations. Cardiovasc Res. 2011; 90:77–87. [PubMed: 21062920]
6. Tandri H, Asimaki A, Abraham T, Dalal D, Tops L, Jain R, Saffitz JE, Judge DP, Russell SD, Halushka M, Bluemke DA, Kass DA, Calkins H. Prolonged RV endocardial activation duration: a novel marker of arrhythmogenic right ventricular dysplasia/cardiomyopathy. Heart Rhythm. 2009; 6:769–775. [PubMed: 19419906]
7. Marcus FI, McKenna WJ, Sherrill D, Basso C, Bauce B, Bluemke DA, Calkins H, Corrado D, Cox MG, Daubert JP, Fontaine G, Gear K, Hauer R, Nava A, Picard MH, Protonotarios N, Saffitz JE, Sanborn DM, Steinberg JS, Tandri H, Thiene G, Towbin JA, Tsatsopoulou A, Wichter T, Zareba W. Diagnosis of arrhythmogenic right ventricular cardiomyopathy/dysplasia: proposed modification of the task force criteria. Circulation. 2010; 121:1533–1541. [PubMed: 20172911]
8. Jain R, Tandri H, Daly A, Tichnell C, James C, Abraham T, Judge DP, Calkins H, Dalal D. Reader- and instrument-dependent variability in the electrocardiographic assessment of arrhythmogenic right ventricular dysplasia/cardiomyopathy. J CARDIOVASC ELECTROPHYSIOL. 2011; 22:561–568. [PubMed: 21114702]
9. Tomlinson DR, Bashir Y, Betts TR, Rajappan K. Accuracy of manual QRS duration assessment: its importance in patient selection for cardiac resynchronization and implantable cardioverter defibrillator therapy. Europace. 2009; 11:638–642. [PubMed: 19174376]
10. De Guillebon M, Thambo JB, Ploux S, Deplagne A, Sacher F, Jais P, Haissaguerre M, Ritter P, Clementy J, Bordachar P. Reliability and Reproducibility of QRS Duration in the Selection of Candidates for Cardiac Resynchronization Therapy. J CARDIOVASC ELECTROPHYSIOL. 2010; 21:890–892. [PubMed: 20233266]
11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1:307–310. [PubMed: 2868172]
12. Bradley EL, Blackwood LG. Comparing paired data: a simultaneous test for means and variances. American Statistician. 1989:234–235.
13. Chiladakis J, Kalogeropoulos A, Arvanitis P, Koutsogiannis N, Zagli F, Alexopoulos D. Preferred QT correction formula for the assessment of drug-induced QT interval prolongation. J. Cardiovasc. Electrophysiol. 2010; 21:905–913. [PubMed: 20233271]
14. Kossmann CE, Brody DA, Burch GE, Hecht HH, Johnston FD, Kay C, Lepeschkin E, Pipberger HV, Pipberger HV, Baule G, Berson AS, Briller SA, Geselowitz DB, Horan LG, Schmitt OH. Recommendations for Standardization of Leads and of Specifications for Instruments in Electrocardiography and Vectorcardiography. Circulation. 1967; 35:583–602. [PubMed: 6021338]
15. Kligfield P, Gettes LS, Bailey JJ, Childers R, Deal BJ, Hancock EW, van HG, Kors JA, Macfarlane P, Mirvis DM, Pahlm O, Rautaharju P, Wagner GS, Josephson M, Mason JW, Okin P, Surawicz B, Wellens H. 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. J Am Coll. Cardiol. 2007; 49:1109–1127. [PubMed: 17349896]
16. Sears GA, Manning GW. Routine electrocardiography: postprandial T-Wave changes. Am Heart J. 1958; 56:591–597. [PubMed: 13571137]
17. Farb A, Devereux RB, Kligfield P. Day-to-day variability of voltage measurements used in electrocardiographic criteria for left ventricular hypertrophy. J Am Coll Cardiol. 1990; 15:618–623. [PubMed: 2137478]
18. Schijvenaars RJ, Kors JA, van HG, van Bemmel JH. Use of the standard 12-lead ECG to simulate electrode displacements. J Electrocardiol. 1996; 29(Suppl):5–9. [PubMed: 9238370]
19. de Bruyne MC, Kors JA, Visentin S, van HG, Hoes AW, Grobbee DE, van Bemmel JH. Reproducibility of computerized ECG measurements and coding in a nonhospitalized elderly population. J Electrocardiol. 1998; 31:189–195. [PubMed: 9682894]

20. Jain R, Dalal D, Daly A, Tichnell C, James C, Evenson A, Jain R, Abraham T, Tan BY, Tandri H, Russell SD, Judge D, Calkins H. Electrocardiographic features of arrhythmogenic right ventricular dysplasia. Circulation. 2009; 120:477–487. [PubMed: 19635971]

21. Sarubbi B, Li W, Somerville J. QRS width in right bundle branch block. Accuracy and reproducibility of manual measurement. International Journal of Cardiology. 2000; 75:71–74. [PubMed: 11054509]

22. Kasamaki Y, Ozawa Y, Ohta M, Sezai A, Yamaki T, Kaneko M, Watanabe I, Hirayama A, Nakayama T. Automated versus manual measurement of the QT interval and corrected QT interval. Ann Noninvasive Electrocardiol. 2011; 16:156–164. [PubMed: 21496166]

23. Kligfield P, Badilini F, Rowlandson I, Xue J, Clark E, Devine B, Macfarlane P, de BJ, Mortara D, Babaeizadeh S, Gregg R, Helfenbein ED, Green CL. Comparison of automated measurements of electrocardiographic intervals and durations by computer-based algorithms of digital electrocardiographs. Am. Heart J. 2014; 167:150–159. [PubMed: 24439975]

Highlights

- Comprehensive statistical evaluation of reproducibility of automated ECG measurements is important for appropriate interpretation of ECG.
- Automated ECG measurements of QRS duration and terminal activation duration in right precordial leads are reproducible to within 25%.
- Precision of automated averaged QRS duration measurement was 99.96%.
- 95% limits of agreement of QRS duration in V1–V3 exceeded 10 ms.

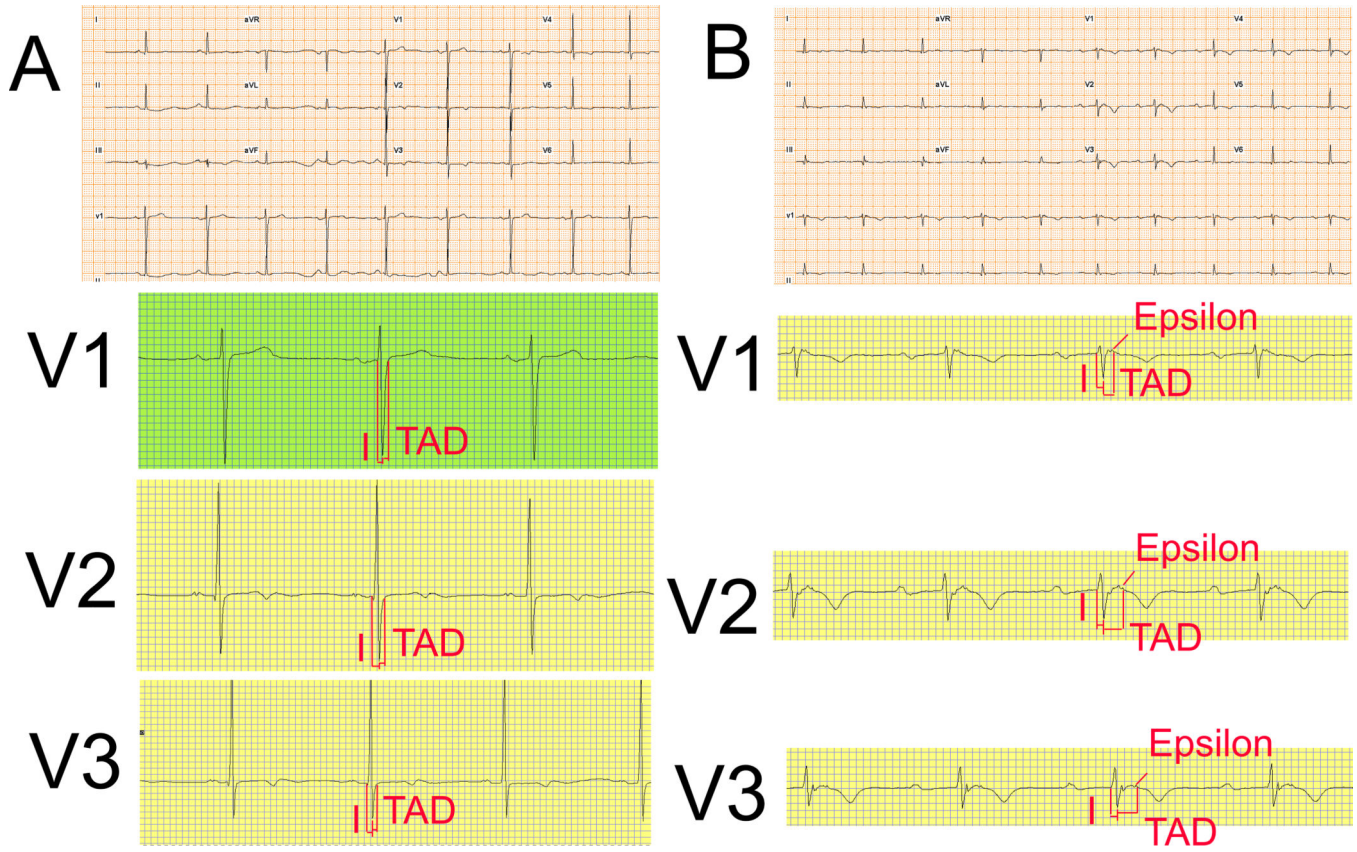

Figure 1.

Example of ECG measurements in a patient without ARVD/C (**A**) and in a patient with definite ARVD/C (**B**). Intrinsicoid deflection (I), terminal activation duration (TAD) intervals and Epsilon waves are shown in leads V1–V3 in ARVD/C patient.

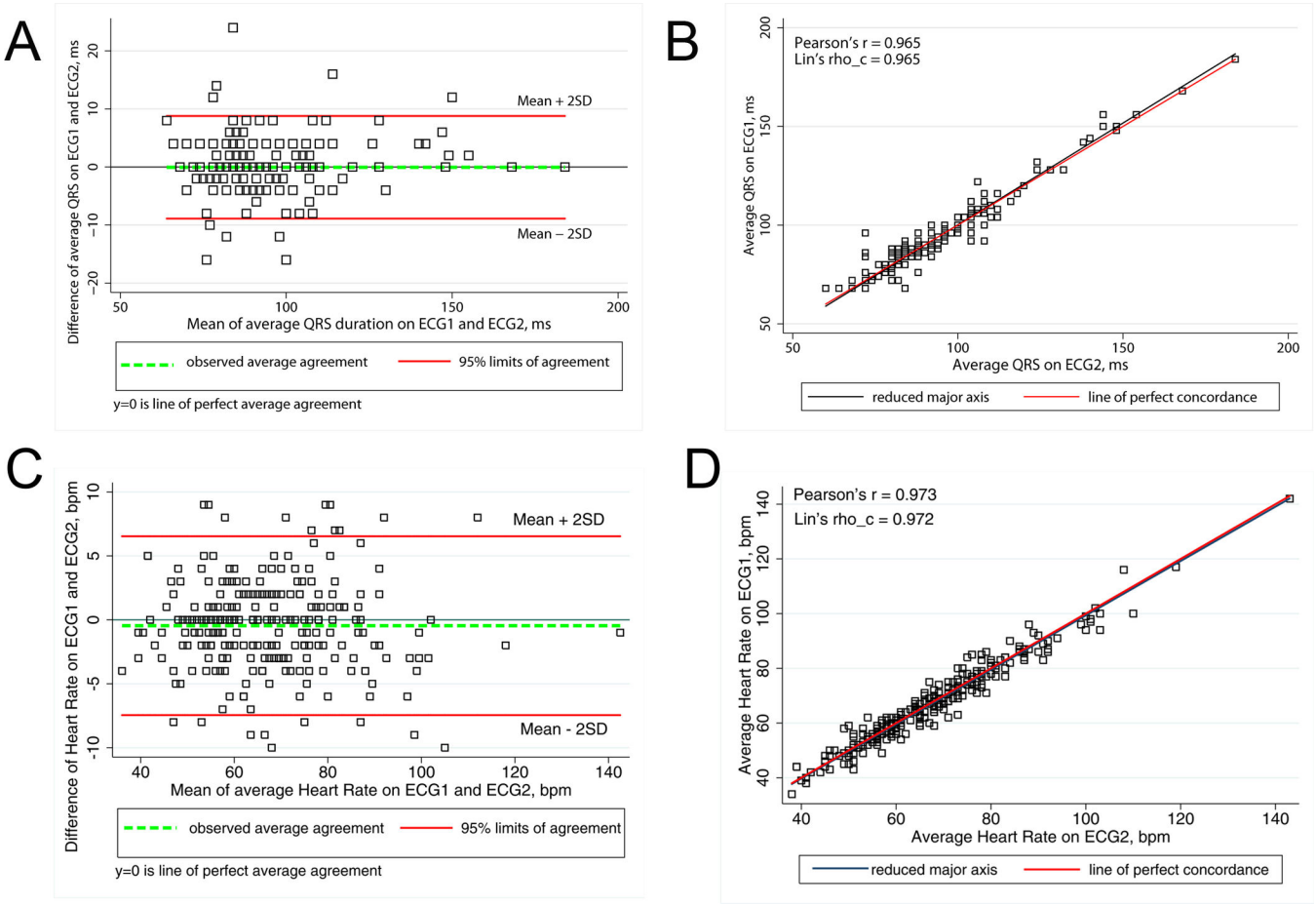

Figure 2.

A. Bland-Altman plots demonstrating agreement of averaged QRS duration measurements on 2 ECGs. The scatterplot presents paired differences (Y axis), plotted against pair-wise means (X axis). The reference line indicates the perfect average agreement, Y=0. The central dashed line indicates the mean difference between the 2 measurements, or mean bias. Upper and lower lines represent the mean ± 2 standard deviations, or 95% limits of agreement. **B**. Concordance scatterplot of the average QRS duration, measured on 2 ECGs. The reduced major axis of the data goes through the intersection of the means and has the slope given by the sign of Pearson's r and the ratio of the standard deviations. The reference line shows the perfect concordance, Y=X. **C**. Bland-Altman plots demonstrating agreement of heart rate. **D**. Concordance scatterplots of heart rate.

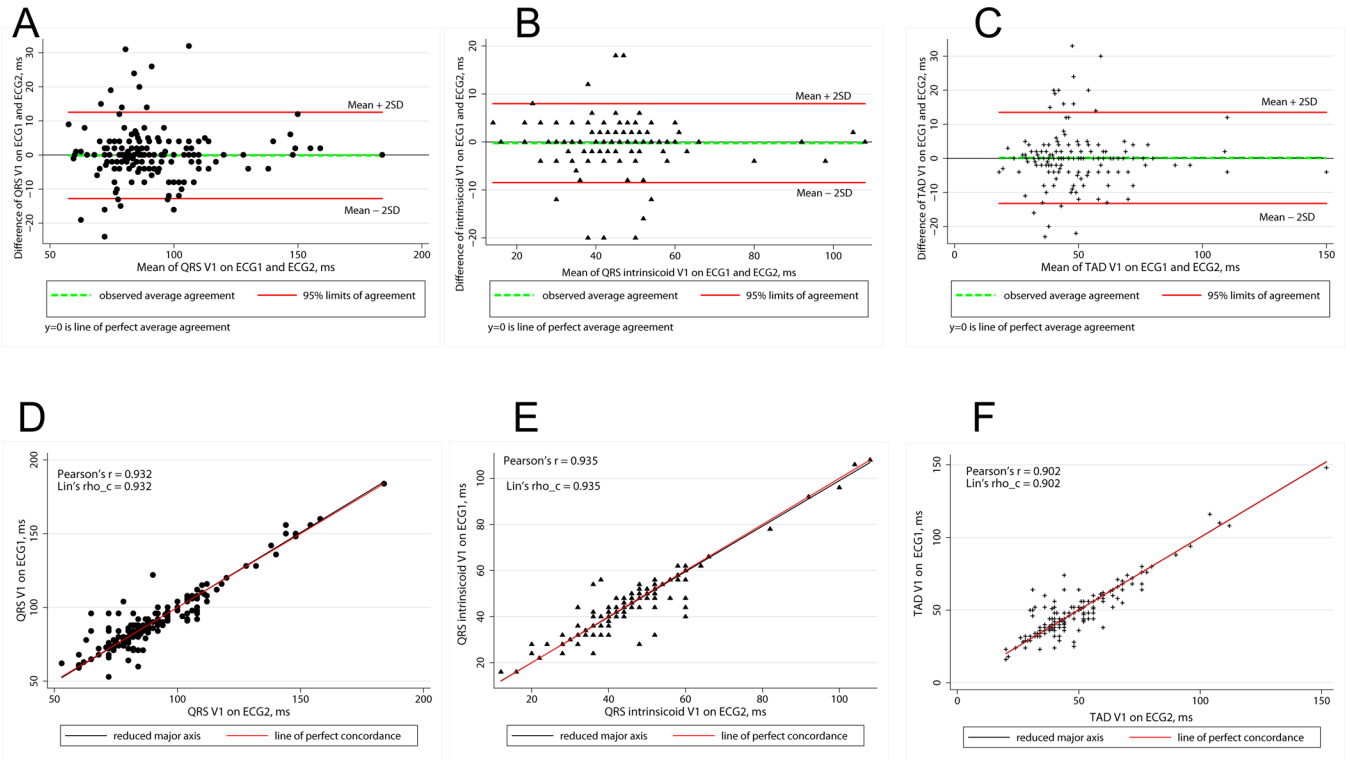

Figure 3.

Bland-Altman plots demonstrating agreement of QRS duration (**A**), intrinsicoid (**B**), and TAD (**C**) in lead V_1. Concordance scatterplots of (**D**) QRS duration, (**E**) intrinsicoid, (**F**) TAD in lead **V1,** measured on 2 ECGs. Definitions given in Figure 2.

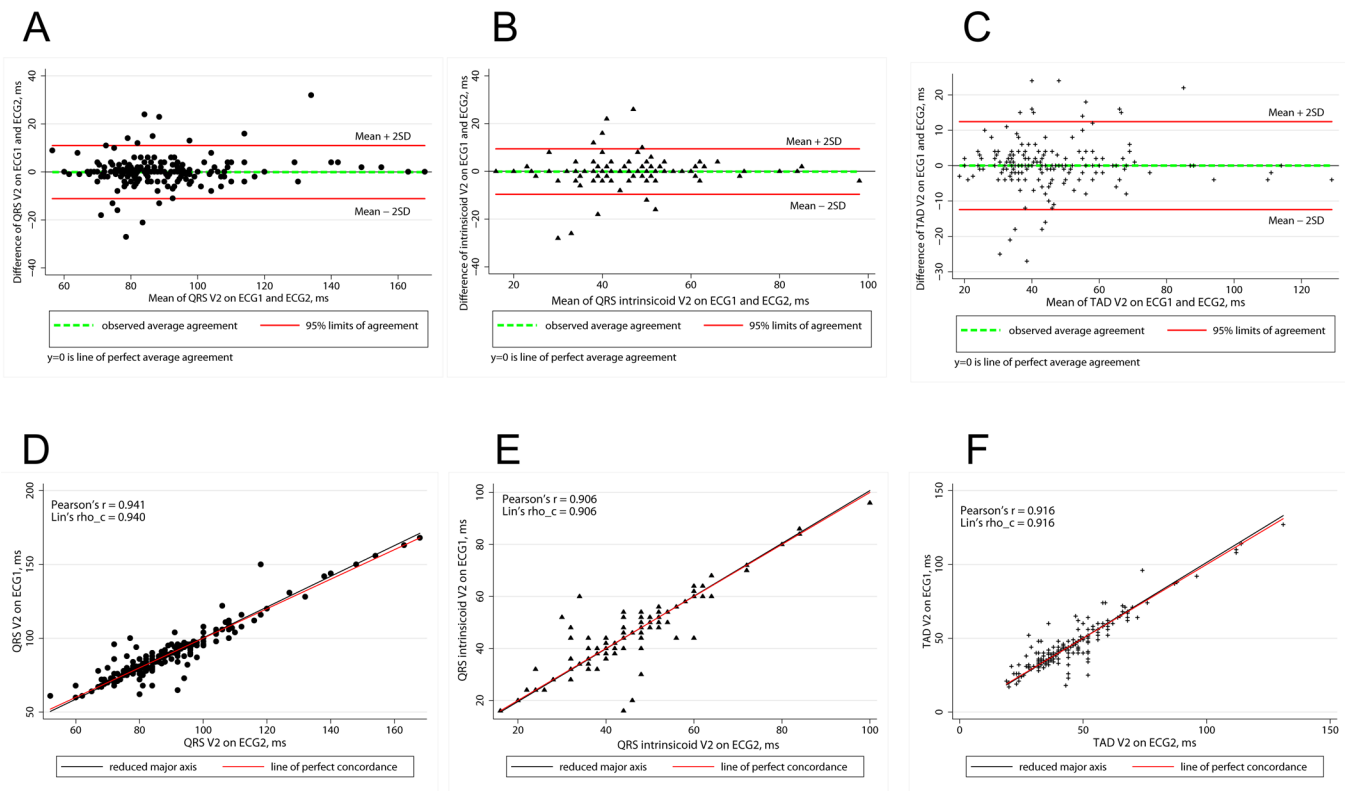

Figure 4.

Bland-Altman plots demonstrating agreement of QRS duration (**A**), intrinsicoid (**B**), and TAD (**C**) in lead **V2**. Concordance scatterplots of (**D**) QRS duration, (**E**) intrinsicoid, (**F**) TAD in lead **V2,** measured on 2 ECGs. Definitions given in Figure 2.

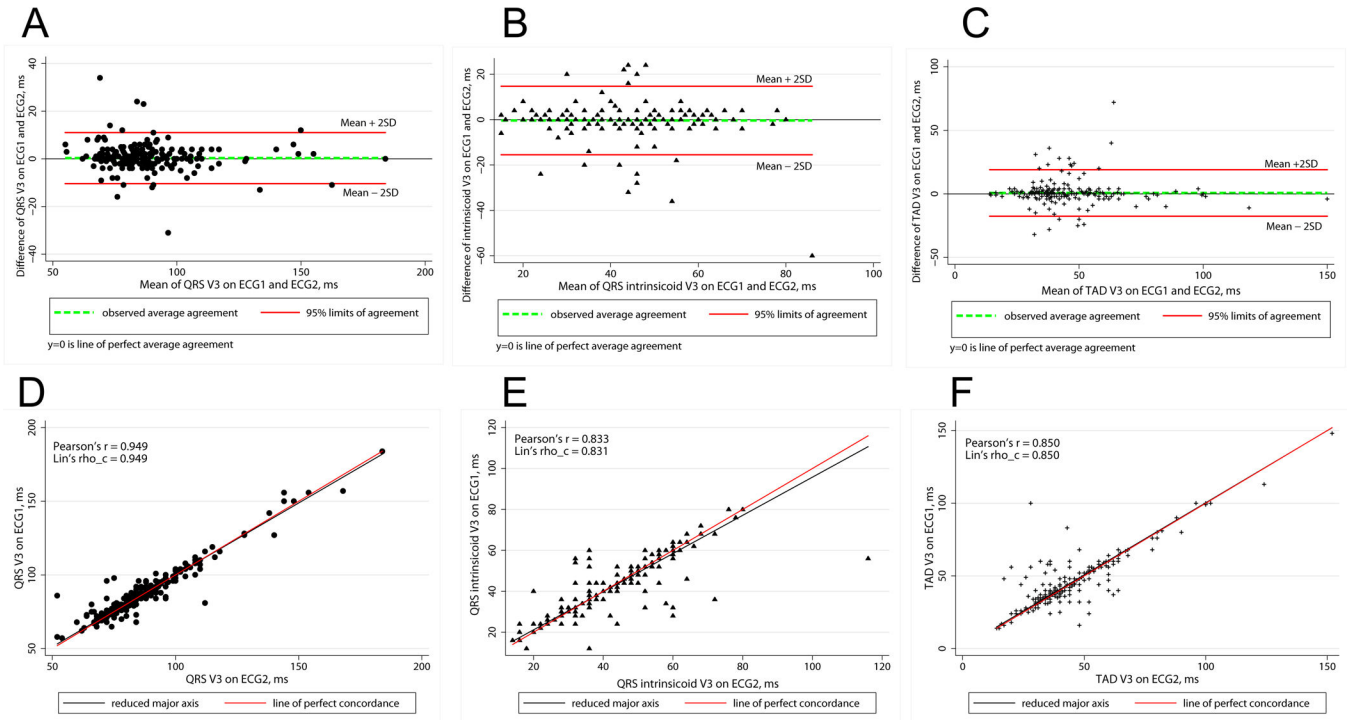

Figure 5.

Bland-Altman plots demonstrating agreement of QRS duration (**A**), intrinsicoid (**B**), and TAD (**C**) in lead **V3**. Concordance scatterplots of (**D**) QRS duration, (**E**) intrinsicoid, (**F**) TAD in lead **V3,** measured on 2 ECGs. Definitions provided in Figure 2.

Table 1

Characteristics of study population

Characteristic	N = 247
Age (SD),y	35.2(15.6)
Male gender, n(%)	144(58.3)
White race, n(%)	227(92)
Definite ARVD/C per Task Force 2010 criteria, n(%)	11(4.4)
Definite idiopathic right ventricular outflow tract tachycardia, n(%)	16(6.5)
Complete right bundle branch block, n(%)	16(6.5)

Table 2

Reproducibility agreement of ECG measurements (n = 247 participants)

Parameter	ECG 1 (mean±SD)	ECG 2 (mean±SD)	Bias	% Bias	95% Limits of Agreement	Pearson, *r*	Lin's, ρ_c [95%CI]	Bradley-Blackwood F (P)
Heart rate, bpm	67.06±15.26	67.52±15.35	−0.46	−0.67	[−7.45; 6.54]	0.97	0.97[0.97–0.98]	2.11(0.123)
Average QRS, ms	93.43±17.19	93.47±16.67	−0.04	−0.04	[−8.80; 8.80]	0.97	0.97 [0.96–0.97]	1.67(0.191)
QRS axis, deg	51.35±42.47	52.74±42.10	−1.39	−2.64	[−29.83; 27.06]	0.94	0.94 [0.93–0.96]	1.20(0.303)
T axis, deg	41.74±39.14	41.91±38.04	−0.17	−0.41	[−21.51; 21.16]	0.96	0.96 [0.95–0.97]	1.35(0.262)
Average QTc, ms	427.24±30.12	429.32±27.81	−2.07	−0.48	[−39.70; 35.55]	0.78	0.78 [0.73–0.83]	3.47(0.033)
Average QT, ms	411.41±45.32	411.99±44.37	−0.58	−0.14	[−29.90; 28.73]	0.95	0.94 [0.93–0.96]	0.70(0.500)
QRS V1, ms	90.74±17.67	90.85±17.43	−0.10	−0.11	[−12.77; 12.56]	0.93	0.93 [0.92–0.95]	0.21(0.814)
QRS V2, ms	88.67±16.56	88.75±15.92	−0.09	−0.10	[−11.13; 10.96]	0.94	0.94 [0.93–0.95]	1.65(0.193)
QRS V3, ms	88.19±16.92	87.87±17.34	0.32	0.36	[−10.39; 11.03]	0.95	0.95 [0.94–0.96]	1.16(0.314)
Intrinsicoid V1, ms	41.74±11.57	41.98±11.72	−0.243	−0.58	[−8.47; 7.98]	0.94	0.94 [0.92–0.95]	0.57(0.564)
Intrinsicoid V2, ms	43.45±11.26	43.55±11.11	−0.097	−0.22	[−9.61; 9.42]	0.91	0.91 [0.88–0.93]	0.18(0.832)
Intrinsicoid V3, ms	42.45±12.79	42.83±13.71	−0.39	−0.91	[−15.48; 14.70]	0.83	0.83 [0.79–0.87]	2.26(0.106)
TAD V1, ms	49.00±15.42	48.87±15.41	0.140	0.29	[−13.23; 13.51]	0.90	0.90 [0.88–0.93]	0.05(0.945)
TAD V2, ms	45.21±15.61	45.21±15.25	0.008	0.02	[−12.42; 12.44]	0.92	0.92 [0.90–0.94]	0.43(0.649)
TAD V3, ms	45.7±17.0	45.0±17.2	0.709	1.58	[−17.61; 19.02]	0.85	0.85[0.82–0.88]	0.73(0.482)
T amplitude V1, μV	−98.0±144.5	−96.2±143.2	−1.81	−1.84	[−74.33; 70.70]	0.97	0.97 [0.96–0.98]	0.37(0.688)
T amplitude V2, μV	211.9±329.9	215.5±327.2	−3.64	−1.69	[−119.0; 111.7]	0.98	0.98 [0.98–0.99]	0.60(0.549)
T amplitude V3, μV	251.4±322.0	244.2±322.7	7.18	2.85	[−125.8; 140.1]	0.98	0.98 [0.97–0.98]	1.39(0.252)
T amplitude V4, μV	305.7±302.0	302.2±297.2	3.49	1.14	[−124.8; 131.8]	0.98	0.98 [0.97–0.98]	1.02(0.361)