

Using multilevel path analysis in analyzing 24-h ambulatory physiological recordings applied to medically unexplained symptoms

JAN H. HOUTVEEN,^a ELLEN L. HAMAKER,^b AND LORENZ J. P. VAN DOORNEN^a

^aDepartment of Clinical and Health Psychology, Faculty of Social Sciences, Utrecht University, The Netherlands

^bDepartment of Methods and Statistics, Faculty of Social Sciences, Utrecht University, The Netherlands

Abstract

A non-clinical group high on heterogeneous medically unexplained symptoms (MUS; $n = 97$) was compared with healthy controls ($n = 66$) on the within-subject relationships between physiological measures using multilevel path analysis. Momentary experienced somatic complaints, mood (tension and depression), cardiac autonomic activity (inter-beat intervals, pre-ejection period (PEP), and respiratory sinus arrhythmia (RSA)) and respiration (rate and partial pressure of CO₂ at the end of a normal expiration) were monitored for 24 h using electronic diary and ambulatory devices. Relationships between measures were controlled for diurnal variation and individual means. Only subtle group differences were found in the diurnal rhythm and in the within-subject relationships between physiological measures. For participants high on MUS, within-subject changes in bodily symptoms were related to changes in mood, but only marginally to the physiological measures. Results of the current path analysis confirm the subordinate role of cardiac autonomic and respiratory parameters in MUS.

Descriptors: Medically unexplained symptoms, Multilevel path analysis, Heart rate, PEP, Respiration, RSA

Several experimental studies have failed to find a relationship between individual differences in medically unexplained somatic complaints on the one hand, and dysregulations in stressphysiological parameters on the other (e.g., see Houtveen, Rietveld, & de Geus, 2003; Wientjes & Grossman, 1994). Although a tendency to report medically unexplained symptoms (MUS) is related to a tendency to experience negative affect (e.g., see Pennebaker, 2000), the underlying mechanism is most likely not based on exaggerated stressphysiological responses (Houtveen, Rietveld, & de Geus, 2003; Grossman, 2003). Nonetheless, the within-subject relationship between negative affect and stressphysiological arousal is unambiguous: A state of mental stress is related to the well-known physiological 'fight-or-flight' response, which is, for example, clearly observable in cardiac autonomic and respiratory measures. Therefore, it remains puzzling why MUS (that are so clearly related to negative affect) do hardly have any stressphysiological manifestations.

In a recent study, we investigated whether there were physiological differences between people with high levels of MUS and controls in a 24-h ambulatory assessment (Houtveen & van Doornen, 2007). In this study, values were mean-aggregated over 4-h time intervals. While there were profound group differences in momentary self-reported levels (i.e., mean values) of negative affect and somatic symptoms, there were no group differences in

the levels of cardiac autonomic and respiratory measures. The absence of differences in means of stressphysiological variables between people high on MUS in comparison to controls raises the question whether a history of chronic stress and negative affect (as frequently found in individuals with MUS) can affect other aspects of cardiac autonomic and respiratory parameters. A more sensitive focus may reveal more subtle differences in the amount of within-subject variability, the diurnal patterns, or the within-subject relationships (i.e., covariances) between physiological variables. This research question is based on the assumption that MUS should somehow (as physiological manifestation of negative affect) be reflected in dysregulations of stressphysiological functioning.

In our previous study (Houtveen & van Doornen, 2007), group differences in the 4-h mean-aggregated values of the stressphysiological measures were tested with repeated-measures ANOVA. This data obtained in the previous study will be re-analyzed in the current study in order to investigate group differences in within-subject dynamics in two groups differing strongly in negative affect and somatic complaints. A statistical method available to test for group differences in the within-subject relationship between repeated measures is multi-group multilevel path analysis (Muthén & Muthén, 1998–2004). Apart from the topic at stake (i.e., MUS), the present study offers the opportunity to illustrate and demonstrate the utility of using multi-group and single-group multilevel path analysis in the field of cardiac autonomic and respiratory psychophysiology. The current multilevel analysis of this dataset was specifically focused

Address reprint requests to: Jan Houtveen, Utrecht University, Faculty of Social Sciences, Clinical and Health Psychology, P.O. Box 80.140, 3508 TC Utrecht, The Netherlands. E-mail: j.h.houtveen@uu.nl

on: a) exploring group differences in the 24-h within-subject relationships between stress-physiological (i.e., cardiac autonomic and respiratory) measures; and b) exploring the within-subject relationship between bodily complaints and psychological measures and between bodily complaints and stress-physiological measures.

To be able to focus on the within-subject relationships of interest, two other sources of variance need to be accounted for. First, large endogenous (circadian and homeostatic) diurnal variation has been demonstrated in baseline values of cardiac autonomic levels (see also van Eekelen, Houtveen, & Kerkhof, 2004a; van Eekelen, Houtveen, & Kerkhof, 2004b) and respiration (Mortola, 2004). If these are unaccounted for, this will lead to high correlations between measures at the level of the individual, but this will provide little insight in the actual mechanisms that are operating at the level of the individual. Secondly, individual differences in physiological baseline values may (as an additional source of variance) mask group differences in the within-subject relationships between these measures. Stressphysiological parameters are evidently influenced by numerous subject factors that have nothing or very little to do with the influence of chronic stress and negative affect. For example, heart rate levels are influenced by genetic factors, height, fitness, smoking, etc., which may mask individual differences in stress-related influences and may therefore be considered between-subjects noise. Hence, to be able to focus on (group differences in) the within-subject relationships, both diurnal rhythm and individual means should be taken into account as additional sources of variation.

Data segments low on physical activity were selected. In the first within-subject model, we investigate whether there are group differences in the diurnal rhythm of the physiological measures, and whether there are group differences in the relationships between these measures while controlling for diurnal rhythm and differences in individual means. When these additional sources of variation are controlled for, changes in heart rate are hypothesized to be influenced by respiratory sinus arrhythmia (RSA) as a measure of cardiac vagal control, and by pre-ejection period (PEP) as a measure of cardiac sympathetic control (Berntson, Cacioppo, & Quigley, 1993; Berntson, Cacioppo, Quigley, & Fabro, 1994). RSA, however, can be influenced by respiration (frequency and depth) without underlying change in cardiac vagal control (as reviewed by Berntson, et al., 1997). An effect of changes in (arterial) carbon dioxide pressure (PCO_2) on RSA has also been demonstrated (Al-Ani, Forkins, Townend, & Coote, 1996; Cooper, Parkes, & Clutton-Brock, 2003; Houtveen, Rietveld, & de Geus, 2002). However, the interrelation of PCO_2 and respiratory parameters in their influence on RSA is a complex one. Including both of them at least will give sight on their independent contributions. In our model, changes in RSA are hypothesized to be influenced by respiratory frequency¹ and by PCO_2 . Changes in respiratory frequency and PCO_2 are theoretically coupled in two directions. First a change in PCO_2 results in a change of blood pH, and to maintain pH homeostasis there are sensors within the brainstem that monitor CO_2 levels and control ventilation to keep blood PCO_2 within the normal physiological

range (e.g., see Richerson, 2004). Additionally, stress/anxiety-related hyperventilation reduces the PCO_2 level (e.g., see Wilhelm, Gevirtz, & Roth, 2001). Therefore, respiratory frequency is predicted to correlate with partial pressure of CO_2 . The hypothesized model including all the within-subject relationships between physiological measures as described above is further explained in the methods section (see also Figure 2). Specific a priori predictions of group differences in this model could not easily be extracted from the literature. This part is, therefore, exploratory.

In the second within-subject model we investigate the diurnal rhythm of momentary bodily complaints (assessed with an electronic diary method), and the relationship between bodily complaints, negative mood, and stressphysiological measures after controlling for diurnal variation and individual differences in means. A relationship between MUS and negative affect has been reported frequently in the literature (Brown & Moskowitz, 1997; Houtveen, Rietveld, & de Geus, 2003; Pennebaker, 2000; Watson & Pennebaker, 1989). For instance, experimental studies have shown that mental stress can cause bodily complaints (e.g., see Houtveen, Rietveld, & de Geus, 2003). Therefore, changes in bodily complaints are hypothesized to be influenced by changes in negative mood. Because previous studies could not consistently demonstrate exaggerated stressphysiological responses in relationship to MUS, changes in bodily complaints are hypothesized not to be influenced by changes in stressphysiological parameters, or only to a minor degree. The hypothesized model including all the within-subject relationships between self-reported and physiological measures as described above is further explained in the methods section (see also Figure 3). Because the control participants reported very few bodily complaints and showed very little within-subjects variation on these, the second model could only be applied to participants high on MUS.

Results of the current exploratory study could provide insight in the existence of different diurnal fluctuations or deviant within-subject relationships between the cardiac autonomic and respiratory systems for individuals high on MUS as compared to controls. These results may also provide further insight in the within-subject relationship between self-reported complaints and mood versus activity of stressphysiological systems. The current study is also intended to be an example of using (multi-group) multilevel path analysis in analyzing the diurnal rhythm and regulatory mechanism of the cardiac autonomic and respiratory systems.

Methods

Participants and Inclusion Questionnaires

A detailed description of how participants were recruited and the inclusion and exclusion criteria have been provided elsewhere (Houtveen & van Doornen, 2007). Briefly, a non-clinical group of participants high on (heterogeneous) MUS (HMUS) were recruited through a website. They were included based on scoring ≥ 27 on the somatization subscale of the Symptom Check List (SCL-90-R) (Arrindell & Ettema, 1981) in combination with the absence of a medically diagnosed physical disease (despite frequent medical testing; based on self-reports assessed by a second questionnaire). The used SCL-90-R somatization cut-off score (27) was the median of the norm-distribution of Dutch psychiatric outpatients and the 95th percentile of the upper part of the

¹Respiratory depth also influences RSA, and inclusion of this parameter would have been of interest to complete our model. A reliable measurement of respiratory volume was, however, not possible with the ambulatory thoracic impedance technique we have used in the current study.

norm-distribution of Dutch normal controls. The control participants low on symptoms (LMUS; cut-off score ≤ 17) were recruited through the internet, advertisements in local papers, or were nearby volunteers. See Table 1 for participant information. The Ethics Committee of the University Medical Centre Utrecht (UMCU) approved the study protocol, and all subjects gave written consent before entering the study.

Ambulatory Measurement Devices

The ambulatory electrocardiogram (ECG) and impedance cardiogram (ICG) were measured using the VU-AMS (version 4.3, TD-FPP, Vrije Universiteit, Amsterdam, The Netherlands; www.psy.vu.nl/vu-ams). Vertical acceleration of the torso was additionally monitored and used as a proxy for gross body movement (motility). For reasons of memory limitations, continuous registration was set to 5 min per 15 min period, resulting in four data segments per hour. Details on the recording methodology, reliability, and validity of the VU-AMS can be found elsewhere (de Geus & van Doornen, 1996; de Geus, Willemsen, Klaver, & van Doornen, 1995; Houtveen, Groot, & de Geus, 2006; Riese et al., 2003; Willemsen, de Geus, Klaver, van Doornen, & Carroll, 1996).

Palm M130 (www.PalmOne.com) Personal Digital Assistants (PDA) were used for (electronic diary) experience sampling of symptoms, mood, and activity during the measurement day. Special software was developed for the generation of alarms (a 10-s auditory signal) during the day, and for the assessments of somatic complaints and mood. A fixed sampling protocol with an interval of 1.5 h was used that continued from awakening till bedtime. This sampling scheme led to approximately 12 alarms per day (based on a 16-h awakening period). Diary prompting was only disabled during sleep, initiated by a button on the PDA. Next, the PDA could be used as a morning alarm, and prompting continued after awakening. All unused buttons were blocked. Alarms without response were repeated (max 3 times with 3-min time interval). The alarm software generated a log-file containing alarm and response times to be used for determination of the compliance. The questionnaire was launched by a start button that was visible for 12 min after prompting. All questions were forced-choice, and they were displayed as sequential screens on the PDA. Participants were not allowed to leaf through the

present or previous diaries. The self-reported dimensions measured on the PDA were: 1) somatic complaints, 2) tiredness, and 3) mood: depression and tension. Somatic complaints were measured by the 12 somatic complaint items used for selection based on the somatization subscale of the SCL-90-R. Tiredness, depression, and tension were measured by 3 items, each adapted from the shortened version of the Profile of Mood States (POMS) (Shacham, 1983). Each item could be rated on a seven-point scale ranging from 1 'not at all' to 7 'very much.'

The TG-951T CO₂ quantitative sensor Kit (Nihon Kohden Corporation, Tokyo, Japan) was used to measure the partial pressure of CO₂ at the end of a normal expiration (PetCO₂). PetCO₂ was measured during the day using a mainstream adapter (once every 1.5 h initiated just after completing the diary questions) and during sleep using an air pump device in combination with a sidestream (nostrils) adapter (assessed once every 15 min). The PDA automatically initiated PetCO₂ assessments. Digitized PetCO₂ values (in mmHg) from the capnometer were sent to the PDA computer for storage through an RS-232 interface cable.

Procedure

Participants who met the inclusion criteria and who agreed to participate were contacted by telephone to receive more information about the study and to make an appointment for the 24-h ambulatory recordings. They were visited at home where they completed additional questionnaires. Next, they were supplied with the VU-AMS, the PDA, and the capnometer. They received detailed verbal and written instructions on how to use all equipment. Participants practised how to complete the questions on the PDA. Next, they were instructed on the CO₂ assessment. Specific instructions were given how to breathe as normally as possible during mainstream CO₂ assessments, which was also practiced under supervision. They got further instructions on how to regularly check the 'all clear' signal of the VU-AMS device (a small blinking light on the side of the device) and how to respond to measurement alarms (e.g., a loose electrode contact) and electronic diary promptings. Twenty-four hour telephone assistance was available. Participants followed their normal day routines. The following day, the researcher came back to collect the devices.

Data Preparation

The recorded heart period time series (recorded as inter-beat intervals, IBI) together with the motility signal were displayed as 5-min segment cardiograms for visual inspection by two independent raters. Segments containing too many artifacts (i.e., > 5) were reduced in length (sub-segments low in artifacts should be at least 2.5 min) or rejected (no sub-segments could be selected). Because individual differences in motility could be responsible for group differences in the physiological measures, only segments (or sub-segments 2.5–5 min) low on motility were selected. 'Low on motility' was based on the mean participant-specific motility value obtained during sleep. In 89% of the cases, a (sub)segment low in motility could be found. Note that low motility (sub)segments were found only during sleep (by definition) and (while awake) during relaxation in sitting or supine position. These (sub)segments were used for segment-specific analysis of the cardiac and respiratory measures (see below).

First, an artifact pre-processing was performed on the selected segments. Artifacts were detected automatically when greater than a user-predefined percentage of the standard deviation or

Table 1. Participant Information

	LMUS (n = 66)		HMUS (n = 97)		Sig.
	M	SD	M	SD	
Male (%)	30.3		22.7		n.s.
Smoking (%)	25.8		25.8		n.s.
Age (years)	33.83	8.10	37.38	8.39	n.s.
SCL90R Som	14.23	1.66	33.30	7.36	<.001
SCL90R Anx	11.03	1.22	22.96	8.1	<.001
SCL90R Dep	18.31	2.13	36.20	9.92	<.001
No. diary measures	11.89	1.46	11.89	0.99	n.s.
No. PCO ₂	25.39	3.09	24.90	4.89	n.s.
No. PEP	45.58	2.15	45.87	1.74	n.s.
No. IBI & RSA	45.88	2.05	46.14	1.55	n.s.
No. Rfreq	45.68	2.70	46.14	1.55	n.s.

Notes: LMUS = low on medically unexplained somatic symptoms; HMUS = high on medically unexplained somatic symptoms; No. = mean number of valid observations per participant; n.s. implies not significant at $\alpha = .05$.

mean value in deviation from the moving mean of a particular segment, and accepted or overruled by visual inspection. IBI artifacts were found and corrected for 0.07% ($SD = 0.06$) of the IBI values. Since artifacts cannot simply be deleted, because the continuity of time would be lost, spuriously short IBI's were summed and missing beats were 'created' by splitting spuriously long IBI's. The IBI mean values were computed from these corrected data for each segment.

Next, uniformly spaced samples were created by interpolation of the corrected IBI data using a Wavelet interpolation algorithm. Discrete Wavelet Transformation (DWT) was performed using a cardinal cubic spline function as base. This method results in identical power values for stationary, relatively short data segments as compared to Fourier transformation (Houtveen & Molenaar, 2001), but it is superior for non-stationary data segments. Since the DWT (like Fourier) suffers from aliasing effects at both ends, the first and last 2.5 sec of the time series were excluded from the derivation of the variances. The High Frequency (HF) power (the variance in the 0.125–0.5 Hz window) was computed as the main measure of cardiac parasympathetic (vagal) control (Berntson et al., 1997).

Changes in the thoracic impedance (dZ) were used to assess respiration. Respiratory-related changes in dZ were obtained by band-pass filtering (0.125–0.5 Hz) of the ICG signal using a DWT filter with a cubic spline function as base. Respiratory frequencies were obtained from the band-pass filtered dZ signal by counting the number of up-going zero crossings and dividing this value by the time of a segment as described elsewhere (de Geus, Willemsen, Klaver, & van Doornen, 1995; Houtveen, Groot, & de Geus, 2006).

Pre-ejection period (PEP) reflects the time interval between the onset of the electromechanical systole (Q-wave onset) in the ECG and the onset of left ventricular ejection at the opening of the aortic valves (B-point) in the ICG. B-points were manually scored by two independent raters using the VU-AMS interactive software, which graphically displays the large-scale ensemble averages ICG signal (i.e., averaged per (sub)segment). One first rated all the B-points and the second evaluated all these ratings. Differences between raters were discussed. The B-points were determined for each ensemble averaged (selected) segment, and the PEP values were determined by summing a fixed Q-to-R interval of 48 ms to the R-B interval time. PEP was computed as measure of cardiac sympathetic control (Sherwood et al., 1990).

End-tidal partial pressure of CO₂ (PCO₂) values were assessed with 1.5 h-intervals during daytime (based on four respiratory cycles) and 15-min intervals during sleep (based on ten respiratory cycles). The Nihon Kohden TG-951T inbuilt algorithm was used for determination of end-tidal plateaus. To reduce the possibility of misdiagnosed plateaus, the largest value of the respiratory cycles within a segment (four during the day; ten during the night) was taken as the most reliable value (i.e., based on the idea that an end-tidal value could not exceed the true arterial partial pressure of CO₂). Visual inspection was performed on the sleep data to reject incidental low values due to mouth breathing. Suspicious values were evaluated in combination with respiration.

Because a time stamp was always added during assessment to a physiological value obtained, mean IBI, RSA, PEP and respiratory frequency (Rfreq) segment values could be coupled to a specific time. Twenty-four hour physiological assessments were converted into 30-min mean-aggregated values, resulting in 48 time-specific values per participant for each physiological mea-

sure. By this method, only a few physiological cells were missing. PCO₂ values during sleep were also converted into 30-min mean-aggregated time-specific values, resulting in (about) 16 values per participant. PCO₂ values during daytime and electronic diary measures of symptoms (somatic complaints and tiredness) and mood (tension and depression) were available (again with a time stamp) with an interval of 1.5 h (for 16 h; about 12 observations per participant). These values were each coupled to one of the 48 time windows. Because daytime PCO₂ and diary values were not available for each 30-min time window, more missing data existed for these measures. The mean numbers of observations used per group are given in Table 1.

The electronic diary data and RSA values were ¹⁰log transformed to obtain normal distributions. Diary and physiological values were Z-transformed (i.e., grand-mean centering combined with standardization across all observations and groups) to improve convergence and to facilitate interpretation of multilevel analysis (Hox, 2002, p 54–58).

Statistical Analyses

The multilevel analyses discussed in the introduction were performed using Mplus v4 (Muthén & Muthén, 1998–2004). We used the two-level option in Mplus and full information maximum likelihood estimation (i.e., maximum likelihood full (MLF) in Mplus), which handles missing data as missing at random. All the crucial models (i.e., the hypothesized models and the improved models) have additionally been tested with the robust chi-square statistic (i.e., maximum likelihood robust (MLR) in Mplus). This chi-square is robust against violations of the normality assumption, but it cannot be directly used for comparison of nested models. The results obtained with MLF and MLR did not lead to different conclusions regarding overall model fit. To compare nested models, we used the chi-square statistic obtained with MLF.

In the first set of analyses, the focus was on: a) group differences in diurnal variation of cardiac autonomic and respiratory measures; and b) group differences in the within-subject relationships between these measures while controlling for possible group differences in diurnal variation and differences in individual means. In the second set of analyses, the focus was on: a) the within-subject relationships between the diary measures (symptoms and mood); and b) the within-subject relationships between the diary and physiological measures while controlling for diurnal variation and differences in individual means. Because people in the LMUS group showed little or no variation on complaints, the second set of analyses was restricted to the HMUS group.

As indicated in the introduction, individuals may differ with respect to their means. To separate this source of variance from the within-subject variability, which we are interested in, we included random means (i.e., intercepts) for all the measures in our model. These random means are regressed on several person characteristics, that is, gender, smoking status, and age. Using α_{ikj} to denote the mean of individual i in group k on variable j , the model at the between-subjects level can be expressed as

$$\alpha_{ikj} = \alpha_{(0)kj} + \alpha_{(1)kj} \text{Gender}_i + \alpha_{(2)kj} \text{Smoker}_i + \alpha_{(3)kj} \text{Age}_i + u_{ikj}, \quad (1)$$

where $\alpha_{(0)kj}$ is the intercept on variable j in group k , $\alpha_{(1)kj}$ is the effect of gender on variable j in group k , $\alpha_{(2)kj}$ is the effect of being a smoker on variable j in group k , and $\alpha_{(3)kj}$ is the regression

parameter for the effect of age on variable j in group k . We do not expect differences between the effects of gender, smoking, and age across the groups, but this expectation will be tested. The residual u_{ikj} is the part of the individual mean that cannot be predicted from these variables. Within each group, the residuals are assumed to come from a normal distribution with mean zero and standard deviation σ_{kj} .

Group differences in diurnal variation of the cardiac autonomic and respiratory measures were tested by performing a multi-group multilevel within-between subject path analysis. To this end, we created 24-h sine and cosine functions to model both amplitude and phase (see also van Eekelen, Houtveen, & Kerkhof, 2004a), and added these as two within-subject variables to the model, such that the model at the within-subject level can be expressed as

$$Y_{ikjt} = \alpha_{ikj} + \beta_{kj} \sin(\omega t) + \gamma_{kj} \cos(\omega t) + e_{ikjt}, \quad (2)$$

where Y_{ikjt} is the observed score of individual i in group k on variable j at occasion t . It can be seen that this score depends on the individuals mean α_{ikj} (for which we have the expression in Equation 1), and on the sine and cosine functions. The regression parameters for the sine and cosine functions were allowed to differ across the two groups as becomes clear from the group index k , but within each group these parameters were deliberately held fixed, such that individuals within the same group were modeled to have the same phase and amplitude. The reason for not making the sine and cosine parameters random is that this would result in arbitrary individual phase shifts, which are unrelated to diurnal rhythm.

To illustrate the model in a simplified manner, we have depicted the within-between subject model for one variable as an example in Figure 1, using the graphic representation employed by Muthén and Muthén (1998–2004). Squares indicate observed variables, while circles indicate latent variables. The black circle at the top of Figure 1 (within-subject level) represents the random mean α_{ikj} , which varies across individuals but not across time. It corresponds to the open circle (identified by α_{ikj}) at the bottom of Figure 1 (between-subject level), where it is regressed on stable person characteristics.

To investigate group differences in the within-subject relationships between the residual cardiac autonomic and respiratory measures, we make use of the residuals e_{ikjt} . Such residuals represent the part of the individual's score that could not be explained by the group's diurnal trend and the individual's mean. Hence, we can interpret it as the individual's value on a physiological measure after the diurnal cyclic effect and the individual's mean are both partialled out. In particular, we were interested in whether and how the residuals of the five physiological measures are related and whether these relationships differ across the groups.

The second set of analyses consisted of performing a multi-level within-between subject path analysis on the self-reported diary measures of the HMUS participants, and to relate the diary measures to the physiological measures. Specifically, we tested whether fluctuations over the day in depression, tension, and the five physiological measures were related to fluctuations in somatic complaints and tiredness, while controlling for diurnal variation and differences in individual means. Throughout, p -values $< .05$ were considered significant.

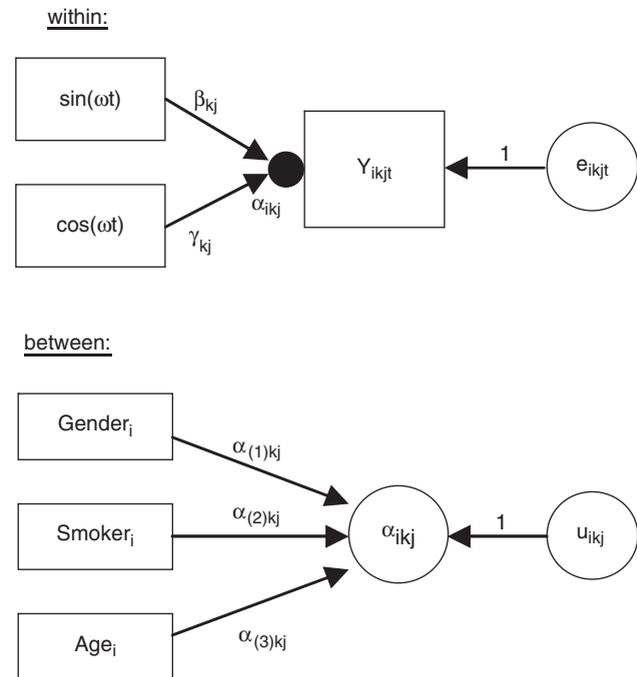


Figure 1. The within-between subject model for a single variable. At the within-subjects level, a measure is controlled for the individual mean (as indicated by the black circle on the left of the square), and the diurnal variation (by regressing on the sine and cosine functions). The individual mean is modeled at the between-subjects level, that is, the large circle corresponds to the black circle in the within-subjects part, and is predicted from gender, being a smoker, and age.

Results

Regarding the diary measures, groups did not differ on their reports of being at work during daytime assessments (LMUS 45% vs. HMUS 47%, $p = .68$), but HMUS participants reported more tension, depression, tiredness, and somatic complaints than LMUS participants ($p < .001$ for all four t -tests). There were no significant group differences on the 24-h mean values of PEP, IBI, RSA, Rfreq and PCO_2 ($ps > .05$). Results of group differences on the 24-h variances (within and between-subject) are presented in Table 2. Regarding the within-subject variances, no significant group differences were found for IBI and PCO_2 , but the HMUS group had significantly smaller variances on RSA and PEP, and a significantly larger variance on Rfreq. At the between-subjects level, there were no significant differences between the variances of these five physiological measures of the two groups.

Multi-Group Multilevel Path Analysis on the Cardiac Autonomic and Respiratory Measures

Multi-group multilevel path analyses were performed to test for group differences in the diurnal variation of cardiac autonomic and respiratory measures, and for group differences in the within-subject relationships between these measures. In order to control for possible differences in individual means, we included a random intercept (i.e., the individual's mean) which was regressed on gender, smoking status, and age. Using a chi-square difference test to test for group differences, we concluded that there were no significant differences between the groups regarding the influence of gender, smoking, and age on the physiolog-

Table 2. Group Differences for the Within and Between-Subject Variances of the Cardiac Autonomic and Respiratory Estimates

Within subject	LMUS	HMUS	$\Delta\chi^2(1)$
PEP	.47	.39	27.5***
IBI	.57	.58	n.s.
RSA	.46	.42	6.3*
Rfreq	.63	.71	11.5***
PCO ₂	.45	.47	n.s.
Between subject			
PEP	.55	.59	n.s.
IBI	.35	.47	n.s.
RSA	.62	.52	n.s.
Rfreq	.27	.35	n.s.
PCO ₂	.58	.51	n.s.

Notes: Values were initially Z-transformed (across all observations and groups). n.s. implies not significant at $\alpha = .05$. * $p < .05$, *** $p < .001$.

ical means at the between-subjects level ($\Delta\chi^2(15) = 14.3$, $p = .50$). Hence, these paths were constrained (to be equal) across the groups.

In order to control for group differences in diurnal variation, we modeled the within-subject relationships between the residuals (i.e., ϵ_{ikjt} in Equation 2) rather than between the variables themselves (i.e., Y_{ikjt} in Equation 2). Results of the multi-group multilevel path analysis are presented in Figure 2 and Table 3. The sine and cosine estimates in Figure 2 have been transformed to polar coordinates (i.e., amplitudes and phases). The hypothesized model (as depicted in Figure 2), which allows between-

group differences in the diurnal variation and in the within-subject relationships between the residuals of the physiological measures, yielded an acceptable fit (CFI = .97, RMSEA = .03, SRMR_{within} = .02, SRMR_{between} = .08), but the χ^2 -test was significant ($\chi^2(57) = 300.5$, $p < .001$; robust $\chi^2(57) = 180.7$, $p < .001$). The fit could be improved by allowing: a) a covariation between ϵ IBI and ϵ Rfreq, b) a covariation between ϵ PEP and ϵ RSA for the LMUS group, and c) a covariation between ϵ PEP and ϵ PCO₂ for the LMUS group. Covariations between these stressphysiological measures could theoretically be explained by a common arousal factor. Adding them indeed increased the fit (CFI = 1.0, RMSEA = .009, SRMR_{within} = .003, SRMR_{between} = .076, $\chi^2(52) = 69.1$, $p = .056$; robust $\chi^2(52) = 51.8$, $p = .48$).

The amplitudes and phases shown in Figure 2 can be compared directly between the two groups. Group differences in diurnal variation were tested by constraining paths and testing $\Delta\chi^2$ values. Except for PEP, the groups differed in the diurnal variation on all other cardiac and respiratory measures (all $\Delta\chi^2(2) > 6.0$, $ps < .05$). Nonetheless, the amplitudes for IBI, RSA, and Rfreq were remarkably similar across the two groups. It was only substantially different for PCO₂: HMUS participants showed less diurnal variation in PCO₂ (.13) as compared to LMUS participants (.21), $\Delta\chi^2(2) = 25.0$, $p < .001$. Regarding the phases, the 24-h sine function was shifted to the left for the HMUS participants on all five physiological measures.

With respect to the within-subject relations between the five physiological measures (controlled for diurnal variation and individual means) we found that ϵ IBI could be predicted from ϵ PEP and ϵ RSA, and ϵ RSA could be predicted from ϵ Rfreq and

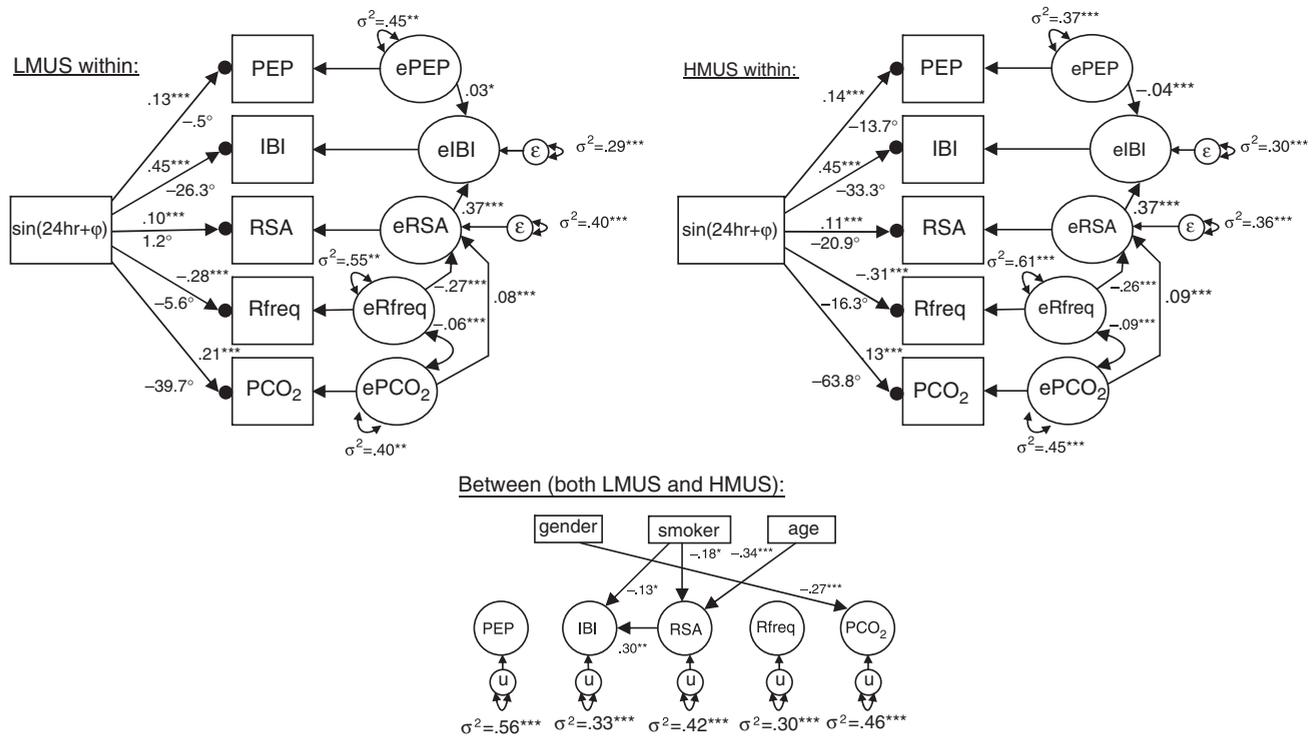


Figure 2. Multi-group multilevel path analysis results of cardiac autonomic and respiratory measures. The hypothesized within-subject model states that the corrected IBI score (i.e., corrected for diurnal variation and individual differences in mean), can be predicted from the corrected PEP and RSA scores, and that the corrected RSA score can be predicted from the corrected Rfreq and PCO₂ scores; Rfreq and PCO₂ are assumed to be correlated. The individual means for IBI at the between-subject level was assumed to be predictable from the individual means on PEP and RSA. Although the between-subject model allowed prediction of all measures from gender, being a smoker, and age, only the significant relationships are depicted here.

Table 3. Group Differences in the Cardiac Autonomic and Respiratory Within-Subject Estimates

Hypothesized	LMUS	HMUS	$\Delta\chi^2(1)$
ePEP → eIBI	.03**	-.04***	14.9***
eRSA → eIBI	.37***	.37***	n.s.
eRfreq → eRSA	-.27***	-.26***	n.s.
ePCO ₂ → eRSA	.08**	.09***	n.s.
eRfreq ↔ ePCO ₂	-.06***	-.09***	n.s.
Added			
eRfreq ↔ ε eIBI	-.07***	-.07***	n.s.
ePEP ↔ ε eRSA	-.04***	.00	19.5***
ePEP ↔ ePCO ₂	-.04***	.00	9.0**

Notes: Estimates are unstandardized and can be compared between groups. Hypothesized relationships are given at the top. Adding the covariances at the lower part did not influence the significance of the hypothesized relationships at the upper part, it only improved the fit of the total model. n.s. implies not significant at $\alpha = .05$. ** $p < .01$, *** $p < .001$.

ePCO₂ as had been hypothesized (see Figure 2). Thus, all the within-subject relationships that were hypothesized proved significant. However, residual variances (denoted as $\sigma^2\epsilon$ in Figure 2) remained relatively large. The proportion of unexplained variance in the LMUS and the HMUS groups, respectively, were .79 and .81 for IBI and .90 and .88 for RSA.

Test results for group differences in the physiological estimates are presented in Table 3. We tested whether the relation-

ships were identical across the two groups through constraining these to be equal and using the chi-square difference statistic. A significant (but small) difference was found for the contribution of ePEP to eIBI, which was positive for the LMUS participants, while it was negative for the HMUS participants. For all the other hypothesized relationships, no significant differences were found across the two groups. The added (unpredicted) covariances between ePEP and ε eRSA, and between ePEP and ePCO₂ were found to be significant only for the LMUS participants, and constraining these covariances across the two groups led to a significant chi-square difference. The added covariance between eRfreq and ε eIBI was not significantly different across the two groups. Despite these subtle group differences, overall these results indicate a common underlying regulatory mechanism between the stressphysiological measures assessed.

Multilevel Path Analysis on the Diary Data for the HMUS Group

For the diary data it was hypothesized that bodily complaints (as measured by Tiredness and Somatic complaints) are affected by negative affect (as measured by Depression and Tension), but not by stressphysiology (as measured by PEP, IBI, RSA, Rfreq, and PCO₂). Again, all measures are controlled for individual differences in means and for diurnal variation. In Figure 3, the standardized results of the multilevel path analysis based on the self-reported diary data for the HMUS participants are depicted. The model yielded a good fit (CFI = 1.00, RMSEA = .00, SRMR_{within} = .001, SRMR_{between} = .02, $\chi^2(8) = 3.2$, $p = .92$; robust $\chi^2(8) = 3.2$, $p = .92$). Although the effects of negative affect on bodily complaints were significant, the proportion of

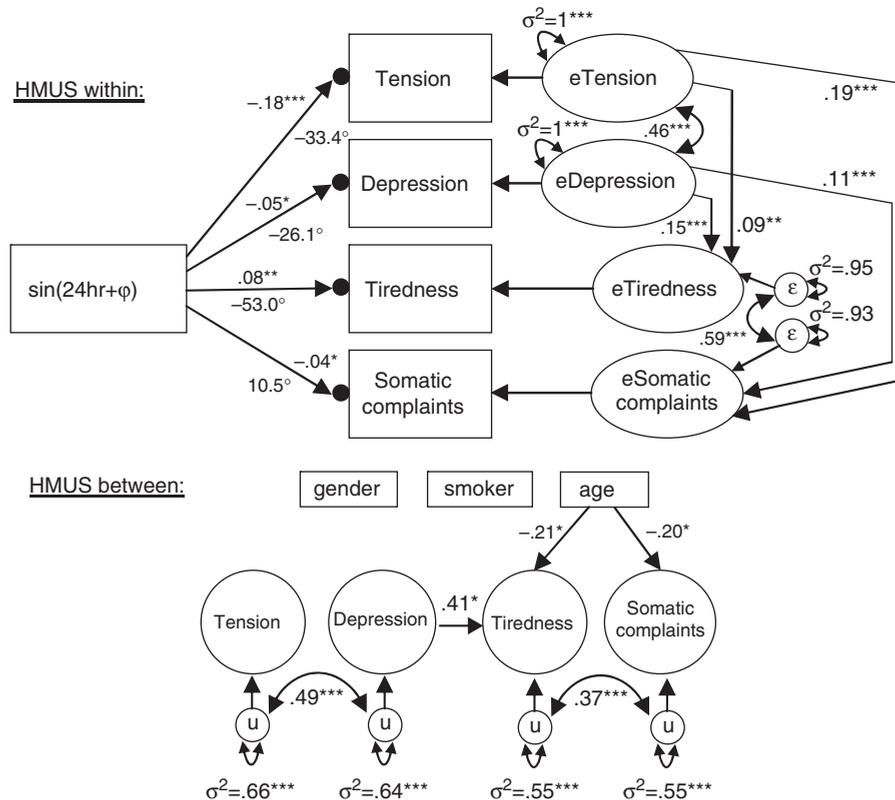


Figure 3. Multilevel path analysis results of the self-reported diary data for the HMUS participants. The hypothesized within-subject model states that the corrected Tiredness and Somatic complaints scores can be predicted from the corrected Tension and Depression scores. The residuals of Tension and Depression and of Tiredness and Somatic complaints are assumed to be correlated at both the within-subjects and the between-subjects level. Although the between-subjects model allowed prediction of all measures from gender, being a smoker, and age, only the significant relationships are depicted here.

unexplained variances in eTiredness and eSomatic complaints (i.e., the $\sigma^2\epsilon$ s) remained large (i.e., .95 for eTiredness and .93 for eSomatic complaints).

Finally, each of the five physiological measures (corrected for diurnal variation and differences in individual means) were added separately to the multilevel path model described as predictors of the bodily complaints. Small but significant effects were found for ePEP on eSomatic complaints (regression coefficient = .10, $p < .05$, $\Delta\sigma^2\epsilon$ somatic complaints = 1%), and for eIBI on eTiredness (regression coefficient = .15, $p < .01$, $\Delta\sigma^2\epsilon$ somatic complaints = 2%). No significant contributions were found for eRSA, eRfreq, and ePCO₂ on either eTiredness or eSomatic complaints.

Discussion

In a previous study, we have shown that a non-clinical group of participants reporting heterogeneous MUS did not deviate from a control group in the 4-h mean-aggregated values of cardiac autonomic (PEP, IBI, and RSA) and respiratory (respiratory frequency and PCO₂) measures (see Houtveen & van Doornen, 2007). The focus of the current study was on the 24-h within-subject relationships between these measures. While the mean-aggregated values were similar, we found subtle but significant group differences for the within-subject variances. Participants of the HMUS group had smaller within-subject variances on RSA and PEP, and a larger within-subject variance on respiratory frequency. The multi-group multilevel path analysis presented in the current study yielded additional (but subtle) group differences in the 24-h diurnal rhythm of the physiological measures (most pronounced for PCO₂), and in the within-subject relationships between these measures (most pronounced for PEP). This clearly demonstrates the additional value of using (multi-group) multilevel path analysis in analyzing the diurnal rhythm and regulatory mechanism of the cardiac autonomic and respiratory systems.

The overall conclusion, however, is that the within-subject relationships between cardiac autonomic and respiratory measures are remarkably similar for people with high levels of MUS as compared to controls (see Table 3). This demonstrates that long-term homeostatic regulatory processes can, to a large degree, compensate for the physiological consequences of chronic stress and negative affect on the body in individuals high on MUS. Hence, while brief periods of mental stress clearly influence cardiac autonomic and respiratory activation levels, a chronic state is only reflected in subtle changes in within-subject dynamics. This confirms our previous statement, based on between-group comparison of mean values, that MUS are probably not mediated by anomalies in stress-physiological responses.

The current study nonetheless demonstrated small group differences in the within-subject variances of PEP and respiratory frequency, in the diurnal rhythms of some of the stress-physiological measures, and in the within-subject relationship between PEP and IBI, PEP and RSA, and PEP and PCO₂. The results found for PEP may indicate a β -receptor down-regulation for individuals high on MUS (see also Vrijkotte, van Doornen, & de Geus, 2004). The question is, however, whether the results of the multilevel path analyses as applied in the current study have truly demonstrated a different diurnal rhythm and/or deviant relationships between stress-physiological measures. The answer is not straightforward. The group differences found in the diurnal

rhythms of the stress-physiological measures and in the within-subject relationships between PEP and the other physiological measures may also be explained by behavioral differences. Although group differences in physical activity were ruled out by selection of data (i.e., (sub)segments low on physical activity) other behavioral differences like sleeping habits and postural differences remained. Exploratory analysis of the electronic diary data showed that, although the mean going to bed and wake-up times did not differ significantly between groups, participants high on complaints slept longer ($t(169) = 2.02$, $p < .05$). Hence, we cannot rule out the effects of sleep schedule.

In addition, the groups may have differed in their time-distribution of posture over the day (i.e., moments of lying-down during the day). Because reliable ambulatory assessment of posture requires complex technical solutions (Bussmann, et al., 2001), posture was not assessed in the current study. However, previous studies have demonstrated that posture can influence respiration and PCO₂ (Yoshizaki, Yoshida, Hayashi, & Fukuda, 1998), and it is considered an important artifact in the assessment of PEP as an index of cardiac sympathetic activity (Houtveen, Groot, & de Geus, 2005; Sherwood et al., 1990). In the current study, especially the *negative* relationship that was found between PEP and IBI for the HMUS group does not warrant a conclusion in terms of a β -receptor down-regulation, because cardiac sympathetic influences are known to result in a *positive* relationship between changes in PEP and IBI (Berntson, Cacioppo, & Quigley, 1993; Berntson, Cacioppo, Quigley, & Fabro, 1994). Finally, the groups differed in the percentage of persons who were regularly engaged in sports, and this may have confounded our physiological results (see also Houtveen & van Doornen, 2007). Thus, the relevance of the small group differences found in the relationships between physiological measures is doubtful.

Multilevel path analysis applied to the self-reported diary measures showed that negative affect was related to bodily complaints: Increased reports of tension and/or depressive mood were (independent of diurnal variation) related to increased bodily complaints. This confirms the well-established relationship between negative affect and MUS (Brown & Moskowitz, 1997; Pennebaker, 2000; Watson & Pennebaker, 1989). As predicted, a strong within-subject relationship between stress-physiological measures and bodily complaints (i.e., a within-subjects concordance) could not be demonstrated. Bodily symptoms were only to a minor degree related to changes in PEP and IBI, but not to changes in RSA, respiratory frequency, and PCO₂. Because bodily complaints may lead to rest behavior, it may again be argued that posture (lying down) was responsible for the small relationship found between PEP, IBI, and complaints. Differences in sampling rates between the diary (1.5 h) and physiological measures (15 min) may also limit the conclusions based on the multilevel path analysis applied to the relationship between self-reported diary and physiological measures. Nonetheless, the absence of a clear relationship between complaints and stress-physiology confirms our previous findings of rather a role for exaggerated perception of *normal* physiological responses to stress in people suffering from MUS (Houtveen, Rietveld, & de Geus, 2003).

In sum, the multilevel analyses as applied in the current study have clearly illustrated the potential of demonstrating within-subject differences that cannot be detected by traditional statistical methods based on mean-aggregation of within-subject data. Despite some small group differences that were found in the current study, we conclude that there is no evidence for deviant

stressphysiological regulatory mechanisms in individuals high on heterogeneous MUS. In addition, a clear within-subject relationship between stress-physiological activity and self-reported complaints was also not established. However, because some pragmatic choices had to be made due to the ambulatory nature

of the study, only a few stress-physiological measures were assessed (i.e., cardiac autonomic and respiratory frequency), and a non-clinical group of participants with complaints was included, the possibility remains that deviant values could be found for other stress-physiological parameters or in clinical (sub-)groups.

REFERENCES

- Al-Ani, M., Forkins, A. S., Townend, J. N., & Coote, J. H. (1996). Respiratory sinus arrhythmia and central respiratory drive in humans. *Clinical Science*, *90*, 235–241.
- Arrindell, W. A., & Ettema, J. H. M. (1981). *Dimensionele structuur, betrouwbaarheid en validiteit van de Nederlandse bewerking van de Symptom Checklist (SCL-90)*. [Dimensional structure, reliability and validity of the Dutch translation of the Symptom Checklist (SCL-90)]. Lisse, The Netherlands: Swets & Zeitlinger.
- Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1993). Cardiac psychophysiology and autonomic space in humans: Empirical perspectives and conceptual implications. *Psychological Bulletin*, *114*, 296–322.
- Berntson, G. G., Cacioppo, J. T., Quigley, K. S., & Fabro, V. T. (1994). Autonomic space and psychophysiological response. *Psychophysiology*, *31*, 44–61.
- Berntson, G. G., Thomas Bigger, J. Jr, Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., et al. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, *34*, 623–648.
- Brown, K. W., & Moskowitz, D. S. (1997). Does unhappiness make you sick? The role of affect and neuroticism in the experience of common physical symptoms. *Journal of Personality and Social Psychology*, *72*, 907–917.
- Bussmann, J. B. J., Martens, W. L. J., Tulen, J. H. M., Schasfoort, F. C., Van Den Berg-Emons, H. J. G., & Stam, H. J. (2001). Measuring daily behavior using ambulatory accelerometry: The activity monitor. *Behavior Research Methods, Instruments, and Computers*, *33*, 349–356.
- Cooper, H. E., Parkes, M. J., & Clutton-Brock, T. H. (2003). CO₂-dependent components of sinus arrhythmia from the start of breath holding in humans. *American Journal of Physiology—Heart and Circulatory Physiology*, *285*(54-2), H841–H848.
- de Geus, E. J. C., & van Doornen, L. J. P. (1996). Ambulatory assessment of parasympathetic/sympathetic balance by impedance cardiography. In J. Fahrenberg & M. Myrtek (Eds.), *Ambulatory assessment. computer assisted psychological and psychophysiological methods in monitoring and field studies* (pp. 141–164). Berlin: Hogrefe & Huber.
- de Geus, E. J. C., Willemsen, G. H. M., Klaver, C. H. A. M., & van Doornen, L. J. P. (1995). Ambulatory measurement of respiratory sinus arrhythmia and respiration rate. *Biological Psychology*, *41*, 205–227.
- Grossman, P. (2003). Challenges to conventional thinking about mind and body. *Journal of Psychosomatic Research*, *55*, 491–492.
- Houtveen, J. H., Groot, P. F., & deGeus, E. J. (2005). Effects of variation in posture and respiration on RSA and pre-ejection period. *Psychophysiology*, *42*, 713–719.
- Houtveen, J. H., Groot, P. F. C., & deGeus, E. J. C. (2006). Validation of the thoracic impedance derived respiratory signal using multi-level analysis. *International Journal of Psychophysiology*, *59*, 97–106.
- Houtveen, J. H., & Molenaar, P. C. (2001). Comparison between the Fourier and Wavelet methods of spectral analysis applied to stationary and nonstationary heart period data. *Psychophysiology*, *38*, 729–735.
- Houtveen, J. H., Rietveld, S., & de Geus, E. J. C. (2002). Contribution of tonic vagal modulation of heart rate, central respiratory drive, respiratory depth, and respiratory frequency to respiratory sinus arrhythmia during mental stress and physical exercise. *Psychophysiology*, *39*, 427–436.
- Houtveen, J. H., Rietveld, S., & de Geus, E. J. C. (2003). Exaggerated perception of normal physiological responses to stress and hypercapnia in young women with numerous functional somatic symptoms. *Journal of Psychosomatic Research*, *55*, 481–490.
- Houtveen, J. H., & van Doornen, L. J. P. (2007). Medically unexplained symptoms and between-group differences in 24-hr ambulatory recording of stress physiology. *Biological Psychology*, *76*(3), 239–249.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Mortola, J. P. (2004). Breathing around the clock: An overview of the circadian pattern of respiration. *European Journal of Applied Physiology*, *91*(2–3), 119–129.
- Muthén, L. K., & Muthén, B. O. (1998-2004). *Mplus user's guide* (Third edition). Los Angeles: Muthén & Muthén.
- Pennebaker, J. W. (2000). Psychological factors influencing the reporting of physical symptoms. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report implications for research and practice* (pp. 301–315). London: LEA.
- Richerson, G. B. (2004). Serotonergic neurons as carbon dioxide sensors that maintain pH homeostasis. *Nature Reviews Neuroscience*, *5*, 449–461.
- Riese, H., Groot, P. F. C., van den Berg, M., Kupper, N. H. M., Magnee, E. H. B., Rohaan, E. J., et al. (2003). Large-scale ensemble averaging of ambulatory impedance cardiograms. *Behavior Research Methods, Instruments, and Computers*, *35*, 467–477.
- Shacham, S. (1983). A shortened version of the Profile of Mood States. *Journal of Personality Assessment*, *47*, 305–306.
- Sherwood, A., Allen, M. T., Fahrenberg, J., Kelsey, R. M., Lovallo, W. R., & Van Doornen, L. J. P. (1990). Methodological guidelines for impedance cardiography. *Psychophysiology*, *27*, 1–23.
- van Eekelen, A. P., Houtveen, J. H., & Kerkhof, G. A. (2004a). Circadian variation in base rate measures of cardiac autonomic activity. *European Journal of Applied Physiology*, *93*(1–2), 39–46.
- van Eekelen, A. P., Houtveen, J. H., & Kerkhof, G. A. (2004b). Circadian variation in cardiac autonomic activity: Reactivity measurements to different types of stressors. *Chronobiology International*, *21*, 107–129.
- Vrijkotte, T. G. M., van Doornen, L. J. P., & de Geus, E. J. C. (2004). Overcommitment to work is associated with changes in cardiac sympathetic regulation. *Psychosomatic Medicine*, *66*, 656–663.
- Watson, D., & Pennebaker, J. W. (1989). Health complaints, stress, and distress: Exploring the central role of negative affectivity. *Psychological Review*, *96*, 234–254.
- Wientjes, C. J. E., & Grossman, P. (1994). Overreactivity of the psyche or the soma? Interindividual associations between psychosomatic symptoms, anxiety, heart rate, and end-tidal partial carbon dioxide pressure. *Psychosomatic Medicine*, *56*, 533–540.
- Wilhelm, F. H., Gevirtz, R., & Roth, W. T. (2001). Respiratory dysregulation in anxiety, functional cardiac, and pain disorders: Assessment, phenomenology, and treatment. *Behavior Modification*, *25*, 513–545.
- Willemsen, G. H. M., de Geus, E. J. C., Klaver, C. H. A. M., van Doornen, L. J. P., & Carroll, D. (1996). Ambulatory monitoring of the impedance cardiogram. *Psychophysiology*, *33*, 184–193.
- Yoshizaki, H., Yoshida, A., Hayashi, F., & Fukuda, Y. (1998). Effect of posture change on control of ventilation. *Japanese Journal of Physiology*, *48*, 267–273.

(RECEIVED July 14, 2008; ACCEPTED June 18, 2009)