Testing the Quality of 12 Lead Holter Analysis Algorithms

R Fischer[1], MF Sinner[2], R Petrovic[3], E Tarita[3], S Kääb[2], TK Zywietz[3]

[1]Department of Biometrics, MHH, Hannover, Germany
[2]Medical Department I, University Hospital Munich, Germany
[3]Biosigna–Institute for biosignal-processing , Munich, Germany

Abstract

Computer assisted analysis of ECGs has recently experienced a remarkable revival due to its increased importance as biomarker in clinical trials. Especially 12-lead 24 h Holter-recording is nowadays routinely applied. But how accurate is the automatic computer analysis of 12 lead Holter recordings? While there are several data bases to evaluate the performance of 2-lead Holter algorithms, e.g., MIT-BIH, AHA etc., to our knowledge, there is no annotated data base for real 12-lead Holter algorithms. We have therefore created a new 12 lead ECG data base containing 50 Holter ECGs from the University Hospital Munich, Campus Großhadern. The following paper describes the data base and its application for evaluating different 12 lead Holter analysis algorithms in detail.

1. Introduction

Since the early application of computer assisted ECG analysis, quality assurance has been a major issue. It is difficult to evaluate and compare the performance of ECG algorithms, unless standardized procedures and data bases are available. Several projects to define standards for ECG analysis were started already in the eighties, for example the CSE project (Common Standards for Quantitative Electrocardiography) [1,2], and continued in the nineties [3,4]. Most of the results have been implemented into international norms published e.g. by the IEC[5,6] or ANSI[7,8]. Today, there are several ECG data bases available for testing (and partly required by the international norms). To name the most common: MIT-BIH[9], AHA (mainly for testing 2-Lead Holter recordings) or the CTS and CSE data bases to validate the accuracy of resting ECG measurements and the diagnostic power. However, the existing data bases have their limitations. Technologically, there has been a big step forward over the last years: Besides the standard ECG applications, such as resting and 2 or 3 lead Holter recordings, today event recorder (24 h times 7 days) and 12 lead Holter recorders (with nearly the same diagnostic power as resting ECG) are routinely applied as well, especially for cardiac safety trials. To our knowledge, there is no consistent data base available for testing real 12 lead Holter recordings. The aim of this work was therefore to develop a data base and the corresponding analysis tools; on the one hand side to allow quality assurance of newly developed algorithms and on the other hand to compare the performance of different algorithms. This will be of significant importance, as we expect that the huge amount of ECG data acquired in clinical trials in future will be analysed automatically rather than manually, as of today.

2. Methods: creating a new data base

The following chapter describes in detail:

- The standard 12 Lead long-term ECG database,
- Annotation of the database,
- The test programs applied to the database, and
- The software to compare annotated reference data with the test results of a tested algorithm.

2.1. Technical description of the data base

To allow consistent testing over the complete data base, all records were recorded with the same standard electrode positions for 12 leads. All records were digitized at 500 Hz with an amplitude resolution of 410 increments per mV using a 12-bit ADC, which allows a range of about 10mV. All records of the database are acquired using the same hardware equipment, the same sample frequency and amplitude resolution. The data have been recorded during routine clinical practice at the

University Hospital Munich, Campus Großhadern. The selected 50 records are a subset of about 600 Long-Term ECGs containing a broad set of pathologies for QRS (e.g. Bundle Branch Blocks or premature ventricular beats (PVC)), rhythm (e.g. supra ventricular premature beats, Bigeminus, Trigeminus), atrial activity (e.g. atrial fibrillation/flutter) and heart rate abnormalities like Bradycardia or Tachycardia. The duration of the original ECG record is 24 hours, but only one hour has been selected to limit the amount of data for the quality test. The collected ECGs contain types of noise, which are typical for Holter ECG recordings:

- Baseline noise (e.g. respiration),
- High frequency noise,
- Line frequency noise,
- Beatlike artifacts (because of temporary pressure on electrodes) and
- High range noise (in case electrodes become loose).

2.2. Annotation of the database

Annotations were made by experienced annotators applying the following steps (see Fig. 1):

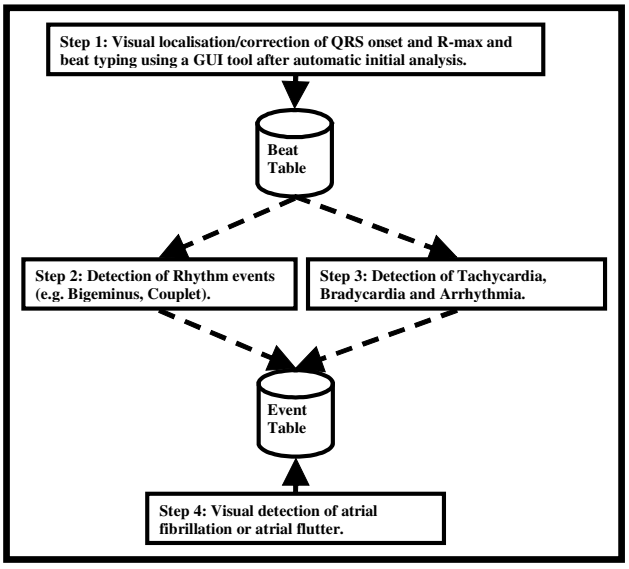

Figure 1: Steps of the annotation.

The automatic initial analysis was done with the HES [10] long-term analysis algorithm for 12 leads. Because some manufacturers use the maximum of R (R-max) as trigger for beat localization, while other manufacturers prefer the QRS onset trigger, the database performs both triggers. The beat localization and typing has been done with a specific proprietary graphical user interface tool, called WINHESPR (see figure 2) allowing to insert/shift and remove QRS-onset and the maximum R beat triggers (R-max) with an accuracy of 1ms. The annotation differentiates between five beat types: Normal beats (=1st dominant type), 2nd dominant type (=Bundle branch block), aberrant beat, premature ventricular contraction (PVC) and supra ventricular premature contraction (SPC).

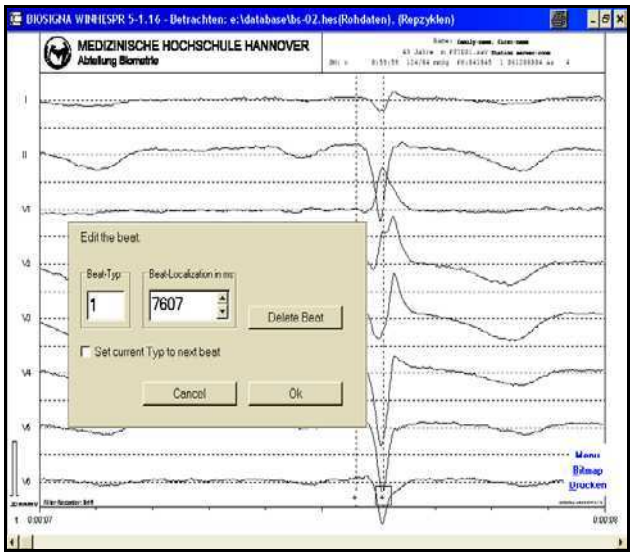

Figure 2: Beat localization and typing using the proprietary tool WINHESPR.

After beat localization/typing an automated detection of rhythm events (HES-Algorithm) was applied to get the following beat related events:

- Bigeminus,
- Couplet, Triplet, Salve,
- Pause1 (RR > 2*mean RR),
- Pause2 (3 * mean RR ≤ RR < 4 * mean RR),
- Asystolie and
- Sequence of SPC

The next step was automated detection of (10s) data intervals for Tachycardia (HR >90 BPM), Bradycardia (HR < 60 BPM) and Arrhythmia (RR differs more than 30%). The last step for annotation was the visual detection of atrial fibrillation/flutter by using the tool WINHESPR.

The results of the annotation steps were stored within a "beat table file" (with the annotation of all beats) and an "event table file" (with all annotated events) per data base record.

2.3. Control software for testing

The core of the quality test is a software tool that automatically compares (within a given tolerance) beat localization and beat typing. The format for the reference

and the test beat table is proprietary, but conversion from/into the widely applied PhysioNet format is implemented to allow the use of other common tools or databases. The reports of the developed software tools are in text format for better readability and simple data import/export.

3. Statistical description of the data base

3.1. Reference annotation

The distribution of premature ventricular contractions (PVC) and supra ventricular premature contractions (SPC) within the data base is shown in fig. 3. We have selected records such that a typical distribution of PVC and SPC is present over all records.

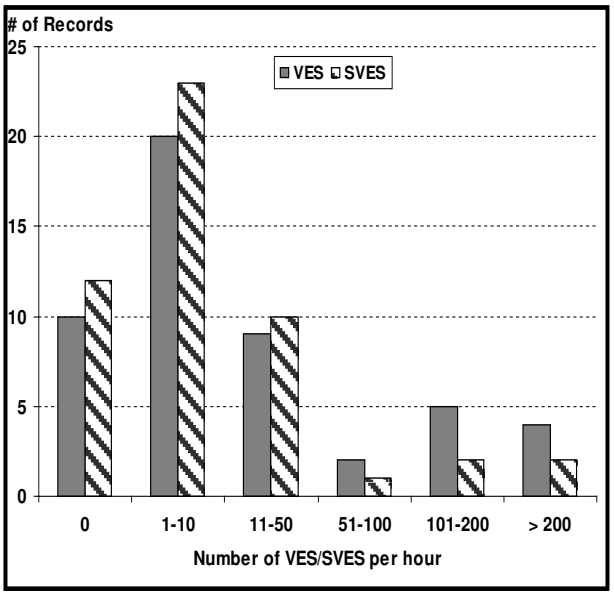

Figure 3: Distributions of PCV and SPC within the 50 record database (each single duration one hour).

Fig. 4 shows the distribution of relative arrhythmia, bradycardia and tachycardia data intervals. Bradycardia is rare (only six of 50 records contain more than 10% bradycardia data intervals), but a few records contain a low heart frequency for more than 50% of the complete recording duration.

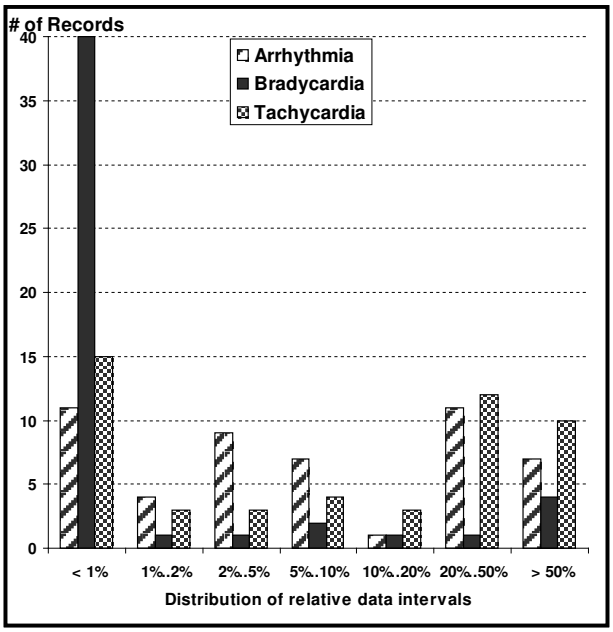

Figure 4: Distribution of data intervals with Arrhythmia, Bradycardia and Tachycardia.

The automated event detection and the visual check for atrial events obtained the following results:

- 18 ECG records contain at least one and up to 34 data intervals of Bigeminus with total durations between 2s and 246s (prevalence 0.3%).
- Six records contain up to 18 couplets and three records contain up to four triplets.
- The RR event Pause1 (= one beat missed) is present in three records with a maximum rate of 14.
- Atrial fibrillation/flutter is present in eight records with a total maximum duration of 26 minutes (prevalence 1.0%).

4. Application of the data base

As a first test, the developed tools were applied to the newly developed CorDIA algorithm (Cor Diagnostic and Interpretation Algorithm) of Biosigna. After running the algorithm with the 50 ECG records of the data base, the results of localization and beat typing were analyzed for performance. Table 1 depicts the localization accuracy for both localization trigger points. The selected threshold for the maximum difference between reference and test localization was set to 30ms. The main conclusion of this table is that the accuracy of the R-max localization is better than the QRS-onset accuracy (mean and minimum value larger, std. dev. smaller). One factor of this result may be the lower signal-to-noise (SNR) value at QRS-onset.

Table 1: Localization Accuracy of all beats and records:

	QRS onset localization	Rmax localization	Diff.
TP beats	265081	265870	789
FN beats	1230	441	-789
FP beats	1174	385	-789
Mean accuracy in %	99,19	99,67	0,48
Min. accuracy in %	89,46	98,04	8,58
St. dev. in %	1,54	0,4	-1,14

The result of the typing test is depicted in Table 2. The first and second columns contain the number of normal beats and bundle branch blocks. Using the entries of the table, the derived performance measurements for PVC and SPC are:

- PVC: Sensitivity 81.4%, Specificity 99.7% and Positive Predictive Value (PPV) 78.3%,
- SPC: Sensitivity 94.7%, Specificity 98.0% and PPV 17.6%.

Table 2: Results of the typing test:

		Reference Typing					
	Total	1st dom.	2nd dom.	SPC	PVC	Abberant	No loc.
Test Typing	1st dom.	256804	67	50	268	12	81
	2nd dom.	0	0	0	0	0	0
	SPC	5027	0	1121	135	3	91
	PVC	340	4	5	2006	15	192
	Abberant	0	0	0	0	0	0
	Artefact	5	0	0	1	1	21
	Undetermined	6	0	0	0	0	0
	No loc.	359	0	7	54	21	0
	Sum	262541	71	1183	2464	52	385

Because the tested CorDIA algorithm provides two localization triggers, this allows the testing tools to check the coherence between localization accuracy and mistyping of beats. Tab. 3 depicts that the R-max localization is very accurate for all three types of PVC detections (true positive, TP; false positive, FP; false negative, FN), while the QRS-onset localization differs significantly. The mean deviation of QRS-onset for the correct detected PVC beats is also accurate, but the mean QRS-onset of the false positive PVC is more than 34 ms early, while the mean QRS-onset of the false negative PVC is about 14 ms late. So in conclusion, in the CorDIA algorithm, the main reason for the false detected PVC beats seems to be the inaccurate QRS-onset localization.

Table 3: Coherence of localization accuracy and beat typing for PVC:

		Mean localisation deviation in ms at	
PVC typing	Beat number	QRS-onset	R-max
TP	2006	-0,04	0,00
FP	364	-34,17	-0,21
FN	404	13,72	-0,31

5. Discussion and conclusions

A new 12 lead Long-term database for quality test has been presented. This database allows ECG equipment manufacturers to test the quality of their automatic ECG analysis algorithms using native 12 lead ECG records without calculation of non-present leads. The manually annotated ECGs contain data of typical noise levels and represent all diagnostic relevant rhythm events. With the help of the data base, manufactures can easily identify the weaknesses of their algorithm and improve systematic the quality and performance of the analysis.

The database, the conversion and comparing tools will be available shortly.

References

[1] Willems JL, Abreu-Lima C, Arnaud P et al, Zywietz Chr, The diagnostic performance of computer programs for the interpretation of electrocardiograms, New Engl J Med, December 19, 1991;325:1767-73.

[2] Willems JL, Arnaud P, van Bemmel JH, Degani R, Macfarlane PW, Zywietz Chr, for the CSE Working Party. Common Standards for Quantitative Electrocardiography: Goals and Main Results, Meth Inf Med, FK Schattauer Verlag 1990;29:263-71.

[3] Zywietz Chr, Willems JL, Arnaud P, van Bemmel JH, Degani R, Macfarlane PW, Stability of computer ECG amplitude measurements in the presence of noise. Computers and Biomedical Research 1990;23:10-31.

[4] Zywietz Chr, Willems JL, European Conformance Testing Services for Computerized Electrocardiography - New Procedures and Standards - J Elect Cardiol 1993; Suppl 26:137-46.

[5] Moody GB, Mark RG. The Impact of the MIT-BIH Arrhythmia Database, Engineering in Medicine and Biology Magazine, IEEE, 2001, Vol 20;45-50.

[6] IEC 60601-2-51 Ed. 1.0. Medical electrical equipment - Part 2-51: Particular requirements for safety, including essential performance, of recording and analysing single channel and multichannel electrocardiographs. 2005.

[7] IEC 60601-2-47 Medical electrical equipment – Part 2-47: Particular requirements for the safety, including essential perf., of ambulatory electrocardiographic systems. 2001.

[8] ANSI/AAMI EC11. Diagnostic electrocardiographic devices, 2ed. 1991/(R)2001.

[9] ANSI/AAMI EC57. Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms. 1998/(R)2003.

[10] HES (Hannover ECG System, developed at Biosigna).

Address for correspondence

Dr. Tosja K. Zywietz
Biosigna Medical Diagnostics
Lindwurmstrasse 109
80337 Munich, Germany
Email: zywietz.tosja@biosigna.de