PSYCHOPHYSIOLOGY SPR   WILEY

Cross-instrument feasibility, validity, and reproducibility of wireless heart rate monitors: Novel opportunities for extended daily life monitoring

Yoram K. Kunkels[1] | Arie M. van Roon[2] | Marieke Wichers[1] | Harriëtte Riese[1]

[1]Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

[2]Department of Vascular Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Correspondence
Yoram K. Kunkels, Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), University of Groningen, University Medical Center Groningen, Groningen, ICPE, PO Box 30.001 (CC 72), 9700 RB Groningen, The Netherlands.
Email: y.k.kunkels@umcg.nl

Abstract

Wired ambulatory monitoring of the electrocardiogram (ECG) is an established method used by researchers and clinicians. Recently, a new generation of wireless, compact, and relatively inexpensive heart rate monitors have become available. However, before these monitors can be used in scientific research and clinical practice, their feasibility, validity, and reproducibility characteristics have to be investigated. Therefore, we tested how two wireless heart rate monitors (i.e., the Ithlete photoplethysmography (PPG) finger sensor and the Cortrium C3 ECG monitor perform against an established wired reference method (the VU-AMS ambulatory ECG monitor). Monitors were tested on cross-instrument and test-retest reproducibility in a controlled laboratory setting, while feasibility was evaluated in protocolled ambulatory settings at home. We found that the Cortrium and the Ithlete monitors showed acceptable agreement with the VU-AMS reference in laboratory setting. In ambulatory settings, assessments were feasible with both wireless devices although more valid data were obtained with the Cortrium than with the Ithlete. We conclude that both monitors have their merits under controlled laboratory settings where motion artefacts are minimized and stationarity of the ECG signal is optimized by design. These findings are promising for long-term ambulatory ECG measurements, although more research is needed to test whether the wireless devices' feasibility, validity, and reproducibility characteristics also hold in unprotocolled daily life settings with natural variations in posture and activities.

KEYWORDS
adolescents, cardiovascular, heart rate, heart rate variability

1 | INTRODUCTION

In the last decennium, the number of heart rate monitors available for clinicians and researchers has increased steadily (El-Amrawy & Nounou, 2015). This opened opportunities for long-term monitoring of cardiac functions such as heart rate (HR) and heart rate variability (HRV) (Kemp & Quintana, 2013; Malik et al., 1996) in daily-life. When feasible and valid, long-term HR(V) monitoring may open-up possibilities for developing indicators of (mental) health processes complementary

PSYCHOPHYSIOLOGY SPR

to those developed for the experience sampling method (ESM, or electronic dairy). ESM monitoring is a scientific method that has shown its potential in research and clinical practice (e.g. Schoevers et al., 2020; Shiffman et al., 2008; Vaessen et al., 2019). It nowadays typically involves filling-out short questionnaires, multiple time a day for weeks/months on a smartphone. The time-series data derived from ESM monitoring have been used to better inform diagnosis, intervention selection, and recently also for early predicting transitions in patients affect state (Kroeze et al., 2017; Smit et al., 2019). Long-term monitoring of HR(V) could potentially add to ESM's potential in clinic practice, as it might supplement a patient's momentary affect data with physiological data.

The current study is performed within the scope of the TRANS-ID (TRANSitions In Depression, www.transid.nl) project, which aims to discover personalized signals that may indicate critical transitions in psychological and physiological symptoms. In TRANS-ID we investigate within single individuals which early warning signals precede depressive symptom change and thereby examine whether psychological symptoms behave according to the principles of complex dynamical systems (Scheffer, 2010). To gain insight into this, we use ESM to capture the micro-level changes of symptoms, emotions, behaviors and daily context over time (Kramer et al., 2014). Moreover, future TRANS-ID studies are planned to investigate whether monitoring patients' HR(V) data can support the study of dynamic processes, such as transitions from a healthy to a clinically affected state in patients. Such combined time-series data collection is required for studies that aim to investigate the use of physiological measurements for predicting transitions in patients' affect state. We previously found support that ESM data can be used to calculate early warning signals to predict transitions in patients' affect state (Smit et al., 2019; Wichers & Groot., 2016), and aim to investigate in future research the usefulness of HR(V) monitoring for predicting transitions in patients' affect state. As long-term continuous wearing of ECG-electrodes is not feasible (e.g. because of skin irritation), and invasive heart rate measurements are not possible in non-clinical settings, the second-best option would be an intensive repeated measurements design. HR(V) is well known to fluctuate with changing posture and activities in ecological real-life monitoring designs (Riese et al., 2004; Vrijkotte et al., 2000). To account for this a highly controlled procedures for data collection in the laboratory, as well as in real-life, was used in the current study.

To test the potential of HR(V) time-series data, a specific study design and a HR(V) monitor suitable for long-term (i.e. four months) monitoring is required. Based on literature search, pilots that include analysis of the raw time-series data with the potential selected HR monitors, and our own expertise, we set the following criteria a monitor should fulfil; (i) feasibility of four months HR(V) monitoring, which requires participants to initialize and operate the monitors themselves on a smartphone; (ii) wireless monitoring, as multiple long wires attached to the

electrodes can be accidentally pulled and detached from the monitor interrupting data collection (Shin et al., 2005; Winokur et al., 2013); (iii) sufficient battery and memory capacity to support long-term assessment; (iv) the ability to upload data to a protected server; (v) good validity and reproducibility of HR(V) measurements and; (vi) access to the raw data.

There are various types of heart rate monitors, for example in cardiology heart rate is typically monitored with an electrocardiogram (ECG) Holter system (Kennedy, 2013). Holter monitors allow for 24 to 48 hr of continuous measurements with high accuracy. Such a higher degree of accuracy comes at a price though, as the large number of ECG spot-electrodes and wires increases the measurement burden. For research purposes, heart rate monitors with substantial less spot-electrodes and wires were successfully developed for robust continuous 24–48 hr ECG measurements (de Geus et al., 1995; Wegner et al., 2020). However, there are still a number of issues hampering long-term measurements (weeks/months) such as limited data storage and battery capacity, wires between the monitor and the electrodes, skin irritation due to wearing ECG electrodes, and monitor costs.

Recently many innovative heart rate monitors were released. One could contend that there are many alternative consumer-grade monitors, including the well-known *Fitbit*, *Polar RS400*, or *Apple Watch*. Indeed, studies have shown agreeable accuracy of such monitors when compared to chest-strapped ECG monitors (Stahl et al., 2016). In other studies, however, wrist-worn monitors were found to provide non-consistent accuracy during motion when compared to a chest strap-based ECG monitor (Wang et al., 2017). In an effort to optimize accuracy, combined with the aforementioned essential criteria, such as access to raw data, we selected the Cortrium C3 ECG monitor (cortrium.com) and the Ithlete photoplethysmography (PPG) finger sensor (myithlete.com). Both ECG and PPG assess *interbeat interval* (IBI) time-series data from which HRV measures can be calculated.

The Cortrium is a wireless 3-lead ECG monitor, which is attached to the chest with three spot-electrodes. The signal is sent via a Bluetooth connection to the user's smartphone and data are saved in real-time. From the smartphone, data can be transferred to any protected server worldwide solving potential data storage issues. The Cortrium also has an internal memory. Such multiple data storage sites can act as a buffer against potential data loss. The renewed interest in heart rate monitoring has also reinvigorated interest in optically based methods, such as PPG. With PPG, blood volume changes are detected by illuminating tissue and measuring changes in light absorption, from which the R-peaks are deduced. Especially for long-term monitoring the PPG method offers the advantage of being ECG spot-electrode free preventing skin irritation which can hamper feasibility and increase non-compliance. Earlier studies established substantial agreement (correlation coefficients between 0.81 and 0.99, and <3% error rate) between PPG and ECG measures

under controlled laboratory conditions (Lu et al., 2009; Teng & Zhang, 2003). The *Ithlete PPG finger sensor* is also controlled by a smartphone application, and offers data storage options via Bluetooth connection and on protected servers. The Ithlete finger sensor uses an infrared light emitting diode as a light source. Investigations into possible negative effects of body mass index (BMI) on accuracy of HR assessment with wrist-worn PPG devices have obtained evidence both for and against such a negative effect. We do not expect BMI to considerably hinder HR accuracy in our study with the Ithlete, as it measures at the tip of the finger, a location which was found to be most sensitive to blood volume fluctuations (Nardelli et al., 2020).

While the Cortrium and the Ithlete offer interesting features that can facilitate longer (e.g. months) intensive monitoring, their feasibility, validity, and test-retest reproducibility has not been established yet. Therefore, in the current study we aim to investigate the feasibility of these monitors during laboratory sessions and long-term ambulatory monitoring. Second, the validity of the wireless Cortrium and Ithlete monitors on HR and HRV measurements are tested against a standard wired ECG reference monitor under standardized laboratory conditions. Thirdly, test-retest reproducibility of HR and HRV assessed with the wireless monitors is tested under laboratory conditions over a period of 2 weeks.

2 | METHOD

2.1 | Participants

For this study, we recruited 64 participants (75% female, mean age = 26 years) from the University Medical Center Groningen (UMCG) and the University of Groningen, through recruitment flyers (see: osf.io/yanqd/). Participants were eligible to participate when the following criteria were met: being (i) 18 years or older, (ii) able to follow the study procedures, (iii) sufficiently proficient in Dutch to fill-out the ESM items and operate a smartphone, (iv) giving written informed consent, and (v) not suffering from cardiovascular diseases, diabetes mellitus, anemia, or using cardio-active medication. Participants received a €25 gift card when completing at least 80% of the measurements and a summary report of their personal data.

Of the original 64 recruited participants, 51 did successfully complete the study. One participant was excluded due to use of cardioactive medication, three other participants did initially agree with study participation but did not to show-up for the first appointment. Lastly, nine participants dropped out during the course of the study. The reasons for dropout were as follows: skin irritation caused by the ECG electrodes (one time), fear caused by observing own heart rate (one time), time constraints (two times), not reacting any more to communication efforts (two times), and no reported reason

(three times). The study protocol was submitted to the ethical review board of the UMCG, who confirmed that formal assessment was not required. The study is registered in the UMCG research register (no. 20160039).

2.2 | Monitor specifications

As reference the *VU-AMS monitor* (www.vu-ams.nl) was used as its validity and reproducibility of measuring cardiovascular indices have been established and are on par with traditional non-ambulatory ECG monitors used in laboratories (de Geus et al., 1995; Goedhart et al., 2007; Willemsen et al., 1996). Recorded signals were ECG (VU-AMS, Cortrium) or PPG (Ithlete). Sample rate was fixed by design at 250 Hz for both the Cortrium and Ithlete devices, and set to 250 Hz for the VU-AMS system to facilitate device comparison, although having a sampling frequency higher than 250 Hz would provide higher resolution data, as described in detail elsewhere, this sample rate is sufficient for the aims of the current study (Greaves-Lord et al., 2010), as the contribution of the rounding error at 250 Hz was found to be small (i.e., error variance = 1.3 ms^2, LF contribution = 0.4 ms^2, HF contribution = 1.4 ms^2). From the ECG and PPG signals R-peaks were triggered (details below) to obtain inter-beat intervals (IBI, in ms) between two successive heartbeats. HRV is calculated as its primary time-domain measure, the root mean square of successive differences (RMSSD, in ms) between two heartbeats (Malik et al., 1996).

2.3 | Study design

A flowchart of the study is shown in Figure 1. Data were collected during two laboratory sessions and 2 weeks of ambulatory measurements. The laboratory measurements were designed to assess the cross-instrument and test-retest reproducibility of the monitors in a controlled laboratory setting. The ambulatory assessments were designed to assess monitor feasibility during ecological valid ambulatory settings. These ambulatory assessments took place in an ESM design; meaning that participants receive an ESM questionnaire five times a day at 3-hr intervals, after which they will conduct ECG/ PPG measurements. The questionnaires were sent via text message, while a reminder text was sent after 10 min if participants had not yet responded. Participants had to complete the ESM questionnaire within half an hour. Filling out the questionnaire took about 2 min. An overview of the included ESM items translated into English is available online (see: osf.io/e8vnh/).

Laboratory: During the first laboratory session participants started with a 15-min intake, which included amongst others, questions about medication use, alcohol use, tobacco
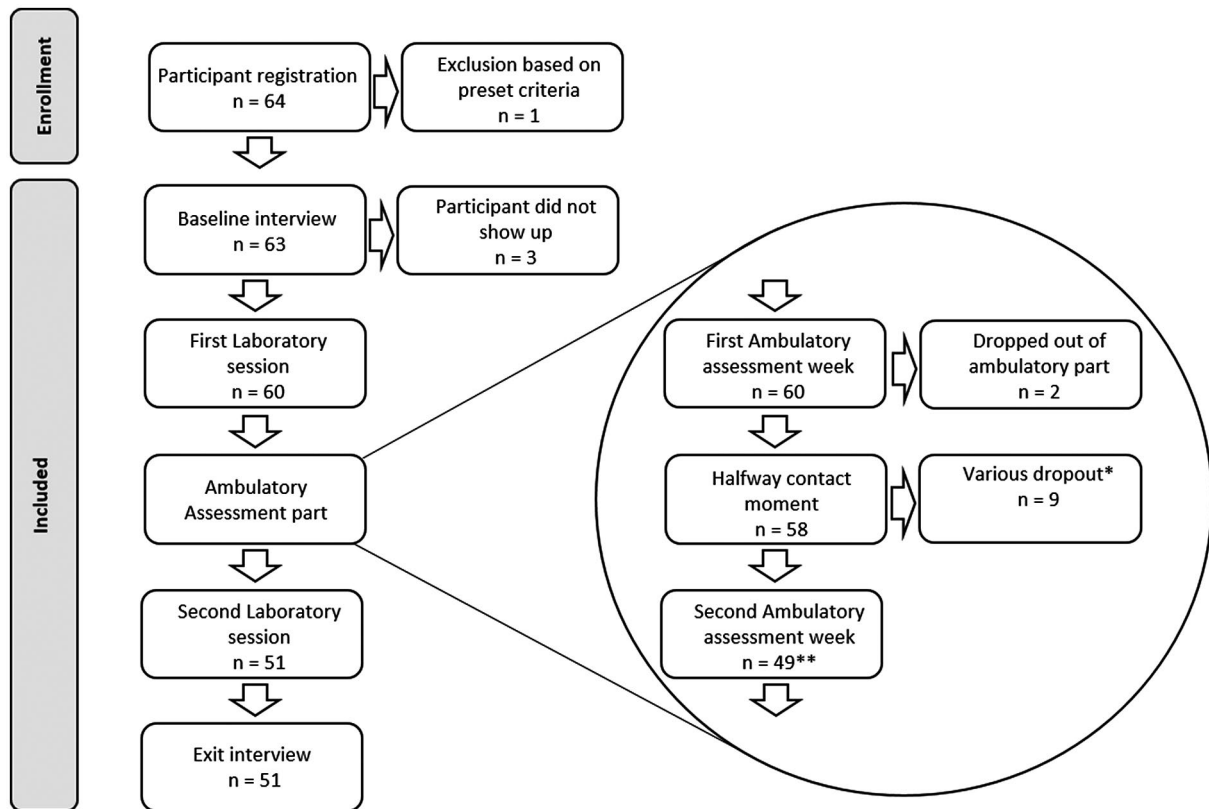
FIGURE 1 Flowchart of the study. *A precise specification of these various dropout reasons is given in the method section. **The second laboratory session included two more participants than the 49 that finished the second ambulatory week is because although they dropped out of the ambulatory assessment part they agreed to participate in the second laboratory session
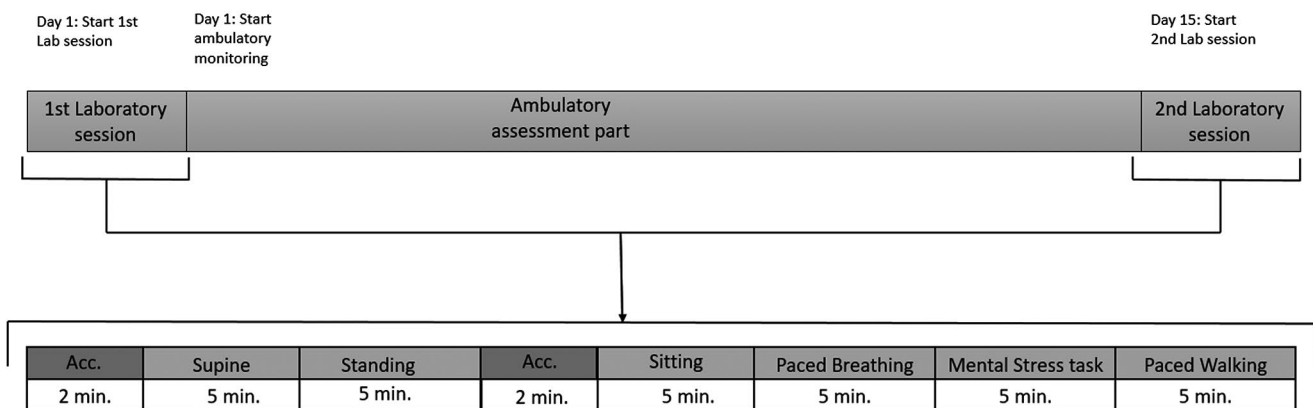

FIGURE 2 Visualization of the study design. Upper part: The study involved two laboratory sessions and ambulatory measurements. Lower part: Enlargement of the laboratory sessions. The two blocks labelled "Acc." indicate 2 min of acclimatization. The other blocks indicate the six laboratory tasks: supine, standing, sitting, paced breathing, a mental stress task, and paced walking

use, and contraception use (see osf.io/yanqd/ for more details). Next, the participant was attached to the *Cortrium, Ithlete*, and *VU-AMS*. Participants wore all three monitors simultaneously during the laboratory sessions. The first laboratory session took approximately 90 min and involved six standardized physical and mental tasks (see Figure 2). After the 14 days of ambulatory measurements, participants returned for the second laboratory session, which involved the same laboratory tasks and an additional structured evaluation interview.

Ambulatory setting: After the first laboratory session, participants continued with 14 days of monitoring themselves with the *Cortrium* and *Ithlete* during their normal daily life. Participants wore the *Ithlete* and *Cortrium* monitors

simultaneously. Participants measured themselves five times a day by filling-out an ESM diary (2 min) and subsequently conduct the ECG/PPG measurements (5 min). ESM was used as a reminder for the heart rate measurement and as a timer for the acclimatization phase. ESM data were not used in the current study.

2.4 | Procedure

Laboratory setting: The lab sessions involved tasks in the following preset order: acclimatization in supine position (2 min), rest in supine position with eyes closed (5 min); standing position with eyes open (5 min); acclimatization in sitting posture (2 min); rest in sitting posture (5 min); paced breathing task in sitting position (5 min); mental stress task (5 min); and paced walking (5 min). Acclimatization after posture changes were used to obtain stationary ECG signals. For the paced breathing task visual stimuli on screen guided participants to pace their breathing with a 0.25 Hz frequency. The mental stress task is a challenging Stroop task. The Stroop task is known to reliably elicit cardiac responses (Eliasson et al., 1983; Freyschuss et al., 1988). To increase mental stress the research assistant delivered critical feedback to the participant such as "That is not good enough". Paced walking was protocolled as walking with the research assistant in a constant normal walking speed through a preset walking route. After the second laboratory session participants participated in an in-house developed evaluation interview. The interview included 52 questions and took approximately 45 min. The difference with the first interview (intake) were the additional evaluation questions about feasibility, ease-of-use, and burdensomeness of the wireless monitors. Participants were also asked about the procedural fidelity, such as reasons for missing heart rate measurements during the 14-day ambulatory setting. An overview of the items used to assess feasibility, ease-of-use, and burdensomeness is available online (see: osf.io/4nuwg/).

Ambulatory setting: At the end of the laboratory session, the research assistant instructed participants to fill-out the ESM diary within 30 min when prompted by a text message on their smartphone in sitting posture. Participants were instructed to remain seated for 2 min to further acclimatize to ascertaining signal stationarity and preventing changes in posture and motion artefacts. Then the ECG/PPG measurements were started in sitting posture while breathing spontaneously and refraining from talking. Assistance from a research assistant was available for participants during the full study period.

2.5 | Data pre-processing

For labelling the VU-AMS data collected during the laboratory sessions, the VU-DAMS software (version 4.3) was used. Each task in the lab session and each assessment in the ambulatory situation was given a label. Labels indicate the start and end of a block of time-series data entered into the pre-processing procedure prior statistical analyses, and reported in the result section. Raw IBI time-series data were pre-processed in R-peak detection software. Data pre-processing steps included converting files, checking file integrity, and correcting for (motion) artefacts. Conceptually there were no differences between preprocessing in either the in-house developed Precar or the Drosan software, although the implementation logically differed due to inherent differences between raw ECG and PPG time-series data. Drosan version 2.52, (Zhang et al., 2019) was used for pre-processing the Ithlete data, and Precar version 3.83 (Greaves-Lord et al., 2010) for pre-processing the Cortrium and VU-AMS data. The CARSPAN program is an in-house developed software package for processing and analyzing IBI time-series (Mulder et al., 1995).

Data pre-processing involved checking the integrity of the time-series data. Missing data were interpolated up to a maximum of 10 s. but in not more than 10% of the total block duration. Otherwise, time-series data in a block was set to missing due to poor data quality. Major reasons for unsatisfactory data quality were poor connection between spot-electrodes and the participants' skin, and motion artefacts. Both ECG and PPG methods are known to be vulnerable for such motion artefacts (Thakor & Zhu, 1991; Trivedi et al., 1997). Data analysts were first trained by analyzing ten example files under supervision of an expert cardiology analyst. Data analysts were allowed to work on the real time-series data files after sufficiently high *intraclass correlation coefficient* (ICC) values (ICC > .95) between the files processed by the analyst and those processed by the expert cardiology analyst were attained.

It was checked whether the data were not too noisy for analysis, whether the R-squared values were at least 0.30, and whether vcIBI values were above 20%. 1.03% of the Cortrium files and 11.00% of the Ithlete files were found to exceed these criteria. When such physiological implausible values were detected, these were followed-up up with a check in the raw data to make sure no R-peaks or artefacts were missed during data pre-processing.

2.6 | Statistical analysis

All statistical analysis and plotting of the data were performed in the statistical programming language R (R Core Team, 2017). Prior analysis, data distributions were checked for dispersion and skewness by visually examining QQ-plots, density plots, and skewness-kurtosis plots (Cullen & Frey, 1999), and testing for normality with a Shapiro-Wilk test (Shapiro & Wilk, 1965). RMSSD values were natural

log transformed to conform to assumptions of linear analyses (Ellis et al., 2008).

First, we described feasibility characteristics of the monitors, such as amount of data collected with each monitor. Descriptive statistics of the evaluation interview were calculated. Second, cross-instrument validity was assessed by comparing both the Cortrium and the Ithlete to the VU-AMS during the laboratory tasks. The variables of interest are mean IBI and ln(RMSSD). Intraclass correlation coefficients (ICC's) were calculated; values closer to one indicate closer adherence to the reference. ICC values were interpreted as follows: <.40 as *poor*, between .40 and .59 as *fair*, between .60 and .74 as *good*, and between 0.75 and 1.00 as *excellent* (Cicchetti, 1994). Third, the test-retest reproducibility was tested. With paired student's *t*-tests changes in mean IBI and ln(RMSSD) values obtained in the first and second laboratory sessions were tested. Absolute reproducibility, which shows the predicted trial-to-trial noise within participants, was assessed by calculating the *standard error of measurement* (*SEM*), also known as the within-subject standard deviation. Furthermore, the *coefficient of variation* (*CV*, in %) was calculated as an indication of reproducibility: lower CV values indicate higher reproducibility (lellamo et al., 1996). For instance, a CV of 20% indicates that around $^2/_3$ of test-retest differences can be found within 20% of the mean score (Atkinson & Nevill, 1998). Missing data were handled through list-wise deletion for the test-retest and cross-instrument parts separately. Bland-Altman plots (Bland & Altman, 1999) were used to visualize agreement between values obtained with the wireless monitors and the VU-AMS. In these plots, the differences of each couple of repeated measurements are plotted against the average of these two measurements. Third, the test-retest reproducibility was tested by comparing the measurements of the Cortrium and the Ithlete during the first laboratory session with the corresponding measurements during the second laboratory session. Additionally, a Welch *t*-test was performed on mean IBI and ln(RMSSD) values of the first and second laboratory sessions for each monitor as Bartlett tests indicated unequal variances.

3 | RESULTS

3.1 | Descriptive statistics

Descriptive statistics for HR and HRV are given in Table 1 (see for a transposed version of Table 1, which allows for easy monitor comparison, https://osf.io/undy2/download). Visual indicators (i.e., QQ-plots, density plots, and skewness-kurtosis plots) and Shapiro-Wilk tests indicated that data were not normally distributed. Therefore, data were natural log-transformed prior statistical analysis. Inspection of the

TABLE 1 Descriptives of IBI and log-transformed RMSSD assessed during the first and second laboratory measurements showing all tested monitors and laboratory tasks. Means (*SD*) and [range] are given

Monitor	Task and Lab session No.	IBI (in ms)	ln(RMSSD) (in ln(ms))
Cortrium	Supine	987.61 (180.40)	4.21 (0.68)
	1st Lab (*n* = 39)	[657.18–1,429.02]	[2.49–5.41]
	Supine	985.47 (167.11)	4.26 (0.72)
	2nd Lab (*n* = 37)	[726.22–1,413.60]	[2.79–5.43]
	Standing	765.54 (134.66)	3.31 (0.54)
	1st Lab (*n* = 38)	[520.08–1,113.66]	[2.26–4.61]
	Standing	746.27 (119.29)	3.27 (0.60)
	2nd Lab (*n* = 37)	[538.96–1,176.21]	[2.21–4.50]
	Sitting	895.38 (147.80)	3.81 (0.64)
	1st Lab (*n* = 39)	[589.67–1,242.34]	[2.20–4.95]
	Sitting	876.96 (134.33)	3.78 (0.62)
	2nd Lab (*n* = 37)	[584.43–1,255.89]	[2.05–4.99]
	Breathing	873.70 (147.02)	3.93 (0.66)
	1st Lab (*n* = 39)	[555.63–1,197.37]	[1.92–5.16]
	Breathing	845.61 (138.61)	3.85 (0.72)
	2nd Lab (*n* = 37)	[570.91–1,221.98]	[2.08–5.39]
	Mental stress	882.86 (149.59)	3.91 (0.60)
	1st Lab (*n* = 39)	[529.28–1,260.23]	[1.92–5.03]
	Mental stress	873.68 (129.91)	3.80 (0.68)
	2nd Lab (*n* = 37)	[511.51–1,183.36]	[1.50–4.98]
	Walking	694.08 (91.60)	3.25 (0.62)
	1st Lab (*n* = 37)	[514.50–907.48]	[1.46–4.64]
	Walking	696.16 (82.20)	3.23 (0.55)
	2nd Lab (*n* = 36)	[526.18–882.94]	[1.96–4.33]
Ithlete	Supine	1,006.91 (195.78)	4.37 (0.61)
	1st Lab (*n* = 23)	[738.68–1,433.23]	[2.86–5.37]
	Supine	1,001.53 (159.51)	4.41 (0.63)
	2nd Lab (*n* = 28)	[796.42–1,414.95]	[2.78–5.40]
	Standing	769.59 (149.77)	3.37 (0.48)
	1st Lab (*n* = 24)	[597.12–1,245.97]	[2.73–4.49]
	Standing	740.24 (106.11)	3.39 (0.48)
	2nd Lab (*n* = 28)	[601.94–1,163.15]	[2.51–4.49]
	Sitting	898.00 (150.12)	3.91 (0.49)
	1st Lab (*n* = 23)	[720.13–1,305.79]	[2.87–4.67]
	Sitting	882.26 (104.52)	3.92 (0.51)
	2nd Lab (*n* = 28)	[722.15–1,199.70]	[2.99–5.15]
	Breathing	879.51 (147.80)	4.05 (0.54)
	1st Lab (*n* = 24)	[660.31–1,226.03]	[2.99–5.18]
	Breathing	851.56 (112.09)	3.94 (0.58)
	2nd Lab (*n* = 27)	[660.21–1,219.78]	[2.28–5.04]
	Mental Stress	897.00 (154.40)	3.97 (0.46)

TABLE 1 (Continued)

Monitor	Task and Lab session No.	IBI (in ms)	ln(RMSSD) (in ln(ms))
	1st Lab (*n* = 23)	[676.07–1,269.69]	[3.00–4.83]
	Mental stress	888.90 (108.80)	3.93 (0.53)
	2nd Lab (*n* = 27)	[730.72–1,185.84]	[2.83–4.97]
	Walking	706.33 (67.21)	4.49 (0.52)
	1st Lab (*n* = 6)	[604.43–814.47]	[3.76–5.16]
	Walking	679.22 (119.87)	4.39 (0.48)
	2nd Lab (*n* = 5)	[539.61–859.38]	[3.82–5.05]
VU-AMS	Supine	991.00 (175.14)	4.22 (0.64)
	1st Lab (*n* = 45)	[659.74–1,432.45]	[2.51–5.39]
	Supine	1,004.81 (173.88)	4.30 (0.71)
	2nd Lab (*n* = 45)	[726.98–1,431.02]	[2.59–5.42]
	Standing	761.78 (145.27)	3.22 (0.53)
	1st Lab (*n* = 45)	[521.67–1,251.65]	[2.27–4.46]
	Standing	750.28 (138.38)	3.22 (0.60)
	2nd Lab (*n* = 45)	[541.23–1,351.38]	[1.89–4.60]
	Sitting	896.43 (151.53)	3.79 (0.61)
	1st Lab (*n* = 45)	[590.93–1,301.51]	[2.20–4.94]
	Sitting	889.33 (147.44)	3.80 (0.59)
	2nd Lab (*n* = 45)	[585.86–1,437.04]	[2.05–4.96]
	Breathing	874.72 (146.82)	3.91 (0.62)
	1st Lab (*n* = 45)	[557.01–1,220.17]	[1.89–5.15]
	Breathing	851.37 (147.68)	3.83 (0.69)
	2nd Lab (*n* = 44)	[572.56–1,344.88]	[2.09–5.38]
	Mental stress	883.39 (146.10)	3.84 (0.54)
	1st Lab (*n* = 45)	[530.77–1,271.88]	[1.90–4.83]
	Mental stress	887.37 (142.97)	3.80 (0.64)
	2nd Lab (*n* = 44)	[512.94–1,371.76]	[1.52–4.99]
	Walking	704.34 (97.58)	3.09 (0.54)
	1st Lab (*n* = 45)	[516.18–985.97]	[1.47–4.70]
	Walking	697.70 (80.62)	3.18 (0.51)
	2nd Lab (*n* = 43)	[527.27–945.73]	[1.88–4.18]

Abbreviations: IBI: interbeat interval, in ms; ln RMSSD: natural logarithm of the root mean square of successive differences between normal heartbeats, in ln(ms).

residuals versus fitted plots indicated that the assumption of equal variances was not violated.

3.2 | Feasibility

Laboratory*:* Of the 51 participants who completed the study, 35 (69%) were able to obtain complete data with all three monitors in both laboratory sessions. During laboratory sessions technical difficulties leading to data loss was applicable for the: VU-AMS monitor in three participants (6%), Cortrium

for six participants (12%), and the Ithlete in 15 participants (29%). In total, 45 participants (88%) completed both laboratory sessions with: the Cortrium (46 hr and 18 min of data), 36 participants (71%) with the Ithlete (29 hr and 40 min), and 48 participants (94%) with the VU-AMS (55 hr and 36 min).

Ambulatory setting: Two of the 60 participants who started with the ambulatory measurements stopped collecting data but agreed to participate in the second laboratory session. Nine participants dropped out completely during the ambulatory part of the study. The remaining 49 participants could maximally obtain 3,430 measurements (49 participants × 14 days × 5 measurements). These 49 participants collected 2,519 measurements (213 hr and 17 min of data, 73,44%) with the Cortrium and 2,182 measurements (176 hr and 20 min of data, 63,61%) with the Ithlete. Three participants experienced technical difficulties (one participant with the Cortrium, two with the Ithlete) leading to a loss of more than 50% of their data.

Evaluation: All 49 participants reported to have missed at least one measurement due to non-adherence to the instructions. Participants specified seven reasons for missing measurements: (i) work (25 times), (ii) spare time activities (22 times), (iii) forgot monitor and/or measurements (16 times), (iv) technical difficulties with monitor(s) (10 times; four out of 49 participants (6.25%) with the Cortrium, six out of 49 participants (9.38%) with the Ithlete), (v) travelling (9 times), (vi) technical difficulties with smartphone or connection (five times), and (vii) skin irritation (four times). Participants reported four reasons to continue with the measurements: (i) having agreed to participate in the study (23 times), (ii) wanting to support research (16 times), (iii) being interested in the study results (12 times), and (iv) being supportive to the researchers (3 times).

The Cortrium was given average scores of 66.78 (*SD* = 20.00) for user-friendliness, 66.39 (*SD* = 22.55) for social acceptability, and 51.65 (*SD* = 27.45) for burdensomeness. The Ithlete was given average scores of 74.29 (*SD* = 23.16) for user-friendliness, 69.65 (*SD* = 23.90) for social acceptability, and 38.51 (*SD* = 25.72) for burdensomeness. The Welch two sample *t*-test indicated that using the Ithlete was evaluated as less burdensome than the Cortrium ($t(100)$ = 2.49, p = .01), and that the monitors did not differ in user-friendliness and social acceptability.

3.3 | Cross-instrument validation

Cross-instrument performance of the Cortrium and the Ithlete against the VU-AMS was tested. Result obtained with the Cortrium are comparable to the VU-AMS (details given in Table 2). Best agreement was found for IBI during the supine task of the first laboratory session (ICC = 1.00, 95% CI = 1.00–1.00, *SEM* = 1.11 ms). Reproducibility, expressed as

PSYCHOPHYSIOLOGY SPR

TABLE 2 Cross-instrument reference method for all laboratory tasks during the first and second laboratory performance of the Cortrium when compared to the VU-AMS session

Lab session	Task		IBI (in ms)	ln(RMSSD) (in ln(ms))
1	Supine (n = 38)	ICC:	1.00	0.91
		95% CI:	1.00–1.00	0.84–0.95
		SEM:	1.11	14.94
		CV:	0.11	17.12
		LOA:	[−3.88–2.26]	[−41.21–41.62]
	Standing (n = 37)	ICC:	1.00	0.89
		95% CI:	1.00–1.00	0.80–0.94
		SEM:	1.13	5.71
		CV:	0.14	17.70
		LOA:	[−3.64–2.63]	[−13.12–18.52]
	Sitting (n = 38)	ICC:	1.00	1.00
		95% CI:	1.00–1.00	1.00–1.00
		SEM:	1.22	0.65
		CV:	0.13	1.14
		LOA:	[−3.93–2.81]	[−1.42–2.17]
	Breathing (n = 38)	ICC:	1.00	0.98
		95% CI:	1.00–1.00	0.96–0.99
		SEM:	1.21	5.77
		CV:	0.13	9.02
		LOA:	[−3.97–2.73]	[−13.66–18.32]
	Mental stress (n = 38)	ICC:	1.00	0.81
		95% CI:	1.00–1.00	0.64–0.90
		SEM:	2.18	13.41
		CV:	0.23	22.45
		LOA:	[−7.22–4.84]	[−29.91–44.43]
	Walking (n = 36)	ICC:	1.00	0.53
		95% CI:	1.00–1.00	0.23–0.73
		SEM:	1.49	10.81
		CV:	0.20	35.20
		LOA:	[−5.65–2.60]	[−22.68–37.24]
2	Supine (n = 36)	ICC:	1.00	0.94
		95% CI:	1.00–1.00	0.89–0.97
		SEM:	1.54	14.20
		CV:	0.15	14.95
		LOA:	[−5.42–3.11]	[−41.47–37.22]
	Standing (n = 36)	ICC:	1.00	0.99
		95% CI:	1.00–1.00	0.98–1.00
		SEM:	1.37	1.33
		CV:	0.17	4.25
		LOA:	[−4.51–3.10]	[−2.73–4.67]
	Sitting (n = 35)	ICC:	1.00	1.00
		95% CI:	1.00–1.00	0.99–1.00
		SEM:	1.42	1.65
		CV:	0.15	3.13
		LOA:	[−4.38–3.51]	[−4.13–4.99]

(Continues)

TABLE 2 (Continued)

Lab session	Task		IBI (in ms)	ln(RMSSD) (in ln(ms))
	Breathing (n = 36)	ICC:	1.00	0.98
		95% CI:	1.00–1.00	0.97–0.99
		SEM:	1.92	5.32
		CV:	0.22	8.89
		LOA:	[−5.76–4.86]	[−12.87–16.64]
	Mental stress (n = 36)	ICC:	1.00	1.00
		95% CI:	1.00–1.00	0.99–1.00
		SEM:	1.41	2.20
		CV:	0.15	3.92
		LOA:	[−4.83–2.97]	[−5.21–7.00]
	Walking (n = 34)	ICC:	1.00	0.88
		95% CI:	1.00–1.00	0.74–0.94
		SEM:	2.27	4.03
		CV:	0.31	13.99
		LOA:	[−8.90–3.71]	[−8.39–13.95]

Abbreviations: 95% CI: 95% confidence interval of ICC; CV: coefficient of variation in %; IBI: interbeat interval, in ms; ICC: intraclass correlation coefficient; ln RMSSD: natural logarithm of the root mean square of successive differences between normal heartbeats, in ln(ms); LOA: lines of agreement; *SEM*: standard error of measurement in ms for IBI mean, ln(ms) for ln(RMSSD).

the CV was 0.11%, indicated that about $^2/_3$ of the differences are within 0.11% of the mean IBI values. Lowest agreement was found for ln(RMSSD) during the walking task of the first laboratory session (ICC = 0.53, 95% CI = 0.23–0.73, *SEM* = 10.81 ln(ms), CV = 35.20%). However, even in this latter case the agreement based on the ICC values should still be interpreted as *fair*.

The results obtained with the Ithlete during most tasks are also comparable to the VU-AMS (details given in Table 3). Best agreement was found for the IBI during the mental stress task of the second laboratory session (ICC = 1.00, 95% CI = 1.00–1.00, *SEM* = 1.39 ms). Reproducibility expressed as the CV was 0.15%: indicating that about $^2/_3$ of the differences are within 0.15% of the mean IBI values. However, during the walking tasks of both laboratory sessions the ln(RMSSD) values calculated from the Ithlete data did deviate substantially from those obtained by the VU-AMS. Lowest agreement was found during the first walking task (ICC = −0.08, 95% CI = −0.10–0.78, *SEM* = 22.93 ln(ms), CV = 32.12%). It should be noted, however, that in both walking tasks the sample sizes were very small (n = 6, and 5, respectively) as motion artefacts resulted in missing data.

In sum, in the walking tasks, the Cortrium outperformed the Ithlete. However, under circumstances without motion artefacts, differences between the Cortrium and the Ithlete were negligible. This is visualized in the Bland-Altman plots given in Figure 3 for IBI data collected during the first laboratory session with the Cortrium and VU-AMS, and the supplementary materials Figures S1 to S7 for the other variables and sessions. Although the absolute mean differences are small,

the Bland-Altman plots showed that the Cortrium tended to underestimated IBI (range: between −2.6 and −0.4 ms). The Ithlete tended to overestimated IBI (range: between −1.3 and 10.7 ms) Again, with small absolute mean differences, both monitors overestimated ln(RMSSD), with values ranging between −0.004 and 0.198 for the Cortrium and values ranging between 0.055 and 1.377 for the Ithlete.

3.4 | Reproducibility

Data assessed during the first and second laboratory session were not different for the Cortrium, the Ithlete, and the VU-AMS (see Table 4 for test-retest statistics). For both the VU-AMS and the Cortrium, we did not find any differences in IBI and ln(RMSSD) during any of the tasks between the first and the second laboratory measurement. For the VU-AMS *good* to *excellent* reliabilities were found (ICC range = 0.64–0.88). For the Cortrium, ICC values indicated *fair* to *excellent* reproducibility (ICC range = 0.53–0.90). For the Ithlete, no difference was found for data assessed in the supine task (ICC range = 0.82–0.86). However, differences were found in the standing, sitting, paced breathing, and mental stress tasks (i.e. lowest ICC values were obtained in the paced breathing task for ln(RMSSD) (ICC = −0.21, 95% CI = −0.71–0.35, *SEM* = 35.21 ln(ms), CV = 54.22%). Not enough observations were attained in the walking task to test for any systematic change due to motion artefacts interfering with R-peak detection. While the Cortrium and VU-AMS did not show differences between the first and second laboratory

TABLE 3 Cross-instrument performance of the Ithlete when compared to the VU-AMS reference method for all laboratory tasks during the first and second laboratory session

Lab session	Task			IBI (in ms)	ln(RMSSD) (in ln(ms))
1	Supine (n = 23)	ICC:		1.00	0.96
		95% CI:		1.00–1.00	0.89–0.99
		SEM:		5.01	8.42
		CV:		0.47	8.81
		LOA:		[−15.22–12.54]	[−16.69–29.99]
	Standing (n = 24)	ICC:		1.00	0.98
		95% CI:		1.00–1.00	0.45–0.99
		SEM:		1.79	1.61
		CV:		0.22	4.77
		LOA:		[−4.47–5.46]	[−0.91–8.00]
	Sitting (n = 23)	ICC:		1.00	0.94
		95% CI:		0.99–1.00	0.71–0.98
		SEM:		10.04	4.80
		CV:		1.06	8.42
		LOA:		[−24.76–30.89]	[−7.71–18.88]
	Breathing (n = 24)	ICC:		1.00	0.98
		95% CI:		1.00–1.00	0.48–1.00
		SEM:		5.03	2.28
		CV:		0.54	3.35
		LOA:		[−14.22–13.67]	[−0.75–11.92]
	Mental stress (n = 23)	ICC:		1.00	0.98
		95% CI:		1.00–1.00	0.95–0.99
		SEM:		6.88	3.13
		CV:		0.72	5.14
		LOA:		[−16.87–21.29]	[−6.50–10.88]
	Walking (n = 6)	ICC:		0.97	−0.08*
		95% CI:		0.78–1.00	−0.25–0.47
		SEM:		10.28	41.17
		CV:		1.38	51.82
		LOA:		[−17.75–39.22]	[−39.91–188.32]
2	Supine (n = 27)	ICC:		1.00	0.98
		95% CI:		1.00–1.00	0.94–0.99
		SEM:		3.58	8.04
		CV:		0.34	7.77
		LOA:		[−10.21–9.64]	[−17.22–27.38]
	Standing (n = 27)	ICC:		1.00	0.79
		95% CI:		1.00–1.00	0.44–0.92
		SEM:		3.24	6.78
		CV:		0.41	20.81
		LOA:		[−9.08–8.91]	[−11.98–25.63]
	Sitting (n = 27)	ICC:		0.88	0.63
		95% CI:		0.76–0.94	0.34–0.81
		SEM:		35.98	18.57
		CV:		3.83	31.54
		LOA:		[−95309–104.40]	[−42.23–60.71]

(Continues)

TABLE 3 (Continued)

Lab session	Task		IBI (in ms)	ln(RMSSD) (in ln(ms))
	Breathing (n = 26)	ICC:	1.00	0.84
		95% CI:	1.00–1.00	0.62–0.93
		SEM:	2.87	11.31
		CV:	0.32	18.54
		LOA:	[−8.26–7.68]	[−22.56–40.14]
	Mental stress (n = 26)	ICC:	1.00	0.82
		95% CI:	1.00–1.00	0.65–0.92
		SEM:	1.39	13.24
		CV:	0.15	21.89
		LOA:	[−4.43–3.29]	[−30.63–42.78]
	Walking (n = 5)	ICC:	0.99	0.20*
		95% CI:	0.70–1.00	−0.1–0.78
		SEM:	6.21	22.93
		CV:	0.87	32.12
		LOA:	[−4.44–29.98]	[−1.42–125.69]

Abbreviations: 95% CI: 95% confidence interval of ICC; CV: coefficient of variation in %; IBI: interbeat interval, in ms; ln RMSSD: natural logarithm of the root mean square of successive differences between normal heartbeats, in ln(ms). ICC: intraclass correlation coefficient; LOA: lines of agreement. *: ICC <= 0.40 indicating poor reproducibility between measurements from the monitor and the reference method (see method section for more details); *SEM*: standard error of measurement in ms for IBI mean, ln(ms) for ln(RMSSD).

sessions, the Ithlete did show differences in four tasks for IBI and ln(RMSSD).

4 | DISCUSSION

In this study we tested the cross-instrument performance of two wireless heart rate monitors, the Cortrium ECG C3 and the Ithlete PPG finger sensor, against a standard wired reference method under controlled laboratory conditions. Moreover, we studied the test-retest reproducibility of these monitors over a period of 14 days, while their ambulatory feasibility was also investigated. We found that both the Cortrium and the Ithlete offer good to excellent cross-instrument agreement with the reference method under five standardized laboratory tasks, namely: supine, standing, sitting, paced breathing, and mental stress. The Cortrium did also perform well in a walking task, whereas the Ithlete showed inferior performance under such circumstances due to its higher sensitivity to motion artefacts. Test-retest analyses showed that results obtained with both the VU-AMS reference and the Cortrium monitor were comparable. Ithlete test-retest results were less robust, although IBI's during supine, standing, sitting, and the mental stress tasks showed good to excellent reproducibility.

Regarding feasibility, during ambulatory measurements, both the Cortrium and the Ithlete delivered at least two thirds of the maximum possible measurements. Participants reported that measurements were missed due to daily interferences, such as work obligations or leisure time activities. As all participants reported to have missed at least one measurement due to such non-adherence to instructions, we can identify non-adherence as an important contributing factor for missing data. Less often monitor related reasons were reported, such as technical difficulties and skin irritation due to ECG spot-electrodes. These findings can be interpreted as that HR(V) data collection with both wireless devices is feasible in highly protocolled ambulatory settings, although more valid data were obtained with the Cortrium than with the Ithlete. Main reasons reported for compliance were having agreed to complete the study and wanting to support research. It seems therefore worthwhile to invest in the participant- researcher relationship to reduce the amount of missing data in a study. Participants reported no differences between the Cortrium and the Ithlete on user-friendliness and social acceptability.

We conclude from the current study that under most of the laboratory tasks, the Cortrium and the Ithlete showed good to excellent agreement with a standard wired ECG reference method when assessing HR(V). It was shown that for measuring HR(V) during tasks that do not involve gross body movements or physical activity, researchers are not limited to standard wired ECG monitors but can also opt for the modern wireless heart rate monitors investigated in this study. These wireless monitors offer a number of advantages of interest to researchers such as: online data storage, no need for battery replacement, giving access to the raw data, and lower monitor costs. There was, however, a difference between the Cortrium
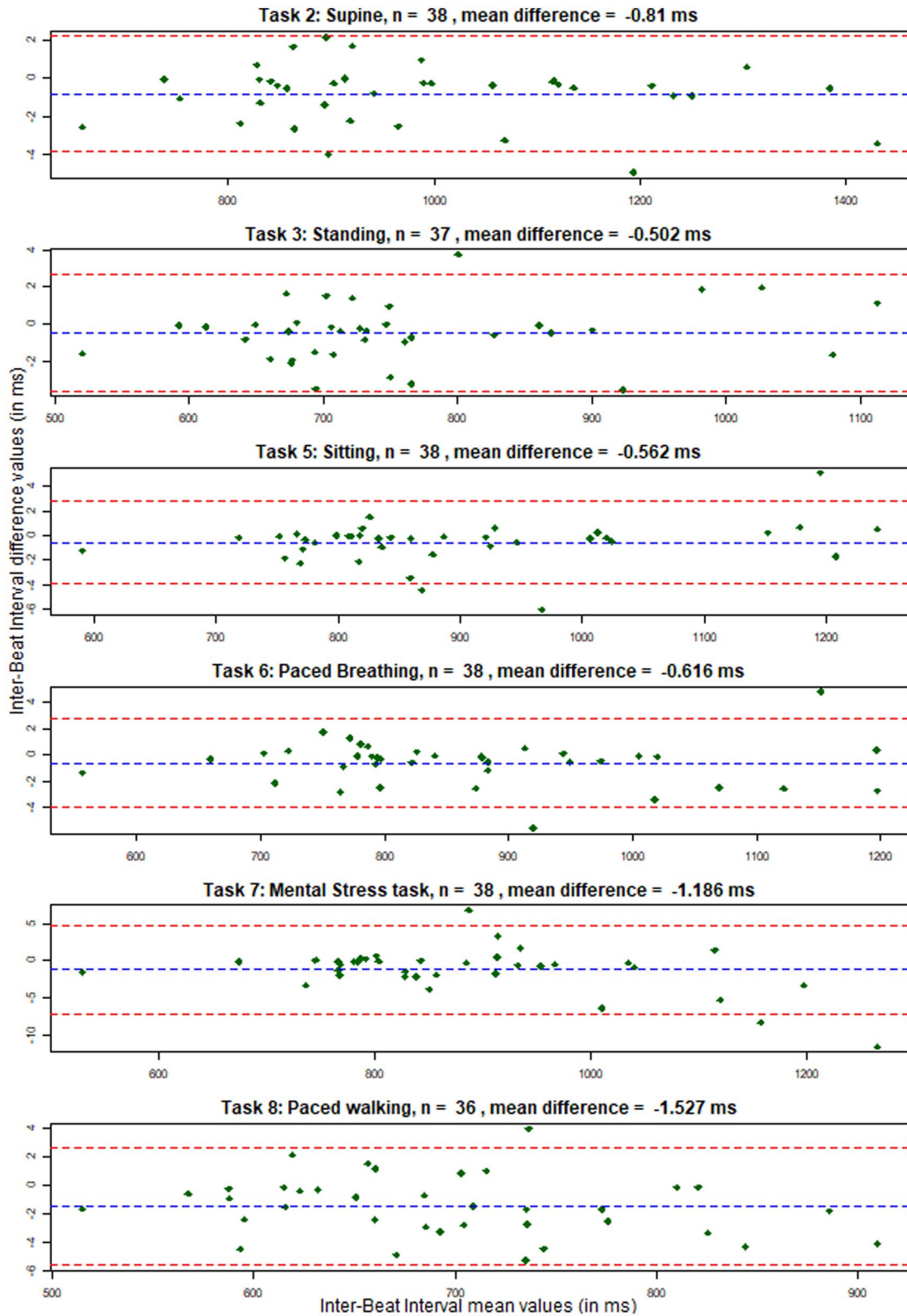
FIGURE 3 Bland-Altman plots of the Inter-Beat Interval data collected during the first laboratory session with the Cortrium versus the VU-AMS device (details on the laboratory session are described in the Method section and depicted in Figure 2). The blue dotted lines represent the mean difference between the Inter-Beat Interval values, while the red dotted lines represent the limits of agreement from negative 1.96 until positive 1.96 times the standard deviation of the differences. On the *x*-axis the Inter-Beat Interval mean values are given while the *y*-axis shows the differences between Inter-Beat Interval values obtained from the two devices

and the Ithlete in sensitivity to motion artefacts, which is associated with the larger amount of missing data obtained with the Ithlete (especially during tasks which include motion such as the walking tasks). Our findings indicate that the Cortrium recordings are fairly robust to motion. As such, the signal of the Cortrium is expected to be not as strongly affected by motion artefacts as consumer-grade wrist-worn PPG monitors, such as Fitbit monitors or the Apple Watch, whose signal is less robust under motion conditions than ECG monitors such as the Polar H7 chest-strap (Wang et al., 2017). Conversely, wrist-worn PPG monitors do offer their own set of characteristics, which could offer advantages regarding feasibility in some research designs. For example, such wearables can offer the ability to provide continuous recordings, as the device can be worn comfortably for long periods. This is due to such wrist-worn wearables often being designed to be worn as a bracelet or a watch. This could also prevent participants forgetting the monitor or the measurements. In our study, the reproducibility of the Ithlete monitor dropped considerably during motion, more so than various wrist-worn PPG monitors under the motion condition (Wang et al., 2017). Such dissimilarities could be due to differences in laboratory tasks, for instance, using a treadmill walking task versus walking a predetermined route alongside a research assistant for a pre-set duration. A future study investigating both types of PPG monitors under similar conditions could elucidate whether wrist-worn PPG monitors definitively outperform finger-worn PPG monitors during motion. The robustness of the Cortrium signal during motion is similar to a chest-strapped ECG monitor, such as the Polar H7. Additionally, it avoids some disadvantages of chest-strapped monitors, such as wearability issues during long-term measurements. While a chest-strap ECG such as the Polar H7 does offer relatively robust signal and was thus considered for use in the current study, it was found less suitable for our long-term monitoring goals due to wearability issues such as obstructions of clothes while putting on the monitor. Additionally, these modern heart rate monitors offer lower prices compared to fully-fledged Holter ECG monitor, as for the price of one VU-AMS system researchers can acquire approximately three to four Cortrium C3ʼs, or 100 Ithlete finger monitors. Such advantages offer researchers new opportunities for designing longitudinal studies wherein HR(V) data is monitored over weeks or months, in large samples within approximately the same budget. Longitudinal studies are necessary when studying dynamic processes, such as transitions from a healthy to a clinically affected state in patients, which unfold over timeframes longer than those studied in short-term research designs. The current study showed that the long-term ambulatory data collection required for such longitudinal studies is indeed feasible, although precautions are to be taken to minimize data loss and to improve adherence to instructions by participants.

Results showed that the PPG-based Ithlete did perform less well in conditions with higher risk of motion artefacts, such as walking. This finding is in line with earlier research showing the vulnerability of PPG measurements to motion artefacts (Trivedi et al., 1997), while extending these earlier findings to both controlled laboratory settings as well as ambulatory settings. Hence, when considering whether to choose the Ithlete or the Cortrium for a scientific study one should consider if heart rate measurements are under conditions free of potential motion artefacts, and whether participants can be recruited easily and inexpensively. Under such conditions the Ithlete could be a sensible choice. However, under other conditions, for example in physical active situations, the Cortrium would be the more sensible choice. Moreover, the amount of data yielded from the Cortrium was higher than that of the Ithlete (83% and 69% of the maximum possible amount, respectively) during the ambulatory measurement period. Therefore, in scientific and clinical contexts wherein minimizing missing data is required, the Cortrium does hold the advantage. This advantage is grounded in the higher robustness of the Cortriums' ECG signal to motion related disruptions of stationarity when compared to the Ithlete's PPG signal. The correspondence between the Cortrium and the VU-AMS is hardly surprising as both monitors measure the ECG signal of lead II, thus delivering R-peaks which are relatively large and easy to detect. It should be noted though that the larger distance between the ECG spot-electrodes for the VU-AMS measurements allow for an even more robust signal, even during 24h monitoring in participants in physical active occupations (Riese et al., 2004; Vrijkotte et al., 2000).

While the current study showed agreeable performance of two wireless heart rate monitors in comparison to a wired ECG monitor, there are some limitations to be noted. First, reliability and validity characteristics of the HR(V) data were obtained from cross-instrument results under a controlled laboratory setting. These findings will thus only generalize to similar laboratory settings only. Reliability and validity of the Ithlete and Cortrium in ecological valid, unprotocolled ambulatory settings remains to be established as the two wireless devices were not tested against the ECG reference method and participants monitored themselves in real-life according to a highly standardized protocol (viz. after stabilization of the signal, in sitting posture). We did show that with both the Ithlete and the Cortrium HR(V) monitor data collection at home is feasible, although more valid data were collected with the Cortrium than with the Ithlete. Second, we only assessed IBI and ln(RMSSD) calculated from the data yielded by the investigated heart rate monitors. There are, however, a multitude of other HR(V) metrics that are of interest to researchers, such as the *high frequency spectral power band* (0.15–0.40 Hz) of IBIs. Future studies could extend the current study to assess performance of the Cortrium and Ithlete regarding these metrics. Although

TABLE 4 Absolute and relative test-retest reproducibility of IBI and log-transformed RMSSD assessed during the first and second laboratory measurements showing all tested monitors and laboratory tasks

Monitor	Task		IBI (in ms)	ln(RMSSD) (in ln(ms))
Cortrium	Supine (n = 35)	ICC:	0.86	0.82
		95% CI:	0.75–0.93	0.67–0.91
		SEM:	61.55	21.76
		CV:	5.97	25.09
		LOA:	[−190.56–150.67]	[−70.59–50.04]
	Standing (n = 34)	ICC:	0.86	0.61
		95% CI:	0.73–0.93	0.35–0.79
		SEM:	48.14	12.60
		CV:	6.03	37.65
		LOA:	[−122.71–144.19]	[−35.16–34.70]
	Sitting (n = 35)	ICC:	0.85	0.70
		95% CI:	0.72–0.92	0.48–0.84
		SEM:	53.85	16.38
		CV:	5.78	30.05
		LOA:	[−142.61–155.91]	[−43.91–46.88]
	Breathing (n = 35)	ICC:	0.86	0.71
		95% CI:	0.74–0.93	0.49–0.84
		SEM:	52.83	21.94
		CV:	5.81	35.23
		LOA:	[−133.04–159.84]	[−61.67–59.95]
	Mental stress (n = 35)	ICC:	0.87	0.70
		95% CI:	0.76–0.93	0.47–0.83
		SEM:	49.61	19.37
		CV:	5.38	32.75
		LOA:	[−146.39–128.65]	[- 52.70–54.71]
	Walking (n = 32)	ICC:	0.90	0.53
		95% CI:	0.80–0.95	0.23–0.74
		SEM:	26.54	11.95
		CV:	3.62	38.09
		LOA:	[−73.84–73.32]	[−30.18–36.05]
Ithlete	Supine (n = 15)	ICC:	0.87	0.70
		95% CI:	0.66–0.95	0.33–0.89
		SEM:	55.60	29.42
		CV:	5.49	29.97
		LOA:	[−178.70–129.54]	[−99.51–63.61]
	Standing (n = 15)	ICC:	0.74	0.38*
		95% CI:	0.38–0.91	−0.17–0.74
		SEM:	46.36	13.28
		CV:	5.97	40.70
		LOA:	[- 126.72–130.27]	[−38.20–35.44]

(Continues)

TABLE 4 (Continued)

Monitor	Task		IBI (in ms)	ln(RMSSD) (in ln(ms))
	Sitting (n = 14)	ICC:	0.72	0.08*
		95% CI:	0.33–0.90	−0.41–0.56
		SEM:	40.76	25.54
		CV:	4.50	45.87
		LOA:	[−142.12–83.86]	[−8370–57.91]
	Breathing (n = 15)	ICC:	0.30*	−0.21*
		95% CI:	−0.27–0.70	−0.71–0.35
		SEM:	83.88	35.51
		CV:	9.31	54.22
		LOA:	[−224.40–240.62]	[−96.80–100.03]
	Mental stress (n = 13)	ICC:	0.81	0.34*
		95% CI:	0.49–0.94	−0.24–0.74
		SEM:	55.68	22.96
		CV:	5.99	37.98
		LOA:	[−176.72–131.96]	[−71.41–55.68]
	Walking (n = 1)	ICC:	Not enough observations	Not enough observations
		95% CI:	Not enough observations	Not enough observations
		SEM:	Not enough observations	Not enough observations
		CV:	Not enough observations	Not enough observations
		LOA:	Not enough observations	Not enough observations
VU-AMS	Supine (n = 42)	ICC:	0.86	0.80
		95% CI:	0.75–0.92	0.64–0.89
		SEM:	66.73	22.73
		CV:	6.35	25.35
		LOA:	[−202.88–167.06]	[−73.98–52.01]
	Standing (n = 42)	ICC:	0.88	0.79
		95% CI:	0.79–0.93	0.64–0.88
		SEM:	50.34	8.98
		CV:	6.30	29.03
		LOA:	[−134.67–144.41]	[−26.61–23.16]
	Sitting (n = 41)	ICC:	0.84	0.68
		95% CI:	0.71–0.91	0.47–0.81
		SEM:	60.70	15.49
		CV:	6.48	29.39
		LOA:	[−172.07–164.42]	[−44.10–41.78]
	Breathing (n = 40)	ICC:	0.86	0.72
		95% CI:	0.75–0.92	0.54–0.84
		SEM:	57.39	21.23
		CV:	6.28	34.53
		LOA:	[−146.44–171.71]	[−51.28–59.43]
	Mental stress (n = 40)	ICC:	0.83	0.80
		95% CI:	0.70–0.91	0.65–0.89
		SEM:	61.06	13.56
		CV:	6.51	24.39
		LOA:	[−177.73–160.74]	[−40.97–34.20]

(Continues)

TABLE 4 (Continued)

Monitor	Task		IBI (in ms)	ln(RMSSD) (in ln(ms))
	Walking (n = 39)	ICC:	0.86	0.64
		95% CI:	0.76–0.93	0.41–0.80
		SEM:	32.00	6.50
		CV:	4.34	25.18
		LOA:	[−92.96–84.41]	[−21.14–14.89]

Abbreviations: 95% CI: 95% confidence interval of ICC; CV: coefficient of variation in %; IBI: interbeat interval, in ms; ln RMSSD: natural logarithm of the root mean square of successive differences between normal heartbeats, in ln(ms). ICC: intraclass correlation coefficient; LOA: lines of agreement. *: ICC <= 0.40 indicating poor reproducibility, and significant differences between measurements during the first and second laboratory sessions. Comparison of data of Lab 1 and Lab 2 tested with paired Student's *t*-test showed all being non-significant (*p* > .05), except the walking task of the Ithlete which did not had enough observations to perform the *t*-test. The Cortrium and VU-AMS did not show significant differences between the first and second laboratory sessions. The Ithlete did show differences in four tasks for IBI means and ln(RMSSD); *SEM*: standard error of measurement in ms for IBI mean, ln(ms) for ln(RMSSD).

no major differences should be expected as these metrics often correlate strongly with each other (Massin et al., 1999; Shaffer & Ginsberg, 2017). Third, in order to provide insights in reasons for missing data, for example what number of measurements is missed due to either monitor issues or non-adherence to instructions, a dedicated ESM question could be included in future studies. This is recommended as it would provide more detail on the amount and reasons for missing data. Fourth, as participants wore the heart rate monitors simultaneously, during laboratory as well as daily life measurements, some burden might have been experienced. However, none of the participants indicated during the evaluation interview that this negatively impacted feasibility, or that it was a reason for missing measurements. Fifth, a method-specific limitation of PPG is that circulation characteristics can result in a phase delay between R-peak and volume pulse start (Lu et al., 2009). However, in our sample of young and healthy participants such variations in delays can be assumed to be negligible (Drinnan et al., 2001) and was not expected to have interfered with the conclusions of the study. Sixth, feasibility results have indicated that there were more technical difficulties with the Cortrium and Ithlete devices than with the reference method during the laboratory sessions. While this can be considered a limitation of feasibility, the reference method is unsuitable for our specific future study goals as, for example, it was not wireless, and not allowing participants to initialize the monitor through their smartphones. However, the devices tested in the current study were selected based on multiple criteria given in the introduction, instead of focusing solely on robustness to technical difficulties.

Lastly, it should be noted that while our study did include a mental stress task, we did not observe the expected cardiovascular response in any of the monitors, but instead showing effects similar to those in the sitting and breathing tasks (see Table 1). This suggests that the used task setup was insufficient to elicit the expected cardiovascular effects of the mental stress task.

In conclusion, two modern wireless heart rate monitors, the Cortrium and Ithlete, are able to provide data quality on par with a standard wired ECG reference method under controlled laboratory circumstances. Although both the Cortrium and the Ithlete performed similarly during non-motion tasks, the Cortrium was more robust during motion. Highly protocolled monitoring with the wireless devices in ambulatory daily-life setting is feasible. Participants highlighted work and spare-time activities as most common reasons to miss a measurement. Overall, we conclude that researchers can benefit from the advantages of modern wireless heart rate monitors such as online data storage and the absence of battery replacements without fully sacrificing data quality.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR CONTRIBUTIONS

Yoram Kunkels: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing-original draft; Writing-review & editing. **Arie van Roon:** Conceptualization; Formal analysis; Investigation; Methodology; Resources; Software; Supervision; Validation; Visualization; Writing-original draft; Writing-review & editing. **Marieke Wichers:** Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Writing-original

draft; Writing-review & editing. **Harriette Riese:** Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Writing-original draft; Writing-review & editing.

ORCID

Yoram K. Kunkels https://orcid.org/0000-0003-2782-2015

REFERENCES

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, *26*(4), 217–238. https://doi.org/10.2165/00007256-199826040-00002

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160. https://doi.org/10.1177/096228029900800204

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cortrium.com. (2019). https://www.cortrium.com

Cullen, A. C., & Frey, H. C. (1999). *Probabilistic techniques in exposure assessment: A handbook for dealing with variability and uncertainty in models and inputs*. Plenum Press.

de Geus, E. J. C., Willemsen, G. H. M., Klaver, C. H. A. M., & van Doornen, L. J. P. (1995). Ambulatory measurement of respiratory sinus arrhythmia and respiration rate. *Biological Psychology*, *41*(3), 205–227. https://doi.org/10.1016/0301-0511(95)05137-6

Drinnan, M., Allen, J., & Murray, A. (2001). Relation between heart rate and pulse transit time during paced respiration. *Physiological Measurement*, *22*(3), 425–432. https://doi.org/10.1088/0967-3334/22/3/301

El-Amrawy, F., & Nounou, M. I. (2015). Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthcare Informatics Research*, *21*(4), 315–320. https://doi.org/10.4258/hir.2015.21.4.315

Eliasson, K., Hjemdahl, P., & Kahan, T. (1983). Circulatory and sympatho-adrenal responses to stress in borderline and established hypertension. *Journal of Hypertension*, *1*(2), 131–139. https://doi.org/10.1097/00004872-198308000-00004

Ellis, R. J., Sollers, J. J. III, Edelstein, E. A., & Thayer, J. F. (2008). Data transforms for spectral analyses of heart rate variability. *Biomedical Sciences Instrumentation*, *44*, 392–397. https://www.ncbi.nlm.nih.gov/pubmed/19141947

Freyschuss, U., Hjemdahl, P., Juhlin-Dannfelt, A., & Linde, B. (1988). Cardiovascular and sympathoadrenal responses to mental stress: Influence of β-blockade. *American Journal of Physiology—heart and Circulatory Physiology*, *255*(6), https://doi.org/10.1152/ajpheart.1988.255.6.h1443

Goedhart, A. D., van der Sluis, S., Houtveen, J. H., Willemsen, G., & de Geus, E. J. C. (2007). Comparison of time and frequency domain measures of RSA in ambulatory recordings. *Psychophysiology*, *44*(2), 203–215. https://doi.org/10.1111/j.1469-8986.2006.00490.x

Greaves-Lord, K., Tulen, J., Dietrich, A., Sondeijker, F., van Roon, A., Oldehinkel, A., Ormel, J., Verhulst, F., & Huizink, A. (2010). Reduced autonomic flexibility as a predictor for future anxiety in girls from the general population: The TRAILS study. *Psychiatry Research*, *179*(2), 187–193. https://doi.org/10.1016/j.psychres.2009.04.014

Myithlete.com. (2019). https://www.myithlete.com

Kemp, A. H., & Quintana, D. S. (2013). The relationship between mental and physical health: Insights from the study of heart rate variability. *International Journal of Psychophysiology*, *89*(3), 288–296. https://doi.org/10.1016/j.ijpsycho.2013.06.018

Kennedy, H. L. (2013). The evolution of ambulatory ECG monitoring. *Progress in Cardiovascular Diseases*, *56*(2), 127–132. https://doi.org/10.1016/j.pcad.2013.08.005

Kramer, I., Simons, C. J. P., Hartmann, J. A., Menne-Lothmann, C., Viechtbauer, W., Peeters, F., Schruers, K., van Bemmel, A. L., Myin-Germeys, I., Delespaul, P., van Os, J., & Wichers, M. (2014). A therapeutic application of the experience sampling method in the treatment of depression: A randomized controlled trial. *World Psychiatry*, *13*(1), 68–77. https://doi.org/10.1002/wps.20090

Kroeze, R., van der Veen, D. C., Servaas, M. N., Bastiaansen, J. A., Voshaar, R. C. O. V., Borsboom, D., Ruhe, H. G., Schoevers, R. A., & Riese, H. (2017). Personalized feedback on symptom dynamics of psychopathology: A proof-of-principle study. *Journal for Person-Oriented Research*, *3*(1), 1–11. https://doi.org/10.17505/jpor.2017.01

lellamo, F., Legramante, M., Raimondi, G., Castrucci, F., Massaro, M., & Peruzzi, G. (1996). Evaluation of reproducibility of spontaneous baroreflex sensitivity at rest and during laboratory tests. *Journal of Hypertension*, *14*(9), 1099–1104. https://doi.org/10.1097/00004872-199609000-00009

Lu, G., Yang, F., Taylor, J. A., & Stein, J. F. (2009). A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. *Journal of Medical Engineering and Technology*, *33*(8), 634–641. https://doi.org/10.3109/03091900903150998

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17*(3), 354–381. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

Massin, M. M., Derkenne, B., & von Bernuth, G. (1999). Correlations between indices of heart rate variability in healthy children and children with congenital heart disease. *Cardiology*, *91*(2), 109–113. https://doi.org/10.1159/000006889

Mulder, L. J. M., van Roon, A. M., & Schweizer, D. (1995). *CARSPAN*. IecProGAMMA.

Nardelli, M., Vanello, N., Galperti, G., Greco, A., & Scilingo, E. P. (2020). Assessing the quality of heart rate variability estimated from wrist and finger PPG: A novel approach based on cross-mapping method. *Sensors (Basel, Switzerland)*, *20*(11), 3156. https://doi.org/10.3390/s20113156

R Core Team. (2017). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. https://www.R-project.org/

Riese, H., van Doornen, L. J. P., Houtman, I. L. D., & de Geus, E. J. C. (2004). Job strain in relation to ambulatory blood pressure, heart rate, and heart rate variability among female nurses. *Scandinavian Journal of Work, Environment and Health*, *30*(6), 477–485. https://doi.org/10.5271/sjweh.837

Scheffer, M. (2010). Complex systems: Foreseeing tipping points. *Nature*, *467*(7314), 411–412. https://doi.org/10.1038/467411a

Schoevers, R. A., van Borkulo, C. D., Lamers, F., Servaas, M. N., Bastiaansen, J. A., Beekman, A. T. F., van Hemert, A. M., Smit,

J. H., Penninx, B. W. J. H., & Riese, H. (2020). Affect fluctuations examined with ecological momentary assessment in patients with current or remitted depression and anxiety disorders. *Psychological Medicine*, 1–10. https://doi.org/10.1017/S0033291720000689

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5, 258. https://doi.org/10.3389/fpubh.2017.00258

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591. https://doi.org/10.2307/2333709

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Shin, K., Hwang, H. T., Kim, Y. H., Kim, J. P., Yeo, H. S., Han, W., & Park, J. C. (2005). WHAM: A novel, wearable heart activity monitor based on Laplacian potential mapping. *Annual International Conference of the IEEE Engineering in Medicine and Biology—proceedings*, 7, 7361–7364. https://doi.org/10.1109/iembs.2005.1616212

Smit, A. C., Snippe, E., & Wichers, M. (2019). Increasing restlessness signals impending increase in depressive symptoms more than 2 months before It happens in individual patients. *Psychotherapy and Psychosomatics*, 88(4), 249–251. https://doi.org/10.1159/000500594

Stahl, S. E., An, H.-S., Dinkel, D. M., Noble, J. M., & Lee, J.-M. (2016). How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport & Exercise Medicine*, 2(1), e000106. https://doi.org/10.1136/bmjsem-2015-000106

Teng, X. F., & Zhang, Y. T. (2003). Study on the peak interval variability of photoplethysmogtaphic signals. *APBME 2003 - IEEE EMBS Asian-Pacific Conference on Biomedical Engineering 2003*, 140–141. https://doi.org/10.1109/APBME.2003.1302623

Thakor, N. V., & Zhu, Y. S. (1991). Applications of adaptive filtering to ECG analysis: Noise cancellation and arrhythmia detection. *IEEE Transactions on Biomedical Engineering*, 38(8), 785–794. https://doi.org/10.1109/10.83591

Trivedi, N. S., Ghouri, A. F., Shah, N. K., Lai, E., & Barker, S. J. (1997). Effects of motion, ambient light, and hypoperfusion on pulse oximeter function. *Journal of Clinical Anesthesia*, 9(3), 179–183. https://doi.org/10.1016/S0952-8180(97)00039-1

Vaessen, T., Steinhart, H., Batink, T., Klippel, A., van Nierop, M., Reininghaus, U., & Myin-Germeys, I. (2019). ACT in daily life in early psychosis: An ecological momentary intervention approach. *Psychosis*, 11(2), 93–104. https://doi.org/10.1080/17522439.2019.1578401

Vrijkotte, T. G., van Doornen, L. J. P., & de Geus, E. J. C. (2000). Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension*, 35(4), 880–886. https://doi.org/10.1161/01.hyp.35.4.880

Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P., & Gillinov, M. (2017). Accuracy of wrist-worn heart rate monitors. *JAMA Cardiology*, 2(1), 104–106. https://doi.org/10.1001/jamacardio.2016.3340

Wegner, F. K., Kochhäuser, S., Ellermann, C., Lange, P. S., Frommeyer, G., Leitz, P., Eckardt, L., & Dechering, D. G. (2020). Prospective blinded evaluation of the smartphone-based AliveCor Kardia ECG monitor for Atrial Fibrillation detection: The PEAK-AF study. *European Journal of Internal Medicine*, 73, 72–75. https://doi.org/10.1016/j.ejim.2019.11.018

Wichers, M., & Groot, P. C. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics*, 85(2), 114–116. https://doi.org/10.1159/000441458

Willemsen, G. H. M., de Geus, E. J. C., Klaver, C. H. A. M., van Doornen, L. J. P., & Carroll, D. (1996). Ambulatory monitoring of the impedance cardiogram. *Psychophysiology*, 33(2), 184–193. https://doi.org/10.1111/j.1469-8986.1996.tb02122.x

Winokur, E. S., Delano, M. K., & Sodini, C. G. (2013). A wearable cardiac monitor for long-term data acquisition and analysis. *IEEE Transactions on Biomedical Engineering*, 60(1), 189–192. https://doi.org/10.1109/TBME.2012.2217958

Zhang, D., Hu, X. U., Li, J., Liu, J., Baks-te Bulte, L., Wiersma, M., Malik, N.-U.-A., van Marion, D. M. S., Tolouee, M., Hoogstra-Berends, F., Lanters, E. A. H., van Roon, A. M., de Vries, A. A. F., Pijnappels, D. A., de Groot, N. M. S., Henning, R. H., & Brundel, B. J. M. (2019). DNA damage-induced PARP1 activation confers cardiomyocyte dysfunction through NAD + depletion in experimental atrial fibrillation. *Nature Communications*, 10(1), 1–17. https://doi.org/10.1038/s41467-019-09014-2

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

Supplementary Material