Exploring the Dynamics of Emotions: Combining the Experience Sampling Method With Continuous Physiological Data

Isabella Huntjens s2609282 (i.c.w.huntjens@student.utwente.nl)

Department of Psychology, The University of Twente

M12 BSc Thesis PSY

Dr. S. M. van den Berg & Dr. P. M. ten Klooster

January 26th, 2024

APA: 7th edition

Abstract

Background. The combination of continuous parallel (objective) data using wearables and the subjective experience of emotions using experience sampling methods (ESM) could offer a powerful tool for providing insights as to how emotions emerge in humans. Currently, a general guideline as how to analyse the combination of these data sources is lacking, causing heterogeneity in the methods used and making it difficult to do cross-study comparisons.

Aims. This study aimed to further investigate the relationship between emotions as they appear according to Russell's theory (1980) and their cardiovascular response by drawing inspiration from Ketonen et al. (2023) and adding on their methodology by comparing several linear mixed models with each other using Akaike's information criterion (AIC).

Sample. The study included five students from the University of Twente.

Methods. A three-day ESM study was conducted, triggering seven semi-random questionnaires a day, while also collecting continuous data with the EmbracePlus bracelet on pulse rate (PR) and the metabolic equivalent of a task (MET) to control for physical activity (PA). For the data analysis, we compared models with unweighted and Gaussian weighted means of PR and MET five minutes before, around, and after the ESM prompt using Akaike's Information Criterion.

Results. Akaike's Information Criterion showed that, in general, the models with the unweighted means for PR and MET five minutes after each ESM prompt performed best compared to the other candidate models. Except for boredom, for this emotion the model with the Gaussian weighted means for PR and MET five minutes around the ESM prompt performed best.

Conclusions. Generally, the models with the unweighted means five minutes after the ESM prompt performed best compared to the other models. This finding can be used to develop a general guideline as to how to analyse experience sampling data in combination with continuous physiological data, which is now lacking. Further research is needed to fully develop such guidelines, since this is just a small component. Such guidelines will facilitate cross study comparison since methodology can be standardized. Limitations, strengths, and implications for future research are discussed.

1 Introduction

Humans have countless experiences in their daily life, we feel, think, experience, and interact. All processes that captivated the interest of psychologists the past decades. The experience sampling method (ESM) is a more recent powerful research tool to investigate the subjective component of these daily life experiences (Maes et al., 2013; Myin-Germeys. I..,& Kuppens. P., 2022). Typically, ESM involves asking participants in the study a set of questions several times a day for a period of several days or weeks. With the advancing technology of present times, these questions are often triggered at random or semi-random points on smartphones or other electronic devices (Vaessen et al., 2021). This offers the opportunity of investigating (fluctuations in) experiences, feelings, thoughts, and activities of participants outside laboratory walls, increasing the ecological validity of the results (Maes et al., 2013). Rather than being dependent on retrospective recall of participants, ESM investigates experiences as they happen in the real world. Since there is ample evidence that humans have limited ability to accurately recall their behaviour or experiences, retrospective based data can be biased (Bradburn et al., 1987). ESM seems to solve this issue since participants are asked to report on the moment they receive the notification, thus there is no or minimal re-call necessary (Christensen et al., 2003).

In addition to the subjective experience, emotions also cause more objective, physiological effects (Purves et al., 2001). One method for detecting physiological effects of emotional states is the cardiovascular response (e.g.. Li et al., 2009; Ketonen et al., 2023; Vaessen et al., 2021). Cardiovascular response can be measured with, among others, heart rate (HR), measuring the beats per minute (Schaffer & Ginsberg., 2017). In addition to the collection of cardiovascular data, it is necessary to control for physical activity (PA) when analysing this data source, since HR will also increase during PA. To do this, the metabolic equivalent of a task (MET) can be collected, this is an indication of PA compared to a resting state of the body (Forsum et al., 2018). It is defined as the amount of oxygen used per kg bodyweight per minute (Jetté et al., 1990). These physiological measures can be increasingly easily and continuously collected with wearable devices like fitness trackers and smartwatches, ensuring minimally invasive data collection, minimally burdening the participants while also providing continual data (De Calheiros Velozo et al., 2022b). These wearables have significantly facilitated the continuous data collection in real world settings, allowing data collection on situations as they are happening in the real world (De Calheiros Velozo et al., 2022b).

Both these active and passive data sources only measure one component of an emotion or feeling, either the subjective cognitive evaluation or the objective physiological manifestation. The combination of these data sources could offer a powerful measurement tool for providing valuable insights on human emotions (De Calheiros Velozo et al., 2022a). Therefore, combining these two data sources could offer an answer to a wider array of questions, hence the increasing interest in this combination in psychological research. However, combining and analysing these two data sources is still a novel and challenging practice. De Calheiros Velozo et al. (2022a) provide a checklist with considerations for study designs to combine ESM data and continual data such as HR and MET values, however, a guideline as to how to analyse and combine this data is lacking. Current guidelines typically emphasize one of these data sources, rather than providing directions on the combination of these active and passive data sources (De Calheiros Velozo.. 2022b). Moreover, the validation of results is challenged because of the wide variety of measures and the heterogeneity of these approaches (Vaessen et al., 2021).

The study by Ketonen et al. (2023) offers an attempt to further our understanding of the combination of these data sources. They investigated if there was an association between the 127 participants' self-reported emotions (ESM data) and their cardiovascular response at specific moments in time. Their study was part of a larger ESM study which send the participants six signals per day for a period of 10 days. Ketonen et al. (2023) continuous physiological data for 72 hours of those 10 days. Ketonen et al. (2023) based the selection of their measured emotions on the multidimensional model of emotions by Russell (1980). This model has two central dimensions: valence and arousal, which together form the arousal-valence matrix. Valence relates to whether an emotion is positive or negative. Arousal relates to whether an emotion is activating or deactivating. With this matrix, four categories of emotions emerge: high arousal and high valence (excitement), low arousal and low valence (boredom), low arousal and high valence (calmness), and lastly, high arousal and low valence (anxiety). However, rather than using the matrix for participants to determine their emotional state by pinpointing it on the matrix, they used it as inspiration for selecting which discrete emotions to investigate by choosing one emotion in each side of the matrix, Ketonen et al. (2023) provided the participants with a Likert scale to measure the selected emotions (e.g.. "To what extent do you currently feel excited, bored, calm or anxious?").

In addition to the investigation into the association between these four emotions and HR and heart rate variability (HRV), another aim in their study was to investigate how to optimally combine the information that both these data sources provide. Ketonen et al. (2023)

used a Gaussian weighted average of HR, HRV and the MET value over a period of 5 minutes before, around and after the ESM prompt, a method often used to smooth out the data (Costa et al., 2012). This method computes the mean over five minutes by giving datapoints closer to the ESM point higher weights compared to those further away from the ESM prompt. Their findings showed that excitement was related to higher HR and lower HRV when controlling for PA with the MET value. In addition, boredom was associated with a lower HR but not HRV, also when controlling for PA. The other two emotions did not show any relationship with HR or HRV.

This thesis aims to further investigate the relationship between self-reported emotions as they appear according to Russell's (1980) theory and the associated cardiovascular response as already found by for example Ketonen et al. (2023) and Marci et al. (2007). More specifically, this research aims to give insight into the optimal way to combine and analyse these data sources. This will be done by means of an ESM study of 3 days to measure self-reported emotions (excitement, calmness, boredom, and anxiety), the same emotions as measured in Ketonen et al. (2023). In addition to the Likert scale, as used in Ketonen et al. (2023), we asked participants to pinpoint their position regarding their emotional state on the arousal-valence matrix upon each ESM promp. All the while continuous physiological data will be collected on HR and MET by means of a fitness tracker. Different linear mixed models will be compared to each other with Akaike's Information Criterion to determine which model performs best per each self-reported emotion. This will tell us what method will give us most power in future studies.

2 Method

2.1 Participants

A non-probability convenience sample, consisting of 5 students (3 were female), aged between 21 and 25 years old (M = 24.4, SD = 2.97), was recruited from the University of Twente, who registered for this experiment through the "SONA System" platform. Participants were required to be fluent in either English or Dutch and not have any cardiovascular health related issues. Participants had to be in possession of an IOS or Android smartphone. Each participant was informed about the study procedures and signed an informed consent before participating in the study (Appendix A). The study was approved by the faculty's ethics committee from the University of Twente under study number 231238 (Ethics (BMS/domain HSS), n.d.).

2.2 Materials

2.2.1 EmbracePlus

To measure HR and MET, the EmbracePlus wristband was used, acquired through the BMS lab at the University of Twente (EmbracePlus. n.d.). Since the EmbracePlus is a relatively new device on the market, at this moment, there is not yet any published literature on the performance of this wristband. However, the EmbracePlus has similar sensors to the Empatica E4 and there is literature on the performance of this Empatica E4, generally describing it as well-performing (McCarthy et al.. 2016). EmbracePlus measures a range of physiological parameters, however, in this research the focus lies on HR and MET. The photoplethysmography (PPG) sensor measures blood pulse rate (PR) and its variability (PRV) by using a sampling frequency of 64 Hz to compute HR and HRV. It enables the researcher to measure PR every minute.

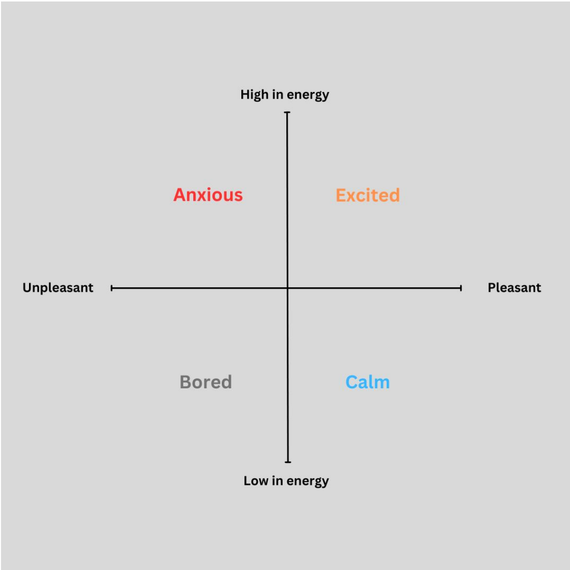
Each participant received this wristband upon set up of the experiment, as well as a charger of the device and a user manual. Upon set up of the experiment, participants downloaded the "care lab" application developed by Empatica to give them insight into their own data. This application also detects when the wristband is not worn correctly or if the battery percentage is low and triggers a push-notification when this happens. This app is connected to the health care platform by Empatica. This platform can be accessed by the researcher on the internet and allows them to access the data real-time, which is directly uploaded from the wristband.

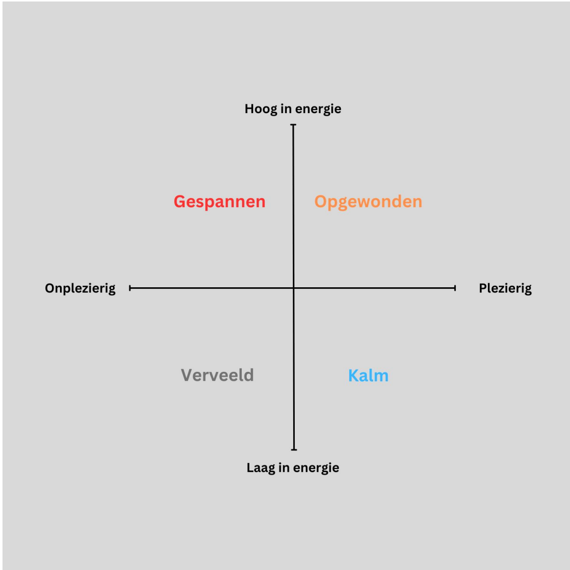
2.2.2 The Twente Intervention and Interaction Machine (TIIM)

To measure self-reported emotions, participants were instructed to install the TIIM app on their smartphone. This app is developed by the University of Twente. TIIM is an application that sends questionnaires to participants at specified time intervals and notifies them when a new questionnaire becomes available for completion and allows the researcher to send reminders when a set of questions is not yet answered. Questionnaires were made available in Dutch and English. To assess emotional states, a set of four statements were given in which the participants had to indicate on a 7-point Likert scale (strongly disagree – strongly agree) how much they agreed with the statement (I feel excited/calm/bored/anxious right now), like done by Ketonen et al. (2023). In addition to these statements participants were presented with a picture of the emotional quadrant (figure 1) and asked to pinpoint their emotional state on the grid. In addition to this, participants were asked how they slept last night (Likert scale). As a last question of the set, they were also asked to indicate what they were doing before they filled in the questionnaire.

Figure 1

The valence arousal matrix in English (left) and Dutch (right)





2.3 Design and Procedure

After the participant signed up to participate in the experiment, a set-up appointment was scheduled with the participant. The purpose of this meeting was to inform the participant about the aim and procedure of the research, as well as addressing any questions the participant had and setting up all the necessary research tools. This set-up appointment lasted around 45 minutes. If the participant decided to still participate in the experiment after the explanation was given, they were asked to sign the consent form (Appendix A). Following the consent process, the fitness tracker was set-up together with the care lab application and an instruction was given on how to properly wear the fitness tracker to enable accurate data collection, in accordance with Empatica's guidelines. It was shown how to charge the tracker in case its battery runs low. In addition to installation of the fitness tracker, the participant was asked to download the TIIM application on their smartphone. Consequently, the participant was asked to sign up to the right study by scanning the QR code provided by the researcher. This enabled the researcher to sign the participant up for the correct study on the TIIM dashboard. The participant was told that the questionnaires would start on the next day in the morning and would last for three days. After this, any remaining questions or concerns of the participant were addressed before scheduling a return appointment for the fitness tracker. During this return appointment the fitness tracker was returned, and remaining questions of the participant were answered. To enable the participant to contact the researcher in case of any issues, the researcher's phone number was shared with the participants during the initial appointment. Participants were told that they could contact the researcher at any time when they deemed necessary.

During the subsequent 3 days, the participant wore the fitness tracker during waking hours and charged it during the night. Following the day of the set-up appointment, the TIIM app started to give push notifications according to a semi-random time schedule (every two hours starting at 09.00, the last notification was given between 21.00 and 23.00), 7 times a day. A reminder notification was sent 20 minutes after the questionnaire became available for the participant. In addition, another notification was sent 20 minutes before the questionnaire closed, which was 40 minutes after the questionnaire opened for the participant. So, participants had one hour to respond to the prompt. The TIIM app does not allow to send individually randomized notification within certain time blocks; hence we randomized the timestamps ourselves and used those to trigger the ESM prompt. After finishing the three days, the fitness tracker was returned, and the researcher granted participants who successfully concluded the experiment two SONA credits.

The study design and procedure aimed was partially inspired by Ketonen et al. (2023). The length of this study is the same as theirs, except that Ketonen et al.'s (2023) study was part of a larger 10 day study and they only prompted the participants five times a day instead of seven times as done in our study. The questions asked during the prompts were largely similar and the discrete emotions under investigation were the same. However, we added a question where participants had to pinpoint their position on the emotional quadrant. Regarding the collection of physiological continuous data, Ketonen et al. (2023) made their participants wear ECG sensors to collect data on HR and HRV. As we were interested in minimally invasive data collection for the participant, we opted for a fitness tracker worn on the wrist.

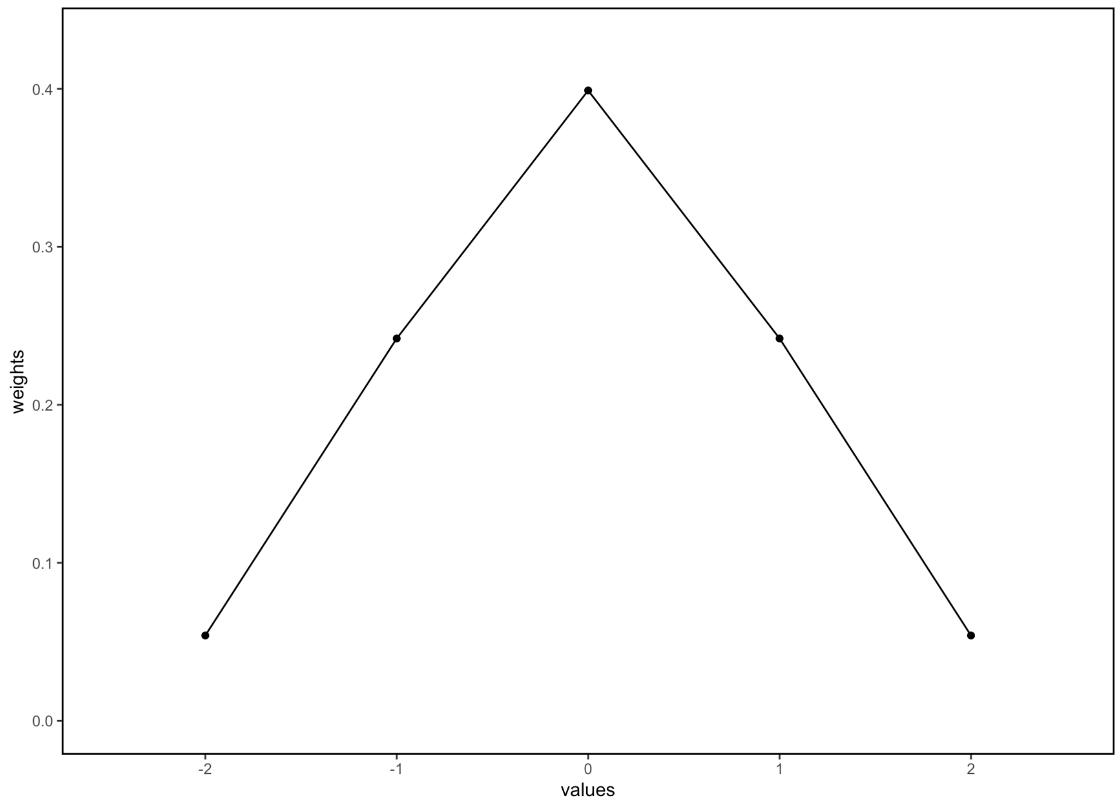
2.4 Data analysis

Data was preprocessed and analyzed using R and Rstudio using the following packages: tidyverse, lme4, tidyr, tibble, dplyr, reshape2, data.table, readxl, zoo, ggplot2, lmerTest, modelr. Data was preprocessed so that data on pulse rate, MET values and the answers to the ESM prompts were in the same data frame, resulting in the final dataset for analysis in long format for each participant separately. Each row represented the ID number, a time stamp with the corresponding pulse rate and MET value of that minute, and if an ESM prompt was filled in that minute, the row also contained data on these answers, otherwise these cells were N/A. For all the questions on the emotions ("I feel excited/bored/calm/ anxious right now"), we gave it a score from 1-7 (strongly disagree – strongly agree). The data on the quadrant consisted of two variables, the valence dimension and the arousal dimension, both running from -100 to 100.

After finishing the above-mentioned steps for each participant, all data was imported into R and combined into one data frame. Since the data on PR and MET was on the exact minute, we rounded the timestamps of the ESM prompt to the closest minute. After this, the mean, standard deviation, minimum and maximum for each of the study variable were computed. In addition, to gain more insight into the emotions over the three days, we plotted the emotional scores across the three days for each participant separately. Following this, we created a mean of five minutes before, around and after each pulse rate and MET measurement using the rollapply() function of the zoo package and created new columns for this. Since we were also interested in the relationships between the self-reported emotions and a Gaussian weighted mean before, around, and after five minutes the ESM prompt, we also created rolling weighted means for pulse rate and the MET value, by using the density function of the standard normal distribution between -2 and +2 standard deviations. More specifically, we used the weights 0.05399097, 0.24197072, 0.39894228, 0.24197072 and 0.05399097, for the five measurements before, during and after the target measurement (figure 2).

Figure 2

The used distribution for the computed weights for the Gaussian weighted means



When all the variables needed were created, we ran linear mixed models with the four emotions and the two dimensions as dependent variables and unweighted and the Gaussian weighted means of PR five minutes before, around and after the ESM prompt as the independent variable. In these models, ID was the random effect to control for individual mean differences in the dependent variable. In addition, we ran these models with an added variable to control for PA (MET). These analyses thus resulted in twelve models per discrete emotion, six with one fixed effect (PR) and six with two (PR and MET).

Ketonen et al. (2023) also performed linear mixed models with the four emotions and MET as independent variables and the unweighted or Gaussian weighted means for PR as the dependent variable to examine whether self-reported emotions could predict physiological arousal. We performed the same models, thus resulting in another 36 multiple linear mixed models with an included random participant effect (ID). Additionally, to check whether these linear mixed models were feasible, the three assumptions were checked by looking at the residuals. The assumptions of equal variances and linearity were checked using the residual vs. fitted values plot. The assumption of normality was checked using the QQ-plot.

Since we were interested in the performance of these models compared to each other, Akaike's Information Criterion was computed for each model. This criterion is a tool for selecting the best model out of a set of candidate models by estimating the predictive accuracy of a model (Forster & Sober, 2011). When comparing models, the lowest number for this value indicates the model that has the highest predictive accuracy. Since this criterion can only be used to compare models with identical response variables, we only computed it for the models with self-reported emotions as dependent variable. Consequently, the models with one fixed effect (PR) were compared to each other and the models with two fixed effects (PR and MET) were compared to each other. The data and R-script used for all the above-mentioned analyses can be found on GitHub (Huntjens, 2024).

3 Results

No participants were excluded from the sample as they all had a response rate of > 60 % surveys that were filled in (respectively 90, 66, 76, 66, and 66 %). The descriptive statistics (mean, SD, observed minimum and observed maximum) of all study variables of all available measurement points can be found in table 1.

Table 1

Descriptive statistics for all study variables of all available measurement points

Variable	Mean	SD	Observed min	Observed max
PR	90.63	18.50	27	239
MET	2.36	1.79	0.95	8.52
Excited	4.43	1.61	1	6
Calm	4.91	1.53	1	7
Bored	2.63	1.49	1	6
Anxious	2.19	1.14	1	6
Valence	27.94	24.29	-30	66
Arousal	-4.83	27.58	-53	61

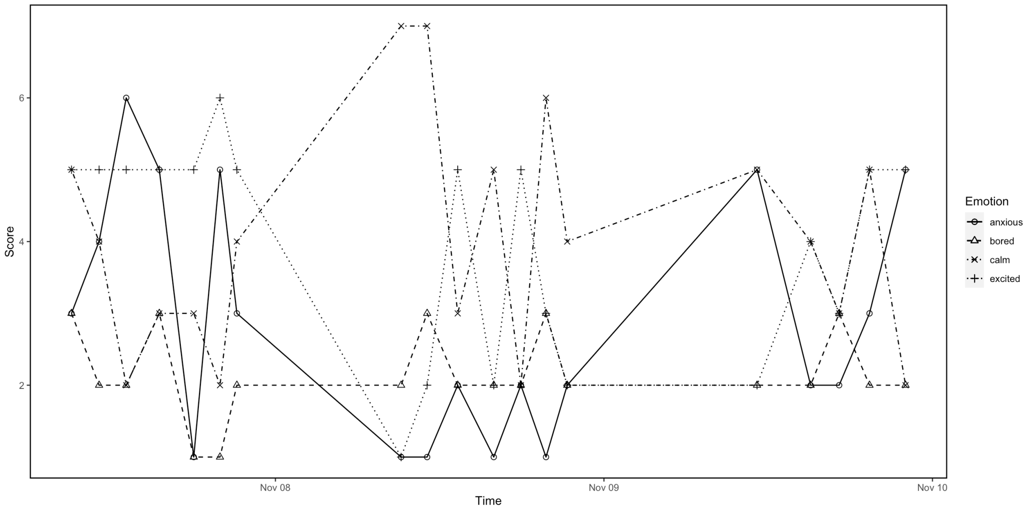
In figure 3, the emotional scores are displayed as they appeared during the three days. Each separate graph (A-E) represents one participant, scores ranged from 1 (strongly disagree) to 7 (strongly agree).

All the assumptions were checked for each model and no deviations from normality, linearity or equal variances were found, so we ran the linear mixed models as planned. All the parameter estimates of these models are presented in the appendix B and C. Table B1 and B2 represent the parameter estimates of the models with either the unweighted or Gaussian weighted mean of PR five minutes before, around, or after the ESM prompt as independent variable and random effect for ID. No significant effects were found for these models. Table B3 and B4 represent the parameters of the models with the added variable to control for PA. So, these models included either the unweighted or Gaussian weighted means for PR and MET five minutes before, around, or after the ESM prompt. A significant negative association was found between PR and valence before ($t(58) = -2.68$) and around ($t(58) = -2.26$) the ESM prompt for the unweighted mean. This result was also significant for the models with the Gaussian weighted means, before ($t(58) = -2.59$) and around ($t(58) = -2.19$) the ESM prompt. In these models, the MET value was also found to be significant. Appendix C represents the parameter estimates for the models where either the unweighted or Gaussian weighted mean of PR was the dependent variable, no significant effects were found in these analyses.

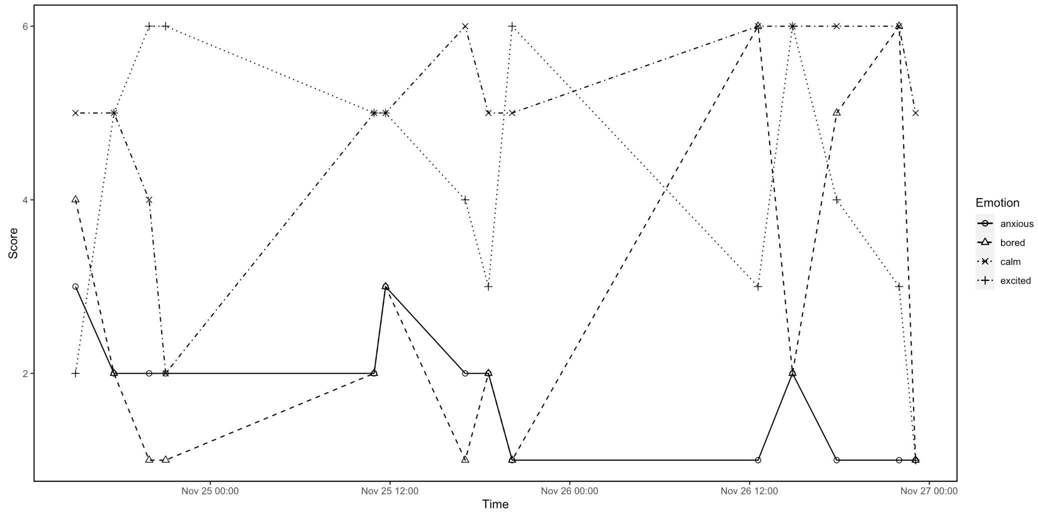
Figure 3

Emotional scores over time per participant

A



B

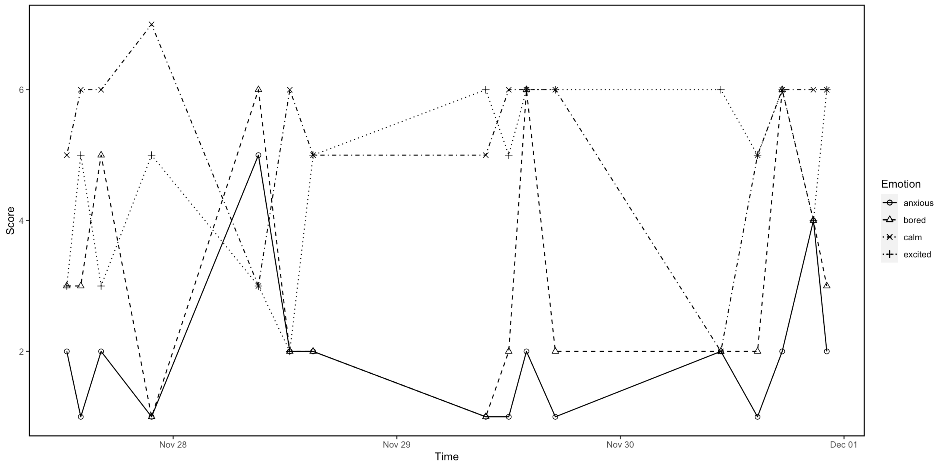


Note. Panel A: participant 1. Panel B: participant 2. Panel C: participant 3. Panel D: participant 4. Panel E: participant 5.

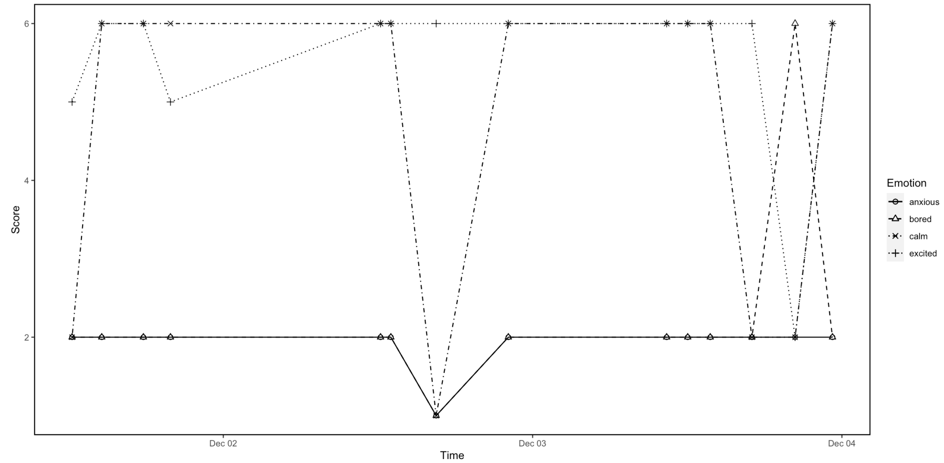
Figure 3

Emotional scores over time per participant

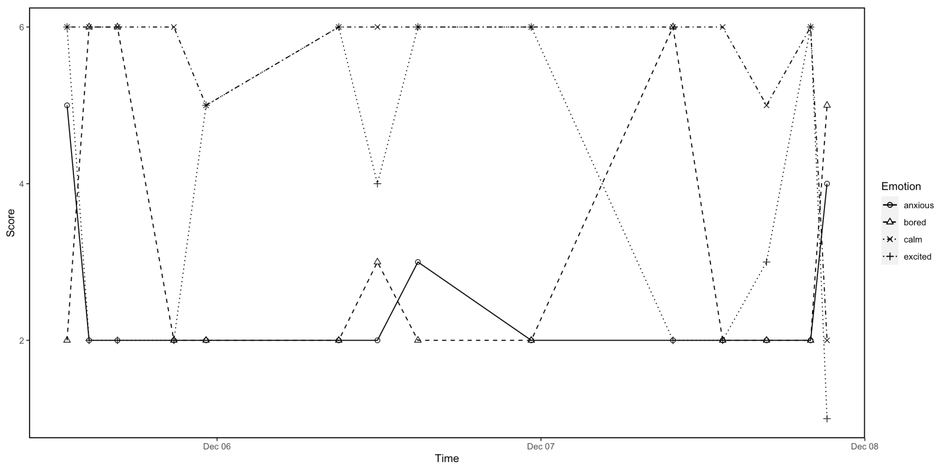
C



D



E



Note. Panel A: participant 1. Panel B: participant 2. Panel C: participant 3. Panel D: participant 4. Panel E: participant 5.

To gain insight into the performance of the different models, Akaike's Information Criterion was computed for each model. Since the AIC can only be used to compare models with the same dependent variables, we only applied this criterion to the models with the emotions as dependent variable (table B1- 4) and then compared each AIC value per each discrete self-reported emotion. The results of this criterion can be found in table 2 for the models with one fixed effect (PR) and the results of the models with two fixed effects (PR and MET) can be found in table 3. The lowest number for each emotion is presented in bold. For the AIC in table 2, there is no clear pattern of what model performs best per emotion. In addition, the values for all six models per emotions are close to each other, suggesting that one is not superior over the other. For the AIC in table 3 however, a clear pattern can be found. Generally, the models with the unweighted mean of PR and MET five minutes after the ESM prompt performed best. Except for boredom, for this emotion, the model with the Gaussian weighted mean five minutes around the ESM prompt performed best.

Table 2

AIC criterion for the models with the unweighted or Gaussian weighted mean of PR five minutes before, around, or after the ESM prompt.

	Unweighted 5 min PR mean			Gaussian weighted 5 min PR mean		
Emotions	Before	Around	After	Before	Around	After
Excited	237.87	**237.05**	237.27	238.09	237.13	237.22
Calm	**235.78**	236.19	236.08	235.90	236.27	236.18
Bored	239.55	239.29	239.26	239.85	239.28	**239.18**
Anxious	**207.02**	207.25	207.21	206.72	207.44	207.56
Valence	567.98	568.94	569.85	**567.99**	568.81	570.02
Arousal	583.23	582.74	582.76	583.48	582.78	**582.58**

Note. AIC criterion is presented for six models per self-reported emotion. Lowest number per emotions is presented in bold.

Table 3

AIC criterion for the models with the unweighted or Gaussian weighted means of PR and MET five minutes before, around, or after the ESM prompt.

	Unweighted 5 min mean of PR and MET			Gaussian weighted 5 min mean of PR and MET		
Emotions	Before	Around	After	Before	Around	After
Excited	485.89	475.81	**473.75**	494.01	477.16	477.43
Calm	484.62	480.3	**477.21**	493.47	480.73	483.54
Bored	486.33	475.86	473.15	494.80	**467.73**	476.66
Anxious	484.16	480.12	**477.06**	492.32	481.07	483.08
Valence	484.67	482.9	**481.18**	491.48	486.68	486.37
Arousal	487.86	484.46	**481.73**	494.62	488.19	486.84

Note. AIC criterion is presented for six models per self-reported emotion. Lowest number per emotions is presented in bold.

4 Discussion

This study aimed at finding the optimal way to compare and analyse ESM data in combination with physiological continuous data by means of a 3-day ESM study to measure self-reported emotions as they appear according to Russel's (1980) theory, while also collecting continuous data on PR and MET. Our aim was to determine the best way to combine and analyse these two data sources by means of comparing the linear mixed models using Akaike's Information Criterion. A total of 108 linear mixed models were performed, with both PR as predictor variable and the self-reported emotions as response variables and vice versa.

A negative association was found between PR and valence before and around the ESM prompt for both the unweighted mean and the Gaussian weighted mean when valence was de dependent variable. For the other models, no significant effects were found. However, it must be noted that these results should be treated with caution, since the sample size was small, and it might not be representative for the population. In addition, the data we collected on boredom and anxiety seemed skewed, indicating that in the three days of data collection, there have been little to no situations where participants experienced high levels of these two emotions. Ketonen et al. (2023) reported this issue only for anxiety.

More relevant to the results question regarding the relative performance of these models, Akaike's Information Criterion was computed for the first 72 linear mixed models (table B1-4) and a comparison was made per each discrete emotion. For the models with just the Gaussian weighted or unweighted mean for PR, no clear pattern was present for this criterion. Numbers seem very close and suggest that one way of analysing is not superior over the other when only including one fixed effect. For the models with two fixed effects, however, there is a clear pattern visible. Generally, the models with unweighted means of five minutes after the ESM prompt seem to have the highest prediction accuracy. Except for boredom, for this emotion the model with the Gaussian weighted means for PR and MET around the ESM prompt performed best. This finding might be explained by considering that boredom is generally associated with a state of lower arousal (Mikulas & Vodanovich, 1993), indicating a lower PR (Raz & Lahad, 2022; Ketonen et al., 2023). When participants are prompted to answer the ESM questionnaire, this might induce arousal and therefore a higher PR, since they now have a task to focus on. This could explain why, for boredom specifically, the model with the Gaussian weighted mean around the ESM prompt has higher prediction accuracy compared to measurements taken before or after the prompt.

The present findings are of interest, given that Ketonen et al. (2023) did not include models with unweighted means. Therefore, it is interesting to observe that models with unweighted means generally perform better compared to those with Gaussian weighted means. Moreover, in context of the models with self-reported emotions as dependent variables, Ketonen et al. (2023) exclusively focused on Gaussian weighted means of five minutes before and around the ESM prompt, disregarding the five minutes after the prompt. Hence, it is noteworthy to observe that for the majority of the measured self-reported emotions, the optimal approach for analysis of these two data sources is to look at five minutes after the ESM prompt. The better performance of the models with the mean of PR five minutes after the prompt compared to before or around the prompt could be attributed to an emotional processing period when completing the ESM questionnaire. When participants are prompted to express their feelings in the questionnaire, it might initiate a cognitive and emotional processing period (Holzman & Bridgett, 2017). This moment of introspection may influence the subsequent physiological reaction. The five minutes after the ESM prompt might capture the peak of this emotional processing, explaining why for these emotions, these models have the highest prediction accuracy.

While our findings contribute to the body of knowledge about the combination of continuous physiological data and ESM data, it is important to acknowledge and discuss the limitations that may have influenced the results. Other than the small sample size already mentioned, there are three noteworthy limitations. Firstly, heart rate variability (HRV), an important cardiovascular response of emotion, was not measured in this study (Purves et al., 2001; Holzman & Bridgett, 2017). We aimed to do so, however, the EmbracePlus did not accurately measure this construct, resulting in a lot of missing datapoints, making it impossible to use this variable in our analyses. Ketonen et al. (2023) found a significant association of HRV and excitement, and we were not able to replicate this result since we could not measure HRV. The addition of this variable in future research might provide valuable insights into the exploration of an adequate method for analysing physiological continuous data in combination with ESM data (Malik, 1996).

Secondly, PA was measured with the MET value. As this is the amount of oxygen used per kg body weight per minute (Jetté et al., 1990) and differs per exercise, this measure might not have been accurate since EmbracePlus is not aware of the exercise and is highly dependent on the fitness, body composition etc. of the participant. In addition, Empatica has not published the algorithm how MET value is exactly calculated, so it is unknown how MET values are computed. Future research could alternatively involve asking participants about their activities upon each ESM prompt and subsequently converting this to a MET value to provide a numerical variable. To convert an activity to a MET value, the compendium of MET values by Ainsworth et al. (2011) can be used for determining the MET value of each activity according to this list, this ensures more accurate values (Louisa et al., 2014).

Thirdly, in this study we have solely focused on five minutes, either before, around or after the ESM prompt. However, the optimal duration remains uncertain; five minutes might be too long, smoothing out important data trends, or too short, possibly failing to capture the complete physiological response to an emotion (Ketonen et al., 2023). Exploring different intervals, shorter (e.g., 3 min) or longer intervals (e.g., 10 minutes), could provide insights into the ideal timeframe for capturing meaningful physiological trends. Future studies could compare these variations to further our understanding on the optimal approach of analysing continuous physiological data together with ESM data.

Despite the identified limitations, this study also has notable strengths. One significant strength is the ecological validity of this study, conducted in real-world and daily life settings. This approach provided valuable insights into both participants' subjective and more physiological objective experience of emotions in their day-to-day lives. The study's design also minimized recall bias by prompting participants to report on their feelings at the exact moment of answering the questionnaire, thus no recall was necessary. Additionally, the study was of non-invasive nature allowing participants to live their life as normal, without disruptions. The fitness tracker worn on the wrist and therefore hardly noticed by the participants. In contrast, the choice of an ECG on the torso, as observed in Ketonen et al. (2023), could pose challenges for participants, highlighting the advantage of the fitness tracker in minimizing participant burden.

Another strong aspect of this study was the addition of the emotional quadrant in our analyses. Since the Likert scale as predefined answers, participants might not feel like one of the answers suits his or her feeling best. The dimensions of valence and arousal allowed the participants to pinpoint where they are on the quadrant, allowing for higher resolution and more nuanced judgement (Klimek et al., 2017). The use of these dimensions contributes to a more nuanced understanding of participants' emotions, since the quadrant acknowledges the multifaceted and dynamic nature of human emotions (Russel, 1980). Future research could focus on the use of this quadrant and how the dimensions interplay, since they might be correlated (Citron et al., 2014). Future studies might benefit from developing a methodology for analysing these two dimensions alongside each other, moving beyond the current approach of treating them as separate variables.

The last strength worth discussing is the addition of the AIC criterion to the analyses, this allowed for comparison of the models to investigate the performance of a Gaussian weighted mean versus an unweighted mean and whether we should focus on the time before, around, or after the ESM prompt. This last finding contributes to the findings of Ketonen et al. (2023) since they have exclusively used the Gaussian weighted mean without comparing the different linear mixed models regarding their performance. By adding models with unweighted means and comparing them to the Gaussian weighted means, it enables us to select the most plausible model and method for analysing this sort of data.

Considering the strengths and limitations identified in our study, it is essential to investigate other possible implications for future research to add on to the body of knowledge on this emerging topic. One implication for future research involves changing the focus from analysing the sample as a whole to a more individual level (N-of-1) analysis (Vieira et al., 2017). ESM provides valuable information about (fluctuations in) emotions as they are subjectively experienced. This might be varying for each participant, so instead of performing a linear mixed model and including a random effect, doing the analyses for n=1 might give valuable within person insights. This individual exploration can give a more nuanced understanding about emotions as they emerge in individual participants. To receive adequate effect sizes, it might be necessary to increase the amount of ESM prompts in the study, this can be done by increasing the length of the study or the number of prompts given each day, both adjustments likely being necessary (Dejonckheere & Erbas, 2022).

This study aimed to enhance our understanding of the combination and analysis of continuous physiological data and ESM data, particularly focusing on self-reported emotions, PR, and MET. The comparison of the linear mixed models revealed that those with an unweighted mean of PR and MET five minutes after the ESM prompt had the highest prediction accuracy compared to the other candidate models. This finding can help develop a guideline for the analysis of these two data sources and enhance standardization of the analyses. The establishment of standardized measures and analyses can help reduce heterogeneity that is present in the methods used across studies (Vaessen et al., 2021). By promoting consistency in analysis approaches cross-study comparison can be facilitated.

References

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Tudor-Locke, C., Greer, J. L., Vezina, J., Whitt-Glover, M. C., & Leon, A. S. (2011). 2011 Compendium of Physical Activities: A Second Update of Codes and MET Values. *Medicine & Science in Sports & Exercise*, *43*(8), 1575–1581. https://doi.org/10.1249/MSS.0b013e31821ece12

Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys. *Science*, *236*(4798), 157–161. https://doi.org/10.1126/science.3563494

Christensen, T. C., Barrett, L. F., Bliss-Moreau, E., Lebo, K., & Christensen, T. C. (2003). A Practical Guide to Experience-Sampling Procedures. *Journal of Happiness Studies*, *4*(1), 53–78. https://doi.org/10.1023/A:1023609306024

Citron, F. M. M., Gray, M. A., Critchley, H. D., Weekes, B. S., & Ferstl, E. C. (2014). Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework. *Neuropsychologia*, *56*(100), 79–89. https://doi.org/10.1016/j.neuropsychologia.2014.01.002

Costa, T., Boccignone, G., & Ferraro, M. (2012). Gaussian Mixture Model of Heart Rate Variability. *PLoS ONE*, *7*(5), e37731. https://doi.org/10.1371/journal.pone.0037731

De Calheiros Velozo, J., Habets, J., George, S. V., Niemeijer, K., Minaeva, O., Hagemann, N., Herff, C., Kuppens, P., Rintala, A., Vaessen, T., Riese, H., & Delespaul, P. (2022a). Designing daily-life research combining experience sampling method with parallel data. *Psychological Medicine*, 1–10. https://doi.org/10.1017/S0033291722002367

De Calheiros Velozo, J., Niemeijer, K., & Vaessen, T. (2022b). Passive Sensing. In I. Myin-Germeys & P. Kuppens (Eds.), *The Open Handbook of Experience Sampling Methodology* (pp. 251–163).

Dejonckheere, E., & Erbas, Y. (2022). Designing an experience based sampling study. In I. Myin-Germeys & P. Kuppens (Eds.), *The Open Handbook of Experience Sampling Methodology* (pp. 33–77).

EmbracePlus | The world's most advanced smartwatch for continuous health monitoring. (n.d.). Retrieved November 2, 2023, from https://www.empatica.com/en-gb/embraceplus/

Ethics (BMS/domain HSS) | Faculty of Behavioural, Management and Social sciences (BMS). (n.d.). Universiteit Twente. Retrieved January 15, 2024, from https://www.utwente.nl/en/bms/research/ethics-domainHSS/

Forsum, E., Janerot-Sjöberg, B., & Löf, M. (2018). MET-values of standardised activities in relation to body fat: Studies in pregnant and non-pregnant women. *Nutrition & Metabolism*, *15*(1), 45. https://doi.org/10.1186/s12986-018-0281-z

Holzman, J. B., & Bridgett, D. J. (2017). Heart rate variability indices as bio-markers of top-down self-regulatory mechanisms: A meta-analytic review. *Neuroscience & Biobehavioral Reviews*, *74*, 233–255. https://doi.org/10.1016/j.neubiorev.2016.12.032

Huntjens, I. (2024). *Icwh2002/BachelorThesis* [R]. https://github.com/icwh2002/BachelorThesis (Original work published 2024)

Jetté, M., Sidney, K., & Blümchen, G. (1990). Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clinical Cardiology*, *13*(8), 555–565. https://doi.org/10.1002/clc.4960130809

Ketonen, E. E., Salonen, V., Lonka, K., & Salmela-Aro, K. (2023). Can you feel the excitement? Physiological correlates of students' self-reported emotions. *British*

Journal of Educational Psychology, *93*(S1), 113–129. https://doi.org/10.1111/bjep.12534

Klimek, L., Bergmann, K.-C., Biedermann, T., Bousquet, J., Hellings, P., Jung, K., Merk, H., Olze, H., Schlenter, W., Stock, P., Ring, J., Wagenmann, M., Wehrmann, W., Mösges, R., & Pfaar, O. (2017). Visual analogue scales (VAS): Measuring instruments for the documentation of symptoms and therapy monitoring in cases of allergic rhinitis in everyday health care: Position Paper of the German Society of Allergology (AeDA) and the German Society of Allergy and Clinical Immunology (DGAKI), ENT Section, in collaboration with the working group on Clinical Immunology, Allergology and Environmental Medicine of the German Society of Otorhinolaryngology, Head and Neck Surgery (DGHNOKHC). *Allergo Journal International*, *26*(1), 16–24. https://doi.org/10.1007/s40629-016-0006-7

Li, Z., Snieder, H., Su, S., Ding, X., Thayer, J. F., Treiber, F. A., & Wang, X. (2009). A longitudinal study in youth of heart rate variability at rest and in response to stress. *International Journal of Psychophysiology*, *73*(3), 212–217. https://doi.org/10.1016/j.ijpsycho.2009.03.002

Maes, I. H. L., Delespaul, P. A. E. G., Peters, M. L., White, M. P., Van Horn, Y., Schruers, K., Anteunis, L., & Joore, M. (2015). Measuring Health-Related Quality of Life by Experiences: The Experience Sampling Method. *Value in Health*, *18*(1), 44–51. https://doi.org/10.1016/j.jval.2014.10.003

Malik, M. (1996). Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task force of the European Society of Cardiology and the north American Society of Pacing and Electrophysiology. *European Heart Journal*, 17(3), 354–381.

Marci, C. D., Glick, D. M., Loh, R., & Dougherty, D. D. (2007). Autonomic and prefrontal cortex responses to autobiographical recall of emotions. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(3), 243–250. https://doi.org/10.3758/CABN.7.3.243

McCarthy, C., Pradhan, N., Redpath, C., & Adler, A. (2016). Validation of the Empatica E4 wristband. *2016 IEEE EMBS International Student Conference (ISC)*, 1–4. https://doi.org/10.1109/EMBSISC.2016.7508621

Mikulas, W. L., & Vodanovich, S. J. (1993). The essence of boredom. *Psychological Record*, *43*(1), 3.

Myin-Germeys, I., & Kuppens, P. (2022). Experience Sampling Methods, an Introduction. In I. Myin-Germeys & P. Kuppens (Eds.), *The open handbook of experience sampling methodology* (pp. 7–18).

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., & Williams, S. M. (2001). Physiological Changes Associated with Emotion. In *Neuroscience. 2nd edition*. Sinauer Associates. https://www.ncbi.nlm.nih.gov/books/NBK10829/

Raz, S., & Lahad, M. (2022). Physiological indicators of emotional arousal related to ANS activity in response to associative cards for psychotherapeutic PTSD treatment. *Frontiers in Psychiatry*, *13*, 933692. https://doi.org/10.3389/fpsyt.2022.933692

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

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

Sylvia, L. G., Bernstein, E. E., Hubbard, J. L., Keating, L., & Anderson, E. J. (2014). Practical Guide to Measuring Physical Activity. *Journal of the Academy of Nutrition and Dietetics*, *114*(2), 199–208. https://doi.org/10.1016/j.jand.2013.09.018

Vaessen, T., Rintala, A., Otsabryk, N., Viechtbauer, W., Wampers, M., Claes, S., & Myin-Germeys, I. (2021). The association between self-reported stress and cardiovascular measures in daily life: A systematic review. *PLOS ONE*, *16*(11), e0259557. https://doi.org/10.1371/journal.pone.0259557

Vieira, R., McDonald, S., Araújo-Soares, V., Sniehotta, F. F., & Henderson, R. (2017). Dynamic modelling of n-of-1 data: Powerful and flexible data analytics applied to individualised studies. *Health Psychology Review*, *11*(3), 222–234. https://doi.org/10.1080/17437199.2017.1343680

Appendix A

Informed consent form

Consent Form for
Finding the optimal way to analyse heart rate data together with data on self-reported emotions

YOU WILL BE GIVEN A COPY OF THIS INFORMED CONSENT FORM

Please tick the appropriate boxes	**Yes**	**No**
Taking part in the study		
I have read and understood the study information dated [DD/MM/YYYY]. or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	□	□
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time. without having to give a reason.	□	□
I understand that taking part in the study involves wearing a fitness tracker and answering a small survey seven times a day for a period of three days.	□	□
Use of the information in the study		
I understand that information I provide will be used for a student research project and perhaps a journal publication.	□	□
I understand that personal information collected about me that can identify me will not be shared beyond the study team.	□	□
Future use and reuse of the information by others		
I give permission for the ESM data and fitness tracker data that I provide to be archived so it can be used for future research and learning.	□	□

Signatures

________________________ ________________________ _________

Name of participant [printed] Signature Date

I have accurately read out the information sheet to the potential participant and. to the best of my ability. ensured that the participant understands to what they are freely consenting.

Isabella Huntjens

Researcher name [printed] Signature Date

Study contact details for further information:

Student: Isabella Huntjens (i.c.w.huntjens@student.utwente.nl)

First supervisor: Stephanie van den Berg (Stephanie.vandenberg@utwente.nl)

Second supervisor: Peter ten Klooster (P.m.tenklooster@utwente.nl)

Contact Information for Questions about Your Rights as a Research Participant

If you have questions about your rights as a research participant. or wish to obtain information. ask questions. or discuss any concerns about this study with someone other than the researcher(s). please contact the Secretary of the Ethics Committee/domain Humanities & Social Sciences of the Faculty of Behavioural. Management and Social Sciences at the University of Twente by ethicscommittee-hss@utwente.nl

Appendix B

Table B1

Parameter estimates of the models with the emotions as dependent variables and the unweighted mean of PR before, around or after the ESM prompt as predictor variable.

Unweighted 5 min PR mean	Excited	Calm	Bored	Anxious	Valence	Arousal
Before	0.02	0.01	0.01	-0.01	-0.38	0.15
	(0.01)	(0.01)	(0.02)	(0.01)	(0.24)	(0.28)
Around	0.02	0.01	0.01	-0.01	-0.30	0.25
	(0.02)	(0.02)	(0.02)	(0.01)	(0.26)	(0.29)
After	0.02	0.01	0.01	-0.01	-0.14	0.24
	(0.02)	(0.02)	(0.02)	(0.01)	(0.27)	(0.30)

Before, around and *after* referring to the ESM prompt time. Fixed effect parameters are presented.

Table B2

Parameter estimates of the models with the emotions as dependent variables and the Gaussian weighted mean of PR before, around or after the ESM prompt as predictor variable

Gaussian weighted 5 min PR mean	Excited	Calm	Bored	Anxious	Valence	Arousal
Before	0.01	0.01	0.01	-0.01	-0.37	0.11
	(0.01)	(0.01)	(0.01)	(0.01)	(0.23)	(0.27)
Around	0.02	0.00	0.01	-0.01	-0.31	0.25
	(0.02)	(0.02)	(0.02)	(0.01)	(0.25)	(0.29)
After	0.02	0.01	0.01	0.00	-0.10	0.27
	(0.02)	(0.02)	(0.02)	(0.01)	(0.26)	(0.29)

Before, around and *after* referring to the ESM prompt time. Fixed effect parameters are presented.

Table B3

Parameter estimates of the models with the emotions as dependent variables and the unweighted means of PR and MET before, around or after the ESM prompt as predictor variables.

Unweighted 5 min mean	Excited	Calm	Bored	Anxious	Valence	Arousal
Before						
PR	0.01	0.01	0.02	-0.01	**-0.62****	0.06
	(0.02)	(0.02)	(0.01)	(0.01)	(0.23)	(0.29)
MET	0.16	0.04	-0.21	-0.11	**6.45****	2.60
	(0.13)	(0.13)	(0.13)	(0.10)	(2.05)	(2.50)
Around						
PR	0.01	0.01	0.02	-0.01	**-0.55***	0.07
	(0.02)	(0.02)	(0.02)	(0.01)	(0.24)	(0.30)
MET	0.22	-0.10	**-0.30***	-0.07	**7.13*****	4.82
	(0.13)	(0.13)	(0.13)	(0.10)	(2.02)	(2.43)
After						
PR	0.01	0.02	0.03	-0.01	-0.44	0.00
	(0.02)	(0.02)	(0.02)	(0.01)	(0.27)	(0.31)
MET	0.20	-0.18	**-0.41***	-0.04	**6.39****	**5.42***
	(0.15)	(0.14)	(0.15)	(0.11)	(2.19)	(2.54)

Note: *** $p \leq 0.001$, **$p \leq 0.01$, *$p \leq 0.05$.

Significant results are presented in bold.

Before, around and *after* referring to the ESM prompt time. Fixed effect parameters are presented.

Table B4

Parameter estimates of the models with the emotions as dependent variables and the Gaussian weighted means of PR and MET before, around or after the ESM prompt as predictor variables.

Gaussian weighted 5 min mean		Excited	Calm	Bored	Anxious	Valence	Arousal
Before							
	PR	0.01	0.01	0.01	-0.01	**-0.57***	0.03
		(0.01)	(0.02)	(0.01)	(0.01)	(0.22)	(0.28)
	MET	0.13	-0.06	-0.19	-0.13	**5.98****	2.21
		(0.12)	(0.12)	(0.12)	(0.10)	(1.92)	(2.34)
Around							
	PR	0.01	0.01	0.02	0.00	**-0.52***	0.08
		(0.02)	(0.02)	(0.02)	(0.01)	(0.24)	(0.29)
	MET	**0.26***	-0.08	**-0.29***	-0.03	**6.20****	**4.85***
		(0.12)	(0.13)	(0.13)	(0.10)	(1.99)	(2.36)
After							
	PR	0.01	0.02	0.03	0.00	-0.41	0.05
		(0.02)	(0.02)	(0.02)	(0.01)	(0.26)	(0.30)
	MET	0.13	-0.19	**-0.37***	-0.08	**6.23****	**4.92***
		(0.14)	(0.13)	(0.14)	(0.11)	(2.02)	(2.35)

Note: **$p \leq 0.01$, *$p < 0.05$.

Significant results are presented in bold.

Before, around and *after* referring to the ESM prompt time. Fixed effect parameters are presented

Appendix C

Parameter estimates of the models with self-reported emotions as predictor variables and the different mean variable variants as dependent variables.

	Unweighted PR mean of 5 min			Gaussian weighted 5 min PR mean		
Predictors	Before	Around	After	Before	Around	After
Excited	0.82	0.98	0.80	0.94	0.97	0.87
	(1.14)	(1.05)	(1.03)	(1.23)	(1.06)	(1.06)
MET	1.01	0.19	1.40	0.78	1.22	1.5
	(0.84)	(0.78)	(0.77)	(0.90)	(0.79)	(0.79)
Calm	0.83	0.38	0.50	0.88	0.26	0.54
	(1.10)	(1.07)	(1.04)	(1.19)	(1.07)	(1.10)
MET	1.02	0.90	1.05	0.77	0.91	1.11
	(0.77)	(0.75)	(0.73)	(0.83)	(0.75)	(0.78)
Bored	0.72	0.92	1.06	0.47	0.97	1.26
	(1.20)	(1.01)	(0.99)	(1.18)	(1.02)	(1.02)
MET	1.27	1.49	**1.71***	1.02	1.51	**1.88***
	(0.83)	(0.76)	(0.75)	(0.89)	(0.77)	(0.78)
Anxious	-0.76	-0.57	-0.68	-1.13	-0.35	-0.41
	(1.39)	(1.35)	(1.32)	(1.49)	(1.36)	(1.40)
MET	1.01	0.83	0.98	0.87	0.75	1.13
	(0.76)	(0.75)	(0.73)	(0.82)	(0.75)	(0.76)
Valence	-0.11	-0.08	-0.05	-0.11	-0.09	-0.05
	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)
MET	**1.82***	1.50	1.44	**1.77***	1.36	1.64
	(0.79)	(0.79)	(0.78)	(0.84)	(0.81)	(0.81)
Arousal	0.02	0.04	0.02	0.01	0.04	0.04
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
MET	1.45	1.16	1.22	1.40	0.99	1.37
	(0.80)	(0.79)	(0.77)	(0.85)	(0.81)	(0.81)

Note: *$p < 0.05$.
Significant results are presented in bold.
Before. around and *after* referring to the ESM prompt time. Fixed effect parameters are presented. Standard error is presented in between the brackets.