Use of ECG Quality Metrics in Clinical Trials

M Vaglio, L Isola, G Gates, F Badilini

A.M.P.S. llc, New York, USA

Abstract

We present a software package aimed to provide a viable mean to manage and automatically classify large amounts of ECGs on the basis of analytical metrics directly computed from the digital waveforms (such as noise content, heart rate regularity, repolarization complexity) and parametric metrics derived from ECG measurements (such as R and T wave amplitudes, QT/QTc intervals).

Single or combined sets of metrics are used to generate ECG quality classes using, for each of the considered metric, predefined thresholds. The number of ECG subsets, as well as the thresholds to be applied, can be determined from the respective built-in distributions derived from an internal database of approximately 300.000 digital ECGs selected from previously conducted clinical trials.

The use of this technique can significantly optimize the quality assessment and the management of digital ECG and secondarily improve the quality of recorded ECG in both clinical trials and common clinical practice.

1. Introduction

In the last decade, regulatory organization such as the Food and Drug Administration (FDA) have enforced the widespread usage of digital ECG which has become the dominant modality of acquisition, processing, measurement, storage and submission of ECGs from clinical trials.

Every clinical trial, including the Thorough QT Study (TQTS), has to cope with the risk of inclusion of ECG data corrupted by noise content and other distorting factors, particularly when the conditions during ECG acquisition are not rigorously controlled. The low frequency (LF) noise is associated with baseline wander or drift in the ECG [1], possibly associated with poor skin-electrode impedance, while the high frequency (HF) noise is typically associated with artefacts originating from non myocardial sources (e.g. skeletal muscle tremor, respiration, other mechanical activity).

Noise content depends on the quality of ECG acquisition at the study sites; however central laboratories do not generally provide any quantitative information of the amount of data corruption, and specifically do not report statistics on the numbers of ECGs incompatible with reliable manual or automated measurements, or even worse ECGs of unacceptable quality. Moreover, annotations typically reviewed by human readers (for example the QT interval) at the central laboratory have inherent within-reader, between-reader as well as day-to-day variability.

Early phase studies (such as phase 1 trials) typically enrol healthy subjects with normal ECG waveforms, and fully automated analysis by ECG manufacturer algorithms that could produce results of equal value to those from central ECG laboratories, is being considered as an option by regulatory agencies. While on one side fully automated measurement approaches would reduce costs, on the other they would necessarily need to rely even more on the strict control of the quality of ECG collected data.

Late phase clinical trials are characterized by additional challenges, since ECGs are acquired directly on the population targeted for a given compound, and are likely to include more abnormalities affecting the morphology of the signal. In these studies, the usage of a robust quality assessment approach could thus provide an added value as a way to screen problematic records before revision by cardiologists at the central ECG laboratory.

In this work, we present an application specifically designed to assess quality of large number of ECG records, employing the usage of several quantitative metrics either analytically computed on the ECG signal or derived from parametric measurements.

2. Methods

2.1. System overview

We have developed a software tool (FDAEcg Suite) aimed to provide e viable mean to manage and automatically classify large amount of ECGs on the basis of built-in quantitative metrics.

FDAEcg is a windows-based application developed in C++ which can simultaneously load large quantities of ECG records from different public domain data standards (such as HL7 XML, also known as aECG, ISHNE and

SCP) or other formats exported by different ECG manufactures such as Mortara, GE, and Philips XMLs.

Once the ECGs have been imported into the system it is possible to display single or multiple ECG records with various modalities: for example the ECGs can be grouped and sorted by protocol (grouping together ECGs from the same study), site identifier (geographical site where the record has been collected), subject identifier and by visit identifier. If available, the annotations (calliper positions, for a QT interval for example) can also be visualized on the ECG tracing.

In addition to sorting the records by demographics, visualization can be done according to one or more of the computed metrics, for example sorting the ECG records from the one with least noise content to the one mostly corrupted. A more detailed description of the available metrics is described in the next paragraph.

2.2. Quality metrics

All the metrics are computed starting from a common signal-processing workflow:

- Computation of the sequence of QRS complexes under sinus rhythm and of the abnormal (non sinus) beats, using cross-correlation criteria.
- For each of the continuous leads, computation of the representative waveforms (median beats), after removal of baseline wander based on a cubic-spline interpolation algorithm [2-4].
- Computation of amplitude and interval parameters from the representative waveforms of each lead. More specifically, PR, QRS, QT intervals, P, R and T wave amplitudes are automatically computed.

2.2.1. Analytical metrics

The following metrics are computed directly from the ECG signal:

- HF noise: a residual ECG signal is first derived by subtracting the computed median beat on each QRST beat complex in the rhythm tracing of each lead. The residual signal is subsequently high-pass filtered (using a 4^{th} order Butterworth high-pass filter, with cut-off frequency of 40 Hz [5, 6]) and the Root Mean Square (RMS) value is computed.
- LF noise: a cubic spline interpolation is first fitted through all the QRS onsets detected on the rhythm data of each lead; the RMS of the cubic-spline curve is then computed.
- Localized HF noise: localized HF noise metric is also computed as the RMS value in a window of ± 50 ms centered around a specific annotation calipers (for example the T-wave offset position). Localized HF noise metrics can be very useful to assess measurement specific confidence indexes.
- Similarly, it is possible to compute a localized LF Noise metric based on the slope of the ECG signal computed between the QRS onset and T-wave offset points.

All noise metrics can be computed either on individual leads (to provide noise content information of a given lead) or on set of leads, by averaging the values on the individual lead noise over the group of leads of interest.

- Heart-rate (HR): this is simply the heart rate of the ECG record. It is computed using the rhythm data and it is based on the computation of the sequence of sinus rhythm QRS with the exclusion of supraventricular and ventricular beats.
- QRS Regularity score: this is the percentage of normal QRS complexes detected by the automatic algorithm versus artefacts and/or abnormal beats.
- Repolarization complexity: this score assesses the complexity of the repolarization segment of a given lead and it is based on the regularity (number of zero-crossing after low-pass filtering) of the first derivative signal.

2.2.2. Parametric metrics

The parametric metrics are associated with specific measurements that are computed on the ECG signal and can be divided in the following subgroups:

- Direct metrics: PR, QRS and QT intervals; P, R and T wave amplitudes.
- Derived metrics: corrected QT intervals using different correction formulae (e.g. Bazett or Fridericia).
- Protocol-related metrics: percentage of successfully measured annotations, according to a predefined measurement protocol**.**

All parametric metrics are computed using an algorithm embedded in a computerized on-screen application which has been previously described (CalECG [7, 8]) and which has been compared with other existing measurement methods [9]. CalECG can automatically measure intervals and amplitudes on both rhythm and representative waveforms, and on individual leads or on global leads (when applicable). For the purpose of this study, interval-related metrics (PR, QRS and QT) are based on a global representative lead whereas amplitude-related metrics are computed on the representative beats of each individual lead.

In order to provide reference values, the metrics

where computed on a large and heterogeneous ECG cohort. These ECGs include standard 12-leads resting ECGs from different clinical trials and Holter 10 seconds extractions automatically obtained with a dedicated application [10].

Table 1. Quality metrics from AMPS internal database of approximately 300.000 digital ECGs .

Quality Metric	Mean ± STD
All Freq. noise on all leads (µV)	66.4 ± 60.1
HF noise on all leads (µV)	4.10 ± 3.89
HF noise around T-wave offset (µV)	5.84 ± 2.15
LF noise on all leads (µV)	21.1 ± 30.9
LF noise around T offset (µV/s)	167 ± 64.0
Heart Rate (bpm)	65.4 ± 11.4
QT (ms)	395 ± 32
QTcB (ms)	407 ± 23
QTcF (ms)	403 ± 21
PR (ms)	171 ± 22
R-wave Amplitude lead II (µV)	1388 ± 497
T-wave Amplitude lead II (µV)	412 ± 179
T-wave complexity lead II	1.47 ± 0.97
QRS regularity (%)	99.9 ± 1.15

2.3. Metrics applications

FDAEcg Suite have many potential implementations; in the following sections two practical examples that have already used are reported.

2.3.1. Focused quality assessment

Clinical trials are characterized by large sets of ECGs recorded at different sites and annotated by technicians and/or cardiologists.

Before submission to the regulatory agencies, the quality metrics computed by FDAEcg Suite allow a focused and optimized review of subsets of ECGs flagged as "potentially unreliable".

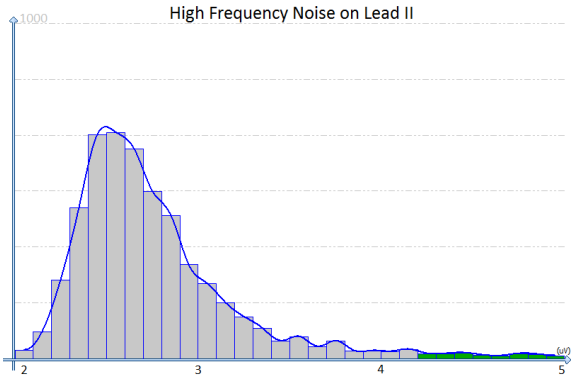

Figure 1: histogram of HF noise on 5000 ECGs; the right tail of the HF noise distribution is selected.

Thus, instead of being based on the entire set, ECG review can be focused on problematic ECGs with a higher likelihood to include measurement errors.

Within FDAECG Suite, the selection of the tails (outliers) within the study distribution of a given metric (Figures 1 and 2), generates an interactive review of the selected ECGs, thus permitting either the correction of specific errors of the identification of study biases.

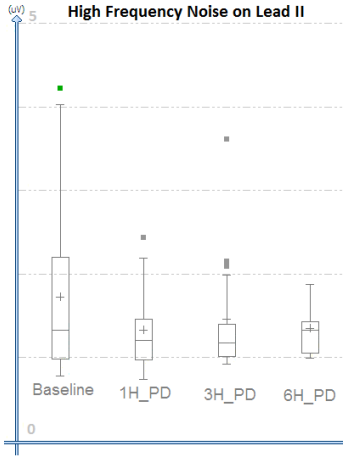

Figure 2: boxplot of HF noise grouped by Visit. The outlier noisy ECG at "Baseline" is selected and can be reviewed, by double-click on the graph selection.

The systematic use of an optimized quality assessment can help highlighting possible issues in the ECG collection process. For example, a given site of ECG collection generating lower than average quality data could be easily identified and immediate action (dedicated training session on ECG acquisition for the given site) could be taken.

2.3.2. Optimized ECG

In addition to this, in the past, cardiologists used to view the entire ECG cohort and manually place the calipers to annotate the ECG measurements (QT, PR, QRS intervals). Since several years a semi-automated approach is being used, where cardiologist review the ECG measurements (QT, PR, QRS intervals) automatically placed by automatic algorithm.

Here an ever more automated approach is presented: the possibility of manually review only a percentage of ECGs while the rest of the cohort will keep the automated measurements.

In this approach, the use of the above metrics can help the cardiologist to define the set of ECGs that need a manual over read, leaving reliable ECGs, where the automatic should perform at its best, with the automatically placed annotations.

The first steps will be the selection of the number of

quality bins (buckets) in which ECGs will need to be classified, typically three will be satisfactory: low, average and good quality. ECGs flagged as low-quality will need manual reading, ECGs flagged as average-quality will need cardiologist over-read and good quality ECGs won't need any further review process.

The core step in this workflow is the correct definition of the set of rules that will be used to flag the ECGs in the correct bucket. This can be achieved with the combination of various quality-metrics providing an overall score that will link each ECG to the right bucket.

Here below results on the implementation of optimized ECG workflow in a clinical study are reported. Ten seconds ECGs were extracted from Holter recordings in triplicate at ten timepoints over a two-days period on 24 subjects after the administration of a compound inducing QT prolongation. ECGs were annotated by a known central lab with manual reading, then they were processed by a fully-automated reading process with CalECG and finally a 5% revision was achieved on the 5% most noisy (HF) ECGs.

Figure 3 shows that there are significant differences between manual and fully-automated readings, but, with as little as 5% revision, the automatic readings are comparable to the manual annotations.

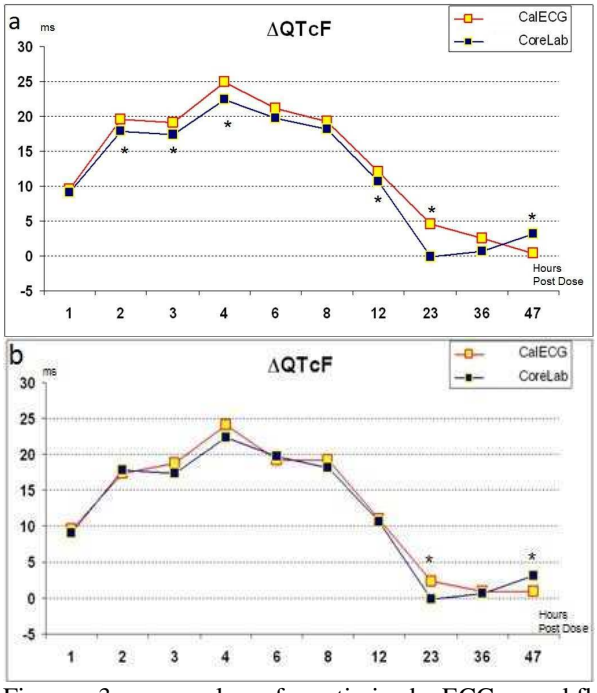

Figure 3: example of optimized ECG workflow implementation: a) comparison of fully automated approach versus central lab b) comparison of optimized workflow with 5% revision versus central lab. Significant different ($p<0.05$, paired t-test) are indicated with *.

3. Conclusions

The usage of advanced quantitative metrics can drastically improve the quality of ECG annotations performed during clinical trials.

Two possible applications of the use of quality metrics within clinical trials have been shown, one applied to the standard clinical trial workflow, the other highly automating the ECG workflow and reducing human review.

References

[1] Pipberger H. Computer analysis of electrocardiograms. In: Waxman and Stacy, Computers in Biomedical Research. New York: Academic Press, 1972;1:382.

[2] Badilini F, Moss AJ, Titlebaum EL. Cubic spline baseline estimation in ambulatory ECG recordings for the measurement of ST segment displacements. Proc Conf IEEE Eng Med Biol Soc 1991;13:584-585.

[3] Badilini F, Pagani M, Porta A. HeartScope: a software tool addressing autonomic nervous system regulation. Computers in Cardiology 2005;32:259-262.

[4] Greville TN. Theory and applications of spline functions. New York: Academic Press, 1969.

[5] Oppenheim AV, Schafer RW. Digital Signal Processing. Prentice-Hall, 1975.

[6] Berson AS, Ferguson TA, Batchlor CD. Filtering and sampling for electrocardiographic data processing. Computers and Biomedical Research 1977;10:605-616.

[7] Badilini F, Sarapa N. Implications of methodological differences in digital electrocardiogram interval measurement. J Electrocardiol 2006;39:152-156.

[8] Sarapa N, Mortara JL, Brown BD, Isola L, Badilini F. Quantitative performance of E-Scribe warehouse in detecting quality issues with digital annotated ECG data from healthy subjects. J Clin Pharmacol 2008;48:538-546.

[9] Panicker GK, Karnad DR, Natekar M, Kothari S, Narula D, Lokhandwala Y. Intra- and interreader variability in QT interval measurement by tangent and threshold methods in a central electrocardiogram laboratory. J Electrocardiol 2009;42:348-352.

[10] Badilini F, Vaglio M, Sarapa N. Automatic Extraction of ECG Strips from Continuous 12-lead Holter Recordings for QT Analysis at Pre-Scheduled versus Optimized Time Points. A.N.E. 2009;14:S22-S29.

Address for correspondence.

Martino Vaglio
A.M.P.S. llc
Via P. Paolo VI, 34
25018 Montichiari (BS) Italy
vaglio at amps-llc.com