Improving the methodology for non-invasive autonomic nervous system recording and its implementation in behavioral research

van Lien, R.

2014

document version
Publisher's PDF, also known as Version of record

CHAPTER 6

Ambulatory measurement of the ECG T-wave amplitude

René van Lien, Melanie Neijts, Gonneke. Willemsen, and Eco J.C.de Geus

This chapter will be submitted to Psychophysiology

Abstract

The preejection period (PEP) can be used to index changes in sympathetic nervous system (SNS) activity, even under naturalistic conditions by non-invasive recording of the impedance cardiogram (ICG) and electrocardiogram (ECG) signal. A remaining obstacle is the necessary interactive visual scoring of 24-hour recordings, which has proven laborious. The ECG T-wave amplitude (TWA) could be an alternative ambulatory SNS measure that suffers less from this drawback and is even more easy to record under naturalistic conditions. Here we report on 24-hour ambulatory monitoring of the TWA in a sample of 564 healthy adults. The TWA could be reliably extracted from the ensemble-averaged ECG in over 90% of the participants. It showed a clear decrease in response to a mental stress task and a stepwise decrease from nighttime sleep to daytime sitting to more physically active behaviors during the 24 hour recording, echoing the expected pattern of SNS activity across these various conditions. In addition, within-participant changes in TWA across the standardized stressors were significantly correlated with the PEP (r=.41). TWA and PEP were also significantly correlated across the unstandardized naturalistic conditions in the majority of the participants (mean r=.35), even after partialling out RSA and IBI. The correlation increased when the analyses were repeated in 96 participants with completely non-ambiguous ICG B-point scoring. We conclude that the TWA seems sensitive to SNS activity, and that ambulatory co-recording of the TWA and PEP may provide a more comprehensive picture of changes in SNS activity across the day than the PEP alone.

Introduction

Functional disturbances of the autonomic nervous system have been frequently linked to several diseases and hyperactivity of the sympathetic nervous system (SNS) may be an important cause for the detrimental effects of stress on cardiovascular health (Lambert et al., 2010; Parati & Esler, 2012). At the moment the preejection period (PEP) is the measure of choice to monitor changes in SNS activity non-invasively in psychophysiological stress research (Berntson et al., 2004; Kelsey & Guethlein, 1990; Kelsey, Ornduff, & Alpert, 2007; Vrijkotte et al., 2004). PEP can be obtained by simultaneous recording of the thoracic impedance cardiogram (ICG) and electrocardiogram (ECG) (Riese et al., 2003; Sherwood et al., 1990; Willemsen et al., 1996) and is defined as the interval from the onset of left ventricular depolarization, reflected by the Q-wave onset (Qonset) in the ECG to the opening of the aortic valve, reflected by the B-point in the ICG signal (Labidi et al., 1970; Lozano et al., 2007; Sherwood et al., 1990; van Lien, Schutte, Meijer, & de Geus, 2013; Willemsen et al., 1996).

The literature supports changes in PEP as a valid index of SNS induced changes in contractility of the left ventricle (Berntson et al., 1994a; Goedhart et al., 2006; Harris et al., 1967; Houtveen et al., 2005; Krzeminski et al., 2000; Kupper et al., 2006; Mezzacappa et al., 1999; Miyamoto et al., 1983a; Nelesen et al., 1999; Newlin et al., 1979; Richter et al., 2009; Schachinger et al., 2001; Sherwood et al., 1986; Smith et al., 1989b; Svedenhag et al., 1986; Vrijkotte et al., 2004; Williams et al., 1993; Winzer et al., 1999). A large advantage of the PEP is that it can be measured outside the confines of a hospital or laboratory setting by using ambulatory monitoring devices (Nakonezny et al., 2001; Sherwood et al., 1998; Wilhelm et al., 2003; Willemsen et al., 1996). This allows examination of individual differences in sympathetic activity in a natural setting, for instance during sleep or during job-related activities with a substantial mental load. These naturalistic conditions may have the largest clinical relevance (Kubiak & Stone, 2012; Trull & Ebner-Priemer, 2013).

However, when ambulatory research moves to an epidemiological scale collecting data in thousands of participants across extended periods of time, the practical feasibility of PEP measurement becomes an issue. Though automated scoring of the PEP has been made more efficient over the years by for example implementing large scale ensemble averaging (Riese et al., 2003) it still remains a time consuming process requiring visual inspection by multiple raters to ensure sufficient reliability of identification of the relevant landmarks (Berntson et al., 2004; Lozano et al., 2007; van Lien et al., 2013). Taken the laborious visual scoring required it would be extremely valuable to have alternative non-invasive measures of cardiac SNS activity that could be assessed through ambulatory monitoring with more ease.

Two such alternative measures are currently in use, the ratio of power in the low to high frequency bands of the heart rate power, (LF/HF ratio, (Pagani et al., 1991; Pagani et al., 1997; Pagani et al., 1986) and salivary alpha-amylase (sAA, (Nater et al., 2009). Unfortunately, the validity of these measures as indices of SNS activity has been strongly questioned. Although its use has become widespread, the LF/HF ratio is theoretically a poor measure of SNS activity (Eckberg, 1997) that does not correlate with other indicators of SNS activity (Goedhart et al., 2008b; Grassi et al., 1999). sAA, particularly when collected by cotton rolls (as is typical), has severe methodological drawbacks and seems to capture parasympathetic activity in addition to sympathetic activity (Bosch et al., 2011). Also, it cannot be sampled with a high resolution during recordings in naturalistic settings or during the nighttime.

In this study we set out to test a third, and seemingly somewhat forgotten, alternative to measure cardiac SNS activity, namely the amplitude of the T-wave in the ECG (TWA). The T-wave is

the asymmetrical wave in the ECG that comes after the QRS-complex and typically lasts approximately 150 msec. It reflects ventricular repolarization (Abildskov et al., 1977; Burgess, 1979; Haarmark et al., 2010; Randall et al., 1977) in which the sympathetic nerves play an important role (Abildskov, 1985). TWA is often defined as the difference between the peak of the T-wave and the isoelectric level during the same heart cycle (Furedy et al., 1984; Kline et al., 1998) where alternative baseline levels are in use, such as the midpoint of the PQ interval (Matyas, 1976) and the isoelectric period between the T-wave offset and the P-wave onset (Contrada et al., 1989). These isoelectric periods represent moments where only a negligible number of fibers in the cardiac conduction system are depolarizing (Goldman 1970).

Decreases in TWA and even TWA inversion were seen to occur after local stimulation of the stellate sympathetic ganglia in (Anitchkov et al., 1961; Yanowitz et al., 1966), intracoronary infusion of epinephrine (Barger et al., 1961), norepinephrine (Russell & Dart, 1986) or the non-selective β-adrenergic agonist isoproteronol (Autenrieth, Surawicz, Kuo, & Arita, 1975) in dogs. In addition, TWA decrease was seen after subcutaneous or intramuscular administration of epinephrine (Hartwell, Burrett, Graybiel, & White, 1942; Katz, Hamburger, & Lev, 1932; Levine, Ernstene, & Jacobson, 1930), and after administration of non-selective (isoproterenol) and cardioselective β-adrenergic agonists (metopolol and alprenolol) in man (Contrada et al., 1989; Contrada et al., 1991; Guazzi et al., 1975; Rau, 1991).

Importantly, such functional TWA decreases could be reversed by β-blockade with propanolol (Contrada et al., 1989; Furberg, 1967; Furberg, 1968; Guazzi et al., 1975; Noskowicz et al., 1968). Additionally, TWA has been shown to be a useful indicator of cardiac SNS activity during laboratory testing of stress reactivity (Furedy et al., 1984; Furedy et al., 1986; Furedy et al., 1996; Heslegrave et al., 1979; Matyas, 1976; Matyas et al., 1976; Scher et al., 1984) since the stress-induced TWA decreases could be blocked by β-adrenergic antagonists (Contrada et al., 1989; Rau, 1991).

The initial enthusiasm for the TWA was tempered when Bunnell (Bunnell, 1980) showed only modest correlation (mean r= .41) between TWA decreases and other, at the time, accepted sympathetic measures (PTT, carotid dP/dt, HR). However, Furedy et al (Furedy & Heslegrave, 1983) rightfully noted that the weak correlations of TWA decreases to criterion measures merely showed that it is not a *perfect* index of SNS activity, and that other SNS indices also did not correlate perfectly amongst themselves. A more serious concern about the TWA is how 'purely' it reflects SNS activity. Increases in SNS activity are often accompanied by increases in heart rate and decreases in vagal activity. Increased heart rate leads to a shortening of the interbeat interval (IBI) which could decrease TWA simply by reducing the rise time of the T-wave. This means that decreases in vagal activity, by increasing heart rate, could also contribute to a reduction in TWA (Schwartz et al., 1983; Weiss et al., 1980).

Dauchot and Gravenstein (Dauchot et al., 1971) and Annila et al. (Annila et al., 1993) indeed found that an acetylcholinergic antagonist (atropine) led to a decrease in the TWA. Contrada et al (Contrada et al., 1989; Contrada et al., 1991) further reported a sudden paradoxical TWA increase during very high doses of isoproterenol infusion and reasoned that (baroreflex-induced) increases in vagal activity might have caused the increase inthe TWA. These findings suggest that changes in cardiac vagal activity also affect the TWA, which would invalidate it as a 'pure' SNS measure. Furedy et al. (1996) have argued that the effects of cardiac vagal activity may partly act by modulating SNS activity through the mechanism of accentuated antagonism (Furedy et al., 1996; Levy, 1977; Levy, Ng, Lipman, & Zieske, 1966; Levy & Zieske, 1969). If TWA reflects ventricular depolarization, it is

unclear how this would work, because accentuated antagonism requires sympathetic and parasysmpathetic nerves to converge on the same neuromuscular synapses, whereas the human ventricle is not innervated by the parasympathetic nervous system. Nonetheless, various studies reported an enhanced decrease of TWA by isoproterenol after atropine infusion compared to isoproterenol alone (Fukudo et al., 1992; Stratton, Pfeifer, & Halter, 1987).

One way to resolve the role of PNS activity in the TWA is to coregister TWA with heart rate variability in the respiratory frequency (RSA) which has been proposed as a proxy for cardiac PNS activity (Berntson et al., 1997; Task Force of the European Society of Cardiology the North American Society of Pacing, 1996). A number of studies reported a TWA decrease in parallel to vagal withdrawal as indexed by a decrease in RSA (Kreibig, Wilhelm, Roth, & Gross, 2007; Palomba, Sarlo, Angrilli, Mini, & Stegagno, 2000; Pan & Li, 2007; Roth et al., 1992). However, these studies employed TWA as an accepted SNS measure and did not correlate TWA to other measures of SNS activity nor directly to RSA. Only Kline et al (Kline et al., 1998) directly correlated RSA and TWA across various stress tasks (Valsava`s maneuver, serial subtraction and cold pressor) in 20 participants and found that after controlling for between-person variance, RSA did not contribute to TWA.

Taken together, the past studies on the TWA are overall suggestive of a role of the SNS in this measure, whereas the additional role of the PNS remains unclear. It is possible that the relative popularity of the PEP rather than poor validity of the TWA has led to its demise in the past two decades. As stated before, the automated scoring of the PEP from the ECG and ICG has a number of practical setbacks and requires visual inspection that can become prohibitive when PEP is used in long term recordings in epidemiology-sized samples. This breathes new life into the attractiveness of the TWA that might not suffer from this setback as it only requires detection of clear landmarks in the ECG. Here we report on 24-hour ambulatory monitoring of the TWA in a sample of 564 healthy adults. First, we assess the feasibility of automated scoring of ECG landmarks required for the TWA using large scale ensemble averaging of the ECG. Large scale ensemble averaging has been successfully applied to the ambulatory impedance cardiogram (ICG) but not yet to the ambulatory ECG. Second, to compare to previous literature, we will score the TWA in an ensemble averaged ECG across a 4 minute resting baseline and 4 minute stress task. These standardized conditions were embedded within the larger 24-hour recording. Third, in the ambulatory recordingwe will test whether the TWA varies in a predictable way within a wake/sleep cycle and across different levels of physical activity. Finally, we will test whether the changes in TWA within a participant across the 24-hour recording are correlated with the changes in PEP. We hypothesize that the TWA will vary in a predictable way between rest and stress conditions and across arousal and physical activity level. We further hypothesize that within-participant changes in TWA will show significant correspondence with changes in PEP, and that these changes survive correction for parallel changes in RSA and IBI. In short, we expect to provide evidence that the TWA can provide meaningful information on SNS activity in ambulatory recordings.

Methods

Participants

Participants were all registered in the Netherlands Twin Register (NTR) and had previously participated in a larger biobank project (Willemsen et al., 2010). A priori reasons for exclusion were participation in an earlier ambulatory recording study (Kupper et al., 2005b; Kupper et al., 2006), heart transplantation, presence of a pacemaker and known ischemic heart disease, congestive heart

failure, or diabetic neuropathy and pregnancy. Ambulatory cardiovascular recordings of 582 participants were collected of which 11 recordings had either a missing or noisy ECG or thorax impedance signal due to equipment failure, and were therefore excluded from the analysis. Seven participants using beta-blockers were excluded from the analysis. The final sample consisted of 277 identical twins (97 men), 234 fraternal twins (96 men), and 53 singleton siblings (21 men) from 297 families. Mean age was 36.9 years (SD 5.4). Zygosity of the twins was determined by DNA typing for 98.6% of the same-sex twin pairs. For the other same-sex pairs, zygosity was based on survey questions on physical similarity and the frequency of confusion of the twins by parents, other family members, and strangers. Agreement between zygosity based on these items and zygosity based on DNA was 96.1% (Willemsen et al., 2013). The Medical Ethical Committee of the VU University Medical Centre Amsterdam approved of the study protocol and all participants gave written consent before entering the study. Participants received a payment of 10 Euros and an annotated review of their ambulatory ECG recording.

Procedure

A detailed description of the general ambulatory monitoring procedure has been provided elsewhere (Kupper et al., 2006). Briefly, all participants were visited in the morning at home or at the work location when this was deemed more convenient. They were fitted with the VU University Ambulatory Monitoring System (VU-AMS) device that records the electrocardiogram (ECG) and the impedance cardiogram (ICG) continuously during a 24-h period (daytime and sleep) through seven disposable, pregelled Ag/AgCl electrodes (de Geus et al., 1995; van Dijk et al., 2013; Willemsen et al., 1996). After visually establishing proper signal quality the recording was started and participants were first interviewed on health, medication, lifestyle and socioeconomic and demographic information after which they filled out a questionnaire on psychological wellbeing. The questionnaire lasted on average 10 minutes and was completed while quietly sitting in a secluded part of the house/work area. The last 4 minutes of this quiet sitting period functioned as a baseline. Next participants were instructed to execute a 2-minute serial subtraction task and a 2-minute Stroop ColorWord conflict task. After a final equipment check the experimenter then departed and participants were left to their daily routine until the next morning when they were asked to remove the VU-AMS device and cables and to mail it all back to the experimenter in a prepared return envelope with a special protective layer. During the daytime and the evening before bedtime participants were asked to give a chronological account of posture, physical activity, physical load, location, and social situation every 60-min period. Participants were asked to refrain from heavy exercise during the recording day.

Physiological recording

The electrocardiogram (ECG) and the impedance cardiogram (ICG) were recorded continuously with the 7-lead version of the VU-AMS device (version 5fs, VU University Amsterdam, www.vu-ams.nl). Cleaning of the skin with alcohol before electrode application ensured that electrode resistance was kept low. ECG electrodes were carefully placed according to a standard protocol to obtain a lead II derivation, which yields the most prominent R-wave peak as well as a clear T-wave amplitude. The first ECG electrode (V-) is placed slightly below the right collar bone 4 cm to the right of the sternum. The second ECG electrode (V+) is placed at the apex of the heart over the ninth rib on the left lateral margin of the chest approximately at the level of the processus

xiphodius. The third ECG electrode (GND) is a ground electrode and is placed on the right side, between the lower two ribs at the right abdomen. The first ICG measuring electrode (V1) is placed at the top end of the sternum, between the tips of the collar bones. The second ICG measuring electrode (V2) is placed at the xiphoid complex of the sternum. The two current electrodes are placed on the back: I- on the spine over the cervical vertebra C4, at least 3 cm (1") above the ICG measuring electrode V-, and I+ between thoracic vertebrates T8 and T9 on the spine, at least 3 cm (1") below the ICG measuring electrode V2.

Data reduction

The ECG and ICG signals were imported into the VU-DAMS software (version 2.3, VU University Amsterdam, www.vu-ams.nl). After automated detection of bad ECG signal fragments (artefacts), R-wave peak detection was done using a modified version of the algorithm by Christov (Christov, 2004). From the R-wave peaks, an interbeat interval (IBI, ms) time series is constructed that was visually displayed for interactive correction of missed or incorrect R-wave peaks. This correction takes on average 2 minutes per 24 hour recording.

Large scale ensemble averaging

Using the activity diary entries in combination with a visual display of an in-built vertical accelerometer signal, the entire 24-h recording was divided into fixed periods coded for posture (e.g. lying, sitting, standing), ongoing activity (e.g. desk work, eating/drinking, meetings, watching TV), physical activity (non, light, medium, and heavy), location (e.g. work, home, outside), and social situation (e.g. alone, with colleagues, with friends). The ECG and ICG signals were ensemble averaged across these periods, time locking both signals to the R-wave peaks. This reduced the data set to, on average, 49 (range: 17-88) ensemble averaged ECG and ICG traces per participant with an average length of 27 minutes (SD= 19). These had less than 0.5% ECG artefacts. An example of a large scale ensemble-averaged ICG and ECG is presented in figure 1. In these ensemble averages, the mean Q-wave onset (Qonset), R-wave onset (Ronset), S-wave peak (Spoint), S-wave offset (Soffset), T-wave peak (Ttop), T-wave offset (Toffset) in the ECG and the B-point, dZ/dt min (C-point), and incisura (X-point) in the ICG were automatically detected and presented for interactive visual scoring. During interactive scoring, undetectable points were deleted which effectively sets them to missing per complex. Interactive ECG and ICG scoring took on average 15 minutes per 24 hour recording, the bulk of which was spent on the ICG. In some participants, Qonset and/or Ronset tended to be systematically missing, which has been reported before (Lozano et al., 2007). In these participants, Qonset and/or Ronset were set to missing. The ICG points, that caused the largest ambiguity, were scored by two raters as recommended (van Lien et al., 2013), and a third rater arbitrated when no consensus could be reached. To remove outliers, Z- scores were calculated for the timing of Qonset, Ronset, Spoint, Soffset, Ttop, Toffset, B-point, C-point and X-point across all participants and conditions and, per point, the top and bottom two percent of Z values of each of these parameters were set to missing.

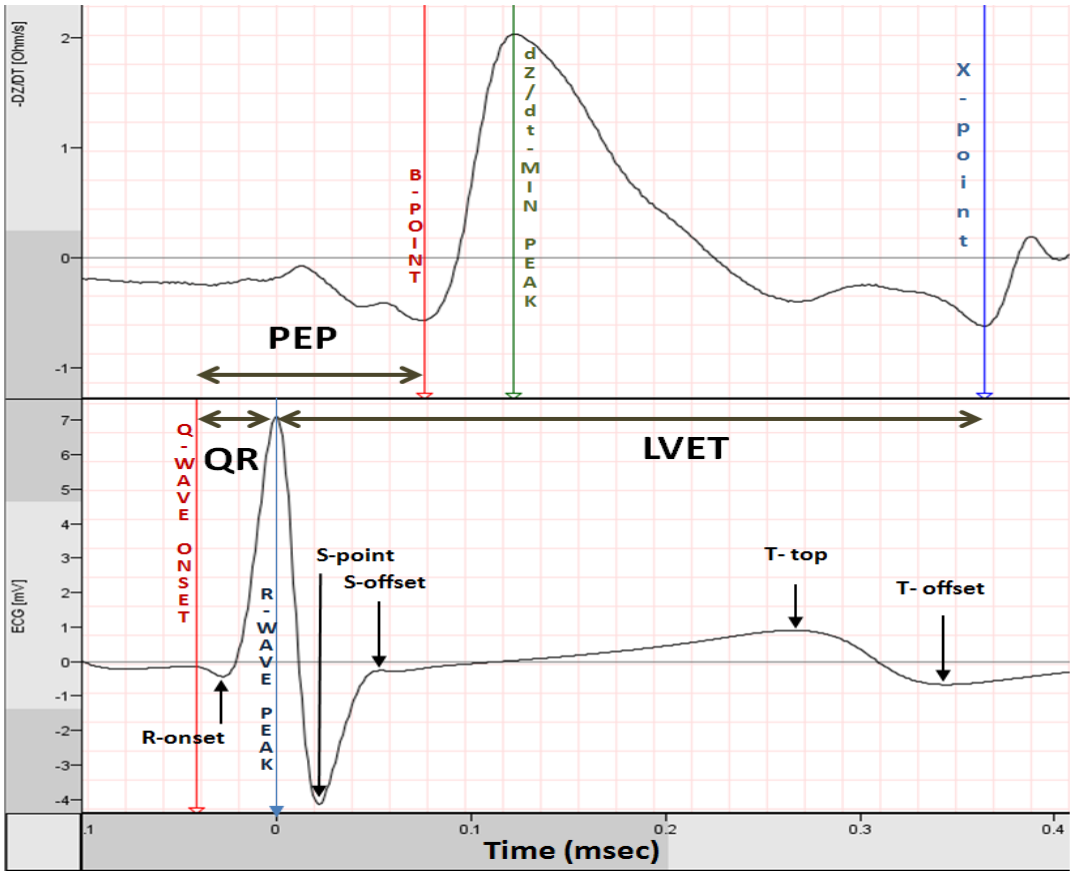


Figure 1. *The impedance cardiogram (top) and the electrocardiogram (bottom) with the four landmarks defining the PEP (Q-wave onset to B-point) and TWA (T-wave top and T-wave offset).*

Cardiac time intervals

From the ECG landmarks indicated in figure 1, we computed the Qonset to R-wave onset (QRonset) interval, the Q-wave onset to R-wave peak (QR), the Ronset to R-wave peak (RonsetR) interval, and the QRS, RS, QT, and ST intervals in milliseconds. From the ICG landmarks we computed the left ventricular ejection time (LVET) as the interval between the B-point and the X-point, and by combining with the ECG, the R-wave peak to B-point (RB) interval and the PEP as the interval from the Q-wave onset in the ECG to the B-point in the ICG signal. As indicated before, a number of participants had no detectable Q-wave onset. For these participants we used the procedure typically used to estimate PEP from the R-wave peak only (Lozano et al., 2007; van Lien et al., 2013) but modified it to exploit the large dataset available. If only Qonset was missing but Ronset was present, PEP was estimated by adding the grand average QRonset interval across 531 participants with valid QRonset (12.6 ms) to the sum of the individual participants' RonsetR + RB intervals. If Ronset was additionally missing we defaulted to the usual estimation of PEP by adding the grand average QR interval across 455 participants with valid QR (43.7 ms) to the individual participants' RB interval.

T-wave amplitude

We computed the TWA (in µV) using different baseline values in the ECG. First we calculated TWA_zero as the difference between the ECG value at the Ttop and the isoelectric level represented by the zero µV line in the ECG. Because ECG signals are prone to drifting due to movement, especially

during ambulatory recordings, the TWA might be under- or overestimated in some participants by using this static baseline. We therefore secondly calculated TWA_Toffset as the difference between the ECG value at the Ttop and the ECG value at the T-offset. We deviated from the usual approach that averages the entire T-offset to P-onset interval with the intention to obtain a stable baseline. We found that at T-offset there is indeed negligible electrical activity, whereas activity is not always low in the entire T-offset to P-onset interval. In addition, we note that the T-offset is more reliably detected than P-onset.

RSA

Combining the ECG with the respiration signal extracted from the thorax impedance signal (dZ), the 'peak–valley' RSA method was used to asses vagal chronotropic effects (de Geus et al., 1995; Grossman et al., 1990; Grossman et al., 1986). In this method, RSA is scored from the combined respiration and IBI time series by detecting the shortest IBI during inspiration and the longest IBI during expiration on a breath-to-breath basis according to the procedures detailed elsewhere (de Geus et al., 1995; Houtveen et al., 2005; van Lien et al., 2011). Breathing cycles that showed irregularities like gasps, breath holding, coughing etc., were considered invalid and were removed from further processing. If no shortest or longest IBI could be detected in inspiration and experiation respectively, the breath was either set to missing or to zero when computing the condition average for RSA. Similar results were found for RSA computed either way and we employed only one (breaths set to missing) in further statistical analyses..

Statistical Analyses

For statistical analyses, the total recording, consisting of the standardized and the ambulatory part, was first converted into six experimental conditions. The first two experimental conditions were the baseline of quietly sitting and the mean of the two stress tasks. The next four conditions came from the ambulatory part of the recording. Because ambulatory activities were not standardized and could differ per participant, all coded periods were aggregated into four comparable conditions that were determined by the posture and the level of physical activity (PA) of the participant: sleep, sitting activities during the day, light physical activity in awake time, moderate physical activity in awake time.

To test content validity of TWA, a mixed model ANOVA with age, sex and BMI as covariates and family as a random factor and baseline versus stress during the experimental part of the recording as the fixed factor was first applied. We expected TWA to be systematically decreased in response to mental stress. Next, a similar ANOVA was applied to the four ambulatory conditions, where we expect the TWA to significantly decrease with stepwise increases in SNS activity from sleep to sitting awake to physical activity.

Criterion validity of TWA in the experimental setting was then tested by computing the correlations between the participants' TWA and PEP reactivity to the stress tasks. To account for the effects of concurrent decreases in RSA and IBI, we also computed the partial correlation between TWA and PEP reactivity, with the effect of IBI, RSA, or joint RSA/IBI reactivity partialled out. Criterion validity of TWA in the naturalistic setting was assessed by computing the within-participant correlations between TWA and PEP across the day on all coded ambulatory periods available for a particular participant. The mean number of coded ambulatory periods was 49. We selected only participants that had complete TWA, PEP, RSA and IBI data in at least 15 coded ambulatory periods

with a minimum of two periods in at least three of the four conditions (sleep, sitting, light and moderate activity). We next plotted the distribution of these within-participant correlations and tested whether they deviate from zero (the expected value if TWA and PEP are not systematically correlated during the recording day).

To take into account parallel changes in vagal activity and heart rate, we recomputed the within-participant TWA-PEP correlations as a partial correlation using the IBI, RSA, or IBI and RSA values during the condition as a covariate. We compared the distributions of the uncorrected and partial TWA-PEP correlations using a test of the difference in correlation coefficients based on the Fisher Z transformation (Preacher, 2002).

Because the TWA-PEP comparison could suffer from poor quality of scoring of the criterion variable (PEP) in some participants, the analyses were repeated in the 96 participants for whom the raters expressed the highest confidence in ICG B-point scoring quality.

Results

Cardiac time intervals from the Large Scale Enembled Averged ECG

The means and standard deviations of the intervals derived from the large scale ensemble averaged ECG are shown per condition in table1. In general the relevant landmarks could be clearly detected in the ensemble averaged ECGs, but as noted before Qonset and Ronset are not always detectable. Overall, the Qonset was missing and estimated in 19.3% of the participants (109 of 564) and Ronset was additionally missing and estimated in 5.8 %, (33 of 564) and this was not different across the ambulatory conditions.

Mixed ANOVA analysis with correction for family relatedness, sex, age and BMI showed a significant main effect of ambulatory condition on most intervals derived from the ensemble averaged ECG; QR ($F(3, 1459)= 32.5$, $P<.001$), RonsetR ($F(3, 1747)=39.7$, $P<.001$), QRS, ($F(3, 1774)=20.6$, $P<.001$), RS ($F(3,1855)= 58.5$, $P<.001$), QT ($F(3, 1719)=1649$, $P<.001$), ST ($F(3, 1674)= 1459$, $P<.001$), RB($F(3, 1893)=132.3$, $P<.001$). Post hoc analyses of the condition effects revealed that the QT, ST, RS and RB intervals showed the expected significant stepwise decrease with increased arousal and physical activity. The effect of condition on QR, RonsetR, QRS, and Qronset was negligible and completely driven by the discrepancy between sleep and physically active periods.

From the covariates considered, sex had the most effect on these intervals. Female participants had shorter Qonset, Ronset, RS and QRS durations and longer ST and QT intervals than males (P's < 0.001). The Qonset, Ronset, QRonset and QRS duration was longer in older participants (p's <0.001). A greater BMI was significantly associated with shorter Qonset ($r=-.111$, $p<.001$), Ronset ($r=-.148$, $p<.001$), ST ($r=-.074$, $p<.001$).

Table 1. *Means and standard deviations (msec) for the intervals derived from the large scale ensemble averaged ECG during 24-hour ambulatory recording.*

Condition		QR	RonsetR	QRS	Qronset	QT	ST	RS	RB
Sleep	N	*427*	*508*	*508*	*416*	*483*	*469*	*533*	*539*
	Mean (SD)	44(4)	32(3)	93(10)	13(4)	411(21)	318(23)	25(3)	62(14)
Sitting activities	N	*443*	*521*	*529*	*432*	*519*	*504*	*552*	*559*
	Mean (SD)	44(4)	31(3)	[a] 91(9)	13(4)	[a]377(22)	[a] 285(23)	[a] 24(3)	[a] 58(16)
Light PA	N	*430*	*504*	*511*	*417*	*500*	*487*	*535*	*542*
	Mean (SD)	43(4)	31(3)	91(9)	13(4)	[b] 367(21)	[b] 276(22)	24(2)	[b] 55(17)
Moderate PA	N	*443*	*521*	*528*	*433*	*514*	*504*	*547*	*557*
	Mean (SD)	43(4)	31(3)	91(10)	13(4)	[c] 356(18)	[c] 264(19)	24(2)	[c] 49(16)

Sample sizes (N) vary depending on the number of participants in which the particular ECG landmark could be scored or in which the condition was missing.
[a] *Significantly different compared to sleep, p <. 05.*
[b] *Significantly different compared to sitting activities, p <. 05.*
[c] *Significantly different compared to light physical activity, p <. 05.*

Response of the TWA to a standardized stressor

The mean values and standard deviation for IBI, LVET, PEP, RSA, TWA_Toffset, and TWA_Zero during baseline and stress task are presented in table 2. Mixed model analyses with correction for family, age, sex and BMI showed a significant main effect of the stress task on IBI ($F(1,777)=180.8$, $p<.001$), LVET ($F(1, 722)=17.0$ $p<.001$), PEP ($F(1,674)=19.4$, $p<.001$), RSA ($F(1, 718)=4.2$, $p<.005$), TWA_Toffset ($F(1,639)=28.3$, $p<.001$) and TWA_Zero ($F(1,659)=41.8$, $p<.001$). Post-hoc analyses on the effect of the mental stress tasks generally showed the expected effect of our manipulation on the ANS measures. The IBI, LVET, PEP, TWA_Toffset and TWA_Zero decreased significantly during the stress task, although RSA did not.

Table 2. *Means and standard deviations for the variables derived from the ECG and ICG during the standardizd baseline and stress task conditions.*

	Baseline	Stress task	N baseline/ N stress task	Reactivity (Δ)
IBI	834 (123)	769 (113)	524/542	*-65
LVET	284 (34)	277(36)	491/510	*-7
PEP	107 (19)	103 (19)	472/488	*-4
RSA	58 (24)	60 (22)	498/530	*2
TWA_Toffset	1.480 (0.548)	1.374 (0.539)	452/474	*-.106
TWA_Zero	.964 (.32)	.877 (.31)	460/485	*-.087

**= Significantly different during stress compared to baseline, p <.05.*

Response of the TWA to increase levels of arousal and physical activity in a naturalistic setting

The mean values and standard deviation for IBI, LVET, PEP, TWA_Toffset, and TWA_Zeroduring the ambulatory recording are presented in table 3. Mixed ANOVA analysis with correction for family relatedness, sex, age and BMI showed a significant main effect of physical

activity on IBI (F3,1923)= 2138, p<.001), LVET (F(3,1907)=691, p<.001), PEP (3, 1893) = 148, p<.001), RSA (F(3,1874)=192, P<.001), TWA_Toffset (F(3, 1711)= 180, P<.001) and TWA_zero (F(3,1759) = 109, P<.001). The overall correlation between TWA_Toffset and TWA_Zero was .97 (p < .001) but as expected the TWA_Toffset, that uses a dynamic baseline, performed slightly better than the TWA_Zero measure, that uses a static baseline. Post hoc analyses of the ambulatory condition effects revealed the expected significant stepwise ordinal decrease of IBI, LVET, PEP, RSA and TWA_Toffset with increased arousal and physical activity. TWA_Zero also showed a significant stepwise decrease across sitting to increased levels of physical activity but in contrast to TWA_Toffset failed to differentiate between sleep and sitting activities. TWA_Toffset was selected as the single TWA measure to be used in further analyses.

Female participants had lower IBI, LVET, RSA and TWA_Toffset values but a longer PEP (p's<.001). RSA was significantly lower in older participants and a larger BMI was associated with shorter IBI (r = -.077, p< .001) and PEP (r = -.153, p< .<001), lower RSA(r = -.113, p< .001) and TWA_Toffset (r = -.113, p< .001) but a longer LVET (r = .074, p< .001).

Table 3. *Means and standard deviations for the variables derived from the ECG and ICG during the ambulatory conditions.*

	N (Range)	Sleep	Sitting activities	Light PA	Moderate PA
IBI	549 - 564	983 (132)	[a] 815 (101)	[b] 743 (93)	[c] 660 (77)
LVET	545 - 561	315 (28)	[a] 281 (29)	[b] 270 (31)	[c] 258 (50)
PEP	539 - 559	106 (14)	[a] 102 (16)	[b] 99 (16)	[c] 92 (16)
RSA	539- 555	63 (24)	[a] 58(20)	[b] 54 (18)	[c] 44 (14)
TWA_Toffset	492 - 516	1.46 (0.55)	[a] 1.38 (0.52)	[b] 1.19 (0.51)	[c] 1.07 (0.49)
TWA_Zero	504 - 525	0.92 (0.32)	0.90 (0.30)	[b] 0.80 (0.29)	[c] 0.75 (0.29)

[a] *Significantly different compared to sleep, p <. 05.*
[b] *Significantly different compared to sitting activities, p <. 05.*
[c] *Significantly different compared to light physical activity, p <. 05.*

Criterion validity of the TWA during the mental stress task

Significant correlations of the reactivity scores were found for TWA_Toffset with PEP reactivity (r=.406, p<.001) and IBI reactivity (r=.694, p<.001). TWA_Toffset decreases were also significantly correlated with RSA reactivity, albeit more modestly (r=.175, p<.001). As expected, IBI reactivity was correlated with both RSA (r=.228, p<.001) and PEP (r=.435, p<.001) reactivity. The average correlation between PEP and TWA_Toffset reactivity remained significant after controlling for RSA reactivity, as shown in the partial correlation coefficient (r_{part} = .417 , p< .001) and after controlling for IBI (r_{part} = .171 , p< .001), and joint IBI/RSA (rpart= .173, p<.001).

Criterion validity of the TWA in a naturalistic setting

Valid levels of TWA_Toffset, PEP, RSA and IBI could be obtained in more than 15 ambulatory conditions in 447 participans. The distribution of the within-participant correlations for TWA_Toffset and PEP for these participants is plotted in the upper panel in figure 2. This panel also gives the correlations between TWA_Toffset and RSA and TWA-Toffset and IBI. The TWA_Toffset showed a mean within-participant correlation of .35 with PEP (p<.001). The number of participants that had a positive within-participant correlation between TWA and PEP meeting a nominal p=0.05 significance

threshold of .11 was 261. In 186 (41.6%) participants no significant within-participant correlation between PEP and TWA was found in the expected direction. The mean within-participant correlation of TWA_Toffset with IBI (.61) and RSA (.38) was also significant (p<.001).

The lower panel in figure 2 shows the partial correlation between PEP and TWA_Toffset after partialling out IBI, RSA and joint IBI/RSA. The mean within-participant correlation between PEP and TWA_Toffset remained significant after controlling for within-participant changes in IBI (r_{part}=.14), RSA (r_{part}=.29) or joint IBI/RSA (r_{part}=.13). The drop in the mean correlation between TWA and PEP before and after taking RSA into account was not significant (p=0.31), but taking IBI (p<.001) or taking both IBI and vagal activity (p<.001) into account significantly reduced the PEP TWA_Toffset correlation.

Optimal ICG signal recording and PEP scoring quality

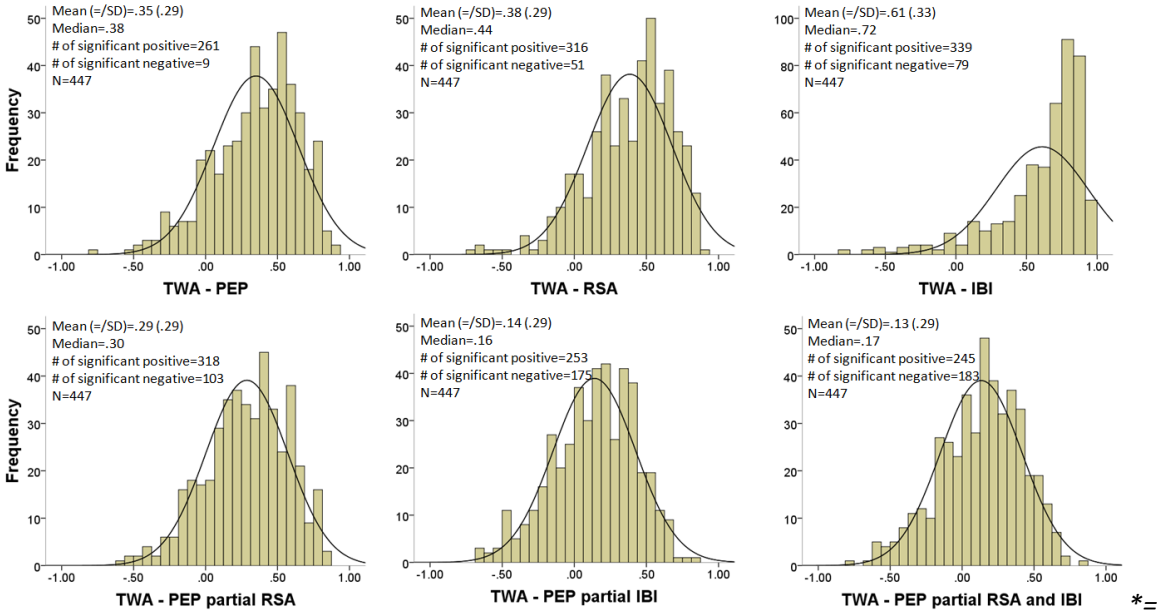
Poor ICG signal quality could have led to low quality PEP scoring and thus an underestimation of the PEP TWA_Toffset correlation. Repeating the analyses in the 96 participants with the highest quality of the ICG B-point scoring yielded higher within-participant TWA_Toffset – PEP correlations although a similar distribution was seen as when including all participants (upper panel in figure 3). The TWA_Toffset showed a mean within-participant correlation of .43 with PEP. The number of participants that had a positive within-participant correlation between TWA_Toffset and PEP meeting a nominal p=0.05 significance threshold of .24 was 76. In 20 participants no significant correlation between PEP and TWA_Toffset was found in the expected direction. The mean within-participant correlation of TWA_Toffset with IBI (.71) and RSA (.43) was also significant (p<.001).

The lower panel in figure 3 shows the partial correlation between PEP and TWA_Toffset after partialling out IBI, RSA and joint IBI/RSA. The mean within-participant correlation between PEP and TWA_Toffset (r=.43) remained significant after controlling for within-participant changes in IBI (r_{part}=.25), RSA (r_{part}=.34) or joint IBI/RSA (r_{part}=.15). The drop in the mean correlation between TWA_Toffset and PEP before and after taking RSA into account was not significant (p=0.47). Taking IBI and joint IBI/ RSA into account again reduced the PEP TWA_Toffset correlation, although this did not reach significance (p's > 0.06).

Discussion

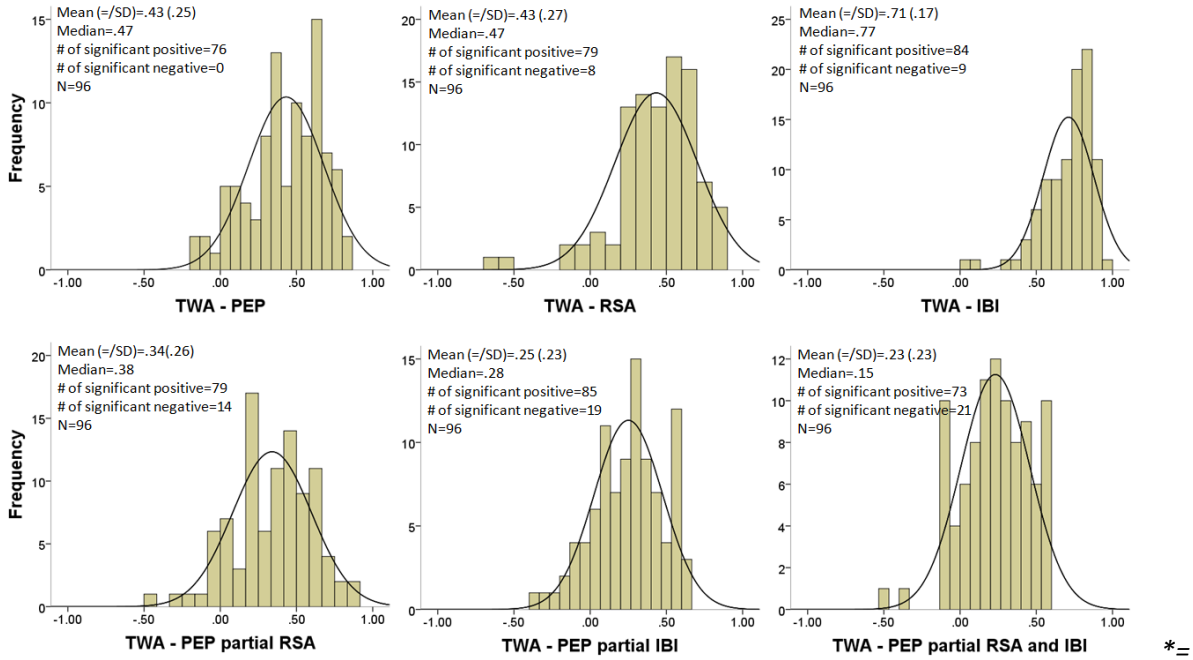
Ambulatory assessment of fluctuations in SNS functioning could greatly help epidemiologists understand the link between psychosocial stress and unfavorable physical and mental health outcomes. The PEP is currently the measure of choice to assess SNS activity in ambulatory studies. The aim of the present study was to test the validity of the ECG T-wave amplitude as an alternative index of cardiac sympathetic nerve system activity. The ambulatory TWA can be derived from three spot electrodes that only present a minimal burden to the participant. If obtained from the ensemble-averaged ECG, the TWA requires less laborious data scoring and may prove more amenable to automation than PEP scoring. PEP scoring requires a detectable Q-onset, which is missing in about 20% of participants, and visual inspection of the ensemble-averaged ICG waveform, which is occasionally ambiguous.

Figure 2. *Distribution of the within-participant correlations of TWA-PEP, TWA-IBI and TWA-RSA in the full sample. TWA-PEP correlations are re-plotted with IBI, RSA and joint IBI/RSA partialled out.*



Significant at p <.05.

Figure 3. *Distribution of the within-participant correlations of TWA-PEP, TWA-IBI and TWA-RSA in the sample only including participants with unambiguous ICG B-point scoring. TWA-PEP correlations are re-plotted with IBI, RSA and joint IBI/RSA partialled out.*



Significant at p <.05.

In our large sample of healthy adult participants, the TWA could be reliably extracted from the ensemble-averaged ECG in over 90% of the participants. It showed a clear decrease in response to a mental stress task in line with earlier research on the TWA (Contrada et al., 1989; Furedy et al.,

1984; Furedy et al., 1996; Heslegrave et al., 1979; Scher, Hartman, Furedy, & Heslegrave, 1986). TWA also showed a stepwise decrease from nighttime sleep to daytime sitting to more physically active behaviors during an ambulatory 24 hour recording, echoing the expected pattern of SNS activity across these various conditions. In addition, within-participant changes in TWA across the standardized as well as the unstandardized naturalistic conditions were correlated with parallel changes in the PEP, which we used as a criterion variable. However, these within-participant correlations were significant in 75% of the participants only, even when selecting the subset of participants with the best quality ICG data. A first conclusion of our data, therefore, is that the ambulatory TWA can be a valuable addition to the epidemiologist's psychophysiology toolbox, but that caution is needed to interpret changes in TWA at the level of a single individual. Below we discuss their validity as a 'pure' index of changes in SNS activity in more detail.

The literature validating the TWA as an SNS index in standardized laboratory conditions precedes the first application of regional cardiac NE spillover (Eisenhofer, Lambie, & Johnson, 1985; Esler et al., 1988) which is probably the only true golden standard of cardiac SNS activity. This is nontrivial, because low or absent correlation to cardiac NE spillover has been the major reason to distrust the LF/HF ratio as an index of SNS activity (Goedhart et al., 2008b). It could be reasonably argued that true validity of the TWA likewise can only be assessed by a comparison of TWA to NE spillover or direct SNS nerve recording (Goedhart et al., 2008b; Grassi et al., 1999). A large amount of pharmacological studies nonetheless bodes well for TWA. There is a clear pattern of decreased TWA with β-adrenergic agonists (Barger et al., 1961; Contrada et al., 1989; Hartwell et al., 1942; Katz et al., 1932; Levine et al., 1930; Russell et al., 1986) and the effect is attenuated or disappears with β-adrenergic antagonists (Contrada et al., 1989; Furberg, 1967; Furberg, 1968; Guazzi et al., 1975; Noskowicz et al., 1968; Rau, 1991), although not all studies have been able to reproduce this pattern (Contrada et al., 1991; Russell et al., 1986; Schwartz, Stone, & Brown, 1976; Taggart et al., 1979).

Admittedly, pharmacological studies have disadvantages in that they engage cardiac and vascular reflex regulation which may prominently include cardiac vagal activity. There has been some debate on whether vagal activity itself might cause a decrease in TWA which would invalidate it as an exclusive SNS index. If the assumption is correct that the TWA reflects ventricular repolarization, the theoretical rationale for such a vagal effect is not strong as the human ventricle is not vagally innervated. Autonomic effects on contractility, for instance, are driven entirely by adrenergic innervation. This provides the basis for using PEP as a "pure" index of cardiac SNS activity. To test for the independent contribution of changes in SNS activity to increases and decreases in TWA our analyses partialled out the parallel changes in RSA. Although RSA was significantly correlated with the TWA, this did not lead to a significant reduction in the within-participant correlation between PEP and TWA reactivity. This suggests that PNS activity is not a confounder of the TWA-PEP correlation. Only when changes in IBI were also partialled out, a decrease in the TWA-PEP correlation was seen. In general, the correlation between changes in TWA and IBI was strong, and it remained significant after partialling out PEP ((r_{part}=.55 in ful set ; r_{part}=.62 in best quality subset). This suggests that the TWA is sensitive to the shortening of the cardiac cycle, independent of SNS activity. However, even after correcting for concurrent changes in IBI a significant relationship between PEP and TWA reactivity was found.

A necessary limitation of this study is the use of changes in the PEP as an index of SNS activity to establish criterion validity of TWA. Although the PEP is the only available measure of cardiac SNS activity in ambulatory recording with established validity, it must also be recognized that the PEP is sensitive to postural or physical activity driven changes in preload that influence contractility

independent of the SNS through the Frank-Starling mechanism (Houtveen et al., 2005). In addition, the PEP is sensitive to changes in mean aortic pressure that can elongate PEP even under conditions of increased SNS activity, as is seen during exposure to cold or static muscle work (de Geus et al., 1993; Krzeminski et al., 2000). Changes in temperature, posture and static or dynamic exercise activities are a necessary element of naturalistic ambulatory monitoring. The PEP itself, therefore, will imperfectly correlate with SNS activity. In ambulatory settings it is the best possible criterion measure to compare to TWA, but certainly not a golden standard.

This limitation is balanced by a number of strengths of this study. First, we had a large sample size allowing us to provide a normative dataset for the main ECG intervals including the QonsetR and RonsetR intervals which we used to improve estimation of the PEP in participants where these landmarks were difficult to score. Secondly, by grace of the large dataset we could repeat our analyses on a high quality ICG data set to avoid ambiguity in the PEP scoring as a potential source of poor cross-measure correlation. Thirdly, although our major aim was to show validity in a naturalistic ambulatory setting, the addition of the standardized stress testing allowed us to also investigate TWA in the same participants independent of confounding posture and activity effects, diurnal variations, temperature fluctuations, ambient noise, ingestive behaviors, etc.

In conclusion, we find support for the usefulness of ensemble-averaged ECG derived TWA in epidemiological research to estimate changes in 24 hour SNS activity. Validity is not sufficiently strong to recommend replacement of the PEP by the TWA; instead recording and analysis of both measures seems prudent and feasible. Prospective follow-up of the physical and mental health of our participants, who are part of a nation-wide longitudinal study, must resolve the clinical value of these ambulatory SNS measures, separately and in combination.