Assessment of ECG Signal Quality Index Algorithms Using Synthetic ECG Data

Aron B Syversen[1], Zhiqiang Zhang[1], Jonathan A Batty[1], Matti Kaisti[2], David Jayne[3], David C Wong[1]

[1] University of Leeds, Leeds, UK
[2] University of Turku, Turku, Finland [3] Leeds Teaching Hospitals NHS Trust, Leeds, UK

Abstract

This study evaluated the performance of several publicly available signal quality indices (SQI) in assessing the quality of synthetic electrocardiogram (ECG) signals with varying categories and levels of noise. We used an existing framework to generate realistic ECG signals with controlled increases in heart rate, power line interference, white noise, and motion artifacts. ECG signals were generated at the threshold of acceptable and unacceptable outputs from each SQI across four categories of noise. The 16 signals were then evaluated by a cardiologist based on four specific criteria and these responses were compared against the SQI outputs. Results showed that the four SQI's were inconsistent with each other; they also frequently disagreed with the cardiologist assessment. When assessing whether the ECG could be used to 'estimate a plausible heart rate', the cardiologist assessment agreed with the SQI outputs in between 9/16 and 15/16 cases. When asked whether the ECG was 'clinically useful', the cardiologist assessment only agreed with SQI's in between 4/16 and 10/16 cases. The findings from this study underscore the importance of users critically analysing the outputs of SQI's as their suitability may be limited to only basic heart rate extraction from ECG signals, rather than more comprehensive clinical applications.

1. Introduction

The increasing use of wearable devices has led to an associated rise in the volume of ECG data being collected. These devices are often used to collect data outside of controlled clinical settings and thus are particularly susceptible to noise. Therefore, the automation of signal quality indices (SQI) is an important task with several applications. Reductions in signal quality could be immediately identified to ensure proper fitting of the device or electrodes. Periods of poor signal quality could simultaneously be highlighted to avoid unnecessary processing.

Researchers have attempted to automate the assessment of signal quality using SQI. Methods vary from the application of feasibility rules on extracted features including heart rate (HR), to more complex models using deep learning. There is limited research on whether SQI are consistent with each other and most have only been tested on datasets collected in controlled clinical conditions.

Previous work has reviewed SQIs to provide an overview of their methods and limitations [1]. Work has also assessed the performance of several SQI tools on different datasets [2]. However, neither of these report on the publication of SQI tools for open-access use across research. One further systematic review published in 2022 concluded that of some 19 SQI tools published between 2012-2022, none published corresponding code [3]. Recently, several open-source SQI tools have become readily accessible.

In this paper, we compare the consistency of several open-source SQIs tools on ECGs with differing types and levels of noise. To do this, we use a new synthetic ECG toolbox that allows us to specify noise mechanisms and control variation in the signal. Outcomes from the SQI are then compared against a clinician's evaluation.

2. Methods

To assess the consistency of SQI tools, we generated a pipeline to control type and level of noise in ECG signals and output an associated SQI. The SQI outputs were then compared against feedback from a cardiologist (Figure 1).

2.1. Synthetic ECG generator

We used Karhinoja et al's framework for generating synthetic ECG signals [4], which is available at `https://github.com/UTU-Health-Research/framework_for_synthetic_biosignals`. The process for generating signals is broken into three segments: beat interval generation, signal generation and noise generation. The beat generation process accounts for average heart rates, breathing modulations and long-term correlations and can be set to specific parameters.

 Page 1 ISSN: 2325-887X DOI: 10.22489/CinC.2024.270

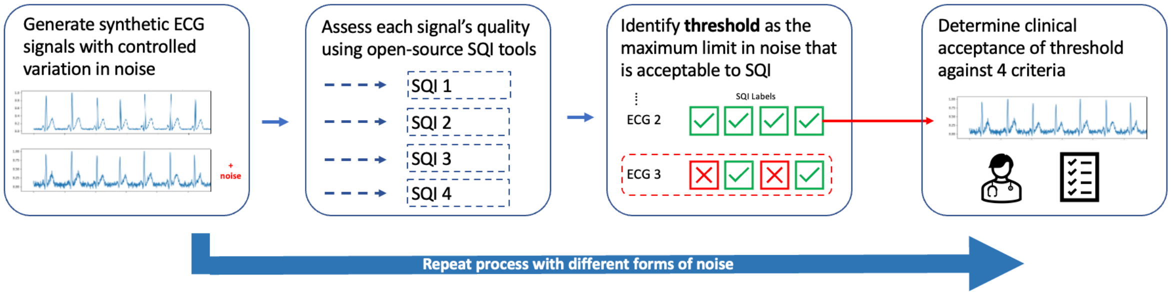

Figure 1. Pipeline to assess SQI outcomes.

Noise was added to the signal through several forms. Point frequency noise was added to the signal to replicate power-line interference (PLI). We used a fixed frequency of 50Hz to model European power supplies and then varied the magnitude in our experiments. White noise, representing electronic thermal noise, was also added with varying magnitude in our experiments. Signals were further augmented through the addition of pre-defined motion artifact types representing muscle artifact noise, hand movement, walking, and baseline wander. Heart rate, although not a form of signal noise, was included as a category of noise to add further common variation to the ECG signals.

2.2. Generating signals

To assess the SQIs, varying amounts of noise were added corresponding to four sources: Heart Rate, White Noise, Power-line Interference and Motion Artefacts.

Each source of noise was investigated independently. The amplitude of noise was initially set using the default parameters of the framework to generate a realistic clean signal of heart rate 80 bpm [4]. For heart rate, white noise and power-line interference the amplitude was increased in set increments; increments were selected empirically so that there was visible change in the signal. The remaining types of noise were fixed at the default values. For white noise and power-line interference the signal-to-noise ratios (SNR) are reported in table 1.

For any set of noise parameters, the resulting ECG signal differs depending on the random number seed - i.e. there is some stochasticity in the synthetic data generation process. To account for this, we generated 100 signals for each set increment of noise. We calculated the binary output of the SQI (acceptable or unacceptable) and reported the proportion of acceptable signals for each set of 100.

A set of parameters was deemed to produce 'unacceptable' ECGs if the proportion was less than 0.5 (i.e. less than 50 of 100 signals were acceptable). Within each noise category, the level of noise was increased until all SQI tools produced 'unacceptable' labels.

For each of the four noise sources, an example ECG was generated using the parameters at which each SQI returned an 'unacceptable' result. These 16 signals were assessed by a cardiologist (author: JB) to determine their clinical acceptability. The cardiologist was blinded to the SQI labels and assessed each ECG signal on four criteria:

1. Can you estimate a plausible HR?
2. Can you locate all QRS complexes?
3. Can you locate all P and T-waves?
4. Is the signal clinically useful?

For criteria 4, 'clinically useful' was defined as 'allow full assessment of heart rate, rhythm and beat-to-beat morphology'. We report the agreement between the SQI output and each of the four criteria responses by the cardiologist.

2.3. Signal Quality Indices (SQIs)

Four SQIs were selected due to open-source code:

SQI1: Orphanidou et al. 2015 Orphanidou et al. produced a four-step algorithm [5]. The first three criteria are feasibility rules: HR-check of 40 and <180 bpm, whether maximum space between subsequent R-peaks is <3 s and whether the maximum to minimum beat-to-beat interval ratio is <2.2. If the signal passes, an adaptive template matching threshold (0.66) is employed to check for the regularity of the signal. Code was taken from [6].

SQI2: Zhao & Zhang 2018 Zhao & Zhang's method to evaluate signal quality combines simple heuristic fusion to extract features and fuzzy logic to evaluate quality [7]. The SQI classifies signals into either 'unacceptable', 'barely acceptable' or 'excellent'. For consistency, we combine the latter two into an 'acceptable' category. We used the SQI as implemented in the neurokit2 package [8].

SQI3: Kramer et al. 2022 Kramer et al. propose a three stage signal quality classification algorithm: whether the signal was stationary, a HR-check of >24 and <300 bpm and a SNR check. Code was taken directly from [9].

SQI4: Elgendi et al. 2023 Elgendi et al. extends *SQI3* with an CNN-LSTM model [10]. Signals are converted into spectrograms using a Short-Time Fourier Transform and are fed into the CNN-LSTM classifier. We used the pre-trained model taken from [10].

3. Results

3.1. SQI comparison

The threshold at which further increases in variation would lead to a SQI label of 'unacceptable' is recorded in Table 1. No amount of white noise was sufficient for *SQI3* to report the signal as unacceptable, reported as 'N/A'. For the *Motion Artifact* source, four different categories of artefact were assessed. An 'N/A' result here indicates that no *Motion Artifact* led to an unacceptable label; only *SQI1* labelled the ECG with 'walking' artifact as unacceptable.

Table 1. Threshold at which ECG becomes 'unacceptable', for each noise source.

Noise Source	Threshold for unacceptable label			
	SQI1	**SQI2**	**SQI3**	**SQI4**
Heart Rate (bpm)	155	925	495	255
White Noise (dB)	4.75	1.41	N/A	1.32
Power Line (dB)	4.13	10.71	0.70	0.70
Motion Artefacts	Walking	N/A	N/A	N/A

SQI1 was the most sensitive to changes in both *Heart Rate* and *White Noise* (Figure 2). Figure 3 presents example ECG signals generated with *white noise* at a threshold that is barely 'acceptable' (threshold = 0.5) for *SQI1* and *SQI2*. These ECGs were also labelled as clinically uninterpretable by the cardiologist.

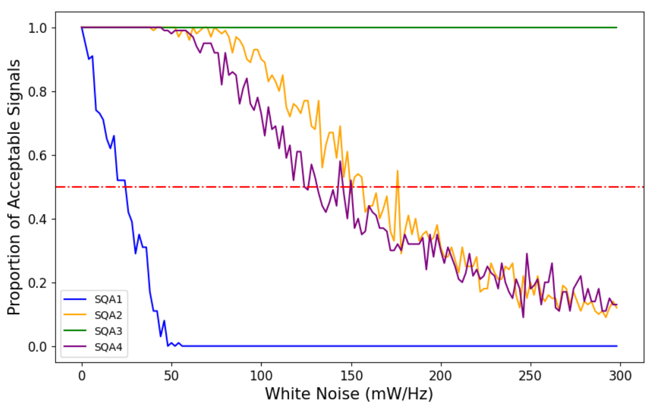

Figure 2. Increase in *White Noise* (original units) plotted against the proportion of 'acceptable' labels for each SQI tool. The highlighted threshold of 0.5 indicates the point at which ECG signals are considered to be unacceptable.

3.2. Comparison with clinical expert

The agreement between the responses from the cardiologist with the SQIs is shown in Table 2. Each cell represents

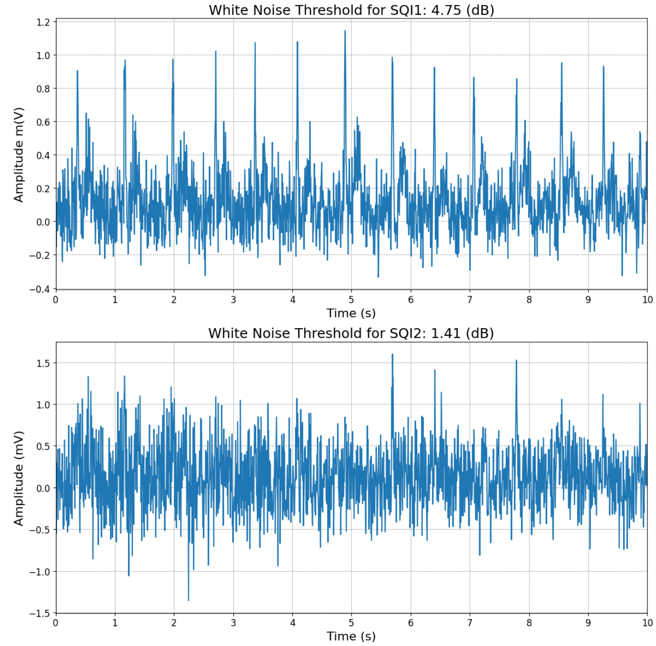

Figure 3. Comparison of ECG signals generated with White Noise set at the threshold between unacceptable and acceptable signals for SQI1 and SQI2. Clinical reviewer reported that the upper signal met criteria 1 and 2 but did not meet criteria 3 and 4. The lower signal did not meet any of the four criteria.

Table 2. The number of cardiologist assessments that agreed with the SQI labels. This is assessed for each of the 16 signals meaning each value can range 0-16 (a score of 16 shows that the SQI label matches the cardiologist criterion label for all 16 generated signals).

	SQI1	**SQI2**	**SQI3**	**SQI4**
Criterion 1	12	9	12	15
Criterion 2	10	11	12	13
Criterion 3	10	7	4	7
Criterion 4	10	7	4	7

the number ECGs in which the cardiologist assessment, for each criterion, agreed with each SQI label.

Overall, *SQI3* had the lowest agreement with only 32 of the 64 total criteria from the cardiologist matching the SQI label across all 16 signals. *SQI2* had slightly higher agreement with 34 criteria matching the SQI label. Both *SQI1* and *SQI4* scored an equal total score for agreement between all four criteria and their SQI labels (42/64).

SQI1 was the most consistent and showed similar agreement with the cardiologist across all four criteria (range 10-12). *SQIs 2, 3 and 4* however, displayed highest agreement with the cardiologist for criteria 1 and 2 but less agreement with criteria 3 and 4. *SQI3* in particular agreed with the cardiologist on criteria 1 and 2 for 12/16 ECGs but only agreed with criteria 3 and 4 for 4/16 ECGs.

Agreement between labelling from the SQIs and the cardiologist also differed by source of noise. For signals with *White Noise*, *SQI1* had the most agreement with the cardiologist. When adding *Power Line interference* and *Motion Artefacts*, all four SQIs showed similar levels of agreement with the cardiologist. For increases in *Heart rate*, *SQI4* had the most agreement with labels from the cardiologist.

We further noted that *SQI4* produced inconsistent results as heart rate increased. Although the initial threshold at which ECGs were 'unacceptable' was at 255 bpm, further increases in heart rate, up to 400 bpm, were deemed to be 'acceptable'.

4. Discussion

This study investigated the performance of four publicly-available SQIs on synthetically-generated ECGs with different modes of noise. The SQI outputs were compared against labels provided by a cardiologist. The experiment yielded several key findings.

We found that *SQI1* and *SQI4* had the highest agreement with the cardiologist assessment [5][10]. This was particularly evident in SQI output when HR was increased; *SQI2* and *SQI3* labelled ECGs as acceptable even when heart rate was implausibly high. Similarly, only *SQI1* showed some sensitivity to changes in white noise. The overall agreement with the cardiologist across the four noise modes was relatively low. This underscores the need for users to critically evaluate the outputs of these SQI tools, rather than relying on them without careful analysis.

There was notable variation in the performance of the SQI tools across the various noise modes. On average SQIs showed highest agreement with the cardiologist when labelling variation in heart rate. SQIs showed very little sensitivity to changes in white noise and showed the lowest average agreement with the cardiologist here.

An important inference can be made from the SQI labels agreement with the cardiologist across the four criteria. The first two criteria related to whether HR can be extracted from the signal. For these, the SQIs showed moderately good agreement with the cardiologist. The final two criteria related to other clinical features of the ECG (e.g. identification of P and T-waves) and exhibited lower levels of agreement. We acknowledge that not all generated signals reflect realistic ECG noise levels. The primary objective was to stress-test the SQI tools by introducing elevated noise levels, rigorously evaluating their limitations beyond typical clinical scenarios.

In conclusion, the tested SQIs were inconsistent with each other and with an expert opinion. This suggests the limited suitability of these SQI tools for clinical applications beyond simply identifying ECG signal segments with acceptable quality for basic heart rate extraction. Future work should consider SQIs that account for the clinical use case and derive SQIs that account for relevant morphology.

Acknowledgments

This work was supported in part by UK Research and Innovation (UKRI) [CDT grant number EP/S024336/1].

References

[1] Kuetche F, Alexandre N, Pascal NE, Colince W, Thierry S. Signal quality indices evaluation for robust ECG signal quality assessment systems. Biomedical Physics & Engineering Express. 2023 Aug 3;9(5):055016.

[2] Rahman S, Karmakar C, Natgunanathan I, Yearwood J, Palaniswami M. Robustness of electrocardiogram signal quality indices. Journal of the Royal Society Interface. 2022 Apr 13;19(189):20220012.

[3] van der Bijl K, Elgendi M, Menon C. Automatic ECG quality assessment techniques: A systematic review. Diagnostics. 2022 Oct 24;12(11):2578.

[4] Karhinoja K, Vasankari A, Jukka-Pekka S, Airola A, Wong D, Kaisti M. Flexible framework for generating synthetic electrocardiograms and photoplethysmograms. arXiv:2408.16291. 2024.

[5] Orphanidou C, Bonnici T, Charlton P, Clifton D, Vallance D, Tarassenko L. Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring. IEEE J Biomed Health Inform. 2015 May;19(3):832-8.

[6] Charlton PH. Signal Processing and Learning for Wearables [Internet]. GitHub; 2023 [cited 2024 March 10]. Available from: https://github.com/peterhcharlton/bsp-book

[7] Zhao Z, Zhang Y. SQI quality evaluation mechanism of single-lead ECG signal based on simple heuristic fusion and fuzzy comprehensive evaluation. Frontiers in Physiology. 2018 Jun 14;9:361542.

[8] Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SA. NeuroKit2: A Python toolbox for neurophysiological signal processing. Behavior Research Methods. 2021 Feb 2:1-8.

[9] Kramer L, Menon C, Elgendi M. ECGAssess: A Python-based toolbox to assess ECG lead signal quality. Frontiers in Digital Health. 2022 May 6;4:847555.

[10] Elgendi M, van der Bijl K, Menon C. An Open-Source Graphical User Interface-Embedded Automated Electrocardiogram Quality Assessment: A Balanced Class Representation Approach. Diagnostics. 2023 Nov 20;13(22):3479.

Address for correspondence:

Aron Syversen
Woodhouse, Leeds LS2 9JT
scabs@leeds.ac.uk