

SUPPLEMENT ARTICLE

Signal quality and patient experience with wearable devices for epilepsy management

Mona Nasser¹ | Ewan Nurse^{2,3}  | Martin Glasstetter⁴ | Sebastian Böttcher⁴  |
 Nicholas M. Gregg¹ | Aiswarya Laks Nandakumar⁵ | Boney Joseph¹  | Tal Pal Attia¹  |
 Pedro F. Viana^{6,7}  | Elisa Bruno⁶ | Andrea Biondi⁶ | Mark Cook³ |
 Gregory A. Worrell¹ | Andreas Schulze-Bonhage⁴  | Matthias Dümpelmann⁴  |
 Dean R. Freestone² | Mark P. Richardson⁶ | Benjamin H. Brinkmann¹ 

¹Bioelectronics Neurophysiology and Engineering Laboratory, Department of Neurology, Mayo Clinic, Rochester, Minnesota

²Seer Medical, Melbourne, Victoria, Australia

³Department of Medicine, St Vincent's Hospital, University of Melbourne, Melbourne, Victoria, Australia

⁴Department of Neurosurgery, Epilepsy Center, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Germany

⁵College of Medicine, University of Illinois, Peoria, Illinois

⁶Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, UK

⁷Faculty of Medicine, University of Lisbon, Lisbon, Portugal

Correspondence

Benjamin H. Brinkmann, Bioelectronics Neurophysiology and Engineering Laboratory, Department of Neurology, Mayo Clinic, Rochester, 200 1st St SW, Rochester, MN 55905.
 Email: Brinkmann.Benjamin@mayo.edu

Funding information

The Epilepsy Foundation of America's Epilepsy Innovation Institute My Seizure Gauge project

Abstract

Noninvasive wearable devices have great potential to aid the management of epilepsy, but these devices must have robust signal quality, and patients must be willing to wear them for long periods of time. Automated machine learning classification of wearable biosensor signals requires quantitative measures of signal quality to automatically reject poor-quality or corrupt data segments. In this study, commercially available wearable sensors were placed on patients with epilepsy undergoing in-hospital or in-home electroencephalographic (EEG) monitoring, and healthy volunteers. Empatica E4 and Biovotion Everion were used to record accelerometry (ACC), photoplethysmography (PPG), and electrodermal activity (EDA). Byteflies Sensor Dots were used to record ACC and PPG, the Activinsights GENEActiv watch to record ACC, and Epitel Epilog to record EEG data. PPG and EDA signals were recorded for multiple days, then epochs of high-quality, marginal-quality, or poor-quality data were visually identified by reviewers, and reviewer annotations were compared to automated signal quality measures. For ACC, the ratio of spectral power from 0.8 to 5 Hz to broadband power was used to separate good-quality signals from noise. For EDA, the rate of amplitude change and prevalence of sharp peaks significantly differentiated between good-quality data and noise. Spectral entropy was used to assess PPG and showed significant differences between good-, marginal-, and poor-quality signals. EEG data were evaluated using methods to identify a spectral noise cutoff frequency. Patients were asked to rate the usability and comfort of each device in several categories. Patients showed a significant preference for the wrist-worn devices, and the Empatica E4 device was preferred most often. Current wearable devices can provide high-quality data and are acceptable for routine use, but continued development is needed to improve data quality, consistency, and management, as well as acceptability to patients.

KEYWORDS

epilepsy, patient experience, signal quality, wearable devices

1 | INTRODUCTION

Wearable devices are increasingly common tools for seizure counting and management in clinical epilepsy, among other health applications.¹ This provides tools to help people with epilepsy and their physicians better understand and ultimately treat their seizures, and track treatment outcomes. Patient underreporting of seizures is a well-recognized problem in epilepsy,^{1,2} and objective seizure counting based on wearable devices may result in more effective management of medications and neuromodulation therapy, and may help in assessing candidacy for surgery. Despite extensive efforts in the field of seizure prediction,³ the unpredictability of seizures remains a significant difficulty for people with epilepsy regardless of seizure frequency.⁴ Noninvasive prediction of seizures through the tracking of temporal cycles,^{5–8} physiological signals,^{9,10} or their combination could prove transformative for many patients.

A key concern with wearable devices is the quality of the data recorded.¹¹ This becomes increasingly important in long-term device use, both for use in clinical trials and for ongoing epilepsy management. Automated analysis of signals with algorithms requires the ability to exclude data epochs corrupted by artifacts or without physiological content. Comparison between devices often focuses on seizure detection performance rather than objective and fundamental signal quality measures.^{12,13} Signal quality measures are simple to compute and allow for direct comparison of similar sensors from data recorded with or without seizures. Signal quality indexes (SQIs) are useful to validate the quality of devices before the commencement of trials and to compare across trials at different centers or with different endpoints. Furthermore, longitudinal analysis of signal quality from individual devices is useful to assess device adherence and maintenance requirements. The use of these methods along with automated data transfer and analysis can provide reports to physicians, people with epilepsy, and researchers to improve clinical care and advance epilepsy research.

Another important aspect of wearable technologies use is device design. Although various studies have surveyed people with epilepsy and their caregivers about their views on various aspects of wearable devices,^{14–16} these studies have not always focused on people with an experience of using wearable devices for epilepsy management. People with epilepsy who have personal experience with wearable devices may be able to provide more insightful feedback on preferences and concerns of long-term device use.

The present study addresses both of these problems of measuring device signal quality and assessing the preferences

Key Points

- Automated measures of signal quality are important for chronic monitoring with wearable biosensors
- Signal quality metrics for ACC, EDA, PPG, and EEG can discriminate good- from marginal-quality data and noise
- People with epilepsy on average have a positive overall response to the use of wearable devices and report a preference for wristbands
- Continued development of devices is needed to improve data quality and consistency, and subject comfort and acceptability

of people with epilepsy while using wearable devices. These results will guide future work on the long-term use of wearable devices for seizure detection and forecasting.

2 | MATERIALS AND METHODS

2.1 | Study design

Commercially available research-grade wearable devices were used by patients undergoing video-electroencephalographic (EEG) monitoring in both in-patient and at-home environments for durations of up to 10 days. Physiological signals utilized were accelerometry (ACC) to evaluate limb acceleration in three axes, electrodermal activity (EDA) to measure skin conductance (which varies with perspiration, reflecting sympathetic tone and psychological arousal), photoplethysmography (PPG) to evaluate microvascular blood volume changes, and EEG to measure cerebral electrical activity. Four commercially available, research-grade wearable biosensors with the ability to record ACC, EDA, and PPG, along with one sensor capable of recording scalp EEG, were assessed with patients with epilepsy undergoing EEG monitoring and healthy volunteer subjects. Patients were recruited at Mayo Clinic, King's College Hospital, Freiburg Hospital, and Seer Medical (Melbourne, Australia). Consecutive patients undergoing stereoelectroencephalographic surgery, and those who were admitted to epilepsy monitoring units or were scheduled for at-home ambulatory monitoring, were recruited for the study unless there were clinical reasons to omit patients, and patients with a range of seizure types

(focal, generalized, and electrographic) were included. From ACC, EDA, PPG, and EEG data, signal quality metrics were assessed retrospectively and benchmarked against data recorded while the device was not being worn. At the end of the recording period, patients were provided with a survey to assess their preferences and comfort with using wearable devices for seizure prediction.

2.2 | Physiological data collection

ACC data were recorded by the Empatica E4, Byteflies Sensor Dot, Biovotion Everion, and Activinsights GENEActiv devices, which were placed on the wrist, chest, arm, and wrist, respectively. For each device, several days of visually assessed good-quality ACC data from at least four patients was identified and compared with data recorded while the device was not worn by the subject. EDA was recorded by the Biovotion and Empatica devices, each from five patients. A few hours of data were recorded with each device placed on the subject over a cloth wrap, not in contact with the subject's skin, to represent poor-quality data. PPG data were collected from patients wearing the Empatica, Byteflies, and Biovotion devices. Segments representing noise with the absence of physiological activity for each device were obtained by a device not worn by a subject, but at rest on a stable surface. EEG data were recorded by the Epitel device. This device is attached to the patient's forehead, close to the hairline, using a provided adhesive sticker with saline gel cutouts over the electrodes, and records a bipolar EEG channel from two electrodes.

2.3 | Data labeling

Ten-minute, nonoverlapping segments of ACC data were scored using a binary scale, identifying segments as good-quality data (on patient) or noise (not on patient). Ten-minute epochs were chosen for ACC to accommodate sleep or sedentary periods, where movements may be infrequent. PPG data were scored over 1-minute nonoverlapping segments by reviewers using a 3-point scale, identifying segments as good quality (physiology easily visible with limited noise or artifact content), marginal quality (some physiology apparent but frequently obscured by noise or artifacts), or poor quality (no physiology apparent and signal dominated by noise or artifact). For both ACC and PPG, segments were reviewed and labeled using custom software written in MATLAB (MathWorks). The Byteflies and Biovotion devices provide PPG data with red, green, and infrared (IR) light sources. To minimize reviewer burden, the IR and green channels for Byteflies and Biovotion devices, respectively, were scored by the reviewers, and these scores were applied to all three PPG

signals to calculate SQIs. The IR PPG signal for the Byteflies device and green PPG for the Biovotion device each showed the greatest physiological signal component on visual inspection.^{17–19} The Empatica does not provide raw PPG values but rather provides blood volume pulse (BVP), a derived signal calculated from the red and green raw PPG signals. The BVP shows pulsatile blood flow features but is similar to the PPG signal, as the slowly varying respiratory component is removed. Therefore, BVP was analyzed using the PPG metrics. For EDA, 1-minute data segments with <4 seconds of artifact were labeled good quality. Marginal-quality data segments had both tonic and phasic EDA components, but each segment contained at least 4 seconds of artifacts (sharp changes¹⁷). The Empatica E4 device returns values <0.05 μ S when not in contact with skin (for example, with a garment under the device), and zero when entirely off the subject. The Biovotion Everion returns 0.047 mS when it is not in contact with skin and not on the patient. Data segments with amplitude at or below these values, and segments with amplitude changes < 0.01 μ S (0.01 mS for Biovotion) within 1 minute, were classified as bad. EEG data were scored using a threshold method. All data with a maximum bandwidth (defined below) of >75 Hz were considered good quality,²⁰ and data with maximum bandwidth > 35 Hz acceptable for recording epileptiform activity.²¹

2.4 | Signal quality indices

2.4.1 | Accelerometry

Prior studies have classified subject activities such as rest, walk, and run based on the frequency components of the ACC data acquired from wireless sensors.²² They showed that the highest amplitude of the frequency spectrum of ACC data for these movements is in frequencies < 5 Hz. To exclude low-frequency noise components, the power spectrum is calculated for frequencies > 0.8 Hz. To avoid ambiguity regarding the directional orientation of movement, the signal quality metrics were calculated on the root mean square (RMS) of the acceleration components of the three axes (ACC_x , ACC_y , ACC_z):

$$RMS = \sqrt{\frac{1}{3} (ACC_x^2 + ACC_y^2 + ACC_z^2)} \quad (1)$$

The signal quality metric for human movement in RMS ACC was defined by the ratio of the narrow spectral band covering the normal physiological range (0.8–5 Hz) to broadband spectral power (power ratio [PR]):

$$SQI_{PR} = \frac{P_{[0.8, 5]}}{P_{[0.8, F_n]}} \quad (2)$$

where F_n is the Nyquist limit frequency for the device. The power ratio of the signal was estimated as the power in the frequency range between 0.8 and 5 Hz divided by the broadband power from 0.8 Hz to the Nyquist frequency. The Empatica E4 records acceleration with a sampling frequency of 32 Hz. Since the sampling frequency of the Byteflies, Biovotion, and GENEActiv devices (48, 51.2, and 1000 respectively) are higher than the Empatica E4, the Empatica E4 Nyquist frequency limit (16 Hz) was used for all devices to allow direct comparison between devices. This power ratio should peak when physiological movements dominate the power spectrum and should take on low values when there is little physiological movement present. The power ratio was calculated for nonoverlapping 4 second segments, and average values over consecutive 10-minute segments are reported. Figure 1A shows the scaled RMS ACC signal in blue and the quality metric in red. The power spectrum of the RMS signal shows higher power in the range of 0-5 Hz (Figure 1B).

2.4.2 | Electrodermal activity

EDA provides time series recording of skin conductance, which is characterized by a slow varying tonic component

(skin conductance level) and a fast-varying phasic component (skin conductance response [SCR]). SCR has abrupt phasic increases in the conductance of the skin and usually has a faster rise time than decay time.²³ The SCR amplitude is defined as the difference between the skin conductance values at peak and the preceding trough. Here a minimum SCR amplitude of 0.01 μS is used to distinguish between physiological data and noise.²⁴ Also, the frequency of non-specific SCR (NS-SCR) has been suggested to measure tonic activity.²⁵ Typically the NS-SCR rate per minute is at least 1 (1-3 during rest and >20 in high-arousal situations).^{25,26}

The EDA signal was recorded using the Empatica E4 and Biovotion Everion in microsiemens and kilo-ohms, respectively. The EDA in microsiemens is related to resistance in kilo-ohms as its reciprocal ($\text{EDA}[\mu\text{S}] = 1/R[\text{k}\Omega] = 1000/R[\text{k}\Omega]$). For analysis, the Biovotion Everion data were converted to millisiemens to have its data in the same dynamic range as the Empatica E4 data. The sampling frequencies for the Empatica E4 and Biovotion Everion are 4 Hz and 1 Hz, respectively. The quality metric for EDA was called rate of amplitude change (SQI_{RAC}), and is the percent amplitude change calculated in concurrent 1-second windows. Sharp changes in signal amplitude of >20% increase or 10%

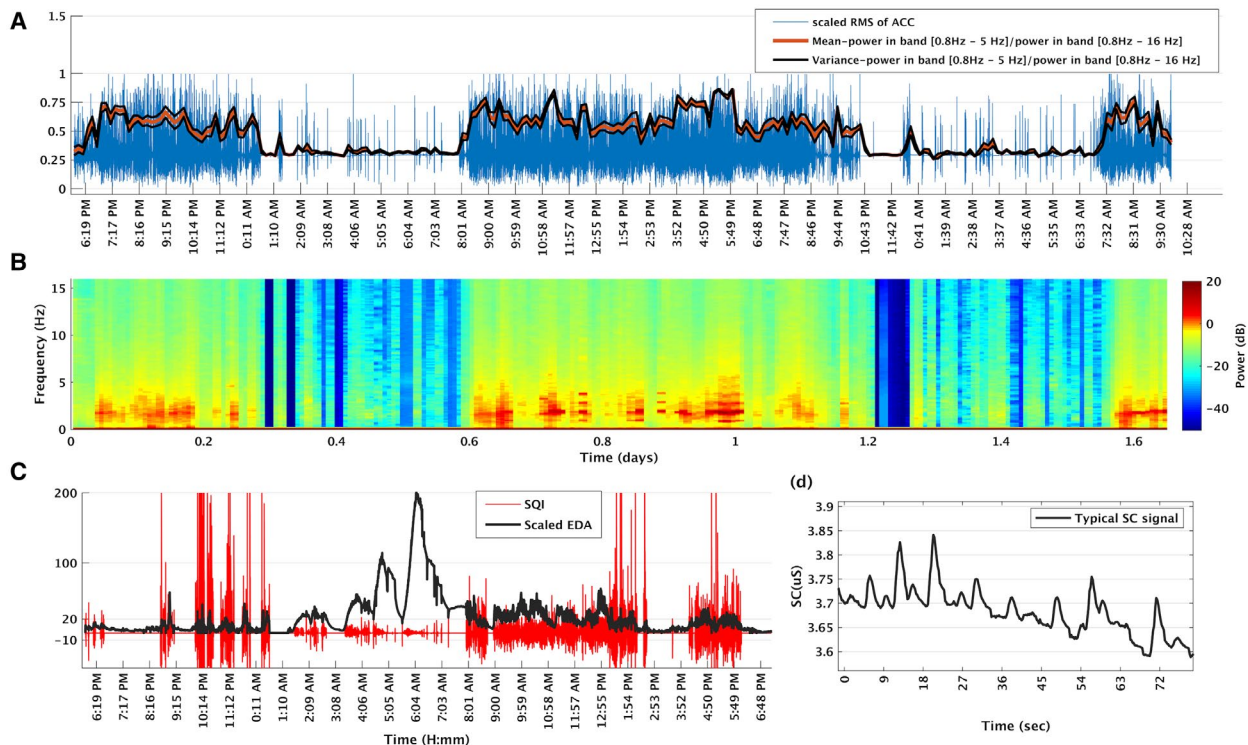


FIGURE 1 A, Scaled accelerometer data of Empatica E4 from a control subject in blue, and average power ratio quality metric over 10-minute segments in red. The mean $\pm$ variance of power ratio signal quality index (SQI_{PR}) is shown in red and black. The subject performed multiple activities, including walking, running, and biking. B, The corresponding power spectrum of the raw accelerometer data. C, The corresponding scaled electrodermal activity (EDA) signal (black) and rate of amplitude change SQI (SQI_{RAC} ; red). Zero SQI_{RAC} shows segments where the amplitude of the EDA signal is <0.05 μS (no skin contact) or the amplitude change is <0.01 μS in a 1-minute segment. D, An 80-second segment of EDA signal from the recording, likely illustrating nonspecific skin conductance response periodic changes consistent with high physical activity.²⁵ ACC, accelerometry; RMS, root mean square; SC, skin conductance

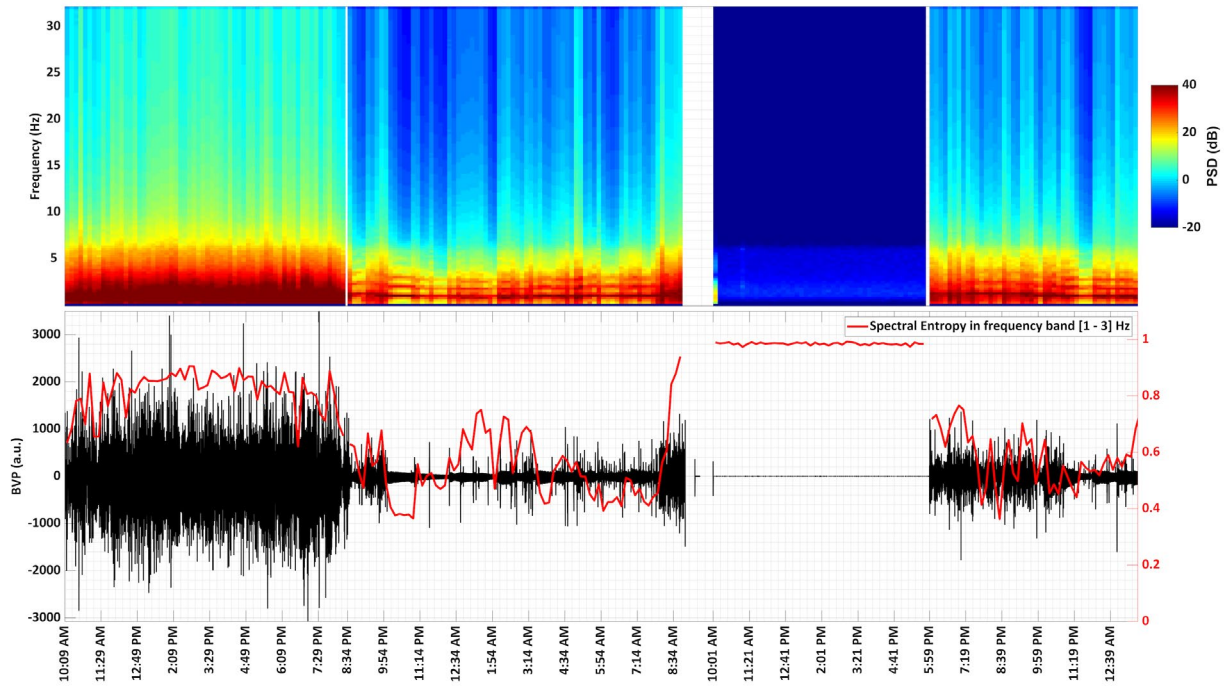


FIGURE 2 Typical on-patient photoplethysmographic (PPG) spectrogram depicted in the upper panel recorded with Empatica E4. Note that most of the signal power distributes among frequencies < 5 Hz and that the signal was not recorded continuously (white spaces in upper and lower panel). The corresponding raw PPG signal is depicted in the lower panel (black line) with the calculated signal quality index spectral entropy on top (red line). BVP, blood volume pulse; PSD, power spectral density

decrease per second are often correlated with subject motion and poor electrical contact with the skin,^{27,28} and these values are interpreted as artifact, or poor-quality data. If the amplitude of EDA changed <0.01 μ S for the E4 signal and 0.01 mS for the Biovotion signal over a 1-minute epoch, the device was likely not recording physiological signals, and the SQI of that minute of data was set to be zero. Figure 1C shows the scaled EDA signals from an E4 recording of a healthy subject, with the calculated SQI_{RAC} . During the day, when the subject is moving, the EDA signal has many sharp increases and decreases in amplitude, which represent motion artifacts. During the night, while the subject is relatively still, the SQI_{RAC} stays within the normal range due to consistent contact between the EDA electrodes and the skin.

2.4.3 | Photoplethysmography

Signal quality of PPG was assessed by calculating spectral entropy, which measures the "peakedness" of the frequency spectrum.^{29,30} Due to its oscillating nature, a good-quality PPG signal exhibits a peaked power spectrum and has lower spectral entropy than a PPG signal corrupted by noise, which exhibits a flatter spectrum with less pronounced spectral peaks, and correspondingly higher spectral entropy. Because most of the physiological PPG signal power lies within $f_1 = 1$ Hz and $f_2 = 3$ Hz (see Figure 2), it is possible to limit

the spectrum for calculating the spectral entropy (SE) to the frequency range of 1-3 Hz:

$$SQI_{SE}(f_1, f_2) = - \frac{\sum_{f_1}^{f_2} \hat{P}(f) \cdot \log_2 \hat{P}(f)}{\log_2(N)} \quad (3)$$

where $\hat{P}(f)$ is an estimate of the power spectral density at frequency f and N is the number of frequency bins in $[f_1, f_2]$. Note that $\hat{P}(f)$ is normalized such that it sums to 1 in the frequency band $[f_1, f_2]$ and, to get a smooth estimate, we took the average of $\hat{P}(f)$ over 15 nonoverlapping 4-second windows (resulting in 1-minute segments) before calculating spectral entropy SQI (SQI_{SE}). Figure 2 shows a typical Empatica E4 on-patient PPG signal with corresponding spectrogram and calculated metric. Values of SQI_{SE} close to 1 indicate noise segments (eg, day 2, 10:01 AM to 5:59 PM). Low values of SQI_{SE} correspond to high-quality data (eg day 1, between 9:54 PM and 11:14 PM). PPG signals corrupted by, for example, motion artifacts show values of SQI_{SE} close to 0.8 (eg, day 1, 10:09 AM to 7:29 PM).

2.4.4 | Electroencephalography

Signal quality of EEG was assessed by calculating the maximum bandwidth, a measure that estimates the highest frequency at which the recorded signal power is significantly

different from background noise.^{31,32} It was calculated as the highest frequency with power greater than a threshold,

$$P_t = 1.5 \times (Q_{75}(P_{Noise}) - Q_{25}(P_{Noise})) + Q_{75}(P_{Noise}) \quad (4)$$

where $Q_n(P)$ is the n th quantile of P and P_{Noise} is the mean of the frequency power in the 200- to 255-Hz band calculated every 10 seconds in a 10-minute window. The power spectrum was calculated from the squared amplitude of the Fourier transform. Noise was defined in this way because it was the highest spectral band below the Nyquist frequency (256 Hz), which can be assumed to have minimal scalp EEG power. Although P_{Noise} could contain a small amount of physiological information, we have approximated the noise in this way in lieu of hardware-based testing of the device.

2.5 | Survey collection

Patients with epilepsy undergoing clinical EEG monitoring were recruited to use wearable devices for the duration of their monitoring. Some patients were willing to wear multiple devices, wearing up to four devices at a time. Monitoring lasted at least 3 days in all cases. Some healthy controls were recruited as well to wear devices for a similar duration. At the end of the recording period, subjects completed a survey on a tablet device, answering questions about basic demographics, device preference (if multiple devices were worn), and impression of comfort. In the case where multiple devices were worn, device impression was only completed for the preferred device. Questions were answered on a 7-point Likert scale.

2.6 | Statistical methods

A t test was used to evaluate metric performances for ACC, PPG, and EDA data. Survey results were assessed by computing the mean response for each question and for each device. A two-way analysis of variance was conducted to

assess the effects of age and gender for each question. For each response, interaction between gender and age group was not found to be significant ($P < .05$). Hence, gender and age groups were analyzed as independent factors. No significant difference was found between recruitment sites, hence these were aggregated for analysis.

3 | RESULTS

3.1 | ACC data analysis

Data from 19 patients with epilepsy were included in the analysis of ACC data extracted from Empatica E4, Byteflies, Biovotion, and GENEActiv devices. Normal data were recorded with the device on the patient, whereas noise segments represent data recorded by a device resting on a stable surface (not worn). The SQI_{PR} (Equation 2) has higher values when the subject is physically active. The mean and standard deviations of the power ratio across 10-minute segments from patients wearing the device are reported in Table 1. Both average and standard deviation together could be considered as a metric to distinguish noise from acceptable data, as well as classifying the physical activities. During sleep, however, the metric may falsely classify the signal as noise when the subject shows no movement. In this case, tracking the metric over a longer time period is suggested (Figure 1A; 10:43 PM to 12:00 AM, second day, average $\pm$ standard deviation [SQI_{PR}] = 0.29 ± 0.003). Shorter time periods may be used during sleep if other signals, such as EDA and temperature, can be associated with ACC data to confirm the device is being worn. For example, in Figure 1A, from 12:44 to 1:19 AM (first day), data are classified as noise based on the ACC metric, and zero EDA signal in Figure 1C supports that. However, from 1:34 to 1:54 AM on the first day, which is classified as noise (device not on patient) using the proposed ACC SQI, the EDA signal has high-quality data, suggesting that the device is being worn while the subject is still. All four devices reported in Table 1 show similar quality of ACC data and are able to capture movement with comparable ability.

TABLE 1 SQI_{PR} for accelerometer data from Empatica E4, Byteflies, Biovotion, and GENEActiv

Device	Location	Patients, n	Duration, d	SQI_{PR} , avg $\pm$ SD (normal)	SQI_{PR} , avg $\pm$ SD (noise)	P
Empatica E4	Wrist	6	18	0.35 ± 0.086	0.29 ± 0.005	<.001
Byteflies Sensor Dot	Chest	4	10	0.32 ± 0.09	0.27 ± 0.006	<.001
Biovotion Everion	Arm	5	12	0.37 ± 0.10	0.29 ± 0.0017	<.001
Activinsights GENEActiv	Wrist	4	12	0.38 ± 0.09	0.30 ± 0.007	<.001

Note: Acceptable data compared with recorded noise.

Abbreviations: avg, average; SD, standard deviation; SQI_{PR} , power ratio signal quality index.

TABLE 2 SQI_{RAC} for electrodermal activity data from Empatica E4 and Biovotion

Device	Location	Patients, n	Duration, d	Amplitude threshold (normal)	SQI _{RAC} , avg $\pm$ SD% (good)	SQI _{RAC} , avg $\pm$ SD% (marginal)	Average percentage of marginal + noise segments per minute	P
Empatica E4	Wrist	5	16	SC > 0.05 μ S	2.05 $\pm$ 3.18 -1.28 $\pm$ 1.67	124.9 $\pm$ 458.2 -30.9 $\pm$ 22.7	35.77	<.05
Biovotion Everion	Arm	5	6	SC > 0.047 mS	1.17 $\pm$ 1.47 -1.08 $\pm$ 1.1	56.11 $\pm$ 50.34 -26.18 $\pm$ 15.58	87	<.001

Note: Acceptable quality data compared with recorded noise. The SQI_{RAC} is between -10% and 20% for good data. Amplitude changes of more than a 20% increase (positive values) or 10% decrease (negative values) per second are caused by artifacts and are often correlated with subject motion.

Abbreviations: avg, average; SD, standard deviation; SQI_{RAC}, rate of amplitude change signal quality index.

3.2 | Electrodermal data analysis

Data from 10 patients were selected for EDA analysis. The average and standard deviation of the SQI_{RAC}, the amplitude change per second, for positive (amplitude increasing) and negative (amplitude decreasing) SQIs separately, are reported for both devices in Table 2. The SQIs were compared with the reviewer labels and showed an average agreement of 86% and 90% between the labels and SQIs for the Empatica and Biovotion devices, respectively. A *t* test showed a significant difference in the SQI_{RAC} for data labeled as good quality compared to low quality (bad and marginal) with $P < .001$. Moreover, the SQIs of the data collected with the electrodes not in touch with the skin were compared with normal data, including good and marginal ($P < .05$ for Empatica and $P < .001$ for Biovotion, reported in Table 2). For the Empatica device, the EDA signal is zero when it is not on the patient's wrist, and the Biovotion device records a constant value of 21 k Ω (0.047 mS) in our experiments, where the device was on the patient but not in contact with skin. As measured by the marginal and noise percentage of data reported in Table 2, the Empatica E4 recorded EDA data with the highest quality on average in our cohort.

3.3 | PPG data analysis

PPG data were analyzed from 15 patients. BVP data from Empatica E4 and PPG data with three light sources from Byteflies and Biovotion devices were analyzed. Table 3 summarizes the mean and standard deviation of the spectral entropy SQI_{SE} for good, marginal, and noise data, respectively. As expected, good-quality data exhibit the lowest spectral entropy of all classes. Noise has the highest spectral entropy, with values close to 1, in correspondence with a flat spectrum in the 1- to 3-Hz frequency band. Marginal-quality data epochs show values in between good and noise SQI_{SE} values. The green PPG for Biovotion and IR light PPG for Byteflies devices were scored by reviewers, and these scores were used to classify the remaining PPG signals. Although the Empatica BVP data are preprocessed, Byteflies and Biovotion devices also provide good-quality data for at least one light source.

3.4 | EEG data analysis

A total of 405 days of Epitel Epilog EEG data from 21 patients were analyzed by calculating the signal's maximum bandwidth, which is an estimate of the highest frequency at which the recorded signal is significantly different from noise. The results show that 21.4% of EEG data were classified as good (maximum bandwidth > 75 Hz), 33.3% classified as

TABLE 3 SE for PPG data (SQI_{SE}) from Empatica E4, Biovotion, and Byteflies

Device	Location	Patients, n	Duration, d	SQI_{SE} , avg $\pm$ SD (good)	SQI_{SE} , avg $\pm$ SD (marginal)	SQI_{SE} , avg $\pm$ SD (noise)	P
Empatica E4	Wrist	5	14	0.56 ± 0.16	0.82 ± 0.10	0.98 ± 0.02	<.001
Biovotion Everion	Arm	5	2				
		Green ^a		0.53 ± 0.13	0.76 ± 0.2	0.97 ± 0.01	<.001
		Infrared		0.79 ± 0.12	0.85 ± 0.06	0.97 ± 0.01	
		Red		0.92 ± 0.09	0.91 ± 0.06	0.97 ± 0.01	
Byteflies Sensor Dot	Chest	5	4.5				
		Infrared ^a		0.59 ± 0.14	0.77 ± 0.11	0.98 ± 0.02	<.001
		Green		0.58 ± 0.13	0.74 ± 0.12	0.98 ± 0.02	
		Red		0.64 ± 0.13	0.80 ± 0.09	0.98 ± 0.02	

Abbreviations: avg, average; PPG, photoplethysmography; SD, standard deviation; SE, special entropy; SQI, signal quality index.

^aData quality was scored by reviewers for one PPG channel, and scores were applied to the remaining channels.

acceptable (maximum bandwidth between 35 Hz and 75 Hz), and 45.3% of data were marginal (maximum bandwidth < 35 Hz).

3.5 | Data validation

The proposed SQIs were computed over a larger set of acquired data including ACC, EDA, and PPG signals. The analysis showed 71% of Empatica EDA data (34 patients, 89 days) and 13% of Biovotion EDA data (10 patients, 7 days) were classified as acceptable quality (rate of amplitude change stays within -10% to 20% limit). Ninety percent of Empatica ACC data (39 patients, 117 days) were classified as acceptable quality, and 91% of Empatica BVP data (28 patients, 71 days) were classified as acceptable quality. Ninety-four percent of Biovotion ACC data (1 patient, 1 day) had acceptable quality. Results of analysis of Biovotion PPG data show that 95%, 91%, and 58% of green, IR, and red light data were classified as acceptable data.

TABLE 4 Demographics of patients' survey respondents

Total patients, N = 70	
n (%) using 1 device	25 (35.7%)
n (%) using >1 device	45 (64.3%)
Female (%)	41 (58.6%)
Age group, n (%)	
<20 y	11 (15.7%)
20-29 y	22 (31.4%)
30-39 y	15 (21.4%)
40-49 y	9 (12.9%)
50-59 y	8 (11.4%)
60-69 y	5 (7.14%)

3.6 | Patient survey analysis

Patients were provided with a survey at the end of data recording to evaluate wearable device acceptability. Table 4 presents the demographics of the patient cohort that completed device acceptability surveys. All respondents completed the entire survey. The majority of respondents identified as female, and the most represented age cohorts are 20-29 and 30-39 years of age.

Survey results are summarized in Table 5. Patients who wore multiple devices were asked which device they preferred, and only responded to the survey for the preferred device. The Empatica E4 was the preferred device more often than other devices. The Biovotion device had the most positive results for both ease of manipulation and usability, followed by the two wrist-worn devices, the Empatica E4 and GENEActiv. GENEActiv had the most positive results for both long-term comfort and comfort during sleep. The Byteflies sensor dots were most preferred for potential seizure prediction. However, all devices generally had positive responses.

Being in the 60- to 69-year-old age group was significantly correlated with more negative usability responses ($P < .05$). Being in the 30- to 39-year-old age group was significantly correlated with more positive long-term comfort responses ($P < .01$). No significant effect was found with gender for response to any question.

4 | DISCUSSION

Automated, objective measures of signal quality are important for large-scale data acquisition and management systems, such as are needed for epilepsy monitoring and management systems. Here we describe and validate quantitative signal quality metrics for ACC, PPG, EDA, and EEG. As a quality

TABLE 5 Patient survey results

Device	Times used, n	Preferred device, n	Preferred device, %	Easy to manipulate	Usable	Long-term comfort	Comfort during sleep	Would use for seizure prediction
Biovotion Everion	9	1	11	1.00 ^a	1.00 ^a	3.00	1.33	1.67
Empatica E4	40	19	48 ^a	1.82	2.25	1.89	2.14	1.75
Byteflies Sensor Dots	14	3	21	2.67	2.67	2.33	2.33	1.33 ^a
Activinsights GENEActiv	15	4	27	1.20	1.20	1.00 ^a	1.20 ^a	1.80
Epitel Epilog	18	2	11	1.50	2.25	2.00	1.75	3.50

Note: Values are mean results from 7-point Likert scale: 1 = strongly agree, 4 = neutral, 7 = strongly disagree.

^aBest result per column.

metric for ACC data, the power ratio of the signal in the frequency range between 0.8-5 Hz over 0.8-16 Hz ($F_n = 16\text{Hz}$), was estimated. According to the results reported in Table 1, the average and standard deviation of the proposed SQI_{PR} (Equation 2) can be used as a metric to determine whether the device is on the patient. However, for short segments classified as noise, observing other signals like EDA and PPG (and many devices also record temperature) is suggested to identify times when subjects may be physically still (for example, during sleep). All devices tested showed significant differences in SQI_{PR} between normal use and device removal, and any of the devices tested would perform well. The rate of amplitude change (%) was defined as a quality metric for EDA data, which for good data is between a 20% increase and a 10% decrease, discarding data with amplitude $< 0.05 \mu\text{S}$ (equal to 0.047 mS for Biovotion) and 1-minute segments with amplitude changes $< 0.01 \mu\text{S}$ for the Empatica E4 (0.01 mS for Biovotion). In some special cases, this range would be slightly different.²⁵ Note that in our study cohort, the arm-worn Biovotion device recorded a greater proportion of marginal and poor-quality EDA data (and a lower proportion of good-quality data) compared to the wrist-worn Empatica device. It is unclear whether this is due to the difference in sensor placement, differences in sensor design, or other factors.

The signal quality of PPG was evaluated by spectral entropy to measure the peakedness of a spectrum. Good PPG signals have lower spectral entropy than a noisy PPG signal. The t test was used to evaluate the metric's performance, and it showed metrics can classify data into good and noise (or device not on the patient) classes with a $P < .05$ for all signals. A direct comparison of the individual green, red, and IR PPG channels from the Byteflies and Biovotion devices to the aggregated PPG signal from the Empatica device is perhaps unclear, although the spectral entropy ranges observed for the green and IR signals from the Biovotion and Byteflies devices gave similar results, suggesting that a similar aggregated measure from these devices may be quite similar to the Empatica BVP. The Biovotion device's IR channel showed

less change in SQI_{SE} amplitude, and the red channel showed almost no change between good, marginal, and poor data compared to the other devices. In general, EDA and PPG together can be used to robustly detect whether a device is being worn with the individual quality metrics. The SQI of ACC data could be used to distinguish between periods of device on versus off the body; however, special care needs to be put toward not misclassifying sleep as device not worn. Signal quality in a modular, ambulatory EEG device has somewhat different characteristics than a typical wired EEG setup, where line noise may dominate as an indicator of poor signal quality. The SQI reported here has been used widely in intracranial EEG³¹⁻³³ and showed good ability to differentiate signal quality in this study.

There are a large number of devices commercially available for epilepsy management, and an increasing number are receiving regulatory approval for detection of specific seizure types.⁸ Automated monitoring of data quality in long-term use will be of great importance for clinical and research applications. Long-term use of wearable devices will also necessitate robust online data transfer to minimize the need for manual data management. Such systems exist for many consumer-grade wearable devices, and will minimize compliance and technically related errors. Online connectivity can also provide the opportunity to give direct feedback to users about their compliance and data quality. This in turn may have a positive effect on increasing compliance. However, many devices currently available have deficiencies in design that limit their acceptability to the user, negatively impacting adherence. It is necessary that people with epilepsy and their caregivers are integrated into the development process of wearable devices, such that interest in the device is not solely based on medical benefits. It is encouraging that all average survey responses were positive to the tested devices, although not reflected in these results were occasional refusals by subjects to use devices. Surveys of larger populations will be required to appropriately power such studies.

ACKNOWLEDGMENTS

This work was funded by the My Seizure Gauge grant provided by the Epilepsy Innovation Institute, a research program of the The Epilepsy Foundation of America's Epilepsy Innovation Institute My Seizure Gauge project. The authors thank Sherry Klingerman, Dan Crepeau, Dominique Eden, William Hart, and Shannon McCollough for technical assistance and coordination.

CONFLICT OF INTEREST

E.N. and D.F. are employees of Seer Medical. M.C. and D.F. are shareholders in Seer Medical. No other authors have conflicts to declare. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

ORCID

Ewan Nurse  <https://orcid.org/0000-0001-8981-0074>

Sebastian Böttcher  <https://orcid.org/0000-0002-3407-8290>

Boney Joseph  <https://orcid.org/0000-0002-1576-9344>

Tal Pal Attia  <https://orcid.org/0000-0003-3456-9628>

Pedro F. Viana  <https://orcid.org/0000-0003-0861-8705>

Andreas Schulze-Bonhage  <https://orcid.org/0000-0003-2382-0506>

Matthias Dümpelmann  <https://orcid.org/0000-0002-1476-7777>

Benjamin H. Brinkmann  <https://orcid.org/0000-0002-2392-8608>

REFERENCES

- Elger CE, Hoppe C. Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection. *Lancet Neurol*. 2018;17:279–88.
- Cook MJ, O'Brien TJ, Berkovic SF, et al. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. *Lancet Neurol*. 2013;12:563–71.
- Kuhlmann L, Lehnertz K, Richardson MP, Schelter B, Zaveri HP. Seizure prediction—ready for a new era. *Nat Rev Neurol*. 2018;14(10):618–30.
- Dumanis SB, French JA, Bernard C, Worrell GA, Fureman BE. Seizure forecasting from idea to reality. Outcomes of the My Seizure Gauge epilepsy innovation institute workshop. *eNeuro*. 2017;4(6):ENEURO.0349-17.2017.
- Karoly PJ, Goldenholz DM, Freestone DR, et al. Circadian and circaseptan rhythms in human epilepsy: a retrospective cohort study. *Lancet Neurol*. 2018;17:977–85.
- Baud MO, Rao VR. Gauging seizure risk. *Neurology*. 2018;91:967–73.
- Karoly PJ, Cook MJ, Maturana MI, et al. Forecasting cycles of seizure likelihood. *Epilepsia*. 2020;61:776–786.
- Gregg NM, Nasser M, Kremen V, et al. Circadian and multiday seizure periodicities, and seizure clusters in canine epilepsy. *Brain Commun*. 2020;2:fcaa008.
- Beniczky S, Conradsen I, Henning O, Fabricius M, Wolf P. Automated real-time detection of tonic-clonic seizures using a wearable EMG device. *Neurology*. 2018;90:e428–34.
- Regalia G, Onorati F, Lai M, Caborni C, Picard RW. Multimodal wrist-worn devices for seizure detection and advancing research: focus on the Empatica wristbands. *Epilepsy Res*. 2019;153:79–82.
- Beniczky S, Ryvlin P. Standards for testing and clinical validation of seizure detection devices. *Epilepsia*. 2018;59:9–13.
- Onorati F, Regalia G, Caborni C, et al. Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors. *Epilepsia*. 2017;58:1870–9.
- Fürbass F, Kampusch S, Kaniusas E, et al. Automatic multimodal detection for long-term seizure documentation in epilepsy. *Clin Neurophysiol*. 2017;128:1466–72.
- Bruno E, Biondi A, Richardson MP. Pre-ictal heart rate changes: a systematic review and meta-analysis. *Seizure*. 2018;55:48–56.
- Simblett SK, Bruno E, Siddi S, et al. Patient perspectives on the acceptability of mHealth technology for remote measurement and management of epilepsy: a qualitative analysis. *Epilepsy Behav*. 2019;97:123–9.
- Simblett SK, Biondi A, Bruno E, et al. Patients' experience of wearing multimodal sensor devices intended to detect epileptic seizures: a qualitative analysis. *Epilepsy Behav*. 2020;102:106717.
- Lee S, Shin H, Hahm C. Effective PPG sensor placement for reflected red and green light, and infrared wristband-type photoplethysmography. In: 2016 18th International Conference on Advanced Communication Technology (ICACT). Pyeongchang, South Korea: IEEE, 2016:556–8.
- Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. *Int J Biosens Bioelectron*. 2018;4:195.
- Tamura T, Maeda Y, Sekine M, Yoshida M, Yoshida M. Wearable photoplethysmographic sensors—past and present. *Electronics*. 2014;3:282–302.
- Sinha SR, Sullivan LR, Sabau D, et al. American clinical neurophysiology society guideline 1: minimum technical requirements for performing clinical electroencephalography. *Neurodiagn J*. 2016;56:235–44.
- de Curtis MD, Gnatkovsky V. Reevaluating the mechanisms of focal ictogenesis: the role of low-voltage fast activity. *Epilepsia*. 2009;50:2514–25.
- Sharma A, Purwar A, Lee YD, Lee YS, Chung WY. Frequency based classification of activities using accelerometer data. In: 2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. Seoul, South Korea: IEEE, 2008:150–3.
- Benedek M, Kaernbach C. A continuous measure of phasic electrodermal activity. *J Neurosci Methods*. 2010;190:80–91.
- Benedek M, Kaernbach C. Decomposition of skin conductance data by means of nonnegative deconvolution. 2010. *Psychophysiology*. 2010;47:647–58.
- Braithwaite J, Watson D, Jones R, Mickey R. A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*. 2013;49:1017–34.
- Dawson ME, Schell AM, Filion DL. The electrodermal system. In: Cacioppo JT, Tassinary LG, Bernston GL, eds. *Handbook of*

- Psychophysiology. 2nd ed. Cambridge, UK: Cambridge University Press; 2000:200–23.
27. Taylor S, Jaques N, Chen W, Fedor S, Sano A, Picard R. Automatic identification of artifacts in electrodermal activity data. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Milan, Italy: IEEE, 2015:1934–7.
 28. Kocielnik R, Sidorova N, Maggi FM, Ouwerkerk M, Westerink JHDM. Smart technologies for long-term stress monitoring at work. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. Porto, Portugal: IEEE, 2013:53–8.
 29. Misra H, Ikbāl S, Boulard H, Hermansky H. Spectral entropy based feature for robust ASR. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Montreal, QC: IEEE, 2004:I–193.
 30. Rezek IA, Roberts SJ. Stochastic complexity measures for physiological signal analysis. IEEE Trans Biomed Eng. 1998;45:1186–91.
 31. Bundy DT, Zellmer E, Gaona CM, et al. Characterization of the effects of the human dura on macro- and micro-electrocorticographic recordings. J Neural Eng. 2014;11:016006.
 32. Nurse ES, John SE, Freestone DR, et al. Consistency of long-term subdural electrocorticography in humans. IEEE Trans Biomed Eng. 2018;65:344–52.
 33. Miller KJ, Sorensen LB, Ojemann JG, den Nijs M. Power-law scaling in the brain surface electric potential. PLoS Comput Biol. 2009;5(12):e1000609.

How to cite this article: Nasser M, Nurse E, Glasstetter M, et al. Signal quality and patient experience with wearable devices for epilepsy management. *Epilepsia*. 2020;00:1–11. <https://doi.org/10.1111/epi.16527>