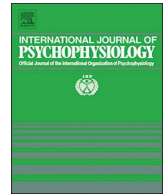




Contents lists available at ScienceDirect

## International Journal of Psychophysiology

journal homepage: [www.elsevier.com/locate/ijpsycho](http://www.elsevier.com/locate/ijpsycho)

## New methods to optimally detect episodes of non-metabolic heart rate variability reduction as an indicator of psychological stress in everyday life

Stephen B.R.E. Brown<sup>a,b,\*</sup>, Jos F. Brosschot<sup>a,c</sup>, Anke Versluis<sup>a,c</sup>, Julian F. Thayer<sup>d</sup>, & Bart Verkuil<sup>b,c</sup>

<sup>a</sup> Leiden University, Department of Health, Medical, and Neuropsychology, Wassenaarseweg 52, 2333AK Leiden, The Netherlands

<sup>b</sup> Leiden Institute for Brain and Cognition, Wassenaarseweg 52, 2333AK Leiden, The Netherlands

<sup>c</sup> Leiden University, Department of Clinical Psychology, Wassenaarseweg 52, 2333AK Leiden, The Netherlands

<sup>d</sup> Department of Psychology, The Ohio State University, Psychology Building, 1835 Neil Avenue, Columbus, 43210 OH, USA

## ARTICLE INFO

## Keywords:

Additional HRV  
Psychological stress  
Worry  
Cardiovascular disease

## ABSTRACT

Cardiovascular disease is the leading cause of death in the western world. Frequent or chronic reductions in heart rate variability (HRV) are a powerful predictor of cardiovascular disease. Psychological stress has been suggested to be an important factor in the development of reduced HRV. Recently, Verkuil et al. (2016) introduced a laboratory-based method to measure additional HRV reduction in everyday life, and reductions in HRV related to psychological stress. In the current paper, we discuss alternative methods to detect additional HRV reductions, in real life data sets without the necessity of laboratory-based calibration, and even in existing data sets. All of these methods use a subset of 24 h' worth of HRV and movement data to do so: either the first 10 min of every hour, the full 24 h, a combination of 10 min from three consecutive hours, or a classification of level of movement. We also present a method to visualize HRV and movement data to be able to detect episodes of reduced additional HRV optically. The method that used the full 24 h' worth of data detected the largest percentage of episodes of reduced additional HRV that actually match with self-reported stress levels, making this method the most promising.

## 1. Introduction

Cardiovascular disease is the leading cause of death worldwide (Alwan, 2011). Reductions in heart rate variability (HRV), that is, reductions in the variation of time between heart beats, have been demonstrated to be powerful predictors of the development of cardiovascular disease (Rosengren et al., 2004; Bosma et al., 1998; Kivimäki et al., 2006; Orth-Gomér et al., 2000; Matthews and Gump, 2002) and precedes the development of a number of risk factors, hypertension, diabetes, high cholesterol, and immunological markers of pathogenic states (Thayer and Lane, 2007; Thayer et al., 2010). In fact, low HRV has been shown to increase the risk of negative cardiovascular events by 32–45% (Hillebrand et al., 2013). Prolonged exposure to psychological stress may be one potential cause of reductions in HRV (McEwen, 2001), due to reduced activation of the vagal nerve, a phenomenon that has been referred to as vagal withdrawal (Porges, 1995). Given the societal costs—not to mention the personal and emotional costs—of cardiovascular disease, studying a powerful precursor like reduced heart rate variability is imperative.

We have recently published a technique to detect episodes of reduced heart rate variability in an ambulatory manner (Verkuil et al., 2016). We tested healthy student participants, who wore an ECG sensor for 24 h while going about their daily business. Every experimental session started with a short laboratory calibration period. During this calibration period, participants' HRV and movement was measured during four categories of simple physical activity that they might also perform during a regular day, that is, standing, cycling, climbing stairs, and lying down. This allowed us to create an HRV baseline during various types of physical activity. Some studies have excluded periods of high physical activity altogether (Pieper et al., 2007; Pieper et al., 2010; Brosschot et al., 2007), and/or have used accelerometer readings to identify epochs of non-movement (Sowder et al., 2010). However, it is crucial to take levels of physical activity into account when studying HRV, because these two variables are related (Rennie et al., 2003). By taking physical activity into account it becomes possible to estimate the amount of so called 'additional physiology', that is, the amount of physiological activity—in this case HRV—that is not purely due to physical activity. This enables researchers to more specifically study the

\* Corresponding author at: Leiden University, Wassenaarseweg 52, office 2A03, 2333AK Leiden, The Netherlands.  
E-mail address: [brownsbre@fsw.leidenuniv.nl](mailto:brownsbre@fsw.leidenuniv.nl) (S.B.R.E. Brown).

<http://dx.doi.org/10.1016/j.ijpsycho.2017.10.007>

Received 6 April 2017; Received in revised form 15 August 2017; Accepted 18 October 2017  
0167-8760/ © 2017 Elsevier B.V. All rights reserved.

relationship between psychological factors (i.e., stress, emotions) and physiological activity. Techniques to estimate ‘additional physiology’ are elaborate techniques and have therefore only occasionally been used in emotion-related ambulatory studies, using heart rate instead of HRV and have not used individualized algorithms (Myrtek and Brügger, 1996; Myrtek et al., 2005; Ebner-Priemer et al., 2007; Myrtek, 2004; Prill and Fahrenberg, 2007). However, it has also become much more feasible to continuously measure more ‘complicated’ parameters such as HRV. In Verkuil et al.’s (2016) study, the HRV and movement data acquired during the laboratory calibration period were used to compute a personalized algorithm that could be used to detect episodes of reduced additional HRV, that is, reductions in heart rate variability beyond what would be expected given the concurrent level of movement, and that are therefore assumed to be related to psychological stress (Myrtek et al., 2005). Stress was indeed measured in this study as participants were prompted every hour to complete some questionnaires on a mobile phone while they were wearing their ECG sensors. These questionnaires included questions on whether participants had experienced stress or worry during the previous hour, so as to be able to relate physiological stress markers to psychological stress markers.

Detecting periods of reduced additional HRV was a two-step process: we first formalized the relationship between HRV and movement by fitting an inverse regression model to the data from the calibration period, separately for every participant. The parameters from the resulting method were then used to detect episodes of reduced additional HRV (see Method for more detail).

The method presented by Verkuil et al. (2016) was successful in identifying episodes of reduced additional HRV, but it relies on a laboratory-based calibration period, which is not necessarily always practical in the context of field studies or experiments with outpatients, for example. Furthermore, although this method allowed us to detect episodes of reduced additional HRV well, the question arises whether there may be more efficient or effective methods to do so. More specifically, it may be possible to circumvent the laboratory calibration and use (part of) the real life data themselves, even allowing for the reanalysis of existing data sets. In the present paper, we will therefore present five alternative methods that can be used to detect episodes of reduced additional HRV and that do not require a laboratory-based calibration phase. To this end, we have reanalyzed the data set of Verkuil et al. (2016).

## 2. Method

### 2.1. Participants

We analyzed the same subset of 32 participants that Verkuil et al. (2016) analyzed (24 females; mean age 21 years).

### 2.2. Materials

All 24-h ECG data and movement data were collected with an ecgMove sensor (Movisens, GmbH, Karlsruhe, Germany). The data were processed offline in Movisens Data-Analyzer (see Verkuil et al., 2016, for details on preprocessing); the Data-Analyzer software uses automated algorithms to detect and remove artefacts in the data. Data were then analyzed in MATLAB™ (MathWorks, Natick, Massachusetts). Analysis scripts are available from the corresponding author.

### 2.3. Analyses

#### 2.3.1. General approach of the methods

In this paper, we will evaluate five alternative methods<sup>1</sup> to the

additional HRV detection method described in Verkuil et al. (2016). Four of these methods have a similar approach: for every individual, an inverse regression model was fitted to quantify the relationship between HRV, expressed as the root mean squares of successive differences, RMSSD (this measure was computed for 30-s intervals throughout the test session and was then averaged over samples spanning different amounts of time, depending on the method used, see below) and movement (expressed as acceleration in g, the averaged acceleration in three axes; this has been demonstrated to be a valid way to measure movement, especially walking, jogging, sitting, and lying down; Lugade et al., 2014). This regression model was fitted to a subset of all data points; for example, Verkuil et al. fitted such a model to the data acquired during the laboratory calibration period. The parameters from this regression model were then used, again separately for every participant, to predict real life HRV levels as a function of movement levels; whenever the actual HRV level was two standard errors below the predicted HRV level, and that difference persisted for 7.5 consecutive minutes, an episode of decreased additional HRV was identified (cf. Verkuil et al. for formulae). If multiple episodes were detected within a given hour, only the first episode was used in further analyses, following Verkuil et al. We will now briefly discuss the five alternative methods that will be presented in this paper.

#### 2.3.2. Method 1: first ten minutes of every hour

The first method used the first 10 min of every hour for which data was collected to compute inverse regression models. So, if data was available for 24 h, 24 separate inverse regression models were computed, and the parameters of these models were used to detect episodes of reduced additional HRV (see above) for every associated hour. The advantage of this method is that data from every hour were sampled, so the regression model that was used to identify periods of reduced additional HRV within a given hour was also based on data from that hour. Of course, that is also the disadvantage of this method: it could be considered “double dipping”, because (part of) the data from hour  $n$  were used to detect additional HRV episodes in hour  $n$ , including in the first 10 min of data, although this seems to be of relatively minor concern. After all, if a given method detected an episode of reduced additional HRV in these first 10 min of an hour, there appears to be no empirical reason to doubt that a person actually experienced stress or worry during those 10 min.

#### 2.3.3. Method 2: full data set

The second method used the full period for which data were collected to compute the inverse regression model. So, if data were

(footnote continued)

not included them in our analyses. However, for completeness’ sake, we will briefly present these methods here. The first of these rejected methods used the first 10 min of all of the data available for a given participant (e.g. 24 h, in the case of the study of Verkuil et al., 2016) to compute an inverse regression model. This method has the advantage of using only 10 min’ worth of data, and therefore it only requires computing one regression model; the disadvantage is that the model is unlikely to be very reliable, because so few—and relatively random—data points are used to construct it.

The second rejected method computed separate inverse regression models for day and night time, on the basis of the first 10 min’ worth of data for 8 AM and 8 PM, respectively. This method takes differing levels of movement during day and night time into account (for example when working or sitting at home). However, every participant has a different circadian rhythm, which complicates estimating when his or her day and night “start”, especially when re-analyzing data from a previous study that did not control for, or record, such variables. Furthermore, there is little control over whether a participant is more or less active during the day relative to the night.

The final rejected method took a “sliding modelling” approach: inverse regression models were computed on the basis of the first ten minutes of every hour, and episodes of reduced additional HRV were then detected on the basis of the model of the previous hour; so the model of hour  $n$  was used to detect additional HRV episodes in hour  $n + 1$ . This method avoids double dipping into data by using a model computed on the first 10 min of an hour to then detect episodes of reduced additional HRV in that same hour. However, HRV and activity levels of the first 10 min of hour  $n$  do not necessarily match those of hour  $n + 1$  better than those levels in the first 10 min of hour  $n + 1$  itself.

<sup>1</sup> We explored three more methods, but as they did not yield reliable results, we have

available for 24 h, a single inverse regression model was computed, based on the entire data set, and its parameters was then used to detect episodes of reduced additional HRV for every hour. The advantage of this method is that the inverse regression models are based on a large number of data points, which makes these models more robust than models based on just 10 min' worth of data. However, the levels of movement and HRV are averaged over the full period and not modelled separately for every hour.

#### 2.3.4. Method 3: first 10 minutes of three consecutive hours

The third method compensated for fluctuations in movement or HRV by using the first 10 min of three consecutive hours to compute the inverse regression model; therefore, every regression model was based on 30 min' worth of data. For the first and last hours of a data set, the regression models were based on the first ten minutes of the first and second and penultimate and final hours, respectively. So, if there were data available for 24 h, 24 separate inverse regression models were computed. This method has the advantage that episodes of reduced additional HRV were identified on the basis of models that were computed from data that take changes in HRV or movement levels in three consecutive hours into account *and* that these models are based on the average of the first ten minutes of three consecutive hours, leading to more reliable model estimations as compared to the first method, that based models on only ten minutes' worth of data instead of 30 min. The advantage of this method, that it uses HRV and movement data from three consecutive hours, is also its disadvantage: although more data is used to compute inverse regression models, movement and HRV levels are autocorrelated over time, which reduces the amount of variation in these levels over three consecutive hours.

#### 2.3.5. Method 4: movement level bins

The fourth method capitalized on utilizing the natural variation in a participant's level of movement throughout the day. To this end, we binned movement data on the basis of quartiles, defined per participant, thereby creating four bins, which classified movement ranging from relatively very low to relatively very high levels of movement. Each bin consisted of 5 consecutive minutes' worth of data, so a participant's level of movement within these 5 min had to fall between two quartiles to be assigned to a specific bin. The choice of five minutes was motivated by a desire to identify all bins for as many participants as possible; cf. the downside to this method, below. We restricted the analysis to the first-occurring 5 min lasting bin of every movement class (so if a participant had four clusters of quartile-1 movement, only the first cluster was used, to keep the amount of data used per quartile equal). All available quartile data (i.e. a maximum of 20 min' worth of data) were used to compute inverse regression models. This method has the clear advantage of taking variations in an individual's movement into account, by representing every movement class in the reverse regression models, which is unlikely to happen when one only uses the data from the first 10 min of every hour. This method also requires the calculation of just a single inverse regression model, as opposed to models for every hour. However, a downside is that in one participant, not all four bins could be identified. Obviously, this participant had movement data that fell between two quartiles, but s/he did not have 5 consecutive minutes' worth of such data. Still, even this participant had 15 min worth' of data that could be used to compute a reverse regression model. This method resembles the laboratory-based method of Verkuil et al. (2016) that used four predetermined physical activity categories: lying down, standing, cycling and climbing stairs, but it is data-driven and does not depend on a laboratory-based calibration session.

#### 2.3.6. Method 5: visualization

We also developed a method that is not dependent on the computation of inverse regression models. This method depends on visualization and subsequent visual inspection of the data. Given the different

measuring scales of HRV (RMSSD is typically measured in tens of milliseconds) and movement acceleration ( $g$  is expressed in considerably smaller values: in Verkuil et al.'s data set,  $g$  ranged from 0.003 to 1.0946), we first multiplied  $g$  values by 1000 to make them numerically more similar to the values of heart rate variability. Both heart rate variability and movement values were then smoothed by replacing every observation by the average of its ten nearest neighbors. So, the observation of value  $n$  was replaced by the average of values  $n-5$  through  $n+5$ . These smoothed values were then plotted for a given participant so that one might then attempt to visually identify moments when HRV decreases while movement does not increase. Although this method is simple and does not depend on the computation of inverse regression models, judging these types of data visually is quite complicated, especially to untrained eyes, and although the results may give a quick indication of when periods of reduced additional HRV may have occurred, we believe the other methods presented here give more accurate results. Furthermore, because no inverse regression model is computed, HRV values cannot be predicted as was done by Verkuil et al. (2016), which complicates applying this technique to future applications like monitoring and detecting periods of reduced additional HRV online. It is important to emphasize that this method merely plots physiological data, so psychological data on actual worrying and stress are not represented in these figures.

#### 2.3.7. Episodes detected during sleep

During the analyses, it became clear that these alternative methods occasionally detected episodes of decreased additional HRV while a participant was sleeping. This can happen when the actual HRV was well below the expected HRV level, but obviously, when one is sleeping, such a phenomenon is not caused by conscious psychological stress. Still, for methods in which the algorithm was calculated during waking periods, these additional HRV decreases are meaningful. Low sleeping HRV has previously been found to be associated with preceding stress (e.g. Brosschot et al., 2007; Hall et al., 2004), and has been hypothesized to reflect an unconscious form of stress-related cognition (Brosschot, 2010; Brosschot et al., 2010). However, since nocturnal HRV is not the primary focus of this study, we have therefore manually removed such episodes for further analyses.

#### 2.3.8. Method comparisons

The goal of the current paper is to compare the efficacy of these five methods in detecting episodes of reduced additional HRV. We have therefore compared these methods in three ways. The first three comparisons mainly serve to demonstrate the variation in number of episodes of reduced additional HRV that each method identified. Firstly, we compared the number of episodes of reduced additional HRV that each method identified to the number of episodes that was detected by the laboratory calibration method as described in Verkuil et al. (2016), to ascertain whether the different methods identified comparable numbers of episodes. This was done to gain insight in the ability of each method to detect such episodes. Secondly, we have computed the overlap in detected episodes between all the methods and Verkuil et al.'s laboratory calibration method to ascertain to what extent the alternative methods identified the same episodes of reduced additional HRV as Verkuil et al.'s laboratory-based "gold standard". To this end, we have identified, for every participant and for every method, what percentage of episodes—based on their exact timing—were identified both by that method and by Verkuil et al.'s laboratory calibration method. In Verkuil et al.' study, participants were prompted once an hour to report whether they had been stressed or worried during the past hour. Crucially, we have compared the onset of participants' self-reported episodes of stress and worry to the episodes of reduced additional HRV that were identified by each of the methods. Whereas the other analyses mainly provide insight into the distribution of the number of physiological episodes of reduced additional HRV that were identified, this key comparison shows to what extent these

**Table 1**

Mean number of additional HRV episodes (SD) identified for a time series of 24 h. Note that the time period used for calibration is also included in the detection periods. Episodes were detected in 31 of the 32 analyzed participants.

| First 10 min of every hour | Full data set (24 h) | Combined 30 min from three hours | Activity class | Verkuil et al. (2016) |
|----------------------------|----------------------|----------------------------------|----------------|-----------------------|
| 9.2 (2.8)                  | 13.7 (3.1)           | 10.5 (3.1)                       | 6.4 (4.0)      | 6.3 (3.6)             |

physiological events actually match up with psychological events of stress and worry. To this end, we have computed, for every method and for every available hour worth of data, what percentage of participants had matching self-reported stress and worry episodes and episodes of reduced additional HRV. Furthermore, we computed the average percentage of such matches within a method. These four comparisons are expected to corroborate the quality of each of the alternative methods presented here and to ascertain which of these methods appears to be the most suitable alternative to the laboratory calibration method that was presented by Verkuil et al.

### 3. Results

We first compared the differences in number of identified reduced additional HRV episodes between the various identification methods; we compared these numbers of episodes to the laboratory calibration method from Verkuil et al. (2016). Some of the alternative methods identified more episodes of reduced additional HRV than others: as can be seen in Table 1, every method identified at least 6 episodes of reduced additional HRV.

As expected on the basis of these observations, a repeated measures ANOVA with estimation method as within-subjects factor revealed a significant main effect of method,  $F(4, 120) = 33.4$ ,  $p < 0.0005$ ,  $\eta_p^2 = 0.53$  (Greenhouse-Geisser corrected; unadjusted degrees of freedom reported), suggesting that the method that used the full 24 h for calibration identified the largest mean number of additional HRV episodes (13.7), while the lab calibration method by Verkuil et al. identified the lowest mean number of additional HRV episodes (6.3). Pairwise comparisons suggest a clear pattern of significant differences between the methods: all methods identified differing numbers of additional HRV episodes (all  $ps < 0.01$ ), except for the lab calibration method (6.3) and the method that was based on classifying activity level (6.4),  $t_{30} = 0.14$ ,  $p = 0.97$ .

#### 3.1. Amount of overlap in identified episodes between methods

Another way to evaluate the suitability of our alternative methods would be to compute the level of overlap in the episodes of reduced additional HRV between these methods and the lab calibration method presented by Verkuil et al. (2016). In other words: to what extent did these alternative methods identify the same episodes as the lab calibration method did? Based on level of overlap with the lab calibration method, the method that used the first 10 min of every hour seemed to be the least suitable with an overlap of just 5.2%. The method based on activity type bins had the highest level of overlap with 35.3%; this may suggest that classifying activity types and using those bins to base regression models on may be a relatively successful approach. This also

**Table 2**

First row: mean percentage of participants with at least one match between self-reported episodes of stress and worry and method-identified episodes of reduced additional HRV. Second row: mean percentage (SD) matches between self-reported episodes of stress and worry and method-identified episodes of reduced additional HRV; note that these percentages are calculated only for the participants that had at least one match – reported in the first row.

|   | First 10 min of every hour | Full data set (24 h) | Combined 30 min from three hours | Activity class | Verkuil et al. (2016) |
|---|----------------------------|----------------------|----------------------------------|----------------|-----------------------|
| Percentage of participants with match   | 51.6%                      | 74.2%                | 61.3%                            | 48.4%          | 38.7%                 |
| Percentage of matches within an episode | 59.0% (22.3%)              | 77.1% (21.2%)        | 67.0% (25.9%)                    | 60.0% (32.6%)  | 51.2% (33.8%)         |

suggests that this method appears to be promising and is eligible for further refinement in future work. The method that combined 30 min from three consecutive hours had a fair amount of overlap with the lab calibration method: 16.2% and the method that used the full data set had 27.1% overlap. This overlap analysis therefore suggests that the methods that used the various activity classes and the full data set found results that were most comparable to the results found by the lab calibration method of Verkuil et al. (2016).

#### 3.2. Using self-reported stress as a criterion

Crucially, we have computed matches between participants' self-reported levels of psychological stress and worrying (as collected by Verkuil et al., 2016) and the episodes of reduced additional HRV. On average, participants reported 3.4 episodes of worrying and/or stress ( $SD = 2.0$ ).

Firstly, we computed the percentage of participants who had at least one such match in their data. These percentages were computed separately for every method. As presented in the first row of Table 2, some methods are characterized by considerably better percentages of matches than others. For example, using the method that utilized the full data set (in this case, 24 h' worth of data) led to 74.2% of matches between participants' self-reported stress and worry episodes and the episodes of reduced additional HRV identified by this method. On the other hand, the lab calibration method by Verkuil et al. was characterized by a match of 38.7% in participants. This would suggest that the method that used the full 24 h' worth of data was better able to detect actual episodes of reduced additional HRV than the lab calibration method used by Verkuil et al., which we used to consider to be a gold standard. The method that combined 30 min of three consecutive hours appears to be a good “runner-up”, with an overall match of 61.3%.

These percentages merely show how many participants had at least one match between self-reported stress or worry episodes and method-detected episodes of reduced additional HRV. Therefore, secondly, it is crucial to ascertain the percentage of matches between self-reported stress or worry episodes and method-detected episodes of reduced additional HRV *within* those participants who had at least one such match. The former percentages give a first indication of the suitability of a given method, as this measure demonstrates, for every method, in how many participants it was able to detect episodes of reduced additional HRV that match self-reported episodes of stress and worry; the latter percentages then show how well any method was able, on average, to detect physiological episodes of reduced additional HRV that actually match self-reported episodes of stress or worry, given that the participant had at least one match. The second row of Table 2 lists these percentages. Again, the method that used the full 24 h' worth of data seems to be most promising, with an average match between self-reported stress and worry and method-identified episodes of reduced additional HRV of 77.1%. The method that used a combination of 30 min of three consecutive hours was also associated with an acceptable percentage of matches, at 67.0%. Taken together, these results suggest that using the full data set to detect episodes of reduced additional HRV appears to be the most promising, as this method is characterized by the best overall match between self-reported stress and worry and additional HRV physiology in the largest number of



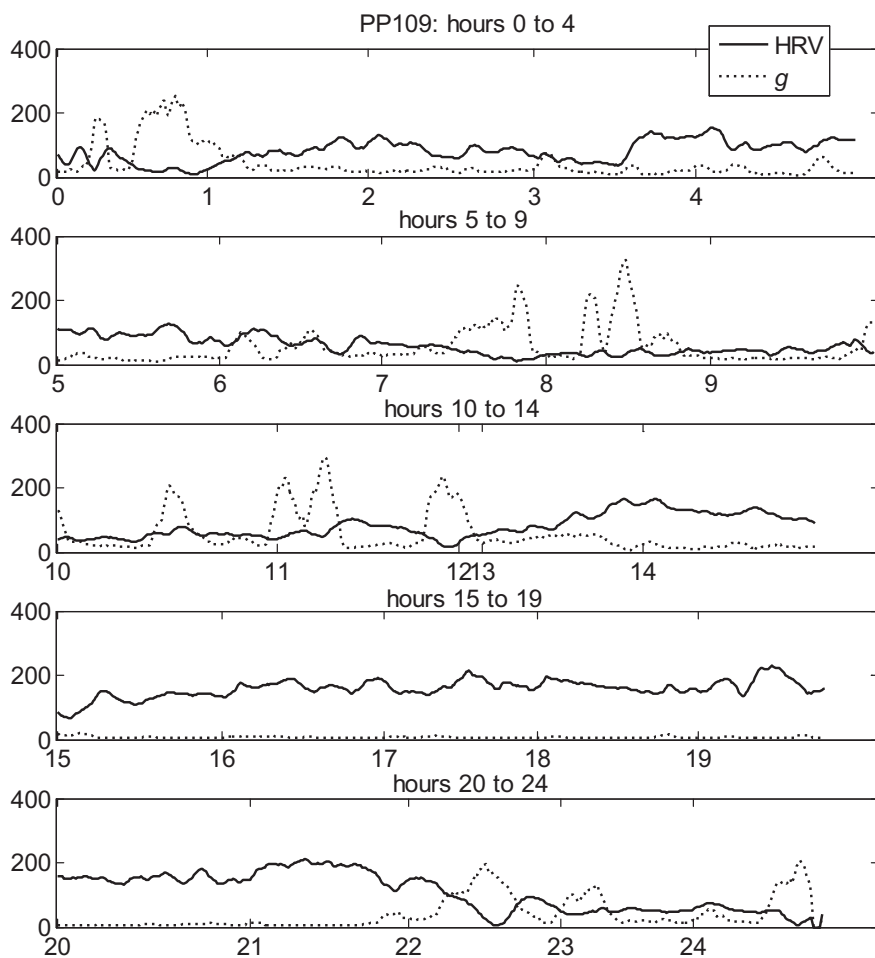


Fig. 1. Visualization method: a plot of HRV and movement level (acceleration in  $g$ ) for one representative participant. Every panel consists five hours' worth of data. Data have been smoothed (see Method).

participants. Interestingly, all of our methods identified more episodes of reduced additional HRV than participants reported stress or worry episodes. On average, the methods identified more episodes than participants did themselves in 87.5% of the cases.

### 3.3. Visualization method

Our visualization method (see Method) allowed us to graph HRV and movement levels for visual inspection to detect episodes of reduced additional HRV without having to compute any inverse regression models. As can be seen in Fig. 1(a visualization of a representative participant), this is insightful (note, for example, the missing data in hour 12) and it clearly characterizes the relationship between HRV and level of movement. For example, in hour 1, it is clear that as the movement level goes up (dotted line), the level of HRV goes down (solid line); the reverse happens in hours 20 to 21. However, it is not that easy to visually identify periods where HRV decreases more than would be expected on the current level of movement (although maybe that does occur halfway through hour 11). We therefore believe that this method is suited to easily visualize and inspect the data, but not to reliably identify periods of reduced additional HRV.

## 4. Discussion

In this paper, we have evaluated a number of alternatives to the method Verkuil et al. (2016) used to detect episodes of reduced additional HRV and associated psychological stress and worry levels. The four alternative methods we present here have all identified different numbers of such episodes in the two data sets we have reanalyzed. Crucially, the method that used the full data set to create an algorithm

to detect episodes of reduced additional HRV correlated best with self-reported episodes of stress and worry. We consider this to be a powerful argument in favor of the quality of this method and we therefore believe it to be a suitable and compelling alternative to the technique used by Verkuil et al.: interestingly, it performed even better than the method that was used by Verkuil et al., as it was characterized by a better match between self-reported stress and worry episodes and method-identified episodes of reduced additional HRV.

Clearly, the Verkuil et al. method was developed in a first attempt to identify episodes of reduced additional HRV in an ambulatory setting, and it has guided the current work. For example, requiring actual HRV to be two standard errors below the predicted HRV for 7.5 min is a carefully considered yet seemingly arbitrary decision. The original reasoning of Verkuil et al. was that two standard errors would provide a pronounced enough difference between the two variables, and that 7.5 min is enough time to qualify an episode as worrying or stress. Clearly, one could use all sorts of other parameters or definitions: three standard errors, 15 min, etc., but there is simply no objective manner to determine, for example, what duration of time spent worrying qualifies as “an episode of worrying”, nor is it possible to objectively determine how strongly predicted and actual HRV should differ from each other during that time. Similar arguments could be raised regarding the use of the first 10 min of an hour as input for our methods: clearly, one could also choose to study 10 min from the middle of every hour, or the last 10 min of every hour, etc. Obviously, one can vary these kinds of parameters almost unlimitedly. Perhaps choosing the first 10 min of an hour was not optimal, as many (stressful) appointments and such take place at the start of an hour – but then again, if one is to face a stressful meeting, would one not worry about it during, say, the last ten minutes of the preceding hour?

The visual method we present here may be useful to aid in data interpretation, but it appears to be complicated to use this method to quantify episodes of reduced additional HRV effectively. We would therefore only recommend to use this method to visualize data sets, but not to consider it as an alternative to the other methods presented in this paper.

One question that arises is why there was such a difference in number of episodes that were detected by the various methods presented here. Clearly, some methods identify more episodes than others; this may simply reflect that some subsets of data are less suitable to create an algorithm to detect episodes of reduced additional HRV. For example, the algorithm based on the activity class bins may simply have been less sensitive in detecting such episodes than the method that used all available data points. Not only did the activity class method use less data points to compute a regression model, but perhaps utilizing different classes of activity simply did not provide enough information to adequately detect episodes of reduced additional HRV. In that context, it is interesting that the lab calibration method by Verkuil et al. (2016), which was also based on including several types of movement, identified similar numbers of episodes as the activity class method. Both of these methods appeared to underestimate the actual number of episodes of reduced additional HRV present in the data. Although there was some variation in the inverse regression parameters obtained for the various methods, this simply reflects that it was more difficult to obtain well-fitting models for a few participants.

Without a doubt, one question is fundamental: why was none of our methods able to detect a 100% match between self-reported stress and worry episodes and method-identified episodes of reduced additional HRV? Even our most successful method only identifies a 77.1% average match between self-reported stress and worry episodes and method-identified episodes of reduced additional HRV. There are two possible explanations for this: of course, there may simply be noise in the data, which precludes even our most promising method from detecting every single stress episode reported by every single participant. Alternatively, the methods may be correct, but participants may simply have under-reported stress and worry, because, importantly, our indicators of psychological stress and worry were self-reported on mobile phones that provided random hourly prompts, and it is possible that participants occasionally failed to report stress or worry episodes, for example because of embarrassment, forgetfulness, or even occasional technical glitches. However, assuming that there were no technical glitches, it is also possible that the method-identified episodes of reduced additional HRV that were not matched by stress and worry reports are actually—at least partly—due to unconscious stress; a notion that is supported by previous work from our lab (Brosschot et al., 2010; Brosschot, 2010).

Clearly, none of the alternative methods we have described here are perfect, but, at the very least, a success rate of almost 80% in identifying episodes of reduced additional HRV and concomitant psychological worry or stress seems to offer a considerable improvement over the success rate of the Verkuil et al. lab calibration method, which identified just a little over 50%. In fact, in 9 of 23 participants, the method that used the full data set had a success rate of 100%.

So, if even our best method is not perfect yet, how could we perfect it in the future? One option might be to use the data for day  $n$  to create a reverse regression model to detect episodes of reduced additional HRV in day  $n + 1$  (or, in any case, use the data from one day to detect in a future day). A similar approach was explored by Johnston (1996); in that study, an underdetection of events was shown, suggesting that further fine-tuning of this technique would be interesting. This has the advantage that “double dipping” in the data is eliminated, because the data that are used to detect episodes of reduced additional HRV in, are not included in the array of data that is used to create inverse regression models. Furthermore, it may be necessary to introduce completely different physiological factors and to incorporate them into regression models in order to adequately identify episodes of reduced additional HRV, like oxygen saturation, a variable that is obviously related to

exercise and that has been studied in the context of heart rate variability and stress too (e.g. Carroll et al., 2009). However, this also complicates data acquisition. The advantage of the current methods is that HRV and movement levels are the only variables required to create the inverse regression models required to detect episodes of reduced additional HRV, and these variables are both measured by the same chest-mounted sensor.

Although more work clearly remains to be done, we have presented several alternative methods to the lab calibration method introduced by Verkuil et al. (2016), that allow one to detect episodes of reduced additional HRV in real life data sets, including existing data sets, provided that data on HRV and movement levels are available. Being able to determine which reductions in HRV are due to physical activity and which are due to psychological factors could also provide more detailed insights into which type of episode is associated with reduced cardiovascular health, including whether and to what extent exertion-induced low HRV actually contributed to CV risk. Excitingly, such a method may even be used to assess stress experiences *online* conditional upon detecting additional HRV reductions in real life. Our most successful alternative method simply uses all available data, which eliminates the need for a lab calibration session. Given the rapid advancement of ambulatory technology, and the need to control for confounding factors such as movement, we foresee that methods such as those proposed here will become widely used. Importantly, such methods may aid in the timely detection of powerful indicators of cardiovascular disease, which, in turn, will hopefully lead to early diagnosis of cardiovascular disease and improved cardiovascular health.

## Acknowledgements

This project was supported by a grant from the Leiden University Fund/Den Dulk-Moermans Fonds (6503/21-6-16) to SB/JB and by a grant from ZON-MW (Netherlands Organization for Health 20 Research and Development; TOP Grant 40-00812-98-11029) to JB; BV was supported by a research grant from the Netherlands Organization for Scientific Research (NWO Veni Grant 451-14-013).

## References

- Alwan, A., 2011. Foreword. In: Mendis, S., Puska, P., Norrving, B. (Eds.), *Global Atlas On Cardiovascular Disease Prevention And Control*. WHO Press, Geneva, Switzerland, pp. VI.
- Bosma, H., Peter, R., Siegrist, J., Marmot, M., et al., 1998. Two alternative job stress models and the risk of coronary heart disease. *Am. J. Public Health* 88, 68–74.
- Brosschot, J.F., 2010. Markers of chronic stress: prolonged physiological activation and (un)conscious perseverative cognition. *Neurosci. Biobehav. Rev.* 35, 46–50. <http://dx.doi.org/10.1016/j.neubiorev.2010.01.004>.
- Brosschot, J.F., van Dijk, E., Thayer, J.F., 2007. Daily worry is related to low heart rate variability during waking and the subsequent nocturnal sleep period. *Int. J. Psychophysiol.* 63, 39–47. <http://dx.doi.org/10.1016/j.ijpsycho.2006.07.016>.
- Brosschot, J.F., Verkuil, B., Thayer, J.F., 2010. Conscious and unconscious perseverative cognition: is a large part of prolonged physiological activity due to unconscious stress? *J. Psychosom. Res.* 69, 407–416. <http://dx.doi.org/10.1016/j.jpsychores.2010.02.002>.
- Carroll, D., Phillips, A.C., Balanos, G.M., 2009. Metabolically exaggerated cardiac reactions to acute psychological stress revisited. *Psychophysiology* 46, 270–275. <http://dx.doi.org/10.1111/j.1469-8986.2008.00762.x>.
- Ebner-Priemer, U.W., Welch, S.S., Grossman, P., Reisch, T., Linehan, M.M., Bohus, M., 2007. Psychophysiological ambulatory assessment of affective dysregulation in borderline personality disorder. *Psychiatry Res.* 150, 265–275. <http://dx.doi.org/10.1016/j.psychres.2006.04.014>.
- Hall, M., Vasko, R., Buysse, R., Ombao, H., Chen, Q., Cashmere, J.D., Kupfer, D., Thayer, J.F., 2004. Acute stress affects heart rate variability during sleep. *Psychosom. Med.* 66, 56–62.
- Hillebrand, S., Gast, K.B., de Mutsert, R., Swenne, C.A., Jukema, J.W., Middeldorp, S., Rosendaal, F.R., Dekkers, O.M., 2013. Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: meta-analysis and dose-response meta-regression. *Europace* 15, 742–749. <http://dx.doi.org/10.1093/europace/eus341>.
- Johnston, D.W., 1996. Determining psychologically significant events in real life. In: Fahnenberg, J., Myrtek, M. (Eds.), *Ambulatory Assessment: Computer-Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies*. Hogrefe & Huber, Kirkland, WA, pp. 133–140.
- Kivimäki, M., Virtanen, M., Elovainio, M., Kouvonen, A., Väänänen, A., Vahtera, J., 2006.

- Work stress in the etiology of coronary heart disease—a meta-analysis. *Scand. J. Work Environ. Health* 32, 431–442.
- Lugade, V., Fortune, E., Morrow, M., Kaufman, K., 2014. Validity of using tri-axial accelerometers to measure human movement - part I: posture and movement detection. *Med. Eng. Phys.* 36, 659–669.
- Matthews, K.A., Gump, B.B., 2002. Chronic work stress and marital dissolution increase risk of posttrial mortality in men from the multiple risk factor intervention trial. *Arch. Intern. Med.* 62, 309–315.
- McEwen, B.S., 2001. From molecules to mind. Stress, individual differences, and the social environment. *Ann. N. Y. Acad. Sci.* 935, 42–49. <http://dx.doi.org/10.1111/j.1749-6632.2001.tb03469.x>.
- Myrtek, M., 2004. *Heart and Emotion: Ambulatory Monitoring Studies in Everyday Life*. Hogrefe Publishing, Ashland, OH.
- Myrtek, M., Brügner, G., 1996. Perception of emotions in everyday life: studies with patients and normals. *Biol. Psychol.* 42, 147–165.
- Myrtek, M., Aschenbrenner, F., Brügner, G., 2005. Emotions in everyday life: an ambulatory monitoring study with female students. *Biol. Psychol.* 68, 237–255. <http://dx.doi.org/10.1016/j.biopsycho.2004.06.001>.
- Orth-Gomér, K., Wamala, S.P., Horsten, M., Schenck-Gustafsson, K., Schneiderman, N., Mittleman, M.A., 2000. Marital stress worsens prognosis in women with coronary heart disease. *JAMA* 284, 3008–3014.
- Pieper, S., Brosschot, J.F., van der Leeden, R., Thayer, J.F., 2007. Cardiac effects of momentary assessed worry episodes and stressful events. *Psychosom. Med.* 69, 901–909.
- Pieper, S., Brosschot, J.F., van der Leeden, R., Thayer, J., 2010. Prolonged cardiac effects of momentary assessed stressful events and worry episodes. *Psychosom. Med.* 72, 570–577.
- Porges, S.W., 1995. Orienting in a defensive world: mammalian modifications of our evolutionary heritage. A polyvagal theory. *Psychophysiology* 32, 301–318.
- Prill, T., Fahrenberg, J., 2007. New methods in ambulatory blood pressure monitoring: interactive monitoring and detection of posture and movement patterns. *Behav. Res. Methods* 39, 390–398.
- Rennie, K.L., Hemingway, H., Kumari, M., Brunner, E., Malik, M., Marmot, M., 2003. Effects of moderate and vigorous physical activity on heart rate variability in a British study of civil servants. *Am. J. Epidemiol.* 158, 135–143. <http://dx.doi.org/10.1093/aje/kwg120>.
- Rosengren, A., Hawken, S., Ounpuu, S., Sliwa, K., Zubaid, M., Almahmeed, W.A., Blackett, K.N., Sitthi-amorn, C., Sato, H., Yusuf, S., investigators, I.N.T.E.R.H.E.A.R.T., 2004. Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries (the INTERHEART study): case-control study. *Lancet* 364, 953–962. [http://dx.doi.org/10.1016/S0140-6736\(04\)17019-0](http://dx.doi.org/10.1016/S0140-6736(04)17019-0).
- Sowder, E., Gevirtz, R.N., Shapiro, W., 2010. Restoration of vagal tone: a possible mechanism for functional abdominal pain. *Appl. Psychophysiol. Biofeedback* 35, 199–206.
- Thayer, J.F., Lane, R.D., 2007. The role of vagal function in the risk for cardiovascular disease and mortality. *Biol. Psychol.* 74, 224–242. <http://dx.doi.org/10.1016/j.biopsycho.2005.11.013>.
- Thayer, J.F., Yamamoto, S.S., Brosschot, J.F., 2010. The relationship of heart rate variability, risk factors and cardiovascular disease. *Int. J. Cardiol.* 141, 122–131. <http://dx.doi.org/10.1016/j.ijcard.2009.09.543>.
- Verkuil, B., Brosschot, J.F., Tollenaar, M.S., Lane, R.D., Thayer, J.F., 2016. Prolonged non-metabolic heart rate variability reduction as a physiological marker of psychological stress in daily life. *Ann. Behav. Med.* 50, 704–714. <http://dx.doi.org/10.1007/S12160-016-9795-7>.