Increasing Patient Safety in Drug Trials with Computer Based Analysis – a Study with 13.000 Resting ECGs

TK Zywietz[1,3], A Khawaja[1], R Petrovic[1], R Fischer[2], J Reinstaedtler[3]

[1]Biosigna – Istitute for biosignal-processing and system-analysis, Munich, Germany
[2]Dept of Biometrics, MHH, Hannover, Germany
[3]Cardinal Health Research Services, Wuerzburg, Germany

Abstract

Today, the pharmaceutical industry relies on ECG recordings (digital Holter or 12 lead resting ECG) to ensure cardiac safety of subjects in drug trials. State of the art is a central ECG analysis in ECG core labs, where the ECGs are analyzed fully manual by specially trained technicians or experienced cardiologists. The ECGs are commonly analyzed within 48 hours after upload to the core lab, independent on whether the ECG is critical or not. In order to reduce the response time in case of critical ECGs, we have developed a 12 lead resting ECG analysis algorithm to quantitatively classify the ECGs (according to typical criteria applied by pharmaceutical sponsors). As bottom line statement, the ECGs are classified into "normal" (N), and "abnormal" (AN). The basic idea was to identify any abnormality and then prioritize the ECGs for over-read in the core lab. The algorithm was tested and validated against a data base of 12.980 ECGs of 1.223 patients, recorded in a drug trial and analyzed in a professional core lab. The following paper describes our methodology, and the results obtained with the above described data base. We also provide an outlook on how our new algorithm will be applied in future drug safety trials.

1. Introduction

Drug safety is the major concern for all pharmaceutical companies on the one hand and obviously for the patients on the other hand. Some drugs may increase the risk of arrhythmias or critical cardiac events. Some time ago, the U.S Food and Drug Administration and the European Medicines Agency therefore published their recommendations for ECG QT/QTc studies (Thorough QT studies, TQT) to evaluate potential arrhythmogenic effects of drugs [1]. The basic assumption behind is that the QT-interval in the surface ECG can be applied as biomarker for the arrhythmogenic potential of drugs. Since all ECGs are currently evaluated manually by experienced staff, QT studies are very time consuming. Consequently, there have been several projects to automate the QT analysis based on computer algorithms [2, 3]. However, many drug safety and efficacy trials apply ECGs in a more general fashion to monitor patients without specifically evaluating the QT-interval. The ECG is used to ensure patient safety by applying a defined set of ECG parameters for evaluation of changes throughout a study (typically around 150 parameters). These are commonly defined by the medical team of the study-sponsor. The major intention in these trials is to detect any change in the ECG during the trial relative to a baseline ECG. The current state-of-the-art is, as for the TQT studies, a central ECG collection with manual over-read. In a Phase III trial, there can be easily 50.000 – 100.000 ECGs for evaluation. The interpretation of ECGs after the recording in central core labs takes typically 48 hours, depending on the requirement of the sponsor. However, this means there will be quite a long time before a patient is re-evaluated at the trial site and may be withdrawn from the study, if clinically necessary. We have extended our HES algorithm for real time analysis of 12 lead resting ECGs to overcome these shortcomings of the traditional core lab analysis procedure. The following chapter describes our methodology, the third chapter the application of our algorithm to a data set of 12.980 ECG data from a clinical trial, and the fourth chapter evaluates the results and provides an outlook.

2. Methods: Pre-classification of ECGs

The major intention of our development was to increase cardiac safety of patients in a trial by reducing the response time in the core lab after an ECG recording and second, to reduce the number of ECGs which must be evaluated manually. The basic idea behind the procedure is that all incoming 12-lead resting ECGs are immediately (in real time) analyzed with our automatic measurement and interpretation algorithm HES [4,5,6] on a core lab server or directly at the trial site. A second

algorithm then checks all sponsor specific classification parameters, which are applied as input, based on the HES measurements and interpretations (ECG pre-classification). If classified as *normal*, the ECG is sent to the core lab and evaluated within the regular 48 hours (common practice for all ECGs), if classified as *abnormal* the ECG is evaluated faster depending on the severity of the ECG abnormality or the change against the base line. The overall procedure is shown in figure 1. As classification criteria, the algorithm applies approximately 150 parameters of the ECG.
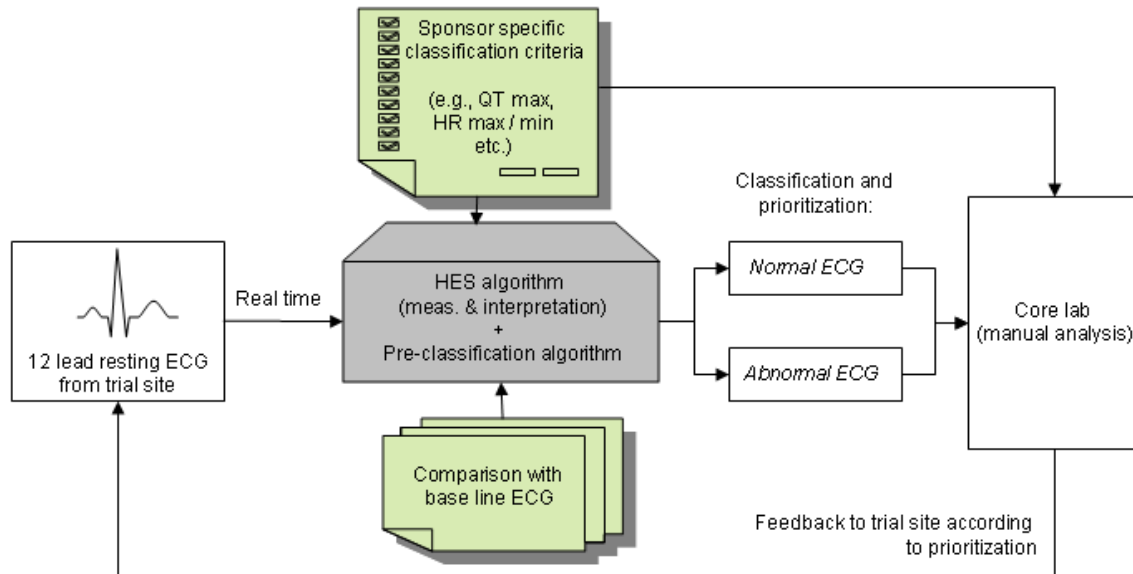

Figure 1. Method of ECG pre-classification for clinical trials.

As input, the pre-classification algorithm uses sponsor-defined thresholds such as, e.g., QTcB > 450 ms, HR > 100 (tachycardia), or HR < 58 = bradycardia, among many others. The pre-classification algorithm follows eight steps:

1. Read the results vector of the HES algorithm
2. Read the sponsor specific thresholds and definitions
3. Check for rhythm abnormalities
4. Check for repolarization abnormalities
5. Check for Hypertrophies or conduction abnormalities
6. Check for abnormal measurements
7. Check for QRS-T abnormalities
8. Check for ECG quality (if the quality is too low, the ECG is classified as abnormal)

If a baseline ECG is available, the pre-classification algorithm also takes changes against the base line ECG into account, which are also defined by the sponsor, e.g., a maximum of ΔQT of 20 ms, etc. (serial ECG comparison). The algorithm itself is trained such that its sensitivity to detect abnormal ECGs is extremely high (near 100%). It must be ensured, that all really critical ECG are also classified as abnormal and quickly checked by the core lab experts. On the other hand, ECGs classified "false positive" as abnormal are uncritical – they would just be analyzed faster than required.

3. Results: Test of algorithm against manually evaluated data from a core lab

We have run and tested our described methodology on a large data set of 12.980 ECGs from a trial with 1.223 subjects with an average age of 43,2 years. 683 subjects were female and 540 male. For all subjects, the first ECG was defined as "baseline ECG". All ECGs were analysed in the core lab according to predefined criteria from the sponsor. We applied the same criteria to our new pre-classification algorithm. The comparison between the core lab analysis and the results from our algorithm contains two parts: The qualitative interpretation and the quantitative comparison of the main ECG measurement parameters. For the quantitative analysis, the data were cleaned such that all ECGs with no measurements by either the core lab or the algorithm were excluded (to eliminate "0" values from the statistical analysis).

3.1. Qualitative classification

The core lab classified the ECGs into four classes: *Normal* ECGs (N – 10.381 ECGs), *Abnormal Not*

Clinical Significant (ANCS – 2.089 ECGs), *Abnormal Clinical Significant* (ACS – 296 ECGs) and *Unable to Evaluate* (UTE – 214 ECGs). The HES algorithm together with the new pre-classification algorithm detected 290 of the 296 ACS (sensitivity of 98,0%). However, three of these records were classified as ACS because of Wolff-Parkinson-White Syndrome, which is not yet implemented in the algorithm. The algorithm classified 2.781 ECG as N, i.e. 27%. By further increasing the sensitivity by reducing the thresholds, we were able to detect all ACS except one (with WPW-syndrome only), but at the cost of many more false positive. In terms of technical quality of the ECG, the algorithm is very critical – for 3.026 ECGs the algorithm proposed to repeat the recording due to bad signal quality or a potential lead reversal.

3.2. Quantitative comparison

It is important to mention, that the manual and automatic measurement procedures currently follow a different approach: In the core lab (manual approach) all interval measurements were done on lead II only, each on three consecutive normal beats. The final measurement values were then built as average from these intervals. The heart rate was also calculated from three consecutive RR-Intervals. (This is common practice and state-of-the-art – manual analysis of all leads and beats would be too time-consuming). The HES algorithm uses the full analysis approach: After localization and classification of all beats (into normal or PVC, premature ventricular contraction), the algorithm averages all normal beats to a so called representative beat. The wave points are calculated on this representative beat. This procedure offers a great advantage, because HF-noise is automatically reduced while maintaining all features of the signal. The intervals are calculated over ALL twelve beats. To derive global wave durations, the algorithm always uses the first on-set and latest off-set of the corresponding interval. Obviously, in comparison to the manual approach, this leads to a systematic difference in the measurements, as discussed below.

Figure 2 shows the correlation between the heart rate (HR) measured by the Core lab versus the automatic measurements of HES. The correlation (linear regression) is very high ($R^2 = 0,96$), i.e. the agreement is very good. The average heart rate is 68 for both, manual and automatic. Also the standard deviation (STD) over the complete data set is nearly the same (see also table 1). However, as expected, there are a few outliers, we have analysed in detail. In case of bigeminy, the algorithm calculates the HR over all beats (including PVC), while the core lab only counts the normal beats. This explains why the HR calculated by HES is higher by a factor of almost 2 in those cases.
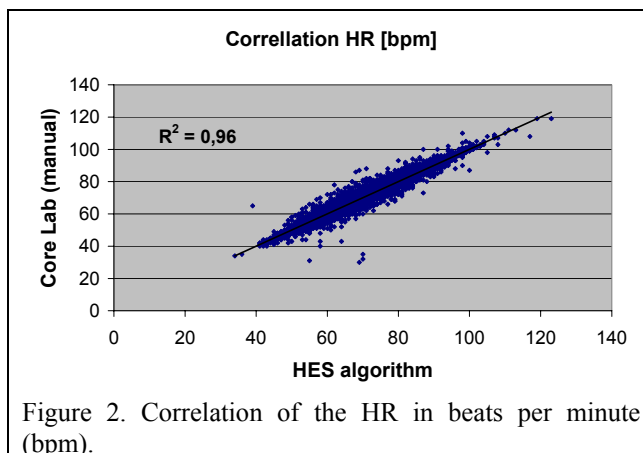

Figure 2. Correlation of the HR in beats per minute (bpm).

The correlation of the QT-interval is lower ($R^2 = 0,75$), but still reasonably good (see figure 3). Since the algorithm uses the first QRS-onset and the latest T-offset over all leads (instead of only lead II), the QT interval is systematically longer (average over all records 18 ms).
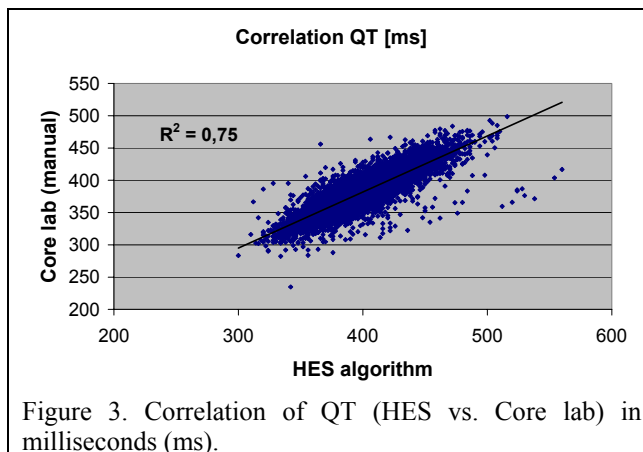

Figure 3. Correlation of QT (HES vs. Core lab) in milliseconds (ms).

Again, the main outliers were analyzed in more detail. The cases where HES calculated QT significantly too long were related to two extreme cases: First, the ECG showed a very prominent U-wave, which was included in the T-wave by HES but not in the manual analysis. Second, the T-wave was extremely flat in all channels and the algorithm wrongly identified the next P-wave as T-offset.

In some cases, the algorithm put T-offset too early. This was related to extreme cases, where the ECG showed a very long PR-interval (>220 ms), in combination with a bradycardia. The algorithm will be further improved also to handle these extreme cases.

The correlation of the PR interval is reasonably high ($R^2 = 0,83$). The mean deviation over all ECGs is

negative 7,7 ms, i.e. the algorithm systematically calculates the PR interval shorter, which is again due to the 12 lead rather than 1 lead analysis as explained above.

	HR [bpm]	QT [ms]	QTcB [ms]	QTcF [ms]	PR [ms]
R^2	0,96	0,75	0,57	0,51	0,83
φ HES	68	389	422	414	155
φ Core Lab	68	380	403	395	161
STD HES	10,4	27,7	20,8	18,3	21,2
STD Core Lab	10,6	27,7	24,5	21,3	22,6
φ Deviation	0,1	17,9	11,1	18,5	-7,7

Table 1. Quantitative comparison of the Core Lab results versus the HES and pre-classification algorithm.

4.   Discussion and conclusions

We have presented a new procedure and algorithm to classify ECGs in clinical drug trials. The results show that abnormal ECGs can be detected with a very high sensitivity (comparable to the manual analysis in a professional core lab). The procedure allows increasing patient safety by analysing ECGs in real time and prioritizing the manual over-read accordingly (pathologic ECGs can be over-read faster instead of "first in, first out"). However, the sensitivity must be further improved by implementing the analysis of the WPW-syndrome. Also, the specificity should be increased while maintaining the high sensitivity, which is always a trade off. Next, the algorithm will be extended to distinguish between Abnormal Clinical Significant and Abnormal not Clinical Significant ECGs.

Although the measurement accuracy is already reasonable, the T-offset detection will be further improved for special pathologies, e.g. for AV-block (long PR) in combination with an extreme bradycardia. In order to quantitatively compare the algorithm and the core lab results in terms of performance, the algorithm and the core lab must apply exactly the same procedure to derive wave durations, as described above.

Besides enhancing patient-safety in trials, the new procedure will also allow to reduce the number of ECGs which must be over-read manually. Since the algorithm is more sensitive than the core lab, one could decide to only manually evaluate ECGs classified as abnormal, but no ECGs classified as normal. In this specific study that would have led to savings of nearly 30% of the manual over-read while at the same time increasing patient safety by reducing the response time of the core lab.

Acknowledgements

References

[1] ICH E14, EMEA, European Medicines Agency, 2005.
[2] Handzel R, Garnett C, Li M, McNitt S, Polonski S, Xia X, Couderc JP, Computers in Cardiology 2008; 35:693-696.
[3] Couderc JP, Zhou M, Sarapa N, Zareba W, Investigating the effect of sotalol on the repolarization intervals in healthy young individuals, J. of Electrocardiology 2008; 41:595-602.
[4] Zywietz C, Borovsky D, Faltinat D, Klusmeier S, Schiemann W. The Hannover ECG System HES. Trends in Computer-Processed Electrocardiogramms, J. H. van Bemmel/J.L. Willems, eds., North-Holland Publishing Company 1977:159-167.
[5] Zywietz C, Borovsky D, Götsch G, Joseph G, Methodology of ECG Interpretation in the Hannover Program. Meth. Inform. Med. 1990;29: 375-385.
[6] Willems JL, Abreu-Lima C, Arnaud P, van Bemmel JH, Brohet C, Degani R, Denis B, Gehring J, Graham I, van Herpen G, Machado H, Macfarlane PW, Michaelis J, Moulopoulos S, Rubel P, Zywietz C, The diagnostic performance of computer programs for the interpretation of electrocardiograms. New England J. of Med. 1991;325:1767-1773.

Address for correspondence

Dr. Tosja K. Zywietz
Biosigna – Institute for Biosignal-processing and System Analysis
Lindwurmstrasse 109
80337 Munich, Germany
Email: zywietz.tosja@biosigna.de