Validation of an automatic diagnosis of strict left bundle branch block criteria using 12-lead electrocardiograms

Xiaojuan Xia M.S.¹ | Anne-Christine Ruwald M.D.^{1,2} | Martin H. Ruwald M.D., Ph.D.^{1,3} |
Nene Ugoeke M.D., M.P.H.¹ | Barbara Szepletowska M.D., Ph.D.¹ |
Valentina Kutyla M.D., Ph.D.¹ | Mehmet K. Aktas M.D.¹ |
Poul Erik B. Thomsen M.D., D.M.S.C., Ph.D.⁴ | Wojciech Zareba M.D., Ph.D.¹ |
Arthur J. Moss M.D.¹ | Jean-Philippe Couderc Ph.D.¹

¹Heart Research Follow-Up Program, University of Rochester, Rochester, NY, USA

²Department of Cardiology, Gentofte University Hospital, Copenhagen, Denmark

³Department of Cardiology, Bispebjerg Hospital, Copenhagen, Denmark

⁴Department of Cardiology, Aalborg University Hospital, Aalborg, Denmark

Correspondence

Xiaojuan Xia, M.S., Heart Research Follow-Up Program, Cardiology Division, University of Rochester Medical Center, Rochester, NY, USA.
Email: heartjxx@heart.rochester.edu

Aims: Strict left bundle branch block (LBBB) criteria were recently proposed to identify LBBB patients to benefit most from cardiac resynchronization therapy (CRT). The aim of our study was to automate identification of strict LBBB in order to facilitate its broader application.

Methods: We developed a series of algorithms to automatically detect and measure parameters required for strict LBBB criteria and proposed a definition of QRS notch detection. The algorithms were developed using training ($n = 20$) and validation ($n = 592$) sets consisting of signal-averaged 12-lead ECGs (1,000 Hz sampling) recorded from 612 LBBB patients from Multicenter Automatic Defibrillator Implantation Trial-CRT. Four trained clinicians independently performed adjudication on 148 different ECGs for comparing automatic and manually adjudicated results, in addition to 13 ECGs for evaluation of intraobserver variability and 32 ECGs for interobserver variability. We assessed the performance of the automated algorithms using manually adjudicated ECGs as references.

Results: Overall sensitivity and specificity for detecting strict LBBB were 95% and 86%, respectively. The mean absolute deviation (MAD) of QRS duration and notch/slur locations for the automated method versus the manual method was below 1 ms, and MAD values were lower than 2 ms for interobserver and intraobserver variability. Sensitivity and specificity for detecting notch and slur locations were 87% and 96% for notches and 78% and 90% for slurs using the automatic method. In addition 95% and 93% agreements for notches and 90% and 88% agreements for slurs were reached for intra- and interobserver.

Conclusion: The proposed algorithms automatically measure QRS features for the diagnosis of strict LBBB. Our study shows good performance in reference to manual results.

KEYWORDS

automated algorithms, cardiac resynchronization therapy, electrocardiography, left bundle branch block

1 | INTRODUCTION

The impact of QRS morphology on the benefit of cardiac resynchronization therapy (CRT) has been studied in meta-analyses and subgroup analyses of clinical trials (Birnie et al., 2013; Peterson et al., 2013; Zareba et al., 2011). These studies showed that patients with left bundle branch block (LBBB) derived substantial clinical benefit from CRT, and no or little clinical benefit was observed in patients with a non-LBBB QRS pattern (right bundle-branch block or nonspecific LV conduction delay: intraventricular conduction disturbances [IVCD]) (Birnie et al., 2013; Goldenberg et al., 2014; Peterson et al., 2013; Zareba et al., 2011). A significant reduction in heart failure events or death was observed in LBBB patients; whereas the non-LBBB patients demonstrated a trend toward a higher mortality when treated with CRT-Defibrillators, and this increase was observed primarily in patients with IVCD delay who received CRT-D compared to implantable defibrillator only (Zareba et al., 2011).

As the normal electrical pathway to the left ventricle (LV) (left bundle branch) is completely blocked, the electrical impulse travels from the right ventricle through the interventricular septum toward the lateral wall of the LV through muscle fibers. This results in a significant delay between activation of the interventricular septum and activation of the LV free wall, due to detoured and slow electrical pathway. CRT applies simultaneous pacing of the septum and LV free wall to correct this delay to resynchronize ventricular contraction. However, about one-third of patients are not correctly diagnosed (Nikoo, Aslani, & Jorat, 2013) with complete LBBB by conventional ECG criteria: QRS duration ≥ 120 ms, QS or rS in lead V_1 , broad notched or slurred R wave in leads I, aVL, V_5 , and V_6 and with absence of Q waves in leads I, V_5 , and V_6 . These standard ECG criteria are also met in patients with a combination of LV hypertrophy, LV dilatation, slowed intraventricular conduction, delayed initiation of LV activation (incomplete LBBB), and left anterior fascicular block, even though they do not have an electrical activation consistent with complete LBBB. These observations emphasize the importance of properly distinguishing the conduction delay due to complete LBBB from other conditions.

In 2011, Strauss, Selvester, and Wagner (2011) proposed a new set of criteria to define complete "strict" LBBB which include QRS duration ≥ 140 ms (men) or ≥ 130 ms (women), QS or rS configurations of the QRS-complex in leads V_1 and V_2 , and mid-QRS notching or slurring in ≥ 2 of leads V_1 , V_2 , V_5 , V_6 , I, and aVL. Different QRS duration thresholds are used for men and women. In addition, the presence of notching is critical for the diagnosis of LBBB. Relevant notches should begin after the first 40 ms of the QRS, but before 50% of QRS duration, when the activation wavefront reaches the endocardium of the LV. Notching should end at approximately 2/3rd through QRS duration when the activation wavefront reaches the lateral aspect of the epicardial LV. These criteria were further confirmed from computer simulations (Galeotti, van Dam, Loring, Chan, & Strauss, 2013) and validated with echocardiography data (Risum et al., 2013). A prospective study showed LBBB patients defined with the new strict criteria had greater improvement in left heart function after 6 months of CRT than did patients with IVCD or

traditionally defined LBBB with no notch or notches in less than two of the leads (Tian et al., 2013). In addition to smaller baseline left ventricular end-diastolic dimension, the strict LBBB morphology was an independent predictor of superresponse to CRT. Also, the strict LBBB morphology was related to a higher event-free survival rate in CRT patients and better echocardiographic response (Mascioli et al., 2012).

Although the new strict LBBB criteria refine the detection of complete block of the left ventricular conduction system and provide insight into patient selection for CRT therapy, its validation and clinical utilities have not been evaluated in a larger cohort of patients. One of the reasons is the lack of quantitative definition of the QRS slur/notch patterns. In this work, we propose a novel computer program to automatically measure and detect complete LBBB based on the strict LBBB criteria.

2 | METHODS

2.1 | Study population and data acquisition

This study involved the ECG recordings from patients enrolled in the Multicenter Automatic Defibrillator Implantation Trial-CRT (MADIT-CRT; Moss et al., 2009). A total of 612 ECGs recorded prior to implantation, which were able to obtain signal-averaged beats, were used from 612 patients with LBBB based on MADIT-CRT LBBB criteria (WHO criteria). Twenty ECGs were randomly selected to form a training dataset. Since four clinicians (ACR, MHR, NU and BS) participated in the reviewing and adjudication process of automatic measurements, 592 ECGs were equally divided into four groups of 148 ECGs. Ten ECGs were randomly selected within each group for measuring intra-observer variability. Another eight ECGs were randomly selected from each group to make up 32 ECGs for measuring interobserver variation. Three out of these eight ECGs were added back into each own group. Therefore, 13 ECGs were used for evaluation of intraobserver variability and 32 ECGs for interobserver variation, overall 185 ECGs were measured by each observer.

The ECG signals used in this study were signal-averaged single beat tracings with standard 12-lead configuration (detail in Xia et al. (2015)). The signal averaging process was performed on the first 20 min of Holter recording when the patients were resting in supine position. A minimum of 130 sinus beats were required for the signal averaging process.

2.2 | Strict complete LBBB detection

The ECG-based strict LBBB criteria involved measurements of QRS duration, QRS morphology in leads V_1 - V_2 and notch/slur detection in six leads. The following subsections describe each of these measurements and associated detection rules, and explain how to combine these results to diagnose the presence of the complete LBBB conduction.

2.2.1 | QRS duration

We implemented the QRS detector developed by Zong, Moody, and Jiang (2003). Several modifications were made to the algorithm to

improve the location of boundaries for wider QRS duration in LBBB patients, that is, ranging from 130 to 200 ms. These adjustments include: elimination of the adaptive threshold procedure required for multiple beats signal, and an adjustment of the detection thresholds of the onset and offset of the QRS complex (Xia et al., 2015).

2.2.2 | Characterization of QRS morphology

Characterization of QRS morphology was designed to automatically assess the shape of QRS based on conventional nomenclature (Fig. 1). The R or S prime waves were defined as the second peak or nadir after the downward or upward signal passing the isoelectric (IE) line (Fig. 1 rsR'). To determine the QRS shape, we implemented a method identifying signal transitions and using the transition points and their polarity to fit the QRS patterns (details see Xia et al. (2015)).

2.2.3 | Notch and slur detection

Detail of notch and slur detections is described in Xia et al. (2015). The algorithm and additional procedures for detection of all notch and slur in six leads are as follows:

1. Noise-level estimation: the standard deviation of signal within a 20-ms window before QRS onset was calculated and used to establish the notch detection threshold, threshold for distinguishing true notch from noise, and criteria for signal filtering in slur detection.
2. Find QRS large fragmentation: ECG signal within QRS complex was traced to extract upward and downward segments (Fig. 2). If one segment did not cross over the IE line, the fragmentation was registered as a notch position (Fig. 2).
3. Notch detection and definition: we developed a peak and nadir detection method based on the algorithm described in Xia et al. (2015) using the noise-adaptive threshold. The threshold ranged from 10 μV for noise level lower than 5 μV , 20 μV when noise level was between low 5 and 25 μV , and 50 μV for noise level above

FIGURE 1 Examples of the nomenclature for QRS complexes

FIGURE 2 Five traced segment: arrows 1–5 show the successive up- and downward inflections of the QRS. Large wave fragment is detected between segment 2 and 3 which does not cross the isoelectric line and therefore labeled as notch

25 μV . All detected peaks and nadirs were included in the pool of candidates for notch detection.

Geometry properties including height, base width, and area were calculated for each notch candidate. The height of a notch was defined as the height of the notch triangle, for example, the triangle formed by the signal lines and green dotted line (baseline) in Fig. 3. The based width was measured by the length of triangle base and the area was calculated using the integral of the signal reference to the baseline.

Then, several criteria were applied to select true notches from the pool (Fig. 4A). They were: (1) height $\geq 75 \mu\text{V}$, width $\geq 5 \text{ ms}$; (2) height $\geq 50 \mu\text{V}$, width $\geq 10 \text{ ms}$; (3) height $\geq 30 \mu\text{V}$, width $\geq 25 \text{ ms}$. In case of the occurrence of two notches (the locations of the peak/nadir) located $< 40 \text{ ms}$ apart, the notch with smaller area was discarded except when the heights and the widths of both notches were $> 50 \mu\text{V}$ and $> 10 \text{ ms}$.

4. Slur detection and definition: slurs were searched on every segment after applying notch detection and the detection method was described in Xia et al. (2015). In brief, the level of noise is estimated using a 20-ms interval inside the PR segment first. Next a Gaussian filter, whose window length is adapted to the noise level, is applied to smooth the QRS signal. Then, the search for slur patterns is performed on each segment of QRS complex. The slur is associated with a concave pattern in the first derivative signal (Fig. 4B). This pattern is detected using three measurements: the minimum amplitude (H1 in panel C of figure 2), the maximum amplitude (H2), and the duration of the concave pattern (D1). Three levels of slur patterns were based on minimizing the detection error: if $H2/H1 > 8$, $D1 > 10 \text{ ms}$; if $5 > H2/H1 > 8$, $D1 > 20 \text{ ms}$, and finally if $3 > H2/H1 > 5$, $D1 > 40 \text{ ms}$.

2.2.4 | Procedures for diagnosis of complete LBBB

First, the global QRS duration was measured and the QRS morphology was characterized from leads V_1 and V_2 . Next, we detected

FIGURE 3 Example of notches patterns in three different QRS configurations for strict LBBB cases: notch in positive QRS A (above isoelectric line) and C (below isoelectric line); notches in negative QRS going down segment B (above isoelectric line) and D (below isoelectric line) for a biphasic QRS. The red arrow points to the notch position which could be a nadir (A and B, above signal above isoelectric line) or a peak (C and D, above signal below isoelectric line). The notch signal and green dotted line form a triangle defining the notch properties, that is, its height, width and area

FIGURE 4 Illustrations of the method for detecting QRS notch (A) and slur (B). Panel B illustrates the detection of QRS notches, the angle between lines fitting the segment prior and after the turning point needs to form an angle superior to 90°, the duration and amplitude of the notches should satisfy one of the conditions. (C) the slur detection relies on the first derivative (bold line) of the QRS wave (light line). The definition of a slur is based on the presence of a concave pattern and its presence is confirmed based on duration thresholds depending on the depth of the concave pattern

the notches and slurs inside the QRS complexes from six leads. Then, the locations of the detected notches/slurs were scanned to identify those in the mid-QRS region, that is, the notches/slurs meeting the strict criteria (Strauss et al., 2011), which are (1) the position of a notch/slur should begin after the first 40 ms of the QRS, (2) the start position of a notch must occur before 50% of QRS duration, and (3) the end position occurs before 2/3rd of the QRS duration. Finally, the number of leads with mid-QRS notch/slur was evaluated.

2.3 | Inter- and Intraobserver variability measurements

Intra- and interobserver variabilities were measured for continuous variables using MAD. The MAD value was calculated as

$$MAD = \frac{\sum_{i=1}^N |X_i - \bar{X}|}{N},$$

Where $\bar{X}$ was the mean of different measurements on the same ECG, $N = 4$ for interobserver, and $N = 2$ for intraobserver. Since multiple notches/slurs could be present in a single ECG, the criteria for associating notches and slurs between automated and manual assessment were based on different levels of time precision: notches within 10 ms and slurs within 20 ms were considered equivalent.

The level of agreement for binary variables (presence /absence of notch or slur) was calculated as the ratio of the number of observers with same findings (presence or absence of a notch/slur) to the total number of observers. For example, if three observers did not report the presence of a notch which was reported by the fourth observer, the level of agreement is 3/4. Therefore, the agreement for each slur/notch ranged from 50% to 100%. The rules for associating notch/slur patterns between observers were the same as the one described above for computing the MAD.

2.4 | Training and validation

The approach to set the definitions of notch and slur patterns and to improve the detection of software used multiple iterative interactions among six clinicians in training process. After a unanimous agreement was reached for each discussed case, the four clinicians conducted the validation.

2.5 | Compare manual and automatic results

Level of sensitivity and specificity were used to assess binary variables. The sensitivity was computed as the percentage of correctly identified patterns by automatic method to the number of patterns detected manually, and the specificity was measured as the

percentage of correctly identified negative patterns (nonexisting) by automatic method to the number of negative patterns detected by manual.

2.6 | Statistic analysis

The differences between automatic and adjudicated results were measured and quantified using Matlab (MATLAB2013b) and R software (R 3.02). Mean and standard deviation values were reported for continuous measurements. Student paired *t* test was used to assess the difference between two measurements on the same subject. One-way repeated measures ANOVA analysis was performed for more than two measurements on the same subject. A *p* value <.05 was considered statistically significant.

3. | RESULTS

3.1 | Study population

Five hundred and ninety-two baseline ECGs were included in the study. Because of high noise level, two ECGs from group 1, four ECGs from group 2, one ECG from group 3, and two ECGs from group 4 were not analyzed. Overall, 583 ECGs recording were used including 13 and 31 ECGs for assessing intraobserver and interobserver and variability, respectively. The patient characteristics are summarized in Table 1.

3.2 | Intraobserver variability

3.2.1 | Continuous variables

The intraobserver variability in continuous variables is reported in Table 2. The absolute differences in locations of the beginning of QRS (QRS onset), the end of QRS (QRS offset), and QRS duration were all below 2 ms. The MAD value for locating the QRS onset within each observer was 0.9 ± 1.0 ms (*p* = .48), while for locating QRS offset was 1.7 ± 1.9 ms (*p* = .27). Difference in QRS duration was 1.5 ± 1.6 ms (*p* = .52). The student paired *t* test showed that none of these measurements were significant different for same observer (Table 2) demonstrating a small intraobserver variability.

Mean absolute deviation values for locating notch/slur in six lead in Table 2 show the largest difference for notch location is

TABLE 1 Patient characteristics (*n* = 583)

Age (years)	64 ± 12
Gender (M/F)	403/180
NYHA functional class	
I	<i>n</i> = 63
II	<i>n</i> = 520
Ischemic	328 (56%)
LVEF (%)	24 ± 5
QRS width (ms)	167 ± 21

TABLE 2 The MAD results for continuous variables measured for interobserver, intraobserver and automatic versus manual. The difference reaching significance (*p* < .05) is in bold

	Interobserver	Intraobserver	Auto versus Manual
N (no. of ECGs)	31	13	583
QRS duration (ms)	1.4 ± 1.7	1.5 ± 1.6	0.9 ± 1.2
QRS onset (ms)	0.6 ± 1.2	0.9 ± 1.0	0.5 ± 0.5
QRS offset (ms)	1.4 ± 2.1	1.7 ± 1.9	0.5 ± 0.8
Notch: I (ms)	0.08 ± 0.23	0.15 ± 0.19	0.01 ± 0.02
aVL (ms)	0.21 ± 0.56	0.39 ± 0.47	0.04 ± 0.04
V ₁ (ms)	0.11 ± 0.64	0.38 ± 0.51	0.00 ± 0.01
V ₂ (ms)	0.16 ± 0.66	0.41 ± 0.54	0.00 ± 0.00
V ₅ (ms)	0.41 ± 0.98	0.70 ± 0.84	0.05 ± 0.02
V ₆ (ms)	0.26 ± 0.76	0.51 ± 0.63	0.02 ± 0.03
Slur: I (ms)	0.68 ± 1.77	1.22 ± 1.5	0.09 ± 0.15
aVL (ms)	0.43 ± 1.59	1.01 ± 1.3	0.13 ± 0.22
V ₁ (ms)	0.38 ± 1.22	0.80 ± 1.0	0.04 ± 0.08
V ₂ (ms)	0.18 ± 0.99	0.58 ± 0.79	0.11 ± 0.19
V ₅ (ms)	0.95 ± 2.0	1.47 ± 1.72	0.20 ± 0.35
V ₆ (ms)	0.32 ± 0.99	0.45 ± 0.58	0.08 ± 0.12

in lead V₅ (0.70 ± 0.84 ms) and the smallest difference is in lead I (0.15 ± 0.19 ms). Similarly, lead V₅ has the largest difference on slur location (0.7 ± 0.8 ms). However, the smallest difference was found in lead V₆ (0.45 ± 0.58 ms). Overall the MAD for notch was 0.42 ± 0.53 ms and 0.92 ± 1.15 ms for slurs.

3.2.2 | Presence of notch/slur

In evaluating the intraobserver variability in assessing the presence of notch/slur, agreement on presence and absence of notch/slur patterns was computed. When there were multiple notches/slurs within a single QRS complex, the same proximity rules (as described in Section "2.5") were applied to pair each pattern. Third column in Table 3 shows the range of agreement on presence and absence of notch over six leads from 97% in lead V₁ to 88% in lead aVL, while the range for slur was from 97% in lead V₂ to 83% in lead I. The average agreements in six leads were 93% and 88% for notch and slur, respectively.

3.3 | Interobserver variability

3.3.1 | Continuous variables

More than twice of the ECGs were used to evaluate interobserver (31) than for intraobserver variability. The absolute differences in finding QRS onset, QRS offset, and QRS duration were also below 2 ms. Among observers, the MAD value for locating the QRS onset was 0.6 ± 1.2 ms (*p* = .64) while for locating QRS offset was 1.4 ± 2.1 ms (*p* = .006). The difference for QRS duration was 1.4 ± 1.7 ms (*p* = .004). One-way repeated measures ANOVA analysis was applied to assess the difference among four observers, measurements for QRS duration and QRS offset were significantly different.

TABLE 3 The performance of automatic classification on the presence and absence of notch and slur, midslur/notch, and true LBBB comparing to manual adjustment. Also, the agreement of the presence or absence for notch/slur among observers (interobserver) and inside observer (intraobserver) is reported

	Sensitivity (%) (auto-manual)	Specificity (%) (auto-manual)	Interobserver agreement (%)	Intraobserver agreement (%)
N (no. of ECGs)	148 × 4	148 × 4	31 (4 observers)	13 (4 observers)
Notch: I	86 ± 11	94 ± 5	93 ± 11	95 ± 12
aVL	89 ± 10	91 ± 8	94 ± 15	88 ± 32
V ₁	86 ± 10	98 ± 3	97 ± 9	98 ± 7
V ₂	84 ± 9	98 ± 2	97 ± 11	94 ± 17
V ₅	86 ± 7	94 ± 5	94 ± 10	90 ± 24
V ₆	88 ± 9	98 ± 2	94 ± 11	94 ± 15
Overall	87 ± 9	96 ± 4	95 ± 11	93 ± 18
Slur: I	76 ± 16	86 ± 14	87 ± 15	83 ± 36
aVL	77 ± 17	92 ± 12	89 ± 21	90 ± 19
V ₁	78 ± 15	91 ± 14	90 ± 14	92 ± 19
V ₂	75 ± 17	93 ± 12	95 ± 12	97 ± 11
V ₅	79 ± 13	87 ± 17	89 ± 15	84 ± 25
V ₆	82 ± 14	88 ± 14	91 ± 14	84 ± 23
Overall	78 ± 15	90 ± 14	90 ± 15	88 ± 22
Midslur/notch: overall	95 ± 5	85 ± 21		
True LBBB				
Group 1	87	100		
Group 2	98	53		
Group 3	97	100		
Group 4	97	93		
Overall	95 ± 5	86 ± 22		

Among four observers over six leads, the largest difference in MAD for notch location was found in lead V₅ (0.41 ± 0.98 ms), and the smallest difference was in lead I (0.08 ± 0.23 ms). Similarly, lead V₅ had the largest difference in slur locations (0.95 ± 2.00 ms), whereas the smallest difference was in lead V₂ (0.18 ± 0.99 ms). Overall the MAD for locating notch was 0.21 ± 0.64 ms and 0.47 ± 1.38 ms for slur. Since not every notch and slur were identified by all four observers, student paired *t* test was applied to assess the differences in locating the notch and slur between pair of observers. Locating slur is significantly different on leads V₁ ($p = .04$) and V₅ ($p = .01$), and locating notch is significantly different on lead V₂ ($p = .01$).

3.3.2 | Presence of notch/slur

Last column of Table 3 presents the interobserver variability in the presence of notch/slur. Over six leads, the agreement on the presence and absence of notch ranges from 98% in lead V₁ and V₂ to 93% in lead I. The highest agreement for slur was found in lead V₂ (95%) and the lowest agreement was in lead I (87%). The average agreements in six leads were 95% and 90% for notch and slur, respectively.

3.4 | Comparing automatic and manual results

3.4.1 | Continuous variables

Lower than 1 ms was observed in absolute difference between automatic and four observers' measurements on locating QRS onset,

QRS offset, and QRS duration. The MAD value for QRS onset was 0.5 ± 0.5 ms ($p = .02$), and was 0.5 ± 0.8 ms ($p < .001$) for QRS offset. The difference in QRS duration was significant but not clinically relevant: 0.9 ± 1.2 ms ($p < .001$).

The MAD results for notch/slur location between automatic and four observers' results were close to zero for all six leads. Similarly to intra- and interobserver results, the lead with the largest MAD for notch and slur locations was V₅, and the smallest MAD was in leads V₁ and V₂. Overall the MAD for notch was 0.02 ± 0.02 ms and 0.11 ± 1.18 ms for slur. None of these measurements between automatic and manual methods were significantly different.

3.4.2 | Presence of notch/slur

The ability of the automatic algorithms to appropriately detect the presence and absence of notch/slur throughout the whole QRS are reported in Table 3 in terms of sensitivities and specificities. The ranges of sensitivity and specificity for notch were 84% (V₂)–89% (aVL) and 91% (aVL)–98% (V₁, V₂ and V₆), respectively; while ranges for slur were 75% in V₂–82% in V₆ and 86% in I–93% in V₂. Overall for all six leads, the sensitivity and specificity for notch and slur were 87%/96% and 78%/90%, respectively.

3.4.3 | True LBBB test results

We tested how well the automatic method detected the presence of mid-slur/notch comparing to manual measurements. The sensitivity

and specificity were 95% and 85%, respectively. Also, the sensitivity and specificity for testing complete LBBB applying strict criteria were 95% and 86%. The range of sensitivity for testing complete LBBB was from 87% to 98% however the range of specificity was much wider: 53%–100%.

4. | LIMITATION

Among the four clinicians involved in the validation process, three clinicians participated in the discussion and training in the training process and the fourth one joined the validation group later with separate training on the definitions of notch/slur and measuring rules. This could cause higher variability between the fourth reviewer and others. Another limitation is that ECG reviewers did not review the same ECGs, only a subset of the data was considered ($N = 32$). Therefore, our results need to be confirmed in another database with two ECG reviewers read all ECGs.

5. | DISCUSSION

Strauss et al. (2011) quantitatively described the QRS morphologies related to a delay between the interventricular septum and left lateral wall contraction. Subsequent studies provided evidences that the presence of a complete LBBB morphology applying Strauss strict criteria derived greater benefit from CRT (Mascioli et al., 2012; Tian et al., 2013). In addition, scar at or near the LV pacing site has been shown to have adverse effect on response to CRT (Wong et al., 2013). In 2008, the QRS Selvester Scoring System, which estimates the size and location of myocardial scar in the LV from the standard 12-lead electrocardiogram was expanded for application in the presence of multiple interventricular conduction abnormalities (Strauss & Selvester, 2009); including complete LBBB based on Strauss's strict criteria (Strauss et al., 2011). Therefore, automation of the detection of complete LBBB based on strict criteria could provide not only a wider application of the strict LBBB criteria on benefit of CRT but also accessible practice of updated Selvester Scoring System for patients with LBBB.

Besides the QRS duration and the lead V_1 QRS morphology, Strauss strict LBBB criteria emphasize the presence of midslur/notch in six leads. There is no standard definition of QRS notch and slur patterns, and clinicians usually have their own interpretation. To evaluate and compare the performance of automatic algorithm with manual measurements, we equally divided 592 signal-averaged ECGs which were labeled as LBBB using WHO criteria into four groups and assigned each group to one of four clinicians. Meanwhile, 13 and 32 ECGs were randomly selected for assessing intraobserver and interobserver variability. The accuracy of slur/notch detection and location were evaluated as binary and continuous variables.

No adjustment was made to correct the automatic results of QRS morphology in leads V_1 and V_2 following the learning phase since it

was classified correctly in 100% cases. Also with average difference below 1 ms, the accuracy of QRS duration was found to be satisfactory. The paired t test shows significant different between automatic and manual measurement on QRS duration, this might be the result of large sample size ($n = 583$) and capture minute differences that may be clinically irrelevant. To evaluate the performance of slur/notch detection, all slur/notch inside the QRS complex were measured and compared. Lead V_5 has the highest MAD value for slur and notch location in all three comparison groups: automatic versus manual, intraobserver and interobserver. This could link to the fact that the QRS morphology changes from rS to R around lead V_5 location in the horizontal plane which resulted in relatively lower amplitude of R or RS wave.

So far, there is no standard definition of slur and notch pattern recorded on surface ECG, we developed these definitions in collaboration with experts in the field (Xia et al., 2015) using a pattern recognition approach consisting of the detection of morphological changes in signal within QRS complex and the geometrical properties of these changes. We developed a set of rules that seem to be sufficient to replicate with satisfactory accuracy on the presence and location of these subtle components inside QRS complex. Not surprisingly, the lowest MAD value for slur/notch were mostly observed in leads V_1 and V_2 , where the amplitude of QRS reached the highest and was less affected by noise. Generally, the measure of the time location of a slur was more challenging than that of notch because the slur patterns have larger time span and it is generally difficult to visually assess the boundaries of such patterns. Therefore, the MAD value for slur were usually $>2\times$ than MAD from notch patterns. Similar results were found for agreement, sensitivity and specificity, however, lead V_2 has the lowest sensitivity which was caused by the misdetection on very small slur and notch patterns.

Interestingly, the specificity of the method to detect slur/notch was higher than its sensitivity, but the opposite result was reported for midslur/notch detection (Table 3). When comparing the result from each observer, the lower overall specificity for midslur/notch was significantly influenced by one observer's result which was accounted by the high false positive detected by automatic method. This result also reminds us that the interpretation of slur/notch, especially slur, varies among clinicians due to lack of standard definition. Therefore, the use of a computer algorithm offers an improved precision by systematically applying consistent rules.

In this study, we developed a series of algorithms to automatically compute all measurements and decision rules for ECG-based strict LBBB criteria. In the meantime, the algorithms and detection procedures are embedded into a software package that was designed with a user interface enabling simple review the automatic measurements and their manual adjustment when needed. More importantly, this method will be integrated into our newly developed QuAReSS software which automatically calculates QRS Selvester Score for patients meeting strict LBBB criteria (Strauss et al., 2011) and therefore ensuring a fully automatic computation of the Selvester score without manual preselection of the tracings (strict complete LBBB pattern).

6 | CONCLUSIONS

We propose a computer-based software for the automated detection of complete left bundle branch block in heart failure patients. Our validation demonstrates that the proposed automated system achieves approximately 95% accuracy that is equivalent to the level of interobserver variability. With such level of accuracy, the proposed automated system could be potentially used to facilitate large-volume screening of patients eligible for implantation of CRT and improve their outcomes.

REFERENCES

- Birnie, D., Ha, A., Higginson, L., Sidhu, K., Green, M., Philippon, F., ... Tang, A. (2013). Impact of QRS morphology and duration on outcomes after cardiac resynchronization therapy. *Circulation. Heart Failure*, 6, 1190–1198.
- Galeotti, L., van Dam, P. M., Loring, Z., Chan, D., & Strauss, D. G. (2013). Evaluating strict and conventional left bundle branch block criteria using electrocardiographic simulations. *Europace*, 15, 1816–1821.
- Goldenberg, I., Kutiyfa, V., Klein, H. U., Cannom, D. S., Brown, M. W., Dan, A., ... Moss, A. J. (2014). Survival with cardiac-resynchronization therapy in mild heart failure. *The New England Journal of Medicine*, 370, 1694–1701.
- Mascioli, G., Padeletti, L., Sassone, B., Zecchin, M., Lucca, E., Sacchi, S., Boggian, G., ... Sinagra, G. (2012). Electrocardiographic criteria of true left bundle branch block: A simple sign to predict a better clinical and instrumental response to CRT. *Pacing and Clinical Electrophysiology*, 35, 927–934.
- Moss, A. J., Hall, W. J., Cannom, D. S., Klein, H., Brown, M. W., Daubert, J. P., ... Zareba, W. (2009). Cardiac-resynchronization therapy for the prevention of heart-failure events. *The New England Journal of Medicine*, 361, 1329–1338.
- Nikoo, M. H., Aslani, A., & Jorat, M. V. (2013). LBBB: State-of-the-art criteria. *International Cardiovascular Research Journal*, 7(2), 39–40.
- Peterson, P. N., Greiner, M. A., Qualls, L. G., Al-Khatib, S. M., Curtis, J. P., Fonarow, G. C., ... Masoudi, F. A. (2013). QRS duration, bundle-branch block morphology, and outcomes among older patients with heart failure receiving cardiac resynchronization therapy. *JAMA*, 310, 617–626.
- Risum, N., Strauss, D. G., Sogaard, P., Loring, Z., Hansen, T. F., Bruun, N. E., ... Kisslo, J. (2013). Left bundle-branch block: The relationship between electrocardiogram electrical activation and echocardiography mechanical contraction. *American Heart Journal*, 166(2), 340–348.
- Strauss, D. G., & Selvester, R. H. (2009). The QRS complex—A biomarker that “images” the heart: QRS scores to quantify myocardial scar in the presence of normal and abnormal ventricular conduction. *Journal of Electrocardiology*, 42, 85–96.
- Strauss, D. G., Selvester, R. H., & Wagner, G. S. (2011). Defining left bundle branch block in the era of cardiac resynchronization therapy. *The American Journal of Cardiology*, 107, 927–934.
- Tian, Y., Zhang, P., Li, X., Gao, Y., Zhu, T., Wang, L., ... Guo, J. (2013). True complete left bundle branch block morphology strongly predicts good response to cardiac resynchronization therapy. *Europace*, 15, 1499–1506.
- Wong, J. A., Yee, R., Stirrat, J., Scholl, D., Krahn, A. D., Gula, L. J., ... White, J. A. (2013). Influence of pacing site characteristics on response to cardiac resynchronization therapy. *Circulation. Cardiovascular Imaging*, 6(4), 542–550.
- Xia, X., Wieslander, B., Strauss, D. G., Wagner, G. S., Zareba, W., Moss, A. J., et al. (2015). Automatic QRS selvester scoring system in patients with left bundle branch block. *Europace*, 18(2), 308–311.
- Zareba, W., Klein, H., Cygankiewicz, I., Hall, W. J., McNitt, S., Brown, M., ... Moss, A. J. (2011). Effectiveness of cardiac resynchronization therapy by QRS morphology in the multicenter automatic defibrillator implantation trial-cardiac resynchronization therapy (MADIT-CRT). *Circulation*, 123, 1061–1072.
- Zong, W., Moody, G. B., & Jiang, D. (2003). A robust open-source algorithm to detect onset and duration of QRS complexes. *Computer in Cardiology*, 30, 737–740.

How to cite this article: Xia X, Ruwald A-C, Ruwald MH, et al. Validation of an automatic diagnosis of strict left bundle branch block criteria using 12-lead electrocardiograms. *Ann Noninvasive Electrocardiol*. 2017;22:e12398. <https://doi.org/10.1111/anec.12398>