2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871559

Evaluating the Empatica E4 Derived Heart Rate and Heart Rate Variability Measures in Older Men and Women*

Kiran K.G. Ravindran[1,2], Ciro della Monica [1,2], Giuseppe Atzori [1], Damion Lambert [1,2], Victoria Revell [1,2] and Derk-Jan Dijk [1,2]

***Abstract*— Wearable heart rate monitors offer a cost-effective way of non-invasive, long-term monitoring of cardiac health. Validation of wearable technologies in an older populations is essential for evaluating their effectiveness during deployment in healthcare settings. To this end, we evaluated the validity of heart rate measures from a wearable device, Empatica E4, and compared them to the electrocardiography (ECG). We collected E4 data simultaneously with ECG in thirty-five older men and women during an overnight sleep recording in the laboratory. We propose a robust approach to resolve the missing inter-beat interval (IBI) data and improve the quality of E4 derived measures. We also evaluated the concordance of heart rate (HR) and heart rate variability (HRV) measures with ECG. The results demonstrate that the automatic E4 heart rate measures capture long-term HRV whilst the short-term metrics are affected by missing IBIs. Our approach provides an effective way to resolve the missing IBI issue of E4 and extracts reliable heart rate measures that are concordant with ECG.**

***Clinical relevance*— This work discusses data quality challenges in heart rate data acquired by wearables and provides an efficient and reliable approach for extracting heart rate measures from the E4 wearable device and validates the metrics in older adults.**

I. INTRODUCTION

Wearable heart rate monitors are being increasingly used to collect physiological data for long term monitoring in healthcare. The incidence of cardiovascular disorders increases with aging and is a common cause of hospitalisation in older adults. Validation of wearables in a relevant population is essential to assess their reliability, ease of use and accuracy.

Wrist-worn photoplethysmography (PPG) devices such as the Empatica E4 wristband (Empatica Srl, Milan, Italy), offer a non-invasive and user-friendly method to collect physiological data including heart rate, body temperature, activity, etc. These devices enable the development and implementation of cardiovascular anomaly detection and machine learning pipelines for remote monitoring of vulnerable populations. E4 is a research grade wearable device that provides access to raw blood volume pulse (BVP) data and several E4 datasets are publicly available [1].

Although the validity of the E4 device has previously been assessed in comparison to electrocardiography (ECG), these studies employed few participants and were not performed in older adults during sleep [2]–[5]. One of the most prevalent issues identified in these studies relates to the significant amount of missing inter-beat intervals (IBI). To the best of our knowledge, the E4 validation studies so far have failed to accurately quantify the incidence and impact of missing data events or provide a robust solution to address this issue. Missing data and data sparsity are of potential concern in the construction and deployment of machine learning solutions in healthcare using wearable physiological monitors such as the E4 [6]. Changes in HR and HRV have been associated with sleep stage transitions, sleep apnoea, etc., and are relevant to sleep medicine and cardiology. In our current work, we propose a simple approach to extract heart rate metrics from the E4 data directly from the blood volume pulse (BVP) data, thereby bypassing the confounds associated with missing IBIs. We also evaluate the heart rate (HR) and heart rate variability (HRV) measures derived from the E4 device and compare them to ECG in older adults during an overnight sleep recording.

II. METHODS

A. Participants and collected data

We collected E4 wristband data alongside gold standard polysomnography (PSG) from 35 older men and women (Age: 70.8 ± 4.9 [$\mu \pm \sigma$] years; Range: 65-83; 14 women) during an overnight sleep study at the Surrey Sleep Research centre (SSRC), Guildford, UK. Sixty percent of the participants had one or more co-morbidities including sleep apnoea, hypertension, type 2 diabetes, etc. A favourable ethical opinion was obtained from the University of Surrey Ethics committee and written consent was obtained from the participants before any data collection. Each recording consisted of 10 hours in bed during which the participants were allowed to sleep. The PSG was recorded using the SomnoHD polysomnography system (SOMNOmedics GmbH™, Germany), including DOMINO software, and the hypnograms (30 second epochs) were independently scored by two experienced scorers in accordance with American Academy of Sleep Medicine (AASM) guidelines [7]. The E4 was deployed on the dominant hand of the participants. For the purpose of the analysis reported in this work, lead II ECG recorded at 256 Hz as a part of the PSG montage was used as the gold standard against which E4 heart rate measures were compared. The ECG derived heart rate and RR interval (RRi) data were exported from the DOMINO software and in our analysis we assume that these extracts are accurate. The E4 device outputs the

*This work was supported by the UK Dementia Research Institute which receives its funding from DRI Ltd, funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK

[1]Surrey Sleep Research Centre, University of Surrey, Guildford, United Kingdom

[2]UK Dementia Research Institute, Care Research and Technology Centre at Imperial College, London and the University of Surrey, Guildford, United Kingdom k.guruswamyravindran@surrey.ac.uk

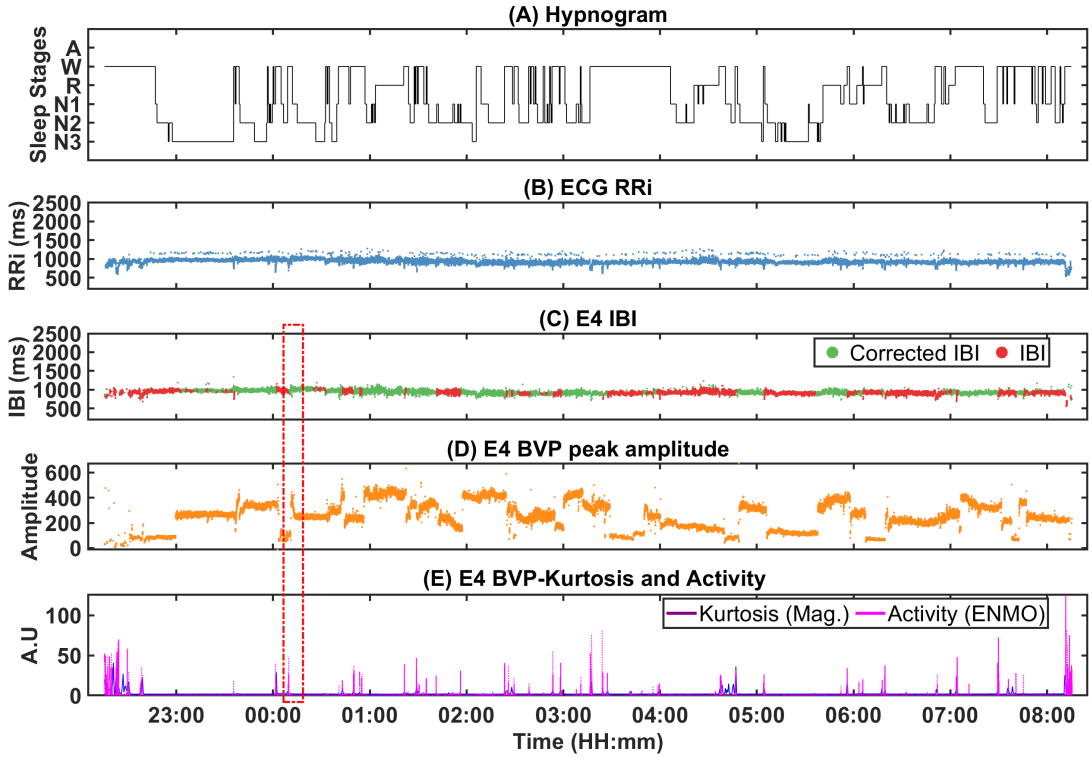

Fig. 1: Example E4 data: A) PSG hypnogram, B) ECG RR interval, C) E4 IBI (true IBI and corrected IBI), D) BVP peak amplitude, and E) activity and BVP Kurtosis. The region marked by a red box shows an example of missing data transition region.

raw blood volume pulse (BVP) signal sampled at 64 Hz. Using BVP data, E4 automatically generates HR and IBI time series [8]. Furthermore, the E4 output includes three axis accelerometer data which provides activity information. We used the Euclidean norm of the three-axis data minus one (ENMO) to quantify the activity of the participant as defined in [9].

B. Data Analysis

We investigated the prevalence of missing IBI data during the recording interval of the PSG across all 35 participants. Presence or absence of IBI detected by E4 was quantified for each 30 s epoch. We also quantified the number of missing data epochs associated with activity. Here presence of activity was defined as an increase in ENMO greater than two standard deviations of the mean or kurtosis (as a measure of signal quality) of the BVP over 10 as suggested in [10], [11]. We performed an IBI detection on the BVP data for the investigation of the missing data problem and to identify approaches to improve the IBI data quality. To extract the IBI from the BVP, we used a robust beat detection algorithm introduced in [12]. The IBIs corresponding to the normal sinus rhythm were extracted by applying a square filter (IBI outside the range of [300, 2500] ms was removed) followed by a double quotient filter on the output of the beat detection algorithm as suggested in [13]. The double quotient filtering was achieved by passing the data twice through the quotient filter. We next introduced a peak BVP amplitude-based filter inspired by the quotient filter to eliminated amplitude anomalies in the detected IBI. Assuming that the vector $X0$ contains lag-0 IBI (or RRi) and $X1$ contains lag-1 IBI, then the IBI to be removed according to the quotient filter is defined as,

$$\frac{X0}{X1} \geq 1.2, \ \textit{or} \ \frac{X0}{X1} \leq 0.8, \ \textit{or} \ \frac{X1}{X0} \geq 1.2, \ \textit{or} \ \frac{X1}{X0} \leq 0.8 \quad (1)$$

The amplitude filtering was performed on the peak amplitude time-series derived from detected IBI and the estimated indices were used to remove amplitude-based artefact beats. This process eliminates any spurious/artefact beats. The IBI data extracted from the above-described approach is referred to as the corrected E4 (cE4) IBI data hereafter.

C. Concordance against PSG ECG

We assessed the concordance of HR and HRV estimated by E4 and cE4 with ECG. This allows us to evaluate the validity of E4 wristband in older adults during sleep. We used Bland-Altmann analysis, intraclass correlation (ICC) and correlation coefficient (ρ) to assess agreement between the E4 estimates and ECG as described in [14]. We also assessed the normality of the difference between the compared devices using the Kolmogorov-Smirnov tests. Paired t-test was used to show significance of difference in concordance. The Bland-Altmann analysis provided an estimate of the bias between the E4 and ECG and agreement interval of the mean differences.

Due to the significant amount of missing IBI in the E4 data, we used Poincaré derived HRV estimates to assess the consensus between ECG and E4 thereby avoiding the interpolation of missing IBIs in the E4 data. The robustness

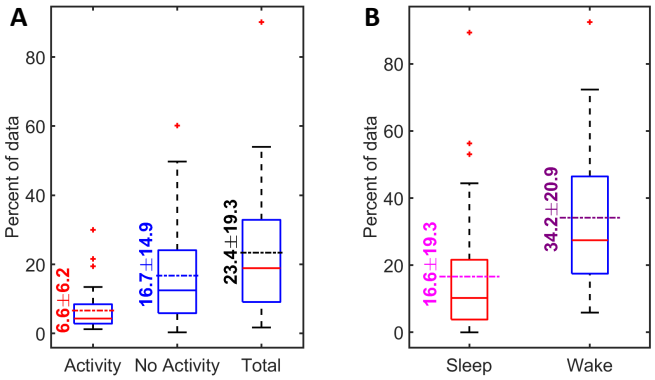

Fig. 2: Quantification of missing IBI in the data. Panel A shows IBI data loss associated with the E4 activity, not associated with the E4 activity and the total missing data. Panel B depicts the total amount of missing IBI in sleep and wake periods.

of Poincaré metrics has been well studied in [15]. The HRV metrics and description used in this work is given below,

- SD1: Short term variability (ms)
- SD2: Long term variability (ms)
- S: Total variability (ms)
- SD1UP: Short term accelerations in heart rate (ms)
- SD1DOWN: Short term decelerations in heart rate (ms)
- SDRR: Standard deviation of RR interval (ECG) or inter-beat interval (PPG) (ms)

We used similar descriptive statistics for both the HR and HRV concordance analysis. The HR analysis was performed on the time-series estimates of 30 s intervals in alignment with the PSG hypnogram and the HRV Poincaré estimates were estimated using IBI data from the complete recording.

III. Results and Discussion

Firstly, we examined the issue of missing IBI in the exported E4 data. The occurrence of missing E4 inter-beat interval data has been previously reported and has been attributed to movement or increased activity [3]–[5]. In these studies, the data were collected during daytime and the extent to which missing data were indeed associated with movement and activity was not accurately quantified. Since the data used in the current work were collected while the participants were in bed, the activity levels were low. We found that a significant portion of the missing IBI data was not associated with increased activity, as shown in Figure2. It can also be seen that about one quarter of the total IBI data was missing with over 15% of the missing data occurring while the participants were asleep. This introduces sparsity into the data even when collected when the participants are in bed. This leads to poor HRV estimates and hinders the direct use of E4 IBI output in machine learning pipelines [15].

On further inspection, the regions of missing IBI occurred after a significant change in the peak amplitude of the beats as estimated via the cE4 approach (Figure 1). It should be further noted that these missing data events were unassociated with activity or BVP signal distortion. From the sub-panel three in Figure 1, it can be seen that the cE4 IBI approach (data

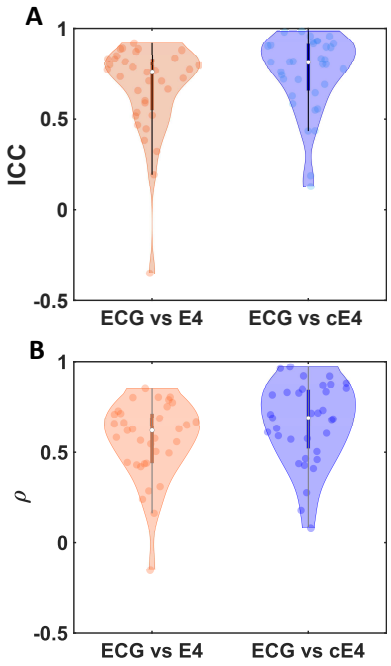

Fig. 3: Concordance between the E4 and cE4 with ECG: Panel A depicts the distribution of ICC and Panel B depicts the distribution of the ρ across all 35 participants. Each data point represents one participant.

in green) was able to robustly capture the IBI irrespective of baseline/ beat amplitude changes. After correction of the E4 IBI, the amount of missing IBI reduced to $2.4 \pm 1.2\%$. We infer from this analysis that the automatic E4 IBI detection approach appear to be sensitive to baseline shifts in the BVP and hence leading to missed beats and incomplete IBI timeseries.

The correlation analysis shows that 30 s E4 HR estimates are accurate compared to the ECG and the proposed E4 correction leads to a significant ($p < 0.001$) improvement in the concordance (both ICC and ρ) as shown in Figure 3. The outliers correspond to participants with cardiac (arrhythmia) and respiratory (severe apnoea) conditions and affects the normality of the data.

The concordance analysis of the HRV parameters showed high agreement and correlation of measures that depict long term heart rate variations (SD2 and SDRR) and total variability while there is poor correlation of short term HRV measures (SD1, SD1UP and SD1DOWN) as shown in Table I. None of the variable differences in the Bland-Altmann analysis were normally distributed. It can also be noted that the HRV measures from the cE4 IBI are highly concordant (high absolute and consistency ICC).

From our analysis, lower concordance of raw E4 HR and automatic IBI derived HRV with ECG can be attributed to, 1) missing IBI information and 2) bounded HR (automatic E4 estimates were found to be bounded between 40 and 196 beats per minute). The proposed data extraction approach uses the raw E4 BVP to estimate the filtered IBI, overcomes the above shortcomings and provides robust data that has a excellent agreement with gold standard ECG.

TABLE I: Concordance of the Heart rate variability (HRV) measures between ECG, E4 and corrected E4 (cE4).

HRV metrics	Concordance measures			
	Difference ($\mu \pm \sigma$)	ρ	**absolute ICC [95%CI]**	**consistency ICC [95%CI]**
ECG vs E4				
SD1	6.15 ± 8.53	0.68 *	0.72 [0.2,0.89]	0.8 [0.6,0.9]
SD2	4.27 ± 23.52	0.97 *	0.96 [0.93,0.98]	0.96 [0.93,0.98]
SDRR	4.15 ± 16.54	0.97 *	0.96 [0.94,0.98]	0.96 [0.93,0.98]
S	2168 ± 3100	0.97 *	0.98 [0.91,0.98]	0.98 [0.97,0.98]
SD1UP	4.65 ± 6.32	0.66 *	0.7 [0.15,0.87]	0.8 [0.56,0.89]
SD1DOWN	4.06 ± 6.0	0.67 *	0.73 [0.26,0.88]	0.8 [0.6,0.9]
ECG vs cE4				
SD1	7.48 ± 6.53	0.81 *	0.77 [-0.12,0.93]	0.89 [0.77,0.94]
SD2	1.01 ± 10.4	0.996 *	0.995 [0.99,0.997]	0.995 [0.99,0.997]
SDRR	2.55 ± 7.09	0.997 *	0.995 [0.99,0.997]	0.996 [0.99,0.998]
S	3140 ± 7210	0.99 *	0.94 [0.87,0.97]	0.95 [0.90,0.98]
SD1UP	5.86 ± 5.0	0.8 *	0.74 [-0.17,0.91]	0.87 [0.74,0.93]
SD1DOWN	4.73 ± 4.3	0.83 *	0.8 [-0.02,0.94]	0.9 [0.8,0.95]

Note $*: p < 0.001$.

IV. Conclusion

Although several previous studies have validated the Empatica E4, to the best of our knowledge, this is the first work that proposes a robust approach to circumvent the issue of missing inter-beat interval data and validates the E4 in an older heterogeneous population. We also show that there is a significant portion of the missing IBI data which is not associated with movement, but rather associated with amplitude shifts in the BVP signal in the E4. Furthermore, the E4 reliably records blood volume pulse data, but the automatically computed inter-beat interval and heart rate estimates are inaccurate and this can be attributed to the missing IBI data. Our approach provides a simple solution to the missing IBI problem in E4 and allows for estimation of complete IBI time-series and robust heart rate measures.

Acknowledgment

This work is supported by the UK Dementia Research Institute, Care Research and Technology Centre at Imperial College, London and the University of Surrey, Guildford, United Kingdom which receives its funding from UK DRI Ltd, funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK.

References

[1] F. Wolling and K. Van Laerhoven, "The quest for raw signals: A quality review of publicly available photoplethysmography datasets," in *Proceedings of the Third Workshop on Data: Acquisition To Analysis*, ser. DATA '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 14–19.

[2] S. Ollander, C. Godin, A. Campagne, and S. Charbonnier, "A comparison of wearable and stationary sensors for stress detection," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 004 362–004 366.

[3] A. A. T. Schuurmans, P. de Looff, K. S. Nijhof, C. Rosada, R. H. J. Scholte, A. Popma, and R. Otten, "Validity of the empatica e4 wristband to measure heart rate variability (HRV) parameters: a comparison to electrocardiography (ECG)," *Journal of Medical Systems*, vol. 44, no. 11, sep 2020.

[4] C. McCarthy, N. Pradhan, C. Redpath, and A. Adler, "Validation of the empatica e4 wristband," in *2016 IEEE EMBS International Student Conference (ISC)*, 2016, pp. 1–4.

[5] N. Milstein and I. Gordon, "Validating measures of electrodermal activity and heart rate variability derived from the empatica e4 utilized in research settings that involve interactive dyadic states," *Frontiers in Behavioral Neuroscience*, vol. 14, 2020.

[6] J. Wiens and B. C. Wallace, "Editorial: special issue on machine learning for health and medicine," *Machine Learning*, vol. 102, no. 3, pp. 305–307, Mar. 2016.

[7] *AASM Scoring Manual Version 2.6l*, American Academy of Sleep Medicine (AASM), 2020. [Online]. Available: https://aasm.org/aasm-updated-version-sleep-scoring-manual/l)

[8] *E4 wristband User's manual*, Empatica, 2018. [Online]. Available: https://empatica.app.box.com/v/E4-User-Manual)

[9] V. T. van Hees, L. Gorzelniak, E. C. D. León, M. Eder, M. Pias, S. Taherian, U. Ekelund, F. Renström, P. W. Franks, A. Horsch, and S. Brage, "Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity," *PLoS ONE*, vol. 8, no. 4, p. e61691, Apr 2013.

[10] N. Selvaraj, Y. Mendelson, K. H. Shelley, D. G. Silverman, and K. H. Chon, "Statistical approach for the detection of motion/noise artifacts in photoplethysmogram," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, Aug 2011.

[11] E. Sabeti, N. Reamaroon, M. Mathis, J. Gryak, M. Sjoding, and K. Najarian, "Signal quality measure for pulsatile physiological signals using morphological features: Applications in reliability measure for pulse oximetry," *Informatics in Medicine Unlocked*, vol. 16, p. 100222, 2019.

[12] J. Lázaro, E. Gil, J. M. Vergara, and P. Laguna, "Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 1, pp. 240–246, 2014.

[13] J. Piskorski and P. Guzik, "Filtering poincaré plots," *Computational Methods in Science and Technology*, vol. 11, no. 1, pp. 39–48, 2005.

[14] S. Haghayegh, H.-A. Kang, S. Khoshnevis, M. H. Smolensky, and K. R. Diller, "A comprehensive guideline for bland–altman and intra class correlation calculations to properly compare two methods of measurement and interpret findings," *Physiological Measurement*, vol. 41, no. 5, p. 055012, jun 2020.

[15] K. K. Kim, H. J. Baek, Y. G. Lim, and K. S. Park, "Effect of missing RR-interval data on nonlinear heart rate variability analysis," *Computer Methods and Programs in Biomedicine*, vol. 106, no. 3, pp. 210–218, Jun 2012.