Heritability and Temporal Stability of Ambulatory Autonomic Stress Reactivity in Unstructured 24-Hour Recordings

Melanie Neijts, MSc, Rene van Lien, PhD, Nina Kupper, PhD, Dorret Boomsma, PhD, Gonneke Willemsen, PhD, and Eco J.C. de Geus, PhD

ABSTRACT

Objective: Measurements of ambulatory autonomic reactivity can help with our understanding of the long-term health consequences of exposure to psychosocial stress in real-life settings.

Methods: In this study, unstructured 24-hour ambulatory recordings of cardiac parasympathetic and sympathetic control were obtained in 1288 twins and siblings, spanning both work time and leisure time. These data were used to define two ambulatory baseline (sleep, leisure) and four stress conditions (wake, work, work_sitting, work_peak) from which six ambulatory stress reactivity measures were derived. The use of twin families allowed for estimation of heritability and testing for the amplification of existing or emergence of new genetic variance during stress compared with baseline conditions.

Results: Temporal stability of ambulatory reactivity was assessed in 62 participants and was moderate to high over a 3-year period ($0.36 < r < 0.91$). Depending on the definition of ambulatory reactivity used, significant heritability was found, ranging from 29% to 40% for heart rate, 34% to 47% for cardiac parasympathetic control (indexed as respiratory sinus arrhythmia), and 10% to 19% for cardiac sympathetic control (indexed as the preejection period). Heritability of ambulatory reactivity was largely due to newly emerging genetic variance during stress compared with periods of rest. Interestingly, reactivity to short standardized stressors was poorly correlated with the ambulatory reactivity measures implying poor laboratory–real-life correspondence.

Conclusions: Ambulatory autonomic reactivity extracted from an unstructured real-life setting shows reliable, stable, and heritable individual differences. Real-life situations uncover a new and different genetic variation compared with that seen in resting baseline conditions, including sleep.

Key words: ambulatory reactivity, heritability, interbeat interval, preejection period, respiratory sinus arrhythmia, twin study.

INTRODUCTION

Exaggerated cardiovascular responses to stress have been associated with an increased risk for cardiovascular disease (1,2), although the effects are modulated by genetic sensitivity and the frequency of stress exposure in daily life (3,4). To assess the propensity toward exaggerated reactivity, many studies have used standard mental stressors under controlled laboratory conditions. The individual differences in stress reactivity detected by such procedures are substantial, have proven to be reliable, and may be associated with long-term adverse cardiovascular outcome (5–7). Nonetheless, there is valid concern about the extent to which laboratory tasks actually translate to stress situations in real life (8–11). The laboratory setting, although

ANS = autonomic nervous system, **DZ** = dyzygotic, **ECG** = electrocardiogram, **HR** = heart rate, **IBI** = interbeat interval, **ICG** = impedance cardiogram, **MZ** = monozygotic, **PEP** = preejection period, **pvRSA** = peak valley respiratory sinus arrhythmia, **RSA** = respiratory sinus arrhythmia, **VU-AMS** = VU University Ambulatory Monitoring System

SDC Supplemental Content

From the Department of Biological Psychology (Neijts, Lien, Kupper, Boomsma, Willemsen, de Geus), VU University Amsterdam, Amsterdam, the Netherlands; EMGO+ Institute for Health and Care Research (Neijts, Lien, Boomsma, Willemsen, de Geus), VU University Medical Center, Amsterdam, the Netherlands; and CoRPS—Center of Research on Psychology in Somatic diseases, Department of Medical & Clinical Psychology (Kupper), Tilburg University, Tilburg, the Netherlands.

Address correspondence and reprint requests to Eco J.C. de Geus, PhD, EMGO+ Institute for Health and Care Research VU Medical Centre & VU University Amsterdam Van der Boechorststraat 7, Room D-452 1081 BT Amsterdam, the Netherlands. E-mail: eco.de.geus@vu.nl

Received for publication November 24, 2014; revision received June 15, 2015.

DOI: 10.1097/PSY.0000000000000227

important for inventorying the stress response in all its facets, cannot adequately capture prolonged activation and subsequent recovery processes (12–14).

Cardiovascular stress research has therefore shifted from using mental stress tasks (e.g., mental arithmetic and reaction time tasks) to also using tasks with a more social evaluative character (e.g., public speaking; see Ref. (15) for an overview of the studies conducted in this area). Since the past two decades, cardiovascular stress research has also increasingly moved from the laboratory to the real-life situation of the participants. The rationale behind the ambulatory approach is that the daily life situation can give a more accurate and ecological valid reflection of the psychophysiological state of the participant and might therefore serve as a better indicator or predictor of disease risk (16). This comes at a price. Compared with the laboratory setting, the ambulatory situation is per definition a much less controlled environment in which confounding factors such as physical activity, posture, and time of day come into play (17). Moreover, there is no standard baseline and there are no standard stressors to which all participants are exposed. Two strategies have been used to deal with this. First, ambulatory recordings have been scheduled during a concrete and relatively comparable stressful event like a school examination (18–20) or an oral presentation (8,21,22). The second strategy is to use prolonged ambulatory recordings that span periods at work and in leisure time. Reactivity is then defined by comparing work levels to leisure time levels, under the assumption that the working day will be enriched for mentally and emotionally engaging events compared with leisure time (12,17,23–26).

A further challenge in ambulatory recording is that it does not allow the same in-depth assessment of physiology as is possible within a laboratory setting. Ambulatory recording has been typically confined to measurements of blood pressure and heart rate (HR), whereas a decrease in vagal tone and increased sympathetic nervous system activity are the key drivers of stress-induced increases in blood pressure and HR. Fortunately, two noninvasive key measures of cardiac sympathetic and vagal control can also be recorded with high fidelity in ambulatory settings: the preejection period (PEP) and respiratory sinus arrhythmia (RSA) (27). When ambulatory HR, PEP, and RSA are assessed in a 24-hour recording across a working day, followed by leisure time and sleep recording, there are various ways to define cardiac autonomic reactivity, for instance, by defining either sleep or leisure time sitting as the baseline, or using periods of peak reactivity while sitting at work or the entire work period. Until now very little systematic study of the reliability and stability of individual differences in such ambulatory autonomic reactivity has been performed.

In the present study, we recorded 24-hour ambulatory HR, RSA, and PEP in more than 1300 participants. During waking hours, participants filled out detailed activity diaries that were used to divide the entire signal of every participant into fixed periods of distinctive activities and postures. The periods were used to define two different resting (baseline) conditions and four conditions reflecting more mentally and/or emotionally engaging episodes (referred to as "stress" from this point forward). Mean sleep levels were used as the ultimate resting baseline condition for each participant. As it is not always possible to measure during sleep, we also used an alternative baseline condition by selecting a period in leisure time during sitting activities which were relaxing in nature (Internet, watching TV, reading). As a first stress level, we used the entire waking period. Since participants were visited on a weekday and more than half of the participants were measured on a working day at the working location, we also extracted the mean working day level including all postures and the mean working day level for sitting activities only as alternative stress levels. We also extracted the most (psychological) arousing activities at work for each participant by searching for periods with low physical activity but accompanied by high HRs.

Two- to 5-year retest data were available for 62 participants which allowed estimation of the temporal stability of these ambulatory reactivity measures. In 576 participants, we also embedded a standard stress testing protocol assessing autonomic reactivity to two mental stress tasks. This allowed for a comparison of ambulatory reactivity to classical short-term stress reactivity. Measurements were done in monozygotic (MZ) and dizygotic (DZ) twins and their family members to be able to estimate the reliability and heritability of the various ambulatory reactivity measures. Significant heritability indicates that a measure is reliable and taps into stable biological differences. It was further tested whether there is amplification of existing or emergence of new genetic variance during the ambulatory stress conditions compared with ambulatory baselines, as has previously been observed for laboratory stressors (28–30). The overarching aim was to identify ambulatory autonomic reactivity measures from unstructured 24-hour recordings for use in stress research that have good temporal stability, significant heritability, and the ability to detect stress-specific genetic variance.

METHODS

Participants

Participants were all registered in the Netherlands Twin Register. The Netherlands Twin Register has been collecting survey and biological data for more than 25 years. In the biennial surveys, data on health, life-style, and personality are assessed (31). A subset of these participants were included in a large cardiac ambulatory monitoring project in which 24-hour recordings were collected in two separate studies (32–34). Study 1 was conducted between August 1998 and June 2003 and included two waves of data collection with partial retest data. This sample was further expanded with a new data collection round, Study 2, that took place between

November 2010 and June 2012. In the latter study, data were collected in a single wave. For both studies, adult twins and siblings without known cardiovascular disease or other relevant health complaints were selected and informed about the study by mail. This was followed by a short telephone interview in which the health status of the participants was verified. A priori reasons for exclusion for all studies/waves were pregnancy, heart transplantation, presence of a pacemaker and known ischemic heart disease, congestive heart failure, or diabetic neuropathy. We excluded data of 8 participants showing many arrhythmias or preventricular contractions. Of the remaining sample (n = 1373), data of participants that were on cardiovascular medication, cardiac therapy, or antidepressant medication were excluded (n = 71). To simplify genetic modeling, we excluded the third member of triplets (n = 1) and used only one pair from families with multiple twins. We further restricted the number of siblings to a maximum of two singleton brothers and two singleton sisters per family, selecting the siblings who were closest in age to the twins and removing data for nine siblings. Three more participants were excluded because the rest of their family members participated in a different wave.

In the end, recordings of 1288 participants were available with 486 MZ twins (210 complete pairs), 517 DZ twins (205 complete pairs), and 285 nontwin siblings. The data of participants belonging to an incomplete twin pair were included because they could still be paired to their nontwin sibling and/or contribute to the estimates of the means and variances. Incomplete pairs occurred because only one of the two participated or because data of the other twin were excluded for reasons mentioned earlier. Mean (standard deviation) age was 33.5 (9.2) years, and 61.6% of the sample was female. Zygosity of the twins was determined by DNA typing for 97.9% of the same-sex twin pairs. For 2.1% of the same-sex pairs, zygosity was based on survey questions on physical similarity and the frequency of confusion of the twins by parents, other family members, and strangers. Agreement between zygosity based on these items and zygosity based on DNA is 96.1% (31). For 62 participants in Wave 1 of Study 1, measurements were repeated in Wave 2 of that study after an average period of 3.3 years (range, 2.1–4.7 years). For more detailed characteristics of Study 1, Study 2, and the retest sample separately, see Table 1. The study protocol was approved by the Medical Ethics Committee of the VU University Medical Center Amsterdam, and all participants gave written consent before entering the study.

Procedure

Participants were visited at home, before starting their normal daily activities. During a short interview, information on health status and current medication use was obtained. They were fitted with the VU University Ambulatory Monitoring System (VU-AMS; VU University, Amsterdam, the Netherlands, www.vu-ams.nl) that records the electrocardiogram (ECG) and impedance cardiogram (ICG) continuously. For participants taking part in Wave 1 and Wave 2 of Study 1, the VU-AMS version 4.6 was used. For Study 2, the 5fs version of the VU-AMS was used. A standard laboratory protocol was embedded within the ambulatory recording protocol of this study. During the home visit, two typical laboratory tasks, each lasting 2 minutes, were executed by the participants in a fixed order. For this, we used the computerized Stroop Color-Word conflict task and a Serial Subtraction task, because both cognitive tasks have proven capable of eliciting a psychophysiological stress response (35–38). For the Stroop task, stimuli consisted of one of four color names that were printed in incongruent colors. In total, 99 presentations of 12 combinations of incongruent stimuli were presented in a random order for 2 minutes. As we only included a 2-minute Stroop conflict task and none of the nonconflicting control tasks, the test was preceded by a 45-second practice session. Participants verbally responded to the stimuli. In the Serial Subtraction task, participants were asked to sequentially subtract backward by 7 aloud as quickly as possible for 2 minutes. Each participant began with the number 1256. When an error was made, the participant was corrected and instructed to continue from that point on. The cognitive tasks were preceded by 4 minutes of quiet sitting in a secluded part of the house/work area.

Subsequently, participants were instructed to wear the VU-AMS device the entire day and night up until the next day, after having worn the device for 24 hours. Instructions were supplied that explained how to respond to potential alarm beeps (e.g., on loose electrode contacts), and telephone assistance was available during waking hours. Participants were requested to keep a paper-and-pencil diary and to write down a chronological account of their activity, posture, location, and social situation over the past period (free recall). For Waves 1 and 2 in Study 1, participants were prompted by an alarm beep to do so every 30 minutes; for Study 2, the diary was filled every 60 minutes. In addition, participants were instructed to write down at which time they had breakfast, lunch, and dinner and they were asked to refrain from vigorous exercise during the ambulatory recording day. The next day, the VU-AMS device was detached and collected by the researcher or returned by mail.

RSA and PEP Measurement

RSA is considered a reliable index of parasympathetic control over the heart, whereas PEP is considered a reliable index of sympathetic control

TABLE 1. Sample Characteristics for Study 1, the Retest Study, Which Was Part of Study 1 and Study 2

	Ambulatory Study 1			Ambulatory Study 2 (2010–2012)
	Wave 1 (1998–2000)	Wave 2 (2001–2003)		Wave 1 (2010–2012)
	New	New	Retest	New
n individuals	367	380	62	541
MZ (*n* belonging to complete pair)	103 (86)	117 (104)	12 (0)	266 (230)
% MZ female	66.0	62.4	50.0	64.3
DZ (*n* belonging to complete pair)	144 (110)	120 (147)	29 (4)	180 (226)
% DZ female	75.9	76.3	65.5	72.9
Siblings (% female)	120 (57.5)	116 (61.2)	21 (57.7)	49 (59.2)
Age, M (SD), y	28.5 (9.6)	32.9 (10.7)	33.6 (9.6)	37.2 (5.4)

MZ = monozygotic; DZ = dyzygotic; M = mean; SD = standard deviation.

All twins and siblings that participated in Study 2 also participated in the laboratory protocol, as this was embedded within the same 24-hour ambulatory measurement.

over the heart (27). The assessment and quantification of respiration and RSA from VU-AMS recordings has been described previously (33). Briefly, the d*Z* signal at the respiration frequency (0.1-0.4 Hz), combined with the interbeat interval (IBI) series, was used to compute "peak − valley" RSA (pvRSA). In this method, RSA is scored by detecting the shortest IBI during inspiration and the longest IBI during expiration on a breath-to-breath basis according to procedures detailed elsewhere (39,40). If no shortest or longest IBI could be detected in inspiration and expiration, respectively, pvRSA was set to zero.

For the assessment of the PEP, a measure of cardiac contractility, both the ECG and the ICG, are used. The ICG signal was ensemble averaged across the diary-coded activity periods (described in the section on "Ambulatory Data Reduction"), time locking the signal to the R-wave peaks (41). The PEP is defined as the time interval between the Q-wave onset of the ECG and the B point of the d*Z*/d*t* signal. The Q wave reflects the onset of left ventricular activity and the B point reflects the opening of the aortic valves. In both VU-AMS versions, the R and B points are scored automatically by the software. In the newer 5fs version of the VU-AMS, the entire ECG signal is stored, so the Q-onset time was available as well. All automated scoring was visually checked by the experimenter. For the calculation of PEP in the two waves of Study 1, a fixed Q-R interval of 48 milliseconds was added to the duration of the R-B interval (42). For Study 2, the true Q-onset point was used when present; otherwise, the grand average of the Q-R interval was summed to the R-B interval of the individual participant. If R onset was additionally missing, we subtracted the grand average Q-onset time from the individual participants' B point (43).

Ambulatory Data Reduction

Using the activity diary entries in combination with a visual inspection of the output of an inbuilt accelerometer (measuring movement), the entire 24-hour recording was divided into fixed periods. These periods were coded for posture (supine, sitting, standing, walking, bicycling), activity (e.g., desk work, dinner, meetings, and watching TV), physical load (no load, light, intermediate, and heavy), and location of the participant (e.g., at home, at work, and public space). Minimum duration of periods was 5 minutes and maximum duration was 1 hour. If periods with similar activity and posture lasted more than 1 hour (e.g., during sleep), they were divided into multiple periods of maximally 1 hour. For each of the coded periods, the mean IBI, RSA, and PEP was calculated. The periods belonging to the standard baseline and stress conditions were additionally coded for the participants who took part in Study 2, with one period representing the resting baseline ("standardized_baseline") condition and one period representing the stress condition, averaged over the two tasks ("standardized_stress").

From the ambulatory autonomic nervous system (ANS) recording, two ambulatory baseline and four ambulatory stress conditions were defined. For the first ambulatory baseline condition, the mean IBI, pvRSA, and PEP value across all sleeping periods was calculated ("sleep"). A period was classified as a sleeping period based on the reported bedtime in the diary, and physical activity was verified by accelerometry. For the second ambulatory baseline, per participant the mean of periods spent sitting while being engaged in recreational activities in the evening, from 6 PM till bedtime, was determined to represent the alternative baseline condition ("leisure"). Periods summed to at least half an hour. For the first ambulatory stress condition, the mean waking level of IBI, pvRSA, and PEP was used, including both sitting and light physical (nonsitting) activities ("wake"). Because participants were explicitly instructed not to engage in vigorous exercise during the recording day, these periods did not include high physical activity. Periods of light physical activity were classified as such based on the activity information obtained from the diary and the accompanying accelerometer signal. The additional ambulatory stress conditions were defined only for participants who reported the testing day to be a working day and who actually reported in the diary to have been at the work location during the testing day. For these participants, all periods in which the participant was engaged in sitting activities or light physical activity at work during a working day between 9 AM and 6 PM were defined as work. Because there were multiple periods that fit this condition, mean IBI, pvRSA, and PEP levels were determined across the entire working period ("work") and across a selection of periods at work when the participants were sitting ("work_sitting"). Finally, an ambulatory stress condition was created from a time frame of at least half an hour of periods with the *highest* HR while the participant was sitting at work between 9 AM and 6 PM. Because this time frame only consisted of few coded periods and we wanted to prevent that relatively high weight would be given to extreme values, the *median* instead of the mean IBI, pvRSA, and PEP values of the periods with the highest HR was selected to represent "work_peak."

Six different ambulatory reactivity measures were calculated by computing absolute difference scores between wake minus sleep (Δwake − sleep), work minus sleep (Δwork − sleep), work_sitting minus sleep (Δwork_sitting − sleep), work_sitting minus leisure (Δwork_sitting − leisure), work_peak minus sleep (Δwork_peak − sleep), and work_peak minus leisure (Δwork_peak − leisure). The standard stress reactivity was computed as the mean of the two mental stress tasks minus the baseline rest condition. For an overview of the ambulatory conditions and ambulatory reactivity measures, see Table 2.

Statistical Analyses

Sample selection, data preparation, and all nongenetic statistical analyses were performed using IBM SPSS 20.0. The distributions of ambulatory IBI and PEP levels and all ambulatory reactivity measures were normal. For ambulatory pvRSA levels, a natural logtransformation was applied to obtain a normal distribution. Significant differences between ambulatory conditions were tested by a mixed-model analysis of variance with age and sex (and with respiration rate for pvRSA only) as covariates and family as a random factor and ambulatory condition as a repeated fixed factor (sleep, leisure, wake, work, work_sitting, and work_peak). A similar analysis of variance was used to test the difference between mental stress and baseline conditions during the standard part of the recording. Temporal stability of the ambulatory measures across the two waves of Study 1 was calculated as an intraclass correlation. Associations between ambulatory reactivity scores among themselves and with the standard stress reactivity were assessed by Pearson correlations. For all statistical testing, effects were considered significant when $p < .01$.

Genetic Analysis

In a twin study, variance is typically decomposed into latent genetic and environmental components. Genetic variance can be further decomposed into shared additive (A) and nonadditive (D) components. Environmental variance can be decomposed in a component that is common in family members (C) or that is unique (E) to the individual. In a study design that only includes twins and siblings, estimates of C and D are confounded and cannot be estimated simultaneously. In that case, the pattern of the twin correlations is used to guide the experimenters' choice to model either an ACE or an ADE model. An ACE model is chosen when the MZ correlations are less than twice as high as the DZ and sibling correlations. Dominance or nonadditive genetic factors may be present when the MZ correlations are more than twice as high as the DZ and sibling correlations, and in that case, an ADE model will be chosen (44). MZ correlations can also be useful to explore the reliability of the trait. Estimation of the ratio of Var(*E*) to the total variance (e^2) provides a first impression of the unreliability of the trait. The *E* factor contains true unique environmental effects on the trait plus measurement error. Because MZ twins are correlated perfectly for genetic and for common environmental factors, $1 - e^2$ is equal to the MZ twin correlation (r_{MZ}). Therefore, r_{MZ} is a lower bound for the test-retest reliability coefficient, because the level of an individual unique environment (*E*) may also lead to stability.

Genetic analyses were performed using structural equation modeling in the software package Mx (45). In structural equation modeling, models

TABLE 2. Ambulatory Conditions for Ambulatory Reactivity Assessment

Ambulatory Condition	Description
Baseline	
sleep	Mean level during sleep
leisure	The mean of periods while the participant is sitting and engaged in recreational activities like Internet, reading, and watching television, but not eating or drinking. Periods sum to at least half an hour. Only periods in the evening, from 6 PM till bedtime, are considered eligible.
Stress	
wake	The mean waking level—including sitting and light physical activity
work	The mean of all periods in which the participant was engaged in sitting or light physical activity between 9 AM and 6 PM during a working day at the work location
work_sitting	Similar to "work," except that the mean of all periods in which the participant was engaged in *sitting activities only* between 9 AM and 6 PM during a working day at the work location was calculated
work_peak	A time frame of periods with the highest heart rate while the participant is sitting between 9 AM and 6 PM during a working day at the work location. Periods sum to at least half an hour. The median of the selected periods is taken to represent work_peak
Reactivity	
Δwake − sleep	Absolute difference score between wake and sleep level
Δwork − sleep	Absolute difference score between work and sleep level
Δwork_sitting − sleep	Absolute difference score between work_sitting and sleep level
Δwork_sitting − leisure	Absolute difference score between work_sitting and leisure level
Δwork_peak − sleep	Absolute difference score between work_peak and sleep level
Δwork_peak − leisure	Absolute difference score between work_peak and leisure level

are fitted to the data and a goodness-of-fit statistic is calculated for each model. Subsequently, the fit of the more parsimonious nested models is compared with the fit of the full model by means of the likelihood ratio test in which the difference in minus twice the logarithm of the likelihood (−2LL) is calculated; this difference has a χ^2 distribution. When the χ^2 test is significant ($p < .01$), the more parsimonious model is considered to fit significantly worse to the data than the fuller model it is tested against. Before the variance was decomposed into genetic and environmental components, saturated models were fitted to the data. In these fully parameterized models, we tested for heterogeneity in male and female variances and family correlations. More specifically, we tested if sex differences were present and if there was evidence for a twin-specific resemblance. The allowed limitations were carried forward in the specification of the genetic models. Overall, we did not find evidence for twin-specific resemblance or for systematic quantitative or qualitative sex differences. We therefore continued estimating all parameters by combining data from males and females. Sex and age (and respiration rate for pvRSA) effects on the mean were regressed out simultaneously with variance decomposition.

Bivariate genetic models were specified to examine the genetic architecture and the change in the genetic influences from baseline ambulatory conditions to ambulatory stress conditions (Figure S1, Supplemental Digital Content 1, http://links.lww.com/PSYMED/A234). More specifically, the heritability of the baseline condition, or the effect of A1, is estimated by the relative contribution of genetic variance to the total variance in the baseline condition and is assessed by the ratio of $a_{11}^2/(a_{11}^2 + c_{11}^2 + e_{11}^2)$. Because the genetic variance during stress conditions can consist of the genetic variance that is shared with baseline activity, and with new genetic variance that emerges during stress, heritability of the ambulatory stress condition is calculated as follows: $a_{21}^2 + a_{22}^2/(a_{21}^2 + a_{22}^2 + c_{21}^2 + c_{22}^2 + e_{21}^2 + {}_{22}^2)$. The effects of the genetic factors that are expressed during baseline can be amplified ($a_{21} > a_{11}$) or deamplified ($a_{21} < a_{11}$) during stress. The significance of stress-specific genetic effects can be assessed by testing if the path coefficient a_{22} is significantly different from zero. The part of the heritability that is due to these new genetic factors (A2) can be calculated as follows: $a_{22}^2/(a_{21}^2 + a_{22}^2 + c_{21}^2 + c_{22}^2 + e_{21}^2 + {}_{22}^2)$. The heritability of ambulatory reactivity was also calculated by procedures described in more detail elsewhere (28). Briefly, it was calculated as a change score within the bivariate model by adding a latent factor with fixed loadings of +1 and −1 on baseline and stress, respectively. The total variance of the ambulatory reactivity score is equal to the sum of the variance during baseline and stress, minus the covariance between baseline and stress conditions. So the genetic part of the variance of ambulatory reactivity is $a_{11}^2 + a_{21}^2 + a_{22}^2 - 2\, a_{11} \times a_{21}$, which is equal to $(a_{21} - a_{11})^2 + a_{22}^2$. Heritability of the ambulatory reactivity score is calculated by $((a_{21} - a_{11})^2 + a_{22}^2)/((a_{21} - a_{11})^2 + a_{22}^2 + (c_{21} - c_{11})^2 + c_{22}^2 + (e_{21} - e_{11})^2 + e_{22}^2)$.

RESULTS

Table 3 gives the mean levels and standard deviations of the three variables of interest for the six conditions and the means and standard deviations of the different reactivities that were calculated based on these conditions. The table shows that all participants have data during waking time. However, only 56.7% of the participants who reported that the testing day was a working day actually reported to be at the work location during the testing day. The proportion of participants who reported that the testing day was a working day but did not spend time in an external working environment (housewives, people working at home) was 13.9%. They did not spend time in an external

TABLE 3. Means and Standard Deviations of Interbeat Interval, Peak Valley Respiratory Sinus Arrhythmia, and Preejection Period in the 6 Ambulatory Conditions

	IBI, ms		pvRSA, ms		PEP, ms	
	n	Mean (SD)	*n*	Mean (SD)	*n*	Mean (SD)
Levels						
sleep	1242	983.62 (132.37)	1242	55.80 (24.63)	1225	106.28 (15.72)
leisure	1075	837.67 (115.93)	1075	48.30 (22.31)	1044	99.20 (16.32)
wake	1288	759.58 (95.46)	1288	40.86 (15.10)	1285	96.74 (15.67)
work	730	749.92 (113.52)	730	41.27 (16.65)	725	97.12 (16.79)
work_sitting	686	784.63 (112.70)	686	44.89 (18.25)	677	99.49 (17.78)
work_peak	633	739.26 (105.45)	633	39.83 (16.49)	624	97.93 (21.01)
Reactivities						
wake − sleep	1242	−223.51 (85.46)	1242	−15.07 (19.00)	1225	−9.42 (12.89)
work − sleep	701	−240.58 (106.43)	701	−14.89 (22.16)	685	−8.77 (14.81)
work_sitting − sleep	657	−205.89 (100.75)	657	−11.48 (21.30)	639	−6.23 (15.17)
work_peak − sleep	609	−254.71 (104.33)	609	−16.99 (22.83)	591	−8.06 (15.83)
work_sitting − leisure	564	−58.70 (84.00)	564	−3.66 (14.77)	544	1.70 (10.03)
work_peak − leisure	523	−106.08 (87.13)	523	−8.91 (16.75)	506	0.02 (11.11)

SD = standard deviation; IBI = interbeat interval; pvRSA = peak valley respiratory sinus arrhythmia; PEP = preejection period.
Reactivity is listed separately.

working environment (housewives, people working at home) was 13.9%.

A significant main effect of ambulatory condition was found for all ANS variables: IBI ($F(5,4422) = 3127.954$, $p < .001$), pvRSA ($F(5,4458) = 335.281$, $p < .001$), and PEP ($F(5,4337) = 235.841$, $p < .001$). Post hoc testing of the reactivity values showed this to mainly reflect the four daytime versus sleep contrasts for all three ANS variables (see Table 3). For IBI and pvRSA, the two contrasts including leisure time as baseline measure were also significant (p values < .001).

There were significant intercorrelations between all six reactivity measures for IBI, pvRSA, and PEP (Table S1, Supplemental Digital Content 1, http://links.lww.com/PSYMED/A234), but the choice of baseline was critical; the four reactivity measures using sleep as a baseline were highly correlated among each other (r values > 0.87) and so were the two reactivity measures using leisure time sitting as a baseline (r values > 0.89). Correlation was less strong between to the reactivity measures based on sleep versus those based on leisure time sitting ($0.30 < r < 0.61$ for IBI, $0.28 < r < 0.58$ for pvRSA, and $0.29 < r < 0.60$ for PEP).

Table 4 presents the levels attained during the standard stress testing protocol that took place at home preceding the ambulatory monitoring in Study 2. In this setting, task levels also significantly declined compared with rest for IBI ($F(1,520) = 602.974$, $p < .001$), pvRSA ($F(1,519) = 33.785$, $p < .001$), and PEP ($F(1,482) = 98.296$, $p < .001$). The standard reactivity measure was very poorly correlated with each of the six ambulatory reactivity measures (Table S2, Supplemental Digital Content 1, http://links.lww.com/PSYMED/A234), particularly for PEP.

Table 5 depicts the temporal stability for the ambulatory levels and reactivity measures. Higher temporal stability is

TABLE 4. Means and Standard Deviations of Interbeat Interval, Peak Valley Respiratory Sinus Arrhythmia, and Preejection Period Levels and Reactivity During the Standard Stress Protocol

	IBI, ms		pvRSA, ms		PEP, ms	
	n	Mean (SD)	*n*	Mean (SD)	*n*	Mean (SD)
rest	522	834.62 (120.27)	522	48.36 (22.22)	493	107.12 (20.75)
task	521	774.32 (111.43)	521	44.75 (17.73)	493	103.06 (21.01)
Δtask − rest	521	−59.82 (55.65)	521	−3.51 (13.87)	483	−4.22 (9.30)

SD = standard deviation; IBI = interbeat interval; pvRSA = peak valley respiratory sinus arrhythmia; PEP = preejection period.

TABLE 5. Temporal Stability of the Ambulatory Levels and Reactivity Measures Over an Average Period of 3.3 Years

	IBI		pvRSA		PEP	
	n	ICC	*n*	ICC	*n*	ICC
sleep	53	0.80	53	0.86	51	0.84
leisure	48	0.68	48	0.82	47	0.87
wake	62	0.79	62	0.82	62	0.89
work	19	0.77	19	0.90	19	0.87
work_sitting	19	0.71	19	0.89	18	0.86
work_peak	16	0.74	16	0.85	15	0.85
Δwake − sleep	53	0.71	53	0.83	51	0.49
Δwork − sleep	17	0.91	17	0.84	16	0.48
Δwork_sitting − sleep	17	0.89	17	0.79	15	0.52
Δwork_sitting − leisure	15	0.36	15	0.58	13	0.76
Δwork_peak − sleep	17	0.40	17	0.68	15	0.73
Δwork_peak − leisure	13	0.44	13	0.85	12	0.65

IBI = interbeat interval; pvRSA = peak valley respiratory sinus arrhythmia; PEP = preejection period; ICC = intraclass correlation coefficient.

more consistently found for the ambulatory levels ($0.71 < r < 0.90$) compared with the reactivity scores, which nonetheless show very good stability over time for all three ANS parameters ($0.36 < r < 0.91$). That the ambulatory levels capture stable individual traits is further reinforced by the MZ twin correlations that can be used as a proxy of the *minimal* test-retest reliability for the ambulatory levels (Figure S2, Supplemental Digital Content 1, http://links.lww.com/PSYMED/A234).

MZ twin correlations were higher than DZ twin correlations, suggesting a role for genetic factors in explaining individual differences in IBI, pvRSA, and PEP levels during rest and stress periods in real life. This was confirmed by genetic structural equation modeling. For all conditions, IBI and pvRSA in both the ambulatory and the standard setting and for PEP in the standard setting only, ACE models were fitted to the data. For PEP in the ambulatory conditions, ADE models were fitted to the data as the MZ twin correlations in these instances were more than twice as high as the DZ twin correlations. Formal testing showed that C and D factors could be dropped from all models and AE models provided the most parsimonious fit for all three ANS measures in all conditions in both settings. Heritability estimates for the IBI, pvRSA, and PEP levels in all baseline and stress conditions and the ambulatory and standard reactivity measures are listed in Table 6. In general, heritability for the baseline levels was lower than heritability of the stress levels but, as can also be judged from the confidence intervals, this difference was only significant for work_sitting compared with leisure and sleep where it went from 51% to 69% and from 53 to 69%, respectively, for IBI. For PEP, heritability of work_sitting (44%) was also significantly higher than leisure (25%) and there was a similar trend for work_peak for which heritability was estimated at 42%.

For all ambulatory reactivity contrasts, there was a common genetic factor that influenced both baseline and stress levels, but in addition, new genetic factors were found to emerge during stress compared with the baseline conditions, yielding significant heritability of the corresponding ambulatory reactivity measures. For IBI, significant heritability of reactivity was found for five reactivity measures, except for leisure compared with work_peak. For the pvRSA, heritability of four reactivity measures was significant, the exceptions being work_sitting compared with leisure and work_peak compared with leisure. For PEP, heritability of ambulatory reactivity was significant for the wake compared with sleep and both the work_sitting and work_peak compared with leisure contrasts. During stress, genes acting at the resting level were *deamplified* for all ambulatory IBI and pvRSA reactivity measures, with the exception of Δwork_sitting − leisure for IBI and pvRSA, and Δwork_peak − leisure for pvRSA only. Deamplification of genetic variance means that the effect of the genes active during rest accounted for a smaller part of the variance during stress. For the ambulatory PEP reactivity measures, no significant deamplification of genetic factors was found when going from rest to stress, and the genetic factors influencing leisure time were even significantly *amplified* by the work_sitting and the work_peak conditions.

For standard stress reactivity, too, a common genetic factor was found to influence both rest and task levels and new genetic factors emerged during task stress for all ANS measures. Although these effects were significant, they were less pronounced in the standard stress environment compared with real life.

DISCUSSION

This study describes a large twin study on changes in ambulatory levels of IBI, PEP, and pvRSA when going from sleep and resting in leisure time to more socially and mentally engaging activities during the working day. We found that ambulatory autonomic reactivity is a stable, heritable individual trait, showing moderate to high temporal stability over a 3-year follow-up period. Depending on the definition of ambulatory reactivity used, heritability ranged from 29% to 40% for IBI, 34% to 47% for pvRSA, and 10% to 19% for PEP, although not all reactivity measures showed heritability. Heritability of ambulatory autonomic reactivity was largely due to new genetic variance specifically emerging during stress, whereas the genetic factors influencing resting baseline levels became less prominent under stress.

These results replicate and extend previous findings in laboratory stress testing. In an earlier laboratory study of

TABLE 6. Heritability of the Levels of Interbeat Interval, Peak Valley Respiratory Sinus Arrhythmia, and Preejection Period During the Baseline and Stress Levels, the Cause of Changes in Heritability From Baseline to Stress (Amplification Versus Emergence), and the Heritability of the Reactivity Measures

	Baseline Level h2 (99% CI)	Stress Level h2 (99% CI)	Amplification/Deamplification of Genes Acting on Baseline Level (99% CI)	Specific h2 due to Genes Emerging During Stress (99% CI)	Reactivity h2 (99% CI)
IBI					
wake − sleep	0.52 (0.41–0.62)	0.55 (0.44–0.64)	Deamplification, a_{21}/a_{11} = 0.63 (0.53–0.74)**	0.18 (0.11–0.25)**	0.38 (0.24–0.51)**
work − sleep	0.53 (0.41–0.63)	0.61 (0.46–0.72)	Deamplification, a_{21}/a_{11} = 0.68 (0.53–0.86)**	0.24 (0.10–0.36)**	0.33 (0.15–0.49)**
work_sitting − sleep	0.53 (0.42–0.63)	0.69 (0.56–0.78)	Deamplification, a_{21}/a_{11} = 0.71 (0.57–0.88)**	0.27 (0.15–0.39)**	0.40 (0.21–0.56)**
work_sitting − leisure	0.51 (0.38–0.61)	0.69 (0.57–0.78)	Deamplification, a_{21}/a_{11} = 0.89 (0.72–1.10)	0.18 (0.05–0.30)**	0.29 (0.08–0.49)**
work_peak − sleep	0.53 (0.42–0.63)	0.60 (0.44–0.73)	Deamplification, a_{21}/a_{11} = 0.60 (0.46–0.76)**	0.26 (0.12–0.39)**	0.38 (0.19–0.55)**
work_peak − leisure	0.51 (0.39–0.62)	0.60 (0.43–0.73)	Deamplification, a_{21}/a_{11} = 0.78 (0.61–0.99)**	0.14 (0–0.29)	0.21 (0–0.43)
task − rest	0.58 (0.42–0.69)	0.50 (0.34–0.63)	Deamplification, a_{21}/a_{11} = 0.79 (0.69–0.89)**	0.06 (0.02–0.11)**	0.38 (0.19-0.53)**
pvRSA					
wake − sleep	0.46 (0.32–0.58)	0.55 (0.43–0.66)	Deamplification, a_{21}/a_{11} = 0.69 (0.52–0.91)**	0.27 (0.16–0.38)**	0.34 (0.19–0.49)**
work − sleep	0.47 (0.33–0.59)	0.52 (0.33–0.66)	Deamplification, a_{21}/a_{11} = 0.52 (0.26–0.81) **	0.39 (0.21–0.54)**	0.43 (0.24–0.60)**
work_sitting − sleep	0.47 (0.33–0.59)	0.55 (0.34–0.71)	Deamplification, a_{21}/a_{11} = 0.58 (0.30–0.88) **	0.40 (0.21–0.55)**	0.47 (0.25–0.64)**
work_sitting − leisure	0.43 (0.27–0.57)	0.53 (0.33–0.69)	Deamplification, a_{21}/a_{11} = 0.89 (0.62–1.26)	0.15 (0–0.32)	0.22 (0–0.47)
work_peak − sleep	0.47 (0.34–0.59)	0.34 (0.09–0.56)	Deamplification, a_{21}/a_{11} = 0.39 (0.07–0.72) **	0.28 (0.05–0.47)**	0.40 (0.15–0.60)**
work_peak − leisure	0.44 (0.28–0.58)	0.33 (0.14–0.54)	Deamplification, a_{21}/a_{11} = 0.86 (0.56–1.18)	0 (0–20)	0.01 (0–0.31)
task − rest	0.50 (0.30–0.65)	0.53 (0.34–0.67)	Deamplification, a_{21}/a_{11} = 0.86 (0.69–1.09)	0.11 (0.03–0.19)**	0.31 (0.09–0.49)**
PEP					
wake − sleep	0.38 (0.24–0.50)	0.41 (0.28–0.53)	Deamplification, a_{21}/a_{11} = 0.88 (0.67–1.16)	0.11 (0.02–0.20)**	0.18 (0.03–0.32)**
work − sleep	0.38 (0.25–0.51)	0.45 (0.26–0.61)	Deamplification, a_{21}/a_{11} = 0.92 (0.61–1.31)	0.16 (0–0.30)	0.21 (0–0.40)
work_sitting − sleep	0.38 (0.25–0.51)	0.44 (0.23–0.61)	Deamplification, a_{21}/a_{11} = 0.99 (0.66–1.42)	0.13 (0–0.30)	0.18 (0–0.41)
work_sitting − leisure	0.25 (0.11–0.39)	0.44 (0.26–0.58)	Amplification, a_{21}/a_{11} = 1.44 (1.07–1.99)**	0.02 (0–0.13)	0.19 (0.02–0.44)**
work_peak − sleep	0.38 (0.24–0.50)	0.39 (0.17–0.57)	Deamplification, a_{21}/a_{11} = 0.86 (0.50–1.30)	0.17 (0–0.34)	0.22 (0–0.45)
work_peak − leisure	0.26 (0.12–0.39)	0.42 (0.23–0.58)	Amplification, a_{21}/a_{11} = 1.43 (1.07–1.92)**	0 (0–0.11)	0.10 (0.01–0.33)**
task − rest	0.30 (0.09–0.49)	0.32 (0.11–0.49)	Deamplification, a_{21}/a_{11} = 0.97 (0.74–1.29)	0.04 (0.01–0.07)**	0.20 (0.04–0.35)**

IBI = interbeat interval; pvRSA = peak valley respiratory sinus arrhythmia; PEP = preejection period; h2 = heritability; CI = confidence interval.

** Significant effect ($p < .01$).

our group, stress-specific heritability for HR was seen in an adolescent and in a middle-aged sample (28). No stress-specific genetic effects for RSA were found in both age groups, but another study in adolescents did report stress-specific genetic effects on HR variability (30). For PEP, stress-specific genetic effects were found in adolescent but not in middle-aged twins (28). Although we did find stress-specific genetic effects on HR, RSA, and PEP in the current study during standard task stress, the influence of stress-specific genetic effects in real life was generally larger. The partly divergent findings may be due to the nature of the stress tasks that were used. De Geus et al. (28) used short mental stress tasks to induce stress, whereas Wang et al. (30) used tasks that more closely approached stress in real life by including virtual reality car driving, a video game challenge, and a social competence interview. Combining this with the evidence from the current ambulatory study, we hypothesize that the use of more ecologically valid stressors tends to increase the contribution of novel genetic factors to individual differences in stress reactivity.

Previous prospective studies on the health effects of reactivity have been limited to short-term reactivity to standard laboratory stress tasks (1,2). Correlations between short-term stress reactivity calculated from our standard rest and stress tasks with ambulatory reactivity were weak, which confirms previous research finding poor generalizability of artificial to more realistic reactivity measures (8–11). This again suggests that the stress induced in the typical laboratory setting may not fully capture the individual differences measured during stress in real life.

An important question is whether our results allow us to select a clear favorite ambulatory autonomic reactivity measure from unstructured 24-hour recordings for use in stress research, using the temporal stability, heritability, and the ability to detect stress-specific genetic variance as prioritization criteria. For IBI, these criteria do not point to an optimal reactivity measure, nor do they disqualify any of the six reactivity measures. Temporal stability was good and of similar magnitude or even better as in previous ambulatory studies (46–48). It was particularly good when compared with the temporal stability of laboratory reactivity for HR (49–57), PEP (49,50,54,56), or HR variability (50,58,59). Five IBI reactivity measures were heritable, and due to genetic emergence, more than half of the variation in stress levels of HR was due to genetic factors. For PEP and RSA, temporal stability was moderate to good, and with few exceptions, all reactivity measures were also significantly heritable. For PEP during work_sitting and work_peak, no new genetic variance emerged compared with leisure time, instead existing variance got amplified. Using MZ correlations as an indicator of reliability, IBI, RSA, and PEP levels also performed well for all six conditions, with no clear "best" baseline or stress condition. We therefore conclude that all six measures as defined here have comparable properties from a psychometric viewpoint.

So that an optimal ambulatory autonomic reactivity measure can be selected, additional criteria, more related to content, need to be considered. Three of these are the mean absolute size of reactivity, the extent of the individual differences (variance), and the avoidance of contrasts incorporating changes in posture and physical activity. Statistical power to detect correlation of reactivity with health outcomes scales with its variance and effect size. Standard deviations for the reactivity values were of near comparable magnitude as those for the IBI, pvRSA, and PEP levels, implying that there is considerable variation in individual responses to daily life situations. This held true for all reactivity measures incorporating either sleep or leisure time as a baseline, although most variation was seen in the reactivity measures that included sleep as a baseline. When higher absolute values of reactivity (reflecting larger autonomic engagement) are considered, reactivity measures based on sleep also seem most favorable because they yield the strongest reactivity. However, sleep versus wake contrasts also contain changes in posture and physical activity which are known to influence RSA and PEP through effects other than true changes in cardiac autonomic control (40,60). Validity of PEP as a readout for cardiac sympathetic control is sensitive to distortion by postural shifts as it is affected by both preload and afterload (40). The contrasts that avoid large influences of posture change and that are characterized by low physical activity may therefore be optimal, that is, Δwork_sitting – leisure and Δwork_peak – leisure. However, these were also the contrasts with lowest absolute reactivity, making them less attractive. The magnitude of reactivity computed for these two work conditions with sleep as a baseline (Δwork_sitting – sleep and Δwork_peak – sleep) was much higher. Work_sitting and work_peak were likely characterized by active mental and/or social engagement with the environment, but in the presence of minimal physical activity. As the postural shift from supine to sitting is less severe than the shift from supine to standing (which could occur frequently during wake and nonsitting work periods), Δwork_sitting – sleep and Δwork_peak – sleep may be the preferred measures for studies seeking to link (genetic variants for) individual differences in ambulatory reactivity to health outcomes.

A limitation of this study is the use of objective landmarks in unstructured 24-hour recordings to delineate periods as stressful without subjective confirmation of stress by the participant. We reasonably assumed that time spent at work would be more enriched than leisure or sleep by mentally and socially engaging activities and that the presence of a high HR within a period of accelerometer-confirmed minimal physical activity was likewise attributable to the effects of such activities. However, the selection of periods of

high HR at work is meaningful only when all participants encounter comparable work stressors at the recording day and at least one stressor with sufficient salience to trigger substantial HR reactivity. Therefore, our physiology-driven definition of peak stress may have led us to inadvertently label people who did not encounter a strong stressor at work as low HR reactors. To resolve this circularity, we should have assessed individual differences in the amount of subjective stress exposure at work, which we did not. On the other hand, subjective experience is known to be only very poorly correlated with physiological reactivity (a topic recently covered in a special issue of *Biological Psychology* (61)). The critical question is whether autonomic reactivity as defined here is a good predictor of health outcomes. This of course remains to be tested—here we have shown that there is stable, heritable variation in ambulatory reactivity that can be meaningfully used in future predictive studies.

A further limitation is that the ambulatory baseline conditions that were defined for this study, leisure time, and sleep, occurred poststress in the real-life assessment. This means that recovery processes instead of a "true baseline" level may have been measured. If recovery processes were indeed involved during leisure time or sleep, this may have contributed to the poor correlation between our standard and ambulatory reactivity measures, as a pretest baseline was used to compute the standard stress reactivity.

In conclusion, ambulatory autonomic reactivity extracted from an unstructured real-life setting shows reliable, stable, and heritable individual differences. Real-life situations uncover new and different genetic variation compared with that seen in resting baseline conditions, including sleep. The contrasts between sitting work levels, including the peak stress period, and sleep baseline seem the most promising ambulatory reactivity measures for research in the field of psychosomatic medicine.

Source of Funding and Conflicts of Interest: This work received financial support from the Netherlands Organisation for Scientific Research. This work has been funded completely in the public domain. and we declare that we have no financial interest in this or any other outcome of our analyses.

REFERENCES

1. Treiber FA, Kamarck T, Schneiderman N, Sheffield D, Kapuku G, Taylor T. Cardiovascular reactivity and development of preclinical and clinical disease states. Psychosom Med 2003;65:46–62.
2. Chida Y, Steptoe A. Greater cardiovascular responses to laboratory mental stress are associated with poor subsequent cardiovascular risk status: a meta-analysis of prospective evidence. Hypertension 2010;55:1026–32.
3. Light KC. Hypertension and the reactivity hypothesis: the next generation. Psychosom Med 2001;63:744–6.
4. Steptoe A, Kivimaki M. Stress and cardiovascular disease. Nat Rev Cardiol 2012;9:360–70.
5. Kamarck TW, Jennings JR, Debski TT, Glickman-Weiss E, Johnson PS, Eddy MJ, Manuck SB. Reliable measures of behaviorally-evoked cardiovascular reactivity from a PC-based test battery: results from student and community samples. Psychophysiology 1992;29:17–28.
6. Kamarck TW, Jennings JR, Stewart CJ, Eddy MJ. Reliable responses to a cardiovascular reactivity protocol: a replication study in a biracial female sample. Psychophysiology 1993;30:627–34.
7. Matthews KA, Katholi CR, McCreath H, Whooley MA, Williams DR, Zhu S, Markovitz JH. Blood pressure reactivity to psychological stress predicts hypertension in the CARDIA study. Circulation 2004;110:74–8.
8. Johnston DW, Tuomisto MT, Patching GR. The relationship between cardiac reactivity in the laboratory and in real life. Health Psychol 2008;27:34–42.
9. Kamarck TW, Debski TT, Manuck SB. Enhancing the laboratory-to-life generalizability of cardiovascular reactivity using multiple occasions of measurement. Psychophysiology 2000;37:533–42.
10. Kamarck TW, Schwartz JE, Janicki DL, Shiffman S, Raynor DA. Correspondence between laboratory and ambulatory measures of cardiovascular reactivity: a multilevel modeling approach. Psychophysiology 2003;40:675–83.
11. Schwerdtfeger AR, Schienle A, Leutgeb V, Rathner EM. Does cardiac reactivity in the laboratory predict ambulatory heart rate? Baseline counts. Psychophysiology 2014;51:565–72.
12. Steptoe A, Cropley M, Joekes K. Job strain, blood pressure and response to uncontrollable stress. J Hypertens 1999;17:193–200.
13. Schwartz AR, Gerin W, Davidson KW, Pickering TG, Brosschot JF, Thayer JF, Christenfeld N, Linden W. Toward a causal model of cardiovascular responses to stress and the development of cardiovascular disease. Psychosom Med 2003;65:22–35.
14. McEwen BS. Protective and damaging effects of stress mediators. N Engl J Med 1998;338:171–9.
15. Chida Y, Hamer M. Chronic psychosocial factors and acute physiological responses to laboratory-induced stress in healthy populations: a quantitative review of 30 years of investigations. Psychol Bull 2008;134:829–85.
16. Zanstra YJ, Johnston DW. Cardiovascular reactivity in real life settings: measurement, mechanisms and meaning. Biol Psychol 2011;86:98–105.
17. Steptoe A, Cropley M, Joekes K. Task demands and the pressures of everyday life: associations between cardiovascular reactivity and work blood pressure and heart rate. Health Psychol 2000;19:46–54.
18. van Doornen LJ. Sex differences in physiological reactions to real life stress and their relationship to psychological variables. Psychophysiology 1986;23:657–62.
19. van Doornen LJ, Knol DL, Willemsen G, de Geus EJ. The relationship between stress reactivity in the laboratory and in real-life: is reliability the limiting factor? J Psychophysiol 1994;8:297–304.
20. Lucini D, Norbiato G, Clerici M, Pagani M. Hemodynamic and autonomic adjustments to real life stress conditions in humans. Hypertension 2002;39:184–8.
21. Houtman IL, Bakker FC. Individual differences in reactivity to and coping with the stress of lecturing. J Psychosom Res 1991;35:11–24.

22. Houtman IL, Bakker FC. Stress and coping in lecturing, and the stability of responses across practice. J Psychosom Res 1991;35:323–33.
23. Riese H, van Doornen LJ, Houtman IL, de Geus EJ. Job strain in relation to ambulatory blood pressure, heart rate, and heart rate variability among female nurses. Scand J Work Environ Health 2004;30:477–85.
24. Steptoe A, Cropley M. Persistent high job demands and reactivity to mental stress predict future ambulatory blood pressure. J Hypertens 2000;18:581–6.
25. Vrijkotte TG, van Doornen LJ, de Geus EJ. Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. Hypertension 2000;35:880–6.
26. Vrijkotte TG, van Doornen LJ, de Geus EJ. Overcommitment to work is associated with changes in cardiac sympathetic regulation. Psychosom Med 2004;66:656–63.
27. de Geus EJC, van Lien R, Neijts M, Willemsen G. Genetics of Autonomic Nervous System Activity. In: Canli T, editor. The Oxford Handbook of Molecular Psychology London: Oxford, University Press; 2014:357–90.
28. de Geus EJ, Kupper N, Boomsma DI, Snieder H. Bivariate genetic modeling of cardiovascular stress reactivity: does stress uncover genetic variance? Psychosom Med 2007;69: 356–64.
29. Riese H, Rijsdijk FV, Ormel J, van Roon AM, Neeleman J, Rosmalen JG. Genetic influences on baroreflex sensitivity during rest and mental stress. J Hypertens 2006;24: 1779–86.
30. Wang X, Ding X, Su S, Li Z, Riese H, Thayer JF, Treiber F, Snieder H. Genetic influences on heart rate variability at rest and during stress. Psychophysiology 2009;46:458–65.
31. Willemsen G, Vink JM, Abdellaoui A, den BA, van Beek JH, Draisma HH, van DJ, van't ED, Geels LM, van LR, Ligthart L, Kattenberg M, Mbarek H, de Moor MH, Neijts M, Pool R, Stroo N, Kluft C, Suchiman HE, Slagboom PE, de Geus EJ, Boomsma DI. The adult Netherlands twin register: twenty-five years of survey and biological data collection. Twin Res Hum Genet 2013;16:271–81.
32. Kupper N, Willemsen G, Posthuma D, de BD, Boomsma DI, de Geus EJ. A genetic analysis of ambulatory cardiorespiratory coupling. Psychophysiology 2005;42:202–12.
33. Neijts M, van LR, Kupper N, Boomsma D, Willemsen G, de Geus EJ. Heritability of cardiac vagal control in 24-h heart rate variability recordings: influence of ceiling effects at low heart rates. Psychophysiology 2014;51:1023–36.
34. Kupper N, Willemsen G, Boomsma DI, de Geus EJ. Heritability of indices for cardiac contractility in ambulatory recordings. J Cardiovasc Electrophysiol 2006;17:877–83.
35. Boutcher YN, Boutcher SH. Cardiovascular response to Stroop: effect of verbal response and task difficulty. Biol Psychol 2006;73:235–41.
36. Renaud P, Blondin JP. The stress of Stroop performance: physiological and emotional responses to color-word interference, task pacing, and pacing speed. Int J Psychophysiol 1997;27:87–97.
37. Sung BH, Izzo JL Jr., Dandona P, Wilson MF. Vasodilatory effects of troglitazone improve blood pressure at rest and during mental stress in Type 2 diabetes mellitus. Hypertension 1999;34:83–8.
38. Tulen JH, Moleman P, van Steenis HG, Boomsma F. Characterization of stress reactions to the Stroop Color Word Test. Pharmacol Biochem Behav 1989;32:9–15.
39. de Geus EJ, Willemsen GH, Klaver CH, van Doornen LJ. Ambulatory measurement of respiratory sinus arrhythmia and respiration rate. Biol Psychol 1995;41:205–27.
40. Houtveen JH, Groot PF, de Geus EJ. Effects of variation in posture and respiration on RSA and pre-ejection period. Psychophysiology 2005;42:713–9.
41. Riese H, Groot PF, van den Berg M, Kupper NH, Magnee EH, Rohaan EJ, Vrijkotte TG, Willemsen G, de Geus EJ. Large-scale ensemble averaging of ambulatory impedance cardiograms. Behav Res Methods Instrum Comput 2003;35: 467–77.
42. Willemsen GH, de Geus EJ, Klaver CH, van Doornen LJ, Carroll D. Ambulatory monitoring of the impedance cardiogram. Psychophysiology 1996;33:184–93.
43. van Lien R, Schutte NM, Meijer JH, de Geus EJ. Estimated preejection period (PEP) based on the detection of the R-wave and d*Z*/d*t*-min peaks does not adequately reflect the actual PEP across a wide range of laboratory and ambulatory conditions. Int J Psychophysiol 2013;87:60–9.
44. de Geus EJC. Quantitative genetics in behavioral medicine. In: Steptoe A, editor. Handbook of Behavioral Medicine London, UK: Springer; 2010. 399–422.
45. Neale MC, Boker SM, Xie G, Maes HH. Mx: Statistical Modeling. 7th ed. Richmond: VCU; 2006.
46. Goedhart AD, Kupper N, Willemsen G, Boomsma DI, de Geus EJ. Temporal stability of ambulatory stroke volume and cardiac output measured by impedance cardiography. Biol Psychol 2006;72:110–7.
47. Goedhart AD, van der Sluis S, Houtveen JH, Willemsen G, de Geus EJ. Comparison of time and frequency domain measures of RSA in ambulatory recordings. Psychophysiology 2007;44:203–15.
48. Vrijkotte TG, Riese H, de Geus EJC. Cardiovascular reactivity to work stress assessed by ambulatory blood pressure, heart rate, and heart rate variability. In: Fahrenberg J, Myrtek M, editors. Progress in Ambulatory Assessment. Computer Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies. Seattle: Hogrefe & Huber; 2001:345–60.
49. Allen MT, Sherwood A, Obrist PA, Crowell MD, Grange LA. Stability of cardiovascular reactivity to laboratory stressors: a 2 1/2 yr follow-up. J Psychosom Res 1987;31:639–45.
50. Burleson MH, Poehlmann KM, Hawkley LC, Ernst JM, Berntson GG, Malarkey WB, Kiecolt-Glaser JK, Glaser R, Cacioppo JT. Neuroendocrine and cardiovascular reactivity to stress in mid-aged and older women: long-term temporal consistency of individual differences. Psychophysiology 2003;40:358–69.
51. Cohen S, Hamrick N, Rodriguez MS, Feldman PJ, Rabin BS, Manuck SB. The stability of and intercorrelations among cardiovascular, immune, endocrine, and psychological reactivity. Ann Behav Med 2000;22:171–9.
52. Hamer M, Gibson EL, Vuononvirta R, Williams E, Steptoe A. Inflammatory and hemostatic responses to repeated mental stress: individual stability and habituation over time. Brain Behav Immun 2006;20:456–9.
53. Hassellund SS, Flaa A, Sandvik L, Kjeldsen SE, Rostrup M. Long-term stability of cardiovascular and catecholamine responses to stress tests: an 18-year follow-up study. Hypertension 2010;55:131–6.
54. Llabre MM, Saab PG, Hurwitz BE, Schneiderman N, Frame CA, Spitzer S, Phillips D. The stability of cardiovascular parameters under different behavioral challenges: one-year follow-up. Int J Psychophysiol 1993;14:241–8.
55. Manuck SB, Garland FN. Stability of individual differences in cardiovascular reactivity: a thirteen month follow-up. Physiol Behav 1980;24:621–4.

56. Sherwood A, Girdler SS, Bragdon EE, West SG, Brownley KA, Hinderliter AL, Light KC. Ten-year stability of cardiovascular responses to laboratory stressors. Psychophysiology 1997;34:185–91.
57. Veit R, Brody S, Rau H. Four-year stability of cardiovascular reactivity to psychological stress. J Behav Med 1997;20:447–60.
58. Bertsch K, Hagemann D, Naumann E, Schachinger H, Schulz A. Stability of heart rate variability indices reflecting parasympathetic activity. Psychophysiology 2012;49:672–82.
59. Dragomir AI, Gentile C, Nolan RP, D'Antono B. Three-year stability of cardiovascular and autonomic nervous system responses to psychological stress. Psychophysiology 2014;51:921–31.
60. Grossman P, Wilhelm FH, Spoerle M. Respiratory sinus arrhythmia, cardiac vagal control, and daily activity. Am J Physiol Heart Circ Physiol 2004;287:H728–34.
61. Friedman BH, Stephens CL, Thayer JF. Redundancy analysis of autonomic and self-reported, responses to induced emotions. Biol Psychol 2014;98:19–28.