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 3

Estimated preejection period (PEP) based on the detection of the R-wave and dZ/dt-min peaks does not adequately reflect the actual PEP across a wide range of laboratory and ambulatory conditions.

René van Lien, Nienke Schutte , Jan H. Meijer, and Eco J.C. de Geus

This chapter was published in:

International Journal of Psychophysiology. 2013; 87, 60–69

Abstract

The current study evaluates the validity of the PEP computed from a fixed value for the Q-wave onset to R-wave peak (QR) interval and an R-wave peak to B-point (RB) interval that is estimated from the R-peak to dZ/dt-min peak (ISTI) interval. Ninety-one participants participated in a 90 minute laboratory experiment in which a variety of often employed physical and mental stressors were presented and 31 further participants participated in a structured 2 hour ambulatory recording in which they partook in natural activities that induced large variation in posture and physical activity. PEP, QR interval, and ISTI were scored and rigorously checked by interactive inspection. Across the very diverse laboratory and ambulatory conditions the QR interval could be approximated by a fixed interval of 40 ms but 95% confidence intervals were large (25.5 to 54.5 ms). Multilevel analysis showed that 79% to 81% of the within and between-participant variation in the RB interval could be predicted by the ISTI with a simple linear regression equation. However, the optimal intercept and slope values in this equation varied significantly across participants and study setting. Bland Altman plots revealed a large discrepancy between the estimated PEP using the R-wave peak and dZ/dt-min peak and the actual PEP based on the Q-wave onset and B-point. We conclude that the PEP estimated from a fixed QR interval and the ISTI could be a useful addition to the psychophysiologist's toolbox, but that it cannot replace the actual PEP to index cardiac sympathetic control.

Introduction

Functional disturbances of the autonomic nervous system have been frequently linked to several diseases (Eckberg, Drabinsky, & Braunwald, 1971; Esler et al., 2000b; Esler, Lambert, & Jennings, 1990; Huikuri et al., 2003; Kleiger et al., 1987; Langewitz, Ruddel, & Schachinger, 1994; Nolan et al., 1998; Nolan et al., 1992; Palatini et al., 2004; Schwartz, La Rovere, & Vanoli, 1992) and hyperactivity of the sympathetic nervous system (SNS) may be an important cause for the detrimental effects of stress on cardiovascular health (Palatini et al., 2004; Schwartz et al., 1992).

Direct recording of action potentials from superficial sympathetic nerves in the muscles and the skin (Wallin et al., 1981; Wallin, Sundlof, & Delius, 1975) or the measurement of organ specific spillover of the post-ganglionic neurotransmitter norepinephrine using radioactive tracers (Esler et al., 1988; Esler et al., 2000b) is extremely valuable for basic research on the sympathetic nervous system. However, when research moves to an epidemiological scale, the expense and invasiveness of these methods becomes prohibitive. Furthermore, these invasive measures restrict research to the confines of a hospital or laboratory setting and are stressful for the participant. This precludes examination of individual differences in sympathetic activity in a natural setting, for instance during sleep or during job-related activities with a substantial mental and emotional load. Nonetheless, it is autonomic control during these naturalistic conditions that may have the largest clinical relevance. It is extremely valuable, therefore, to have non-invasive, unobtrusive measures of sympathetic nervous system activity.

At the moment the preejection period (PEP) is the measure of choice to monitor changes in cardiac sympathetic activity non-invasively. Under conditions of stable preload and afterload, changes in PEP reflect changes in contractility (Newlin et al., 1979) which are influenced by sympathetic but not parasympathetic activity in humans. The extant literature supports changes in PEP as a valid measure of changes in β-adrenergic inotropic drive to the left ventricle. Laboratory studies manipulating β-adrenergic tone in within-participants designs by epinephrine infusion (Houtveen et al., 2005) amyl nitrite inhalation (Mezzacappa et al., 1999; Svedenhag, Martinsson, Ekblom, & Hjemdahl, 1986; Svedenhag, Martinsson, Ekblom, & Hjemdahl, 1991) and adrenoceptor blockade (Nelesen et al., 1999), exercise (Harris et al., 1967; Schachinger et al., 2001; Winzer et al., 1999), emotional stress (Krzeminski et al., 2000; Miyamoto, Nakazono, Hiura, & Abe, 1983b; Smith et al., 1989b) or monetary reward (Berntson et al., 1994a; Newlin et al., 1979; Sherwood et al., 1986) have shown a dose-dependent shortening of the PEP. Between-participant differences in PEP level are stable over time (Richter et al., 2009), show comparable heritability to plasma catecholamine levels (Goedhart et al., 2006; Vrijkotte et al., 2004), and reliably reflect interindividual differences in cardiac sympathetic activity assessed by dual blockade (Kupper et al., 2006; Williams, Puddey, Beilin, & Vandongen, 1993).

PEP can be obtained by simultaneous recording of the thoracic impedance cardiogram (ICG) and electrocardiogram (ECG) (Riese et al., 2003; Willemsen et al., 1996) and is defined as the interval from the onset of left ventricular depolarization, reflected by the Q-wave onset in the ECG, to the opening of the aortic valve, reflected by the B-point in the ICG signal (Labidi, Ehmke, Durnin, Leaverton, & Lauer, 1970; Lozano et al., 2007; Sherwood et al., 1990; Willemsen et al., 1996). Figure 1 displays the ECG and ICG signals with the relevant landmarks. Throughout, the term 'actual' PEP is used to refer to the interval between the ECG Q-wave onset and the ICG B-point. To improve signal quality, PEP is usually scored from the ICG waveform after ensemble averaging over multiple beats, time locked to the R-wave peak. This improves automated detection of the crucial landmarks in the

ECG and in the ICG but even after ensemble averaging substantial errors in positioning of the Q-wave onset and B-point remain (Berntson et al., 2004; Labidi et al., 1970; Lozano et al., 2007; Willemsen et al., 1996). For this reason, visual inspection of the automatically detected Q-wave onset and B-point is needed and, to ensure sufficient reliability, scoring is often repeated by multiple raters. The latter visual inspection can be time-consuming and presents an obstacle to the assessment of PEP in epidemiological studies with thousands of participants or in ambulatory studies collecting data across extended periods of time. In addition, when signal quality of the ICG is compromised, for instance, during unsupervised activities in ambulatory recordings, reliable visual scoring of the B-point is very hard, even when employing multiple raters, leading to the exclusion of a substantial portion of the participants.

Two practical solutions have been proposed to sidestep the difficult detection of the Q-wave onset. The first is to score the more easily detected R-wave onset and add a fixed value for Q-wave duration of 15 ms (Berntson et al., 2004). The R-wave onset was used by Berntson et al. (Berntson et al., 2004) in 30 healthy participants, of which 10 showed no clear Q-wave in a lead II axis ECG derivation. In these participants scoring of Q-wave onset defaulted to the R-wave onset, and it was shown that using the R-wave onset for *all* participants significantly reduced the error variance in the individual differences in the PEP. This suggests that a PEP based on the R-wave onset was more reliable than the actual PEP based on the Q-wave onset, although this was established under resting conditions only. The second solution is to extend this reasoning, and use the R-wave peak instead of the R-wave onset as it is an even more sharply defined landmark in the ECG. This makes the further assumption that the R-wave onset to R-wave peak interval is also reasonably constant. Current practice is to estimate the Q-wave onset by subtracting a fixed value of 48 ms from the time of the R-wave peak (Brydon et al., 2008; Willemsen et al., 1996). To our knowledge the validity of this practice has not been verified.

To assist in the detection of the B-point in the ICG the physiological connection between the timing of the B-point and the dZ/dt-min peak can be exploited. The dZ/dt-min peak (in the literature variously called C-point or Z-point) is a maximum defined by a zero-order crossing in the first derivative of the ICG and can be detected with much more fidelity than the B-point, which is often a (subtle) inflection defined by a zero-order crossing in the second derivative of the ICG rather than a true minimum. Changes in cardiac contractility, the main concept that PEP aims to asses, are reflected in the time it takes the left ventricle to build up sufficient force to open the aortic valve (B-point) but also in the time it takes to reach peak ventricular ejection (dZ/dt-min peak) . Theoretically, the timing of these events should be highly correlated. Empirically, two groups have independently confirmed that the interval between the R-wave peak and the dZ/dt-min peak or the initial systolic time interval (ISTI, see figure 1) is a significant predictor of both the R-wave peak to B-point (RB) interval as well as the actual PEP (Berntson et al., 2004; Lozano et al., 2007; Meijer et al., 2007; Meijer et al., 2008). Meijer et al. (Meijer et al., 2007) found high correlations between ISTI and PEP in the supine position at rest and after light exercise in both old and young healthy adults and a group of older Parkinson patients (Meijer et al., 2007; Meijer et al., 2008; Meijer, Smorenberg, Lust, Verdaasdonk, & Groeneveld, 2010). Lozano et al. (Lozano et al., 2007) used the ISTI in quadratic curve fitting to estimate the RB interval. In 26 young adults their equation ($RB = -31.59 + 1.233 * ISTI + 0.0032 * ISTI^2$) accounted for over 90% of the variance in the actual RB interval during rest, a mental arithmetic task and a speech preparation task in three separate samples.

Taken together, the above suggests that adequate estimation of the PEP could be achieved by the detection of the two most salient features in the ECG or ICG, the R-wave peak and the dZ/dt-

min peak respectively. The estimated PEP is then defined as the sum of a fixed QR interval and an RB interval derived from the ISTI by regression. As this sidesteps the difficult detection of the Q-wave onset and the B-point, PEP estimation could be achieved in a near-complete automated fashion relying less on laborious visual inspection, even in noisy or degraded signals. In spite of these obvious advantages, use of an estimated PEP instead of the actual PEP has not become commonplace in the recent literature. A fixed Q-wave onset to R-wave peak (QR) interval of 48 ms is sometimes employed, but the ISTI has not been used to estimate the RB interval in spite of the encouraging results with this measure (Berntson et al., 2004; Lozano et al., 2007; Meijer et al., 2007; Meijer et al., 2008). Possibly the evidence from the studies published so far is considered to be insufficient to consider estimated PEP as a valid alternative to the actual PEP, as these studies had relatively small samples, were completely confined to a laboratory setting, and used only a few of the many conditions generally employed in psychophysiological studies.

To address these concerns, the current paper reports on two further studies that assessed the estimated and actual PEP scored by multiple raters in participants undergoing a wide variety of controlled experimental stress manipulations in a laboratory setting, and in participants that underwent a supervised ambulatory protocol containing activities that resemble natural daily activities but that induced a large variance in posture and physical activity. First, it was explored whether the QR interval can be approximated by a fixed interval even in very diverse laboratory and ambulatory conditions. Next, multilevel analysis was used to derive an equation to estimate the RB interval from the ISTI and it was tested whether a set of fixed regression coefficients applied to all participants in both laboratory and ambulatory recordings. Finally, Bland Altmann plots were used to test whether the estimated PEP from the R-wave and dZ/dt-min peaks and fixed QR interval adequately predicts the actual PEP across a large set of participants and a wide range of conditions.

Methods

Participants

In total, 91 undergraduate students (20 male, 71 female) with a mean age of 21.7 years (SD = 3.2) and a mean body mass index (BMI) of 22.2 (SD = 2.9) participated in the laboratory study. The participants in the ambulatory study were 31 undergraduate students (11 male, 20 female) having a mean age of 22.0 (SD = 1.9) and a mean BMI of 23.4 (SD = 4.3). Participants to both studies did not report any psychiatric diseases or cardiovascular problems and none were using cardioactive medication (e.g. antihypertensives) or any other kind of medication. They gave their written consent prior to participation. All participants were volunteers and received study credits or a 10 € gift voucher for their participation.

General Procedures

Laboratory study

The participants were asked to refrain from smoking and alcohol- or caffeine-containing beverages the evening before the test day and on the morning of laboratory testing. The experimental sessions took place between 9 a.m. and 4 p.m. and lasted approximately 90 minutes. ECG and ICG leads were attached to the participants using 6 pregelled Ag/AgCl spot electrodes (Ultratrace, Cosmed, USA) after which they were seated in front of a 19” monitor in a dimly lighted,

electrically-shielded, sound-attenuated cabin. The experimental session commenced with some general instructions and a brief period of rest in which optimal signal quality was ensured. To elicit variation in SNS activity, various experimental physical and mental stressors were presented in a fixed order as outlined in figure 2. The participant was monitored by a webcam to ensure safety as well as compliance with experimental instructions.

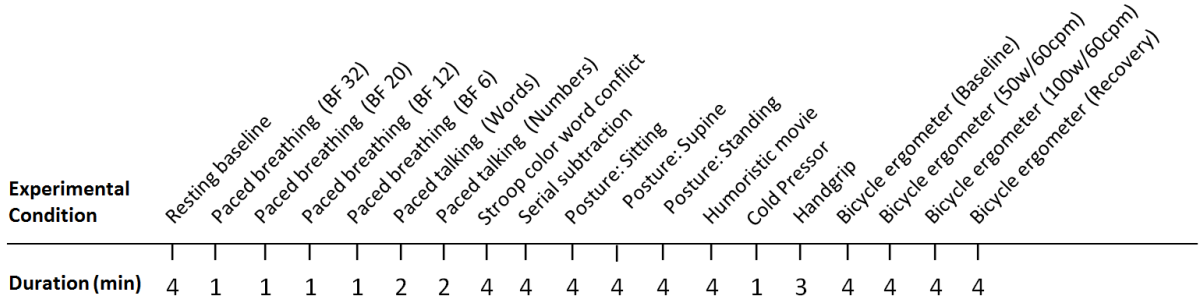

Figure 2. *Experimental protocol of the laboratory study.*

Resting baseline: Participants sat quietly for 4 minutes with their eyes open. The computer screen depicted an empty sandy palm beach or a calm lake in a lush green forest. Paced breathing: Participants were asked to breath in concert with a visual metronome on the computer screen at 32 and 20 cpm, two frequencies that are above the typical spontaneous breathing frequency (14 cpm), at 12 cpm, and at 6 cpm, a rate much lower than the typical spontaneous breathing frequency. Each breathing frequency was maintained for 1 minute. Paced talking: Participants were asked to first read out color words (2 minutes) and then numbers (2 minutes) that were presented on the computer screen in a frequency corresponding to the trial speed in the actual Stroop/Subtraction tasks. Stroop word color conflict: Participants were shown the names of colors printed in conflicting ink colors (e.g. the word "blue" in red ink) on the computer screen and asked to verbally identify the color of the ink rather than the word as fast as possible. Mental load is created by the interference between the discrepant ink color and the color name. The experimenter would correct the participant over the intercom in case of a wrong answer (‘WRONG”). The experimental condition lasted 4 minutes. Serial Subtraction: The participants were presented with a starting number 1248 and were asked to continuously subtract 7 from this number (speaking out loud) until the 4 minute test period ended. Whenever a participant gave the wrong answer, the experimenter stated the correct number from which the participant then had to restart over the intercom (‘WRONG, the correct number was ...’). Posture: Participants were asked to sit down quietly for 4 minutes, followed by lying down quietly for 4 minutes (Supine) and finally to stand quietly for 4 minutes (Standing). The postural transitions were kept below 5 seconds. Humoristic movie: Watching a humoristic movie (4 minutes). The participant got a choice of 4 short humoristic movies. They choose one on a participant of their own liking and sat to watch it for 4 minutes. Cold pressor: The cold pressor test used in this experiment consisted of a 3–5 °C ice bath, composed of tap water and melted ice held in a small plastic container. The container was placed adjacent to the participant on a table. The participant was asked to submerge the dominant hand up to the wrist joint and to hold the fingers in a relaxed position. After exactly 60 sec the hand was removed from the bath. In this condition the experimenter was present in the cabin throughout to ensure compliance. Hand grip: During practice and instruction, maximum grip strength in the dominant hand is established in two separate attempts with a hand grip dynamometer. During the actual hand grip task, participants maintain

isometric contraction at 30% of their maximum voluntary contraction for a period of 3 minutes in sitting position. A dial on the dynamometer indicates deviations of the target force, upon which the experimenter indicated to restore force. In this condition the experimenter was present in the cabin throughout to ensure compliance. Bicycle ergometer: Participants were asked to sit quietly on the bicycle ergometer for 4 minutes before the final exercise condition. Then 4 minutes of biking on the bicycle ergometer at a resistance of 50 Watt with 60 rotations per minute and 4 minutes at 100 Watt with 60 rotations per minute followed. In the second half of the sample, a 4 minute recovery period while still sitting on the bike was added, that followed immediately after biking on the bicycle ergometer at 100 Watt (N = 46 participants).

Ambulatory study

The participants were asked to refrain from smoking and alcohol- or caffeine-containing beverages the evening before the test day and in the morning before coming to the laboratory. All standardized experimental sessions took place approximately 2 hours after wakening and lasted 2 hours. The ECG and ICG leads were attached to the participants using 5 pregelled Ag/AgCl spot electrodes (Ultratrace, Cosmed, USA), after which they were seated in front of a dimly lighted, electrically-shielded, sound-attenuated cabin to measure a resting baseline. Next, a sequence of protocolized normal daily activities in a supervised ambulatory setting was performed in the lab, outdoors, and in the university sports centre. The protocol aimed to create variations in posture and intensity of physical and mental activity in close resemblence to normal daily activies. Figure 3 outlines the experimental protocol for the ambulatory study .

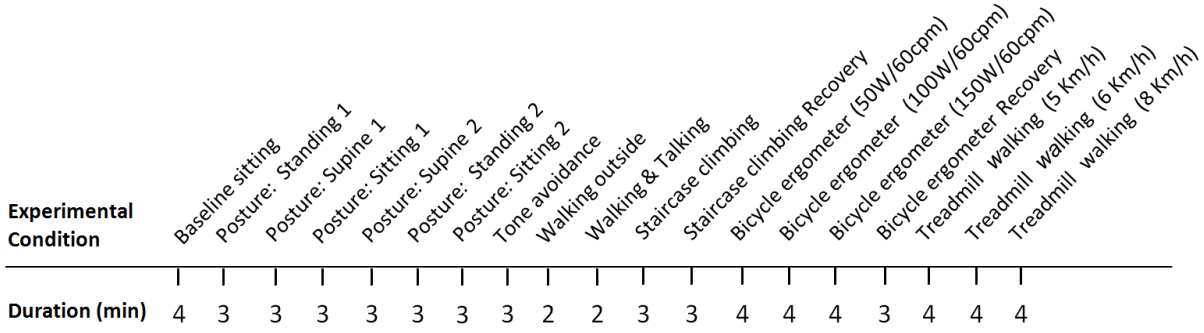

Figure 3. *Experimental protocol of the ambulatory study.*

Resting baseline: 4 minutes of quietly sitting in a chair in a laboratory room. Posture: Several postures were measured in a laboratory room; 3 minutes of standing; 3 minutes in supine position on a stretcher; 3 minutes of sitting in a chair; 3 minutes in supine position on a stretcher; 3 minutes of standing; 3 minutes of sitting in a chair. Tone Avoidance task: The tone avoidance task is four-choice reaction time task where participants have to push the appropriate button within 550 ms to avoid a loud noise (1000 Hz, 85 dB). A cross appears in one of te corners of the screen and the opposite diagonal button on a corner of a sqare keypad has to be pushed. This task is know to elicit strong mental load and was performed in a laboratory room. Standardized physical daily activities: These activities were performed on the way to the VU sports centre where the standardized physical activities were performed; 2 minutes of quietly walking outside; 2 minutes of walking outside while talking; 3 minutes of climbing stairs (7 stores up and down); 3 minutes of quietly sitting as a recovery.

Standardized physical activities: The participants performed the bicylce ergometer test and treadmill test at the VU-sport centre. Three bicycle ergometer tests were done, each for 4 minutes at 60 cycles per minute at 50W, 100W and 150W respectively followed by 3 minutes of sitting quietly as a recovery. Three 4 minute treadmill conditions were done at 5 km/h, 6 km/h and 8 km/h respectively.

In both studies, participants reported to the VU University Amsterdam on the day of testing and were given a brief description of the procedures and an informed consent was signed. A short standardized interview was used to obtain demographic information, smoking behaviour, exercise behaviour, regular medication use, and to verify that they had no current psychiatric complaints (Beck depression scale < 4) or cardiovascular disease. Height, and weight were measured using standard procedures.

Physiological recordings

Laboratory study

The ECG and ICG signals in the laboratory study were continuously recorded at a sample rate of 1000HZ with use of the ECG100C and NICO1000C modules of the BioPac data-acquisition system (BioPac systems INC, Santa Barbara, CA). Cleaning of the skin with alcohol before electrode application ensured that electrode resistance was kept low. The first ECG electrode (V-) was placed slightly below the right collar bone 4 cm to the right of the sternum. The second ECG electrode (V+) was 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 was placed on the right side, between the lower two ribs at the right abdomen. The first ICG measuring electrode (V_1) was placed at the top end of the sternum, between the tips of the collar bones. The second ICG measuring electrode was placed at the xiphoid complex of the sternum, where the ribs meet. The two current electrodes were 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 vertebraes T8 and T9 on the spine, at least 3 cm (1") below the ICG measuring electrode V_2. This electrode placement was used in many previous studies that attest to high intra day (Kupper et al., 2006) and day-to-day reliability (Vrijkotte et al., 2004) of the systolic time intervals derived from this placement as well as good temporal stability across a two year period (Goedhart et al., 2006) and the ability to discriminate low and high chronic work stress (Vrijkotte et al., 2004).

Ambulatory Study

The electrocardiogram (ECG) and the impedance cardiogram (ICG) of the ambulatory study were recorded continuously with the 5-lead version of the VU-AMS5fs device (VU University, Amsterdam, www.vu-ams.nl). This device was developed to study autonomic nervous system activity in naturalistic settings (de Geus et al., 1995; Willemsen et al., 1996). Cleaning of the skin with alcohol before electrode application ensured that electrode resistance was kept low. A single dedicated ECG electrode (V+) was 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 first ICG measuring electrode (V-) was placed at the top end of the sternum, between the tips of the collar bones . This electrode also functions as the first ECG (V-) electrode. The second ICG measuring electrode was placed at the xiphoid complex of the sternum, where the ribs meet. The two current electrodes were placed on the back: I- on the spine over the cervical vertebra C4, at least 3 cm (1") above the ICG measuring

electrode V_1, and I+ between thoracic vertebraes T8 and T9 on the spine, at least 3 cm (1") below the ICG measuring electrode V_2.

Signal Analyses and Data Reduction

The Acknowledge algoritm was used on the laboratory data to detect all relevant landmarks in the ECG including the Q-wave onset and the R-wave peak. The algorithm incorporates the open source OSEA QRS detector and beat classification library provided by EP Limited (http://www.eplimited.com). The QR interval was tested for outliers at the single beat level by using a within-participant criterion of 2 standard deviations, and visual inspection with random sampling was used to confirm proper operation of the Acknowledge algoritm across experimental conditions. The mean QR interval was then computed per condition across all beats with a valid Q-wave onset.

For detection of the Q-wave onset in the ambulatory data, the ECG was imported into the VU-AMS5fs software. For each experimental condition the interbeat interval (IBI, ms) was scored from the R-wave peaks in the ECG. The IBI time series was visually inspected and missed or incorrect R-wave peaks were interactively corrected; bad ECG signal fragments were removed. The mean Q-wave onset for each ambulatory condition was visually scored in the R-wave peak locked ensemble averaged ECG across all valid beats in that condition by two raters. Post-scoring, the raters chose a consensus for the points that did not overlap, and these were retained for the analyses.

The procedures to score the ISTI were identical for both the ambulatory and laboratory study. First, the ICG and ECG signals were imported into the VU-AMS software. After obtaining the corrected IBI time series, interactive visual scoring of the R-wave peak locked ensemble averaged ICG signal from all valid beats was used to mark the B-point and the dZ/dt-min peak in each condition (see figure 1 for an example). The actual PEP was computed as the interval from the Q-wave onset in the ECG to the B-point in the ICG signal. The ISTI was computed as the time interval between the R-wave peak in the ECG and the dZ/dt-min peak in the ICG (Meijer et al., 2007; Meijer et al., 2010). To allow computation of the inter rater reliability, two raters independently scored the B-point and dZ/dt-min peak. Post-scoring, the raters chose a consensus for the points where their judgement did not overlap, and these were retained for the analyses.

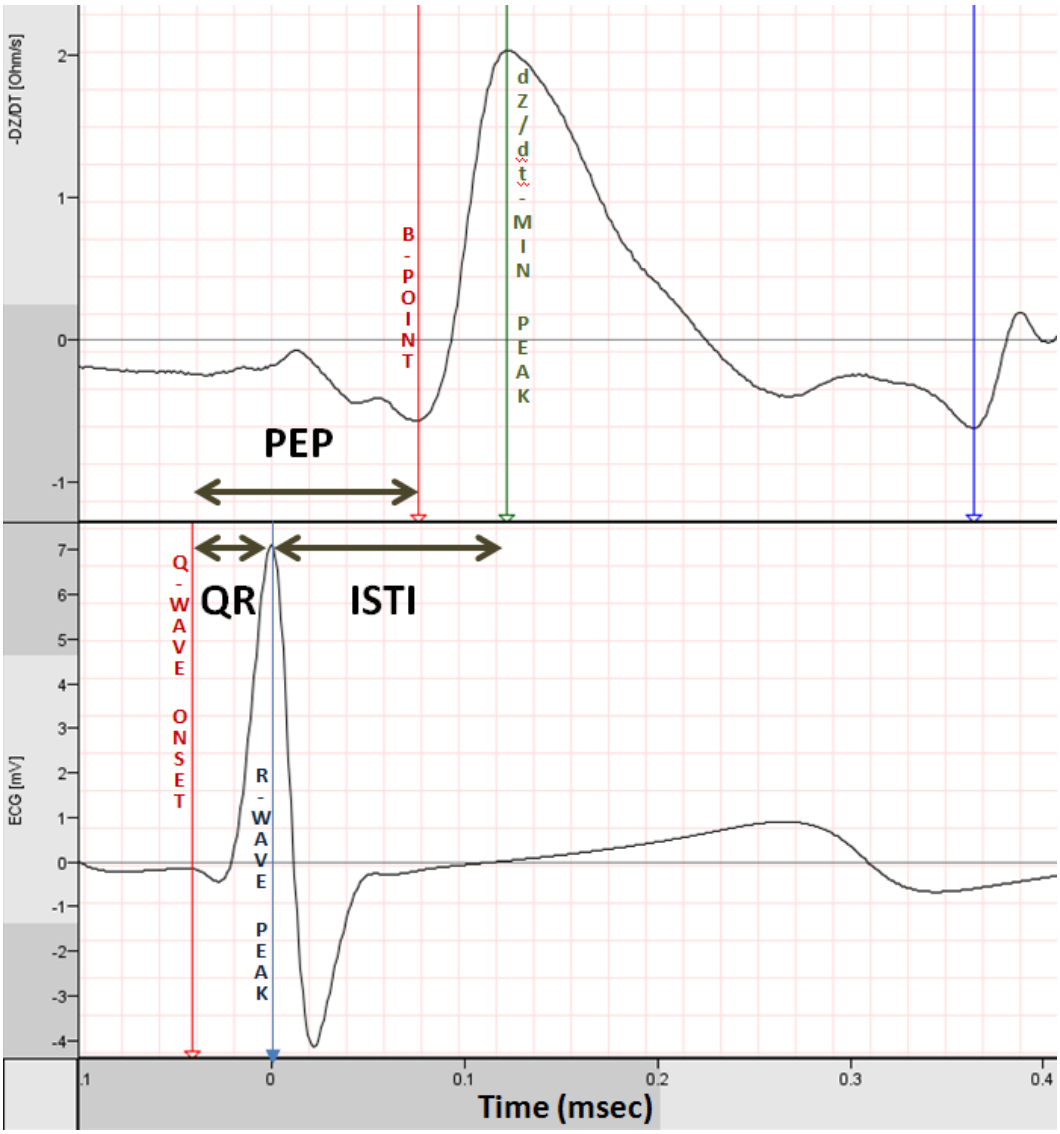

Figure 1. *The impedance cardiogram top) and the electrocardiogram (bottom) with the four landmarks defining the PEP (Q-wave onset to B-point) and the ISTI (R-peak to dZ/dt-min peak).*

Statistical Analyses

For both laboratory and ambulatory studies a mixed ANOVA was used to test for the effect of condition (fixed effect) on IBI, ISTI and the actual PEP to verify succesful manipulation of autonomic tone within participants. Mixed ANOVA was further used to test whether the QR-interval changed over conditions. Post-hoc T-tests were performed for the four variables on *a priori* defined contrasts. In the laboratory the pre-test resting baseline was compared to paced breathing, paced talking, stroop colour conflict, serial subtraction, stress recovery, orthostatic manipulations, humoristic movie, and the handgrip and cold pressor tests, and the ergometer baseline level was compared to the bicycle ergometer (50W and 100) and the bicycle ergometer recovery conditions. In the ambulatory study, the resting baseline condition at the start of the recording was compared to all other conditions.

To test whether the QR interval can be approximated by a fixed interval, a oneway ANOVA compared the QR intervals across study setting (laboratory vs ambulatory) in the resting sitting

baseline condition. Next the grand averaged QR interval across all participants and all laboratory and ambulatory conditions was subtracted from the actual QR interval values observed and the distribution of the difference scores was used to compute the mean error as well as its the 95% confidence intervals.

To compute interrater reliability of the Q-point, B-point and dZ/dt-min peak scoring, intraclass correlation coefficients (ICCs) were computed between rater 1 and rater 2 using a random effects model (absolute agreement) on the values of for the RB interval and the ISTI. This was done separately for each of the conditions across all participants (per-condition inter rater reliability) and for each of the participants across all conditions (per-participant inter rater reliability).

Multilevel analysis was then used to establish the optimal regression equation to predict the RB interval from the ISTI. Multilevel analysis is a general method of analyzing data with a hierarchical or clustered structure (Snijders & Bosker, 1999). Here different conditions were clustered within participants. RB was considered to be a function of ISTI and a random error term:

$$RB_{ij} = B0_j + B1_j * ISTI_{ij} + \varepsilon_{ij} \qquad (1)$$

with i indexing the lower level of repeated samples and j indexing the higher level of the participants. Regression equation (1) defines the relation between RB and ISTI within each of the participants. Coefficient B0j and B1j are the participant-specific intercept and slope which, i.e. $B0_j = g00 + \varepsilon0_j$ and $B1j = g10 + \varepsilon1_j$, where g00 and g10 are the fixed mean intercept and slope across all participants, whereas the random coefficients $\varepsilon0_j$ and $\varepsilon1_j$ vary across participants. Previous research suggested that a single equation can be used for all participants (Lozano et al., 2007), which means that models with a random slope and intercept should not provide a significant better fit than the more parsimonious model with a fixed intercept and slope. This was explicitly tested by comparing the explained variance and the fit of the model with freely estimated slopes and intercepts to that of the model with fixed values. Explained variance in the RB interval was computed with the following formula suggested by Blackwell et al: (unrestricted error – restricted error) / unrestricted error (Blackwell, de Leon, & Miller, 2006). The deviance fit test, or likelihood ratio test, was used to compare the fit of two regression models. This test is based on the difference between the deviance statistics of the two models, which has a chi-square distribution with degrees of freedom equal to the difference in the number of parameters estimated in the models being compared. If the equation with a random slope/intercept fitted better, sex, age and BMI were added as potential predictors in the level 2 model to see whether these variables could account for the individual variance in the intercept and slope. The multilevel analysis was repeated separately for laboratory and ambulatory conditions, which allowed a comparison of the parameter estimates obtained in different study settings.

Finally, the estimated PEP was computed for each condition for each individual in both studies by summing the grand averaged QR interval to the RB interval estimated from the ISTI using the slope and intercept parameters from the best fitting model in the laboratory study. A Bland-Altman analysis was used to test the absolute agreement between the estimated PEP and the actual PEP across all laboratory and ambulatory conditions.

Results

Effects of the Experimental Manipulations

The means and standard deviations for IBI, PEP, QR interval and ISTI are presented per experimental condition in table 1 (for the laboratory study) and table 2 (for the ambulatory study).

Table 1. *Means and standard deviation for IBI, actual PEP, QR interval and ISTI (all in ms) in the laboratory study. Change scores are given from resting baseline for all experimental conditions, but change scores for bicycle ergometer are compared to bicycle ergometer baseline values. The bicycle recovery was compared to the bicycling at 100W/60cpm condition.*

Experimental conditions	IBI	Δ IBI	PEP	Δ PEP	QR	Δ QR	ISTI	Δ ISTI
	Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)	
Resting Baseline	845 (111)	NA	107 (18)	NA	38 (7)	NA	127 (17)	NA
Paced breathing (BF32)	803 (114)	-42*	101 (18)	-6*	39 (7)	1*	121 (17)	-6*
Paced breathing (BF20)	805 (112)	-40*	105 (17)	-2*	38 (7)	0	123 (17)	-4*
Paced breathing (BF12)	803 (104	-42*	105 (17)	-2*	38 (7)	0	126 (16)	-1
Paced breathing (BF6)	825 (93)	-20*	107 (18)	0	39 (7)	1*	127 (16)	0
Paced talking (Words)	780 (106)	-65*	109 (18)	2	38 (8)	0	128 (18)	1
Paced talking (Numbers)	757 (97)	-88	107 (21)	0	39 (8)	1*	123 (19)	-4*
Stroop color word conflict	758 (97)	-87*	104 (19)	-3*	39 (7)	1*	121 (19)	-6*
Serial subtraction	748 (101)	-97*	104 (20)	-3*	39 (7)	1*	120 (19)	-7*
Posture: Sitting	859 (108)	14	108 (18)	1	38 (7)	0	128 (16)	1
Posture: Supine	963 (126)	117*	95 (19)	-12*	38 (6)	0	121 (21)	-6
Posture: Standing	723 (83)	-123*	117 (16)	10*	38 (7)	0	138 (16)	11*
Humoristic movie	875 (123)	29*	108 (17)	1	39 (7)	1*	124 (17)	-3
Cold pressor	836 (117)	-9	110 (19)	3	40 (7)	2*	126 (17)	-1
Handgrip	840 (119)	-5	112 (17)	5*	40 (8)	2*	128 (17)	1
Bicycle ergometer (Baseline)	833 (109)	NA	115 (17)	NA	39 (8)	NA	133 (17)	NA
Bicycle ergometer (50W/60cpm)	598 (78)	-235*	77 (20)	-38*	43 (9)	4*	94 (19)	-39*
Bicycle ergometer (100W/60cpm)	504 (80)	-329*	66 (17)	-49*	43 (10)	4*	82 (180)	-51*
Bicycle ergometer (Recovery)	715 (116)	211*	90 (25)	24*	38 (7)	-1*	118 (22)	36*

**= significantly different compared to the appropriate baseline, in Bonferroni-corrected post hoc tests ($p < 0.05/18 = .0028$).*

Table 2. *Means and standard deviations for IBI, actual PEP, QR interval and ISTI (all in ms) in the ambulatory study. Change scores are given from resting baseline for all experimental conditions.*

Experimental conditions	IBI	Δ IBI	PEP	Δ PEP	QR	Δ QR	ISTI	Δ ISTI
	Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)	
Baseline sitting	784 (98)	NA	112 (17)	NA	43 (6)	NA	118(19)	NA
Posture: Standing1	674 (84)	-110*	117 (16)	5*	41 (6)	-2*	123 (17)	5*
Posture: Supine1	900 (117)	114*	108 (21)	-4	44 (6)	1	120 (19)	2
Posture: Sitting 1	787 (97)	3	113 (17)	1	43 (5)	0	122 (18)	4*
Posture: Supine2	899 (120)	115*	114 (25)	2	44 (6)	1	127 (19)	9*
Posture: Standing 2	691 (79)	-92*	119 (14)	7*	41 (6)	-2	125 (18)	8*
Posture: Sitting 2	808 (105)	24*	117 (13)	5*	44 (6)	1	123 (15)	5*
Tone Avoidance Task	755 (99)	-29*	110 (15)	-2	43 (7)	0	115 (17)	-2
Walking outside	597 (65)	-187*	86 (13)	-36*	43 (7)	0	88 (16)	-30*
Walking & Talking	587 (69)	-197*	85 (11)	-37*	43 (7)	0	87 (13)	-31*
Staircase Climbing	430 (43)	-354*	66 (9)	-46*	41 (6)	-2	61 (14)	-57*
Staircase Climbing Recovery	642 (124)	-142*	80 (16)	-32*	42 (6)	-1	82 (15)	-36*
Bicycle ergometer (50W/60cpm)	557 (124)	-227*	88 (18)	-24*	43 (7)	0	90 (17)	-28*
Bicycle ergometer (100W/60cpm)	463 (76)	-321*	73 (14)	-39*	41 (5)	-2	75 (19)	-43*
Bicycle ergometer (150W/60cpm)	397 (48)	-387*	63 (12)	-49*	40 (6)	-3	57 (15)	-61*
Bicycle ergometer recovery	505 (86)	-279*	79 (19)	-33*	40 (5)	-3	77 (23)	-41*
Treadmill walking (5km/h)	511 (83)	-273*	80 (14)	-32*	42 (6)	-1	78 (19)	-40*
Treadmill walking (6km/h)	486 (77)	-298*	76 (13)	-36*	42 (6)	-1	79 (17)	-39*
Treadmill walking (8km/h)	400 (51)	-384*	72 (19)	-40*	41 (5)	-2	67 (26)	-51*

**= significantly different compared to the baseline, in Bonferroni-corrected post hoc tests ($p < 0.05/18 = .0028$).*

Laboratory study

Mixed ANOVA showed a significant effect of experimental condition on IBI ($F(17, 154.5) = 110$, $p < .001$), PEP ($F(17, 140.2 = 73$, $p < .001$), and ISTI ($F(17, 144.8 = 38$, $p < .001$). As expected IBI, PEP, and ISTI were found to decrease significantly over baseline levels during conditions known to increase cardiac sympathetic activity, i.e. during mental stress invoked by the Stroop color word and serial subtraction conditions and during dynamic exercise on the bicycle ergometer. IBI, PEP, and ISTI also clearly evidenced the well-known decrease in cardiac sympathetic activity during recovery from

exercise. Significant changes in IBI, PEP, and ISTI also suggest an increase in cardiac sympathetic activity during forced breathing at high frequencies and during the reading aloud of numbers and words at a forced speed. Lying down and standing up caused the expected decrease and increase respectively in heart rate, but the ISTI and PEP showed the expected reversed pattern, which is caused by preload (ventricular filling pressure) and afterload (systemic vascular resistance) effects (Houtveen et al., 2005). The cold pressor test, the handgrip test and the funny movie did not lead to major changes in IBI, PEP or ISTI.

The mean QR interval ranged from 37.9 to 42.7 ms across conditions with a median value of 38.7 ms and a mean of 39.1 ms. The standard deviation was substantial and averaged 7.5 ms across all conditions with the largest variance in QR interval seen during exercise.

Ambulatory study

Mixed ANOVA showed a significant effect of experimental condition on IBI ($F(18, 29.7) = 194$, $p < .001$), PEP ($F(18, 27.9) = 103$, $p < .001$) and ISTI ($F(18, 55.2) = 66.7$, $p < .001$). As expected IBI, PEP and ISTI were found to decrease significantly over baseline levels during conditions known to increase cardiac sympathetic activity, i.e. daily physical activities and during dynamic exercise on the bicycle ergometer and treadmill. IBI, PEP and ISTI also clearly evidenced the well-known decrease in cardiac sympathetic activity during recovery from exercise. Lying down and standing up caused the expected decrease and increase respectively in heart rate while PEP and ISTI again showed the expected reversed pattern.

The mean Q-wave onset to R-wave peak interval ranged from 39.7 to 43.8 ms across conditions with a median value of 42.3 ms and an mean of 42.0 ms. As in the laboratory condtions, the standard deviation was substantial and averaged 6.4 ms across all conditions. Overall interrater reliability of the Q-point scoring was .89.

Using a fixed Q-wave onset to R-peak interval

Mixed ANOVA showed a significant effect of experimental condition on the QR interval in the laboratory study ($F(17, 145,1) = 2.7$, $p < .001$) but not in the ambulatory study. Baseline values were significantly different across study settings ($F(1, 119) = 8.9$, $p = .003$). Of note, the QR intervals in both settings were at least one standard deviation lower than the often employed fixed value of 48 ms. To obtain the best possible value for a fixed QR interval we computed the weighted average of both studies of $((39.1 * 91) + (42 * 31))/122 = 39.8 \approx 40$ ms. We substracted this from the observed QR intervals in both settings to get an impression of the error made by using a fixed interval. Figure 4 depicts the distribution of differences between the actual QR intervals minus the fixed estimate of 40 ms. The mean error is acceptably close to zero (-0.03 ms) but the 95% confidence intervals (-14.5ms – 14.5ms) are substantial.

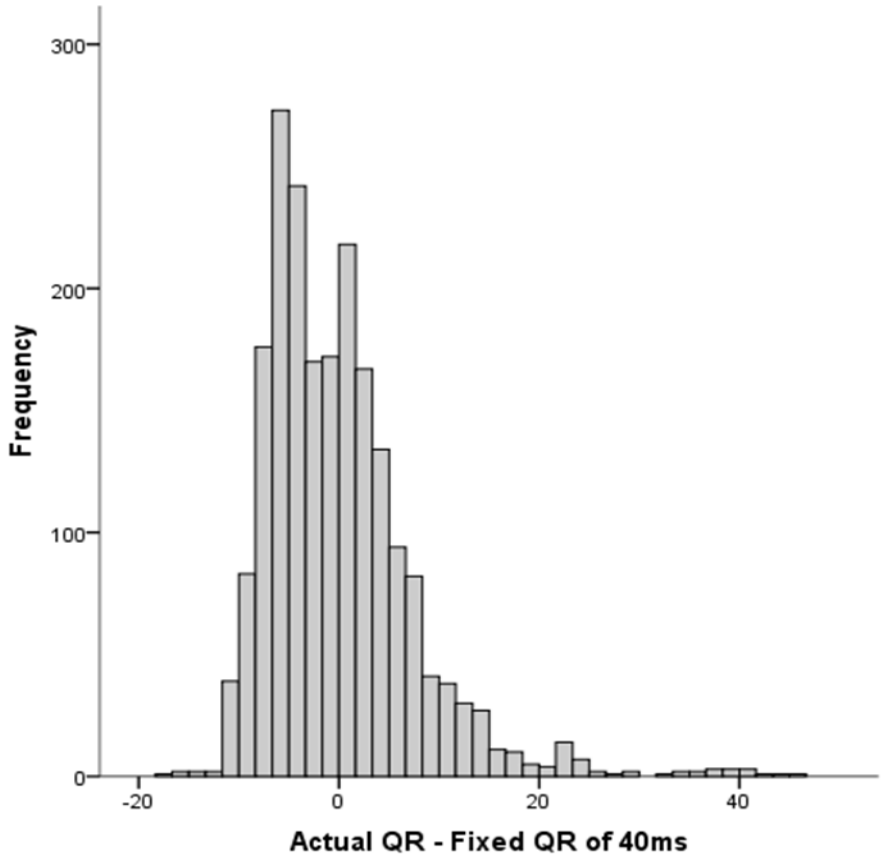

Figure 4. *Distribution of the difference between the measured QR interval and a fixed QR value of 40ms.*

Estimation of the R-wave peak to B-point interval using the ISTI

Mean per-condition interrater reliability for the B-point (using the RB interval, i.e. correlating the RB interval of rater 1 with the RB interval of rater 2 across all participants for each condition) was .92 in the laboratory conditions (from .85 during bicycle ergometer at 100W/60cpm to .99 in bicycle ergometer recovery) and .74 in the ambulatory conditions (from .21 during lying to .99 during the ergometer bike test at 150W/60cpm). For the dZ/dt-min peak (using the ISTI) it was 0.99 in the lab (from .96 during lying to 1.00 in 8 of the other conditions) and .85 (from .48 during walking at 6 Km/h to 1.00 during baseline sitting) in the ambulatory study. The lower per-condition interrater reliability for the B-point compared to dZ/dt-min peak was confirmed in a lower per-participant interrater reliability (e.g. correlating RB/ISTI of rater 1 with RB/ISTI of rater 2 across all conditions for each participant). In the laboratory study, mean per-participant interrater reliability was .92 for the B-point (ranging from .73 to .98) versus 1.00 for the dZ/dt-min peak (ranging from .93 in the 1.00). In the ambulatory study it was .85 for the B-point (ranging from .64 to .98) versus .92 for the dZ/dt-min peak (ranging from .89 to .99).

Table 3 gives the results of the multilevel analysis that was used to examine the relationship between the RB interval and the ISTI. The linear model using ISTI as a predictor of RB with a random intercept and a fixed slope explained 72.2 % of the variance in the RB interval and had a significant better fit than the null model without ISTI, $\chi^2(1) = 2155$, $p < .001$. Allowing individual differences in the regression of ISTI on PEP slightly improved prediction further. This extended linear model with a random intercept and a random slope, *RB = -46.39 + (0.9 * ISTI)*, explained 79 % of the total variance in the RB interval and had a better fit than both previous models, $\chi2(2) = 2276$, $p < .001$ and $\chi2(1) =$

121, p < .001 respectively. Adding a quadratic term improved the model (χ2 (1) = 28, p < .001) but added only .7% to the explained variance. Sex and age and BMI were also added but did not significantly contribute to the individual differences in the slopes or intercepts (model not shown).

Table 3. *Multilevel results for predicting the RB interval from the ISTI in the laboratory study.*

	Null model (without ISTI)		Random intercept and fixed slope for ISTI		Random intercept and random slope for ISTI		Random intercept and random slope with quadratic term.	
		SE		SE		SE		SE
Fixed effects								
Intercept (B0)	64.15	1.47	-43.74	1.82	-46.39	3.32	-82.35	7.44
ISTI (B1)			0.89	0.013	0.90	0.03	1.54	0.12
ISTI squared (B2)							-0.003	0.001
Random effects								
level 1 residual	238.03	8.74	59.21	2.17	50.05	1.9	48.40	1.8
Level 2 residual								
intercept	181.21	28.27	68.64	10.7	771.18	147.72	1085.90	193.8
slope					0.04	0.01	0.06	0.011
Log likelihood	13329.11		11173.93		11053.28		11025.91	
Extra degrees of freedom			1		1		1	
Explained variance (%)			72.2%		79%		79.7%	

When we repeated the multilevel analysis for the ambulatory study (table 4), the linear model with a random intercept and a random slope explained 81 % of the variance in RB and again fitted the data bettter than a model with fixed parameters for all participants (67.4% explained variance), but as can be seen from the standard errors of the estimates, the slope and intercept were significantly different from the laboratory study: *RB = -15.04 + (0.70*ISTI).* This does not bode well for the generalisability of a single set of parameters to estimate the RB interval from the ISTI across different study settings.

Table 4. *Multilevel results for predicting the RB interval from the ISTI in the ambulatory study*

	Null model (without ISTI)		Random intercept and fixed slope for ISTI		Random intercept and random slope for ISTI		Random intercept and random slope with quadratic term.	
		SE		SE		SE		SE
Fixed effects								
Intercept (B0)	52.119	1.59	-14.91	2.07	-15.04	2.25	3.73	5.28
ISTI (B1)			0.69	0.02	0.70	0.03	0.27	.11
ISTI squared (B2)							0.002	0.001
Random effects								
level 1 residual	472.53	19.13	101.35	6.25	89.85	5.7	86.83	5.51
Level 2 residual								
intercept	181.21	28.27	68.64	10.7	771.18	147.72	1085.90	193.8
slope					0.02	0.01	0.02	0.01
Log likelihood	5044		4226		4190		4175	
Extra degrees of freedom			1		1		1	
Explained variance (%)			79%		81%		82%	

Figure 5 depicts the crucial test of whether the estimation of the PEP based on the detection of the R-peak and dZ/dt-min peaks adequately reflects the actual PEP across a wide range of laboratory and ambulatory conditions. Estimated PEP was calculated as the weighted average QR of the laboratory and ambulatory data plus the RB estimated from the regression equation from the best multilevel fit on the laboratory (exploratory) data, with the ambulatory data acting as the confirmatory set. Hence, PEP was estimated as: *40 + (-46.39 + (0.9 * ISTI))*. The mean difference between the actual PEP and the estimated PEP in the laboratory study was 8.3 ms, and in the ambulatory study -3.6 ms. The 95% confidence intervals were very large, ranging from –19.9 to 36.5 ms for the laboratory study and from -25.3 to 17.9 ms in the ambulatory study.

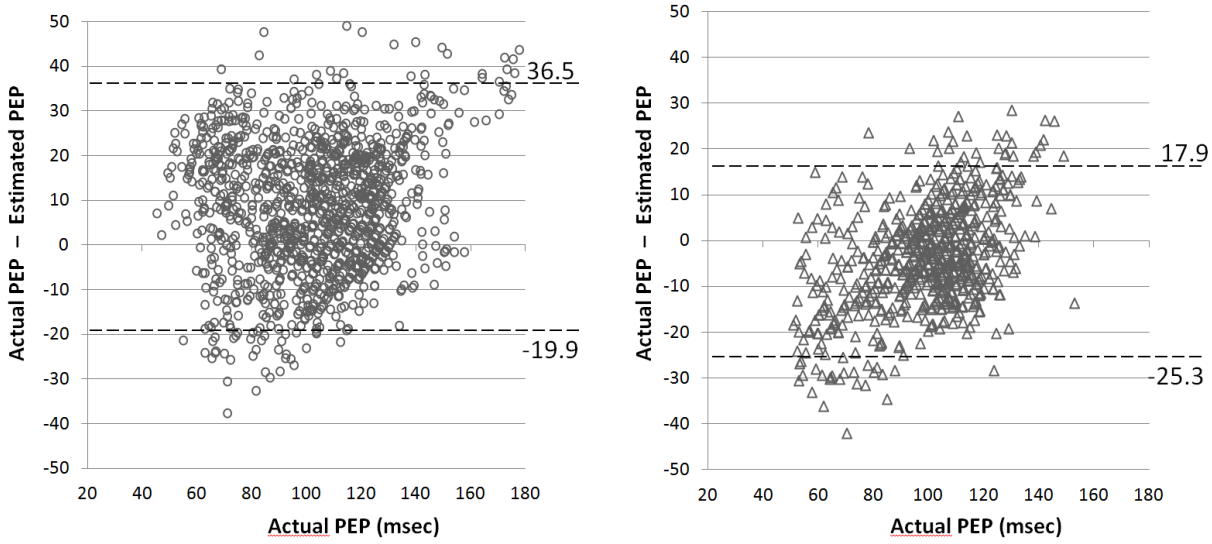

Figure 5. *Bland-Altman plots of the difference between the actual PEP based on the ECG Q-wave onset and the ICG B-point and the estimated PEP based on the ECG R-wave and ICG dZ/dt-min peaks. The difference is plotted as a function of the absolute value of the actual PEP for the laboratory study (left) and the ambulatory study (right). Dotted lines represent the 95% confidence intervals around the mean difference.*

Discussion

Two studies in different samples and settings evaluated the validity of an estimated PEP computed as the sum of a fixed QR interval and an RB interval predicted from the R-wave and dZ/dt-min peaks. These peaks represent the two most clear landmarks in the ECG and ICG respectively and are detected with greater ease and reliability than the Q-wave onset in the ECG and B-point in the ICG that represent the actual physiological events defining the PEP as index of sympathetic control over cardiac contractility. We found substantial discrepancies between the estimated PEP using the R-wave and dZ/dt-min peaks and the actual PEP based on the Q-wave onset and B-point in both study settings. In the laboratory study, at least 84% of the differences between the estimated PEP and the actual PEP exceeded 3.5 ms which was the mean reactivity found to two often used tasks in laboratory stress studies. This is an unacceptable large error of estimation.

About half of the error is due to the use of a fixed QR interval. We found an mean QR interval of 38 ms across a wide range of laboratory conditions. The QR interval in the ambulatory study was

around 42 ms, and the value was significantly larger than in the laboratory. The mean value across both settings was 40 ms which is in good agreement with previous studies. For instance, Goldberger and Bhargava (Goldberger & Bhargava, 1983) reported a QR interval of around 37 ms at rest which decreased by a few ms during exercise. Of note, the QR intervals in both laboratory and ambulatory setting were at least a standard deviation lower than the value of 48 ms often employed to compute PEP. The 48 ms estimate seems to originally derive from a paper that used a 250 Hz sampling of the ECG (Willemsen et al., 1996) which means that precision of the Q-wave detection was only 4 ms.

Apart from the QR interval the estimation of the RB interval from the ISTI also strongly contributed to the difference between estimated and observed PEP. Here, our findings are in clear contrast to previous reports showing the ISTI to be a significant predictor of both the R-wave peak to B-point (RB) interval as well as the actual PEP (Lozano et al., 2007; Meijer et al., 2007; Meijer et al., 2008). Specifically, Lozano et al. (Lozano et al., 2007) found that the equation $RB = -31.59 + 1.233 * ISTI + 0.0032 * ISTI^2$ accounted for 95% of the variance in the actual RB interval during rest, a mental arithmetic task and a speech preparation task in a discovery sample of 26 healthy participants after exclusion of 7 participants without 'a clear B-point upstroke'. The same equation also predicted the RB with high precision in two new samples of 9 adolescents and 15 middle-aged participants, after exclusion of 1 and 4 additional participants respectively based on the difficult B-point. Our results disagree with these earlier findings as different intercepts and slopes were found (even when using the quadratic solution) in both the laboratory ($RB = -84 + 1.58 * ISTI -0.003 * ISTI^2$) and ambulatory ($RB = 3.7 + 0.27 * ISTI + 0.002 * ISTI^2$) setting compared to the previously suggested regression parameters. Also, the explained variance in the observed RB by the estimated RB was only 79% and 81%.

Our study was intended to resemble real data collections as closely as possible making it differ from the earlier study by Lozano (Lozano et al., 2007) in four important respects. First, more participants were included in total (41 vs. 122) creating more room for between-participant variance. Secondly, *no* participants were excluded; we simply scored the B-point as good as we could in all participants. Thirdly, we used multilevel analysis to estimate the regression equation, which takes into account within-participant as well as between-participant variation, whereas Lozano and colleagues based their regression on between-participant variation only. Finally, and potentially most importantly, we used a laboratory and ambulatory setting and used a wider range of conditions in both settings. All conditions were selected as being regularly used in experiments on cardiac autonomic function. We even included conditions that are known to invoke apparently paradoxical changes in the PEP caused by strong afterload and preload effects (e.g. standing and cold pressor test), because the estimated PEP should behave as the actual PEP even in these conditions. With the more complete sampling from the universe of conditions in which PEP is measured in current research, it became more clear that a single regression equation relating ISTI to RB is not working as well as was expected from the first proof-of-principle study (Lozano et al., 2007). This is not unexpected or unprecedented and in no way discredits the original study. However, it does lead us to conclude that a regression equation based on ISTI predicts the RB interval with insufficient precision.

At first sight, these results are not encouraging for researchers involved in large scale data collections either in terms of many participants or in terms of prolonged (e.g. 24 hour) recordings. They show that for valid PEP scoring the detection of the Q-wave onset and B-point remains mandatory. PEP scoring is typically done after ensemble averaging the ICG waveform over all beats in a one-minute period, time locked to the R-peak. Scoring is ideally done by automated algorithms, but current practice is to always visually inspect the correct positioning of the B-point as it is hard to

detect algorithmically with sufficient accuracy in all participants. Two (or more) independent raters ideally score the PEP to reduce subjective rater bias. Multiple raters are particularly required in ambulatory studies, where ICG signal quality can be poor, for instance due to physical activity which creates bimodal ICG waveforms, making it difficult to detect the crucial B-point. This was illustrated by the lower interrater reliability for B-point scoring in the ambulatory (0.86) compared to the laboratory (0.92) setting in the current study. A practical solution is to ask participants not to engage in too many physical activities during their participation in ambulatory recording but this reduces ecological validity, which is a main asset of the ambulatory approach. The requirement of visual inspection by independent raters makes PEP scoring very time consuming and presents a major obstacle in large epidemiological studies or in prolonged ambulatory monitoring studies that generate many hours of data. A current strategy to deal with this is the use of large scale ensemble averaging across longer chunks of time, e.g. 30 minutes (Riese et al., 2003) which decreases work load for the raters and slightly increases reliability of signal scoring but also comes with a reduction in temporal precision (i.e. having one-minute PEP values across the entire 24-hour period). By far the best solution would be to increase the reliability of the automated detection of the Q-wave onset and B-point.

In that regard, our results also contain a clear positive message. The ISTI and a fixed QR interval of 40 ms do present meaningful estimates of the expected location of the Q-wave and B-point scoring. By a priori focusing the detection algoritms in a window around these expected locations should greatly help automated detection. In addition, the ISTI is fairly strongly correlated to the PEP. Of note, the within-participant correlations between ISTI and PEP were significant for *all* 118 participants and in *all* of the 38 different experimental laboratory and ambulatory conditions (data not shown). Changes in cardiac contractility are reflected in the time it takes the left ventricle to build up sufficient force to open the aortic valve (reflected in the B-point) but also in the time to takes to reach peak ventricular ejection (dZ/dt-min peak), which is reflected by the ISTI. Hence the information between PEP and ISTI strongly overlaps empirically and theoretically, and ISTI might itself be considered as a measure of cardiac sympathetic control based on physiological grounds. Clearly, before considering ISTI as an additional indicator of cardiac sympathetic responses, extensive testing against criterion measures or pharmacological validation is needed.

We conclude that for valid PEP scoring the detection of the Q-wave onset and B-point remains mandatory. PEP estimated from the R-wave and dZ/dt-min peaks should not be used to replace the actual PEP, but it could be a useful addition to the psychophysiologist's toolbox because it can help detection of the Q-wave and B-points.