Effects of public green space on acute psychophysiological stress response: a systematic review and meta-analysis of the experimental and quasi-experimental evidence

Accepted for publication with Environment & Behavior

Lærke MYGIND[1], Eva KJELDSTED[1], Rikke HARTMEYER[1], Erik MYGIND[2], Matt P. STEVENSON[1,2], Daniel S. QUINTANA[3,4], & Peter BENTSEN[1,2]

[1]Steno Diabetes Center Copenhagen, the Capital Region of Denmark, Health Promotion, Gentofte, Denmark

[2]University of Copenhagen, Department of Geosciences and Natural Resource Management, Fredensborg, Denmark

[3]University of Oslo, NORMENT, KG Jebsen Centre for Psychosis Research, Oslo, Norway

[4]Oslo University Hospital, Division of Mental Health and Addiction, Oslo, Norway

[*]Corresponding author: Lærke Mygind, Steno Diabetes Center Copenhagen, the Capital Region of Denmark, Niels Steensens Vej 6
DK-2820 Gentofte, Denmark
E-mail: laerke.mygind.groenfeldt@regionh.dk Tel: +45 30 13 30 28 (direct)

Please cite as Mygind, L., Kjeldsted, E., Hartmeyer, R., Mygind, E., Stevenson, M.P., Quintana, D.S., Bentsen, P. (2019) Effects of public green space on acute psychophysiological stress response: a systematic review and meta-analysis of the experimental and quasi-experimental evidence. *Environment & Behavior XX(XX): XX-XX. Doi: XXXX*

Acknowledgements

We wish to acknowledge and thank the international board of experts, Dr. Ulrich Dettweiler, Dr. Niels Ejbye-Ernst, and Dr. Annette Bischoff, who have provided valuable supervision in the development of the review protocol and literature search string. Likewise, we are thankful to our project group, Dr. Peter Elsborg, Dr. Camilla Roed Otte, Mads Bølling, Julie Hellesøe Christensen, Paulina Sander Melby, and Marc Künkel Pedersen, for aiding with screening the literature and data extraction. Our sincerest thanks to the authors of the reviewed literature who have kindly shared their data with us and contributed to the development of the research field. Lastly, we would like to commend the anonymous reviewers for their rigorous and insightful feedback which has contributed greatly to the final manuscript. The research is funded by a grant from The Danish Outdoor Council, Denmark (ID#105821).

Abstract

Contact with nature is widely considered to ameliorate psychological stress, but the empirical support for a causal link is limited. We conducted a systematic review to synthesize and critically assess the evidence. Six electronic databases were searched. Twenty-six studies evaluated the difference between the effect of natural environments and that of a suitable control on the acute psychophysiological stress response. Eighteen studies were rated as being of moderate quality, 4 low quality, and 4 high quality. Meta-analyses indicated that seated relaxation ($g = .5$, $p =.06$) and walking ($g = .3$, $p =.02$) in natural environments enhanced heart rate variability more than the same activities in control conditions. Cortisol concentration measures were inconsistent. While intuitively and theoretically sound, the empirical support for acute stress-reducing effects of immersion in natural environments is tentative due to small sample sizes and methodological weaknesses in the studies. We provide guidelines for future research.

Keywords: biomarker, green exercise, mental health, relaxation, restorative environments, social ecology/human ecology

While factors that increase the risk of mental disorders are multifarious and complex (Tost, Champagne, & Meyer-Lindenberg, 2015), contact with nature has been suggested to ameliorate chronic stress and improve mental health (Hartig, Mitchell, Vries, & Frumkin, 2014; Tost et al., 2015). Cumulative exposure to events or environments perceived as stressful are suggested to over-activate neurobiological responses that are normally activated during adaptation to a threat. These responses are essential and generally protective, but when activated repeatedly under circumstances of chronic or overwhelming adversity, they can become pathogenic (Lupien, McEwen, Gunnar, & Heim, 2009; Shonkoff, Boyce, & McEwen, 2009). The phenomenon is sometimes referred to as *allostatic load* (McEwen, 1998) and is linked to a wide array of physical and mental health impairments, as well as physiological and emotional dysregulation (Lupien et al., 2009; Shonkoff et al., 2009). Restoration through contact with nature has frequently been mentioned as a theoretically plausible pathway to improve mental health (Hartig, 2008; Hartig et al., 2014; Maller, Townsend, Pryor, P. Brown, & St Leger, 2006; Tost et al., 2015). However, the empirical support for the effects of contact with nature on psychological stress remains unclear due to a scarcity of well-controlled studies (Tost et al., 2015).

Research Background and Hypotheses

We performed a systematic review and meta-analysis to quantitatively synthesize and critically assess the existing research literature investigating the effects of immersion in natural environments on psychophysiological stress response indicators. This review focuses on acute psychological stress response, measured by psychophysiological indicators, and not on the physiological dysregulation of other systems (e.g., cardiovascular or immune), caused by accumulated or chronic stress (i.e., allostatic load). In the following, we refer to the acute psychological stress response simply as the “stress response”.

The promise and challenge of psychophysiological measures. While stress response may be investigated using self-report measures or a mix of methods, we focused on psychophysiological measures for two reasons. First, psychophysiological measures are not influenced by response bias stemming, for example, from biophilic enculturation. Second, psychophysiological measures merit attention because they are vulnerable to specific methodological challenges, such as measurement reactivity and intrusiveness, that can render measurements useless (Clay-Warner & Robinson, 2015). Applications of psychophysiological measurement instruments outside of labs, for example, during or directly after immersion in natural environments, present a plethora of potential challenges distinct to the type of measurement. As such, psychophysiological measures offer both unique methodological opportunities and challenges. To fulfil the potential of psychophysiological stress indicators and thereby provide stronger assessments of the effects of immersion in nature on the stress response, special precautions should be taken during data generation. Therefore, systematic and thorough quality assessment, tailored to research using psychophysiological measures in the specific research domain, is essential.

While no universally recognized standard for stress response evaluation exists, we included only psychophysiological measures that provide valid indicators of neurobiological stress system activation or relaxation, e.g., cortisol, noradrenaline, or vagally mediated heart rate variability (HRV). These responses, among many others, are triggered or, in the case of vagally mediated HRV, reduced as part of the two main, interdependent stress response pathways: the hypothalamic-pituitary-adrenal (HPA) axis and the autonomic nervous system (ANS) (Marques, Silverman, & Sternberg, 2010). The ANS consists of two branches: the parasympathetic nervous system (PNS) and the sympathetic nervous system (SNS). In response to a perceived stressor, the HPA axis is responsible for the release of glucocorticoids and cortisol, whereas parasympathetic withdrawal and SNS activation allow the release of adrenaline and noradrenaline (Marques et al., 2010). These neural and neuroendocrine pathways can provide insights into the transmission of stress responses through the body.

Other measures sometimes used as stress response indicators, e.g., blood pressure or heart rate (HR), are influenced by multiple cardiac factors. Although related to stress response, HR constitutes an imprecise indicator since a multitude of physiological cardiac functions influence HR. In contrast, the specific input of the PNS, responsible for rest and relaxation, may be approximated by exploring fast changes in beat-to-beat intervals, i.e., the time domain (e.g., root mean square of successive differences [RMSSD]) or spectral measures (e.g., high frequency [HF] HRV) of HRV (Akselrod et al., 1981; Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996). The utility of vagally mediated measures of HRV as measures of stress response was recently reviewed and found to be valid (H.-G. Kim, Cheon, Bai, Lee, & Koo, 2018). Measures of vagally mediated HRV can be used as an indicator of adaptability and resilience to stressors (e.g., Chalmers, Quintana, Abbott, & Kemp, 2014) or as an indicator of acute psychophysiological stress response. We focus on the latter use of the indicator.

Literature review. While previous reviews provided valuable overviews over aspects of the effects of exposure to natural environments on stress response (Bowler, Buyung-Ali, Knight, & Pullin, 2010; Haluza, Schönbauer & Cervinka, 2014; Tsunetsugu, Park, & Miyazaki, 2010), none, to our knowledge, offer a contemporary, comprehensive overview in which studies are subjected to systematic and tailored quality appraisal. In a recent meta-analysis, Twohig-Bennett and Jones (2018) reviewed the benefits of nature on various health outcomes, including psychophysiological outcomes. Although the review had several strengths and displayed an impressive generalized overview of the evidence, the results should be interpreted with caution. The pooled measures in the review were based on heterogeneous exposures and study designs: exposures ranged from accessibility to nature, i.e., the proximity or opportunity to engage with nature, to various types of sustained and intentional engagement with natural environments. Observational and intervention studies were collapsed to generate mean differences between the highest and lowest green space exposure groups. While systematic quality assessments were conducted, the quality items used were generic, e.g., “Is the hypothesis/aim/objective of the study clearly described?” and “Are the interventions of interest clearly described?” (Twohig-Bennett & Jones, 2018, p. 632), to match the diversity of interventions, study designs, and outcomes included in the review. The authors found that 58.3% of the observational studies and 77% of the intervention studies scored 9 or more out of 11 items in their respective quality appraisal tools. As such, the usefulness of the quality assessments, particularly when used to assess the rigor in studies based on psychophysiological data, requires reevaluation. This type of data is highly sensitive to a number of exogenous sources of measurement disturbance, such as physical activity or diet, and requires strict experimental procedures to avoid confounding.

Research questions and hypotheses. In the present review, we focus exclusively on experimental and quasi-experimental research involving direct and immersive experiences of nature. To increase intervention homogeneity, we include only public green space, excluding private gardens, indoor nature, views of nature, and virtual nature. Based on quality assessments inspired by the review by Ohly et al. (2016), we provide methodological recommendations for future research tailored to the research designs and adapted to psychophysiological outcomes. Ohly et al. (2016) developed and applied a framework specifically for use with research using similar designs and exposures, although it was focused on cognitive performance outcomes. Synthesis and critical appraisal is warranted to quantify potential benefits that may be used to inform practice and policy, as well as future research in the field by formulating methodological recommendations. Therefore, this systematic review will address the following two key questions:

1) (How much) are psychophysiological indicators of the stress response improved by exposure to natural environments relative to indoor or outdoor control environments across populations?
2) a) How strong is the evidence when subjected to systematic quality assessment, and b) what recommendations for psychophysiological research can be drawn from these assessments?

Method

The systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009) (for PRISMA checklist, please see supplementary material A). The results presented in this review are a subsample of a systematic review with a wider scope (results from this larger systematic review are reported elsewhere, (Mygind et al., 2019; Mygind, Kjeldsted, Hartmeyer, Mygind, & Bentsen, 2018)). While the purpose of the larger systematic review was to investigate the effects and associations of nature-based interventions and programs on mental, social, and physical health, this review centers specifically on experimental and quasi-experimental studies utilizing psychophysiological stress markers. This strategy was based on the assumption that relevant papers would be framed within this broader scope and that themes included in the search string might be investigated using psychophysiological stress indicators. The search string included terms relating to the broader themes of mental health, well-being, psychological stress, life stress, and psychological resilience (for the specific search string, see supplementary material B). It was not feasible to include the search term *stress* due to the resultant extent of irrelevant hits relating to the verb stress, even in conjunction with the nature exposure search terms. The review protocol of the larger systematic review can be accessed on the PROSPERO register of systematic reviews (id: CRD42017057988).

Eligibility Criteria

Studies were included if they were based on within- or between-subject experimental or quasi-experimental designs with control groups and compared the effects of direct and immersive experiences of natural environments against those of a suitable control. As mentioned above, we included only public green space. Immersive experiences were operationalized as types of interactions with nature that were direct, intentional, and sustained. This, for example, did not include motorized activities or transport to and from work through natural environments. We included studies in which any type of control group or condition had been used, regardless of assignment approach. For within-subject comparisons, control conditions were considered suitable when the type of activities were held constant across the two environments. For between-subject comparisons, a "no-treatment" or "treatment-as-usual" condition was required, depending on the population. For example, a population diagnosed with hypertension or depression should be assigned to either the intervention or the usual treatment provided to the population. Since the present systematic review was nested within a larger systematic review, the initial search strategy did not specifically focus on psychophysiological stress indicators. As such, we did not predefine the dependent variables.

Publications were included in this review if they had been subject to peer review, presented one or more empirical original studies, were communicated in the English language, and had been published between January 2004 and May 2017. The latter of these criteria was chosen to build on existing reviews. We did not apply any restrictions pertaining to population characteristics. The eligibility criteria are summarized in Table 1.

[TABLE 1]

Information Sources

An electronic literature search was performed in ERIC, PsycINFO, Scopus, the Web of Science, SPORTDiscus, and Dissertation Abstracts and was finalized by the beginning of May 2017. A generic search string was adapted to the individual electronic databases. We obtained additional literature through snowballing, i.e., manual reference list checks (Greenhalgh & Peacock, 2005), of all eligible publications. When means and standard deviations (SDs) were not explicitly reported in the publications, authors were contacted. Authors were also contacted in relation to unclear quality appraisal items, e.g., randomization method.

Study Selection

All identified literature was screened by two individual reviewers (LM, EK, RH, EM) by reading titles and abstracts. Full-text eligibility was likewise determined by two independent reviewers. Disagreements were settled through discussion between the two reviewers. If an agreement could not be made, a third reviewer (LM, PB) made the final decision.

We identified 26 individual studies that met all eligibility criteria. These studies were, as mentioned above, a subsample of a larger systematic review. The results of the full screening procedures are included in Figure 1 and described in detail in Mygind et al. (2019).

Most publications from the larger review were excluded because the investigated health outcomes were outside the scope of this review, e.g., social well-being, physical activity, or cognitive functioning. Furthermore, as described in the introductory section, studies in which stress was measured by self-report (e.g., Morita et al., 2007; Takayama et al., 2014) were considered outside the scope of the review. Likewise, we excluded studies including only HR, blood pressure, or pulse as indicators of a stress response (Kjellgren & Buhrkall, 2010; Mao, Cao, et al., 2012; Sahlin et al., 2016). Fluctuations in these measures are related to the neurobiological stress activation system, but they are also influenced by a range of other cardiac influences. Due to the questionable construct validity, we excluded these studies.

Publications that were not subjected to formal peer review were excluded (Bertone, 2015; Thompson, 2014). Studies based on nonexperimental designs, for example, one-group before and after designs (Ochiai et al., 2015; Yu, Lee, Kim, Yoon, & Shin, 2016), were also considered ineligible for inclusion at this stage: the absence of a control group or condition provides poor internal validity and an inappropriate basis for conclusions related to effects. Two papers compared a sitting condition indoors with a walking condition outdoors (Gatersleben & Andrews, 2013; Toda, Den, Hasegawa-Ohira, & Morimoto, 2013). For the purpose of this review, we did not consider it appropriate to compare a physically active outdoor condition with a sedentary indoor condition, and the studies were considered ineligible on this basis.

[FIGURE 1]

Data Items and Extraction Process

Data were procured from the literature by a single reviewer using a consistent data extraction procedure. Data extracted from the literature included study information (i.e., publication year, authors, and country in which the study was performed), study sample (sample size, sex, participant characteristics, and age), study design (research design (Ryan et al., 2013)), activity, length of exposure, characteristics of natural and control conditions, outcome measures, and reported results. Means and SD scores were extracted from papers when possible.

Risk of Bias in Individual Studies

The risk of bias assessment was inspired by Ohly et al. (2016) who combined three quality indicator guidelines (Centre for Reviews and Dissemination, 2009; Critical Appraisal Skills Programme, 2013; Effective Public Health Practice Project, 2013) to assess studies investigating the effects of exposure to nature on attention restoration outcomes.

Quality indicators included rigor and transparency of 1) the study design (e.g., reporting of power calculation); 2) potential confounders (e.g., baseline balance in psychophysiological measures between groups or pre-exposure to conditions); 3) intervention integrity (e.g., clear description of the intervention and control and consistency of the intervention); 4) data collection methods (e.g., blinding of outcome assessors); 5) analyses (e.g., usage of appropriate statistical methods); and 6) external validity (i.e., participants' representativeness of the population). Furthermore, we added a unique quality indicator that relates specifically to the use of psychophysiological outcome measures. Psychophysiological measures are extremely sensitive to influences of stable variables, e.g., smoking or habitual consumption of alcohol, and transient variables, e.g., physical activity and food consumption before and during the experiment. Strategies should be employed to control for these confounders by asking the participants to abstain from or follow rules regarding these behaviors or thorough statistical control (Quintana & Heathers, 2014). Taking actions, per protocol or statistical control, to regulate confounders was therefore included in the quality assessment scheme.

Ten percent of the full sample was assessed by two reviewers to validate the assessments. Only minor discrepancies were found and solved through discussion.

Summary Measures and Synthesis of Results

Eleven studies presented comparable psychophysiological measures, i.e., HF HRV and cortisol, and types of activities, i.e., walking and seated relaxation, as well as the necessary statistical information, i.e., mean and SD. To ensure intervention and outcome homogeneity, four individual meta-analyses for each combination of activity and outcome were conducted.

Random effects meta-analyses were performed in RevMan 5.3 (The Cochrane Collaboration, 2014) to calculate pooled effect sizes. The principal summary measure was standardized difference in means (Hedges' *g*). Hedges' *g* is used to measure the magnitude of the difference between the intervention and control post-exposure means and is standardized by dividing the difference between the means by the pooled SD of the means. Effect sizes and confidence intervals were illustrated by forest plots, i.e., graphical summaries of the estimated results for the individual study and pooled findings from the meta-analyses. Individual references and study information are included in the left-hand column, and the right-hand column lists the measures of the effect. Here, individual studies as well as the pooled effect sizes are plotted with confidence intervals represented by horizontal lines. The forest plot includes a vertical line representing no effect. Differences between post-exposure means were deemed significant at a *p*-value lower than .05. The extent of between-study heterogeneity for each meta-analysis was tested using the I^2 statistic, which indicates the fraction of variance in the effect size estimates that is caused by heterogeneity.

The individual meta-analyses were divided into subgroups according to whether individual studies had reported that psychophysiological outcomes 1) had been balanced at baseline, i.e., no statistical difference between groups; 2) had not been balanced at baseline, i.e., a significant difference between groups; or 3) were missing explicit statistical information to determine baseline balance. This information is essential for determining whether the potential for change in the outcome measures was comparable between groups or, for within-subject studies, before exposure to the different environments. For example, lower levels of cortisol before an exposure could result in a lower or higher responsivity to the exposure, which in turn provides a poor background for comparison of post-exposure means. Therefore, balanced baselines are preferable. Funnel plots were generated to visualize the risk of small-study bias. Funnel plots illustrate the relationship between the magnitude of the difference, i.e., effect sizes, in the individual studies versus variability in the magnitude of the difference, i.e., standard error of the effect sizes. Information pertaining to study quality was included in the funnel plots.

When meta-analyses did not indicate significant differences in effects, we used the two one-sided test (TOST) procedure to assess the evidence for statistical equivalence (Lakens, 2017). The TOST procedure requires a determination of the smallest effect size of interest against which an assessment of data insensitivity or statistical equivalence can be made. The smallest effect size of interest can be determined theoretically or from existing empirical work. Here, we defined the smallest effect size of interest to be small based on a theoretical consideration: when accessible on a population level, small intervention benefits for individuals can be powerful preventive measures (Rose, 2001). Given that natural environments are accessible in some form to most people, we determined the smallest effect size of interest to be small and set equivalence test bounds at .2. Equivalence tests were conducted using the TOSTER spreadsheet (version 0.4.6) (Lakens, 2017).

Results

Study Characteristics

Of the 26 included studies, most were from Asian countries (*n* = 18), with substantial representation from Japan (*n* = 15). Six studies were from European countries, and one study was derived from the USA. Individual studies are presented in Table 2.

Research designs. Three experimental studies were designed as between-subject, randomized controlled trials, but only D.K. Brown, Barton, Pretty, and Gladwell (2014) reported the randomization procedure used and applied a type of randomization (i.e., computer-randomized numbers) that was in accordance with the Cochrane recommendations

(Ryan et al., 2013). Accordingly, we categorized the two remaining studies (Calogiuri et al., 2015; Mao, Lan, et al., 2012) as quasi-randomized trials. Most studies were experimental and designed as within-subject, randomized, crossover trials (*n* = 16). Among the crossover trials, only Gidlow, Jones et al. (2016) described the manner by which randomization was performed. Three studies were quasi-experimental and utilized within-subjects designs without crossover of the sequence of exposures to reduce potential order bias (Aspinall, Mavros, Coyne, & Roe, 2015; Li et al., 2011, 2016). Last, four studies were quasi-experimental and designed as controlled before-and-after studies in which a group of participants was pragmatically assigned to an intervention or a control group.

Participants. Half of the studies included university students (*n* = 13), and 85% of these studies included males. Six studies included healthy adult participants (Beil & Hanes, 2013; D. K. Brown et al., 2014; Calogiuri et al., 2015; Gidlow, Jones, et al., 2016; Li et al., 2011; Tyrväinen et al., 2014), one study included healthy students in junior high school (Hohashi & Kobayashi, 2013), and one study included students in elementary school (Dettweiler, Becker, Auestad, Simon, & Kirsch, 2017). Two studies included middle-aged individuals with prehypertension or stage one hypertension (Li et al., 2016; Song, Ikei, Kobayashi, et al., 2015), and one study included elderly hypertensive patients (Sung, Woo, Kim, Lim, & Chung, 2012). Last, one study included individuals with major depressive disorder (W. Kim, Lim, Chung, & Woo, 2009).

Activities and duration of exposure. The two most common types of activities included walking and seated relaxation for short durations of time (from ten to 30 minutes). In five studies, participants performed both activities (Park, 2009; Park et al., 2007; Park, Tsunetsugu, Kasetani, Kagawa, & Miyazaki, 2010; Tsunetsugu et al., 2007; Tyrväinen et al., 2014). One study explored the acute effects of 30 minutes horseback riding in nature versus simulated horseback riding indoors (Matsuura et al., 2011).

A few interventions were of longer duration (e.g., one day that included two walks (Li et al., 2011, 2016) or an eight-week walking program (D. K. Brown et al. (2014)). Other types of more complex interventions included forest therapy, where the accumulated effects of cognitive behavior therapy and nature exposure were compared to treatment as usual (W. Kim et al., 2009; Sung et al., 2012); cognitive behavior therapy with no nature exposure (W. Kim et al., 2009); and control groups that were provided with no treatment at all (Han et al., 2016). The interventions ranged from two days (Han et al., 2016) to eight weeks (Sung et al., 2012). Dettweiler et al. (2017) investigated how curriculum-based teaching in a natural environment affected psychophysiological measures in comparison to the effects of typical, classroom-based teaching.

Outcomes. Psychophysiological outcomes included serum and salivary cortisol (*n* = 13), HF HRV (*n* = 12), time domain HRV indicators (e.g., standard deviation of the normal-to-normal [SDNN] interbeat intervals and RMSSD) (*n* = 2), salivary amylase (*n* = 3), adrenaline, noradrenaline and dopamine (*n* = 2), and cortisol awakening response (*n* = 1). Hemoglobin concentration in the prefrontal area of the brain was used as an indicator of stress response (*n* = 1), although descriptions of the utility of the measure in the study was scarce. Last, distinct parts of the cerebral cortex in the brain were used as indicators of stress, arousal and frustration (*n* = 1). The latter two indicators are less well-established measures with limited background research to back up their utility as stress response measures.

Risk of Bias within Studies

Overall, most studies (69%, *n* = 18) were categorized as being of moderate quality, four studies of low quality, and four of high quality (see Figure 2). Scores ranged from 27.8% to 83.3%. The quality assessment of individual studies is presented in Table 3. The sum of these indicators was intended to reflect the overall quality of the studies in the context of this review's focus, which may not be identical to the individual study aim. Quality indicators that were frequently unclear included statistical power calculation, randomization procedures, blinding of participants (in relation to research question and aims) and outcome assessors (in relation to assignment of participants), and sample representativeness. While outcomes, i.e., means and SDs, were not consistently reported in all studies, this information was in several cases supplemented through direct contact with the authors of the individual studies.

[TABLE 2]

[FIGURE 2]

[TABLE 3]

Results of Individual Studies

Individual studies most frequently (*n* = 13) reported that natural environments improved psychophysiological outcomes more than control environments and conditions (see Table 2). Among these 13 studies, two were rated as being of low quality (Aspinall et al., 2015; Park et al., 2010) and eleven were rated as being of moderate quality (Dettweiler et al., 2017; Han et al., 2016; W. Kim et al., 2009; Lee, 2014; Lee, Park, Tsunetsugu, Kagawa, & Miyazaki, 2009; Lee et al., 2011; Li et al., 2011, 2016; Mao, Lan, et al., 2012; Song, Ikei, Igarashi, Takagaki, & Miyazaki, 2015; Sung et al., 2012). On average, these studies scored 50%.

Four studies reported mixed findings of alternate psychophysiological outcomes (Calogiuri et al., 2015; Lee, Park, Ohira, Kagawa, & Miyazaki, 2015; Park et al., 2007; Tsunetsugu et al., 2007). For example, Tsunetsugu et al. (2007) reported that walking in natural environments reduced cortisol levels more than walking in the urban comparison environment, while no significant differences could be observed in HF HRV. Of these studies, one was rated as being of high quality (Calogiuri et al., 2015) and three were of moderate quality (Lee et al., 2015; Park et al., 2007; Tsunetsugu et al., 2007). These studies scored 59% on average.

Two studies showed that individual indicators, measured over several time points, provided mixed results: indicators were at some time points different in favor of the natural environment, while no statistically significant differences could be observed at other time points (Park, 2009; Park et al., 2008). These studies were deemed as being of low quality and scored 33% on average.

Finally, seven studies reported no significant differences between natural and control environments and conditions (Beil & Hanes, 2013; D. K. Brown et al., 2014; Gidlow, Jones, et al., 2016; Hohashi & Kobayashi, 2013; Matsuura et al., 2011; Tyrväinen et al., 2014; Yamaguchi, Deguchi, & Miyazaki, 2006). Three of these were rated as high quality (Beil & Hanes, 2013; D. K. Brown et al., 2014; Gidlow, Jones, et al., 2016) and four were rated as moderate quality (Hohashi & Kobayashi, 2013; Matsuura et al., 2011; Tyrväinen et al., 2014; Yamaguchi et al., 2006). Studies indicating no difference scored, on average, 64% out of the quality appraisal items.

Synthesis of Results

In the following sections, we present the meta-analyses of the effects of walking and seated relaxation in the natural and control environments on cortisol and HF HRV, respectively.

Effects of walking in natural versus control environments. Six studies explored the effects of walking in natural versus control environments on serum and salivary cortisol (Gidlow, Jones, et al., 2016; Mao, Lan, et al., 2012; Park et al., 2007, 2010; Tsunetsugu et al., 2007; Tyrväinen et al., 2014). Four experimental studies, for which all necessary data were available, were included in the meta-analyses, and data for 72 unique participants were pooled. The studies were rated as being of moderate (Mao, Lan, et al., 2012; Park et al., 2007; Tsunetsugu et al., 2007) to high (Gidlow, Jones, et al., 2016) quality. Only two studies reported that baseline cortisol levels were balanced (see Figure 3). For example, Mao, Lan, et al. (2012) presented descriptive statistics to demonstrate that the intervention and control groups had similar baseline serum cortisol concentrations, whereas Gidlow, Jones, et al. (2016) used a statistical approach that controlled for baseline differences but did not report any baseline comparison statistics. The pooled effects of walking in natural versus control environments on serum and salivary cortisol measures were not statistically significant ($g = -.27$ [-.85 to 0.3], $p = .35$). Equivalence tests indicated that we could not reject an effect as large or larger than .2 ($Z = .3$, $p = .95$), indicating data insensitivity rather than an absence of a “true” effect. The heterogeneity statistics suggested that variance within and between studies may have been an issue, although conventionally speaking, not at a significant level ($I^2 = 58\%$, $p = .07$). Figure 4 illustrates the relationship between effect sizes and variability in effect sizes and suggests a larger variability among the studies that included small samples and that were of moderate quality. The number of studies included in the funnel plot is limited and should be interpreted with caution.

Seven studies explored the effects of walking in natural versus control environments on HF HRV (D. K. Brown et al., 2014; Gidlow, Jones, et al., 2016; Lee, 2014; Park, 2009; Park et al., 2010; Song, Ikei, Kobayashi, et al., 2015; Tsunetsugu et al., 2007). Five experimental studies, encompassing a total of 130 unique participants, were included in the meta-analysis (Gidlow, Jones, et al., 2016; Lee, 2014; Park, 2009; Song, Ikei, Kobayashi, et al., 2015; Tsunetsugu et al., 2007). HF HRV was higher after walking in natural environments than after walking in control environments ($g = .31$ [.06 to .55], $p = .01$), a finding associated with a small effect size (see Figure 5). The studies were of low (Park, 2009), moderate (Lee, 2014; Song, Ikei, Kobayashi, et al., 2015; Tsunetsugu et al., 2007), and high quality (Gidlow, Jones, et al., 2016). Higher quality studies with larger sample sizes displayed the least within-study variability and limited between-study disagreement in effect sizes (see Figure 6). A subgroup analysis including only the four studies that reported that participant levels of HF HRV were balanced at baseline (see Figure 5) provided similar results, although with a slightly larger effect size (*g* = .39 [.01 to .69], *p* = .01). There was no evidence of heterogeneity between studies (all studies: I^2 = 0%, *p* = .55, studies with balanced baselines: I^2 = 0%, *p* = .57).

[FIGURES 3-6]

Effects of seated relaxation in natural versus control environments. Eight studies explored the effects of seated relaxation in natural versus control environments on serum and salivary cortisol (Beil & Hanes, 2013; Lee et al., 2009, 2011; 2015, Park et al., 2007, 2010; Tsunetsugu et al., 2007; Tyrväinen et al., 2014). Estimates were pooled across 74 unique participants and six small-scale, experimental studies of high (Beil & Hanes, 2013) and moderate quality (Lee et al., 2009, 2011, 2015; Park et al., 2007; Tsunetsugu et al., 2007), of which three reported balanced baseline cortisol levels (see Figure 7). Salivary cortisol levels were lower after seated relaxation in natural environments compared to after seated relaxation in control environments (*g* = -.72 [-1.19 to -.25], *p* = .003), with an effect size associated with a medium effect (see Figure 7). Variance between studies was observed, but not at a statistically significant level (I^2 = 48%, *p* = .09). A subgroup analysis including only the three studies that reported that participant levels of cortisol were balanced at baseline did not indicate a differential effect of seated relaxation in natural environments compared to seated relaxation in control environments (*g* = -.61 [-1.26 to .04], *p* = .07, I^2 = 48%). The funnel plot (see Figure 8) hinted a linear tendency for the studies with the largest variability in effect sizes to also report the largest effect sizes in favor of the natural environments. As such, the individual study effect sizes were not distributed symmetrically around the pooled effect size.

Six studies investigated how seated relaxation in natural and control environments affected HF HRV (Lee et al., 2011, 2015; Park, 2009; Park et al., 2008, 2010; Tsunetsugu et al., 2007). Four experimental studies encompassing a total of 48 unique participants were included in the meta-analysis (Lee et al., 2011, 2015; Park, 2009; Tsunetsugu et al., 2007). The quality of the individual studies was low (Park, 2009) or moderate (Lee et al., 2011, 2015; Tsunetsugu et al., 2007). HF HRV tended to be higher after seated relaxation in natural environments than after seated relaxation in control environments ($g = .51$ [-.01 to 1.03], $p = .06$), a finding associated with a medium effect size although, conventionally speaking, not significant (see Figure 9). Equivalence tests indicated that an effect as large or larger than .2 ($Z = -1.12$, $p = .99$) could not be rejected, indicating data insensitivity rather than an absence of a "true" effect. A subgroup analysis of only the three studies with balanced baseline HF HRV values showed unclear results ($g = .30$ [-.17 to .78], $p = .21$). There was no evidence of heterogeneity between studies (all studies: $I^2 = 36\%$, $p = .12$, subgroup of studies with balanced baseline values: $I^2 = 0\%$, $p = .43$). Although based on only four studies, the funnel plot (see Figure 10) illustrated a linear tendency in which larger within-study effect size variability was associated with larger effect sizes favoring the natural environments, rather than a random distribution around the pooled effect size estimate.

[FIGURES 7-10]

Discussion

Summary of Results

Across 26 experimental and quasi-experimental studies, most studies ($n = 13$) indicated that outcomes were improved more in the immersive nature-experience groups and conditions than in the control groups or conditions. These studies were rated as being of low to moderate quality, with an average score of 50%. Comparatively, studies ($n = 7$) that reported no differences between the immersive nature-experience groups and the control groups or conditions were rated as being of moderate to high quality, with an average score of 63%. While this difference is suggestive and does not, in itself, imply inefficacy, the body of literature may be skewed towards positive findings being based on predominantly low to moderate quality studies (Schwarzer, Carpenter, & Rücker, 2015).

Meta-analyses indicated that walking in natural environments, in comparison to walking in urban environments, enhanced vagally mediated HRV ($p = .02$). The effects of seated relaxation on vagally mediated HRV were, conventionally speaking, not significant. Equivalence tests indicated data insensitivity rather than a null effect, potentially as a result of small sample sizes and the quality of the studies.

The pooled results for cortisol were unclear. Across all studies, seated relaxation in natural environments provided favorable effects compared to the effects associated with urban environments. However, upon exclusion of studies in which baseline values were significantly unbalanced, no such effect could be observed. No effect could be observed for walking. The effect sizes were heterogeneous in all three cases, although according to conventions not at a significant level. Equivalence testing indicated that there was no statistical equivalence to indicate a null effect of walking in natural environments compared to the effects associated with the control conditions. As such, the cortisol findings were unsuited as a basis of quantifying to what extent psychophysiological indicators of stress were improved by exposure to natural environments relative to indoor or outdoor control environments across populations. At present, the most consistent results indicate small effects of walking in natural environments relative to walking in control environments on vagally mediated HRV.

Some of the observed heterogeneity could potentially be explained by methodological limitations that will be addressed more thoroughly in the strengths and limitations section below.

Results Compared to Previous Reviews

In 2014, Haluza et al. reviewed 17 studies exploring the physiological effects of exposure to natural environments. They included studies that had been published before January 2012 and a vast number of physiological outcomes, some of which were psychophysiological stress markers (e.g., salivary cortisol or amylase) and others that were not (e.g., white blood cell count). Since January 2012, 14 new papers utilizing psychophysiological stress markers specifically were published and included in this systematic review. As such, there appears to be a rapid growth in the quantity of this type of research.

The meta-analyses by Twohig-Bennett and Jones (2018) differ from the present meta-analyses in a number of ways. Most notably, statistically significant reductions in salivary cortisol (MD: -.05, $p < .001$) and enhancements of vagally mediated HRV (MD: 91.87, $p < .001$) were observed by Twohig-Bennett and Jones (2018) by comparing high versus low green space exposure groups. Furthermore, no heterogeneity was observed for the cortisol measures. Based on these results, nature interaction was interpreted to ameliorate stress response. However, some important differences in the included literature and meta-analytic approaches should be considered.

Twohig-Bennett and Jones (2018) included both observational and experimental studies in their pooled analysis of cortisol measures, and short- and long-termed exposures were collapsed. The meta-analysis included one study (Toda et al., 2013), which was excluded from the present systematic review since the control condition was considered unsuited for comparison: a sitting condition indoors was compared with a walking condition outdoors introducing a probability that physical activity confounded the results. Furthermore, information pertaining to balance of the pre-exposure outcome levels was not presented, which introduces a risk of bias due to unequal potential for change. For example, upon excluding unbalanced studies from our meta-analyses, we did not identify any significant differences in cortisol measures following seated relaxation in natural and control conditions. Given the mix of acute and accumulated responses to diverse interventions and inclusion of both experimental and observational studies, as well as a poorly controlled study, it seems possible that the statistical homogeneity derived in Twohig-Bennett and Jones (2018) is influenced by confounding variables and does not equate clinical homogeneity. We maintain that the heterogeneity in the cortisol measures in response to an acute effect of direct contact with natural environments calls for further investigation.

Previous reviews reported benefits of exposure to natural environments for self-reported, distinct positive emotions, e.g., tranquility, and negative emotions, e.g., anxiety (Bowler et al., 2010; Haluza et al., 2014), but not experiences of stress responses specifically. However, self-reported constructs related to stress responses seem to be improved. Bowler et al. (2010) found that these self-reported constructs were more consistently improved than physiological indicators such as cortisol or diastolic and systolic blood pressure. Bowler et al. (2010) argued that the self-reported measures could be influenced by prior beliefs of the participants and that the lack of blinding of the participants to the research question could be problematic. These issues would not pertain to psychophysiological measures. However, these measures are susceptible to other types of bias, and we discuss these further in the following sections.

Sources of heterogeneity. Pooled effects of walking and seated relaxation on cortisol favored exposure to natural environments (only significantly for seated relaxation) but were either significantly or close to being significantly heterogeneous. Participant characteristics were similar across the individual studies, and as such, unlikely to have contributed greatly to the observed heterogeneity between study effect sizes: all participants were male university students, except for the sample of older, healthy individuals included in the studies by Gidlow, Jones, et al. (2016) and Beil and Hanes (2013). However, inconsistent baseline cortisol levels between studies and small sample sizes could have introduced heterogeneity. The possibility of small-study effects was supported by visually asymmetric funnel plots suggesting that the smaller, low-to-moderate quality studies more often reported findings associated with larger effect size variability but also larger effect sizes in favor of the natural environments. While the number of studies included in the funnel plots was limited and served only as tentative indications, small-study effects have previously been found to skew meta-analyses towards exaggerated effect sizes (Schwarzer et al., 2015). Based only on small-scale, low- or moderate-quality studies, the estimated effect sizes related to seated relaxation for both vagally mediated HRV and cortisol could be overestimated and should be interpreted with caution.

In comparison, a slightly larger number of studies (*n* = 4) with a pooled total of 92 participants, in which vagally mediated HRV was used as an outcome measure, included balanced baseline values before walking in natural and control environments. Here, a small and significant effect size was observed. The same tendency could be observed for vagally mediated HRV after seated relaxation, although it is possible that the smaller sample sizes in congruence with low or moderate study quality resulted in extensive within-study variability and the, conventionally speaking, nonsignificant pooled effect size. As reported above, no evidence of between-study heterogeneity was observed for vagally mediated HRV for either the seated or the walking condition.

Cortisol as a measure of acute stress response. Studies have suggested that there might be a time lag in cortisol response to a psychological stressor of approximately 18 minutes in comparison to salivary amylase, which was immediately affected (Takai et al., 2004). While the cortisol response to a stressor may not be directly compared to the relaxing factor, it is possible that the timing of data collection is particularly important for cortisol measures. In the studies included in the meta-analysis, it was described in detail how salivary cortisol samples were collected using oral swabs from Salivette (Gidlow, Jones, et al., 2016; Lee et al., 2009, 2011, 2015; Park et al., 2007; Tsunetsugu et al., 2007) or Salimetrics kits (Beil & Hanes, 2013) collected before and after exposure (and immediately frozen in all studies). However, the exact timing from the end of exposure to collection of saliva samples was not reported. As such, varying timing of the measurements could result in heterogeneous effect sizes between studies. In comparison, vagally mediated HRV is theorized to provide an instantaneous, peripheral indicator of autonomous nerve activity. Although this could be a plausible explanation of the observed heterogeneity, it does not match the empirical results of the study conducted by Gidlow, Jones, et al. (2016). Here, two post-exposure cortisol measurements were performed, one that was immediately after exposure and an additional measure 30 minutes after exposure. Neither indicated a differential effect of the natural and urban environments. Corresponding results from vagally mediated HRV supported that there was no differential impact of the natural environment. In summary, the heterogeneity of the acute responses cannot be explained with reference to timing only, although it may be a contributing factor to be considered in future studies.

Strengths and Limitations of the Reviewed Studies

This review has taken steps toward quantifying the acute effects of natural environments on selected psychophysiological stress markers. The studies included in this systematic review were based on small sample sizes and were heterogeneous in terms of types and durations of exposure, psychophysiological outcome measures, baseline levels, level of detail in describing natural and control environments and conditions, and statistical methods used. The samples were skewed towards male students.

Distinguishing factors for the four most highly rated studies (Beil & Hanes, 2013; D. K. Brown et al., 2014; Calogiuri et al., 2015; Gidlow, Jones, et al., 2016) included clear and detailed descriptions of participant inclusion criteria and recruitment, as well as intervention and control conditions, which also made it possible to determine the consistency of the intervention between and within groups. Additionally, high-quality papers reported all outcomes and accounted for the loss of participants. Furthermore, the use of random assignment of participants to groups or sequences and the reporting of randomization procedures distinguished the high-quality papers (Beil & Hanes, 2013; D. K. Brown et al., 2014; Calogiuri et al., 2015). The authors reported attempts to blind participants to the research aims (Beil & Hanes, 2013; Calogiuri et al., 2015; Gidlow, Jones, et al., 2016), and in one case included power calculations in a prepublished study protocol (D. K. Brown et al., 2014). Prepublication of study protocols would greatly enhance the confidence with which results could be interpreted, as this reduces researcher biases. Future research would benefit from consideration of the quality items applied in this review and could take inspiration from the included high-quality papers.

Recruitment strategies and sociodemographic background variables were rarely described in much detail, making it difficult to determine the representativeness and generalizability of the sample. Furthermore, studies lacked descriptions of the natural and, especially, control environments. While the natural environments were often described to some extent in terms of type of vegetation and elevation, the control conditions were frequently only described as *urban*, perhaps with a picture to illustrate the environments (e.g., Mao, Lan, et al., 2012; Park et al., 2008).

Consequences of intervention blackboxing. *Blackboxing*, that is, the simplification of complexity pertaining to the characteristics and processes of the exposures, including the control conditions, limits the interpretational value of the comparisons and therefore the possibility of convincingly attributing stress-reducing or stress-recovering effects to the natural environment. In other words, if the response to natural scenery is compared with a bustling downtown area, the observed effects may be caused by a stressful experience as much as a stress-reducing influence of spending time in nature. Significantly, Gidlow, Jones, et al. (2016) did not find a difference in cortisol levels between what was described as a pleasant urban environment and natural environment. While the comparative nature of the evidence is a condition for this type of research, it also stresses the importance of thorough description of both natural and control environments. There is no environmental vacuum or placebo against which to measure the effect of exposure to an environment. The cursory descriptions of the environments and conditions hinders the identification of the composite features of immersive nature experiences that reduce stress, the judgment of the suitability of the control conditions, the interpretation of positive or absent effects, and the explanation of variations in effect sizes across studies. Some environmental characteristics that should be described and, if possible, measured include the type and density of vegetation and buildings, ground elevation, noise level, people density, ambient temperature, and humidity.

Blackboxing raises both interpretational and data sensitivity issues. Arguably, the cooling from trees and lower ambient temperatures can be seen as constituting elements that set natural and synthetic environments apart. Simultaneously, it introduces a potential spurious factor and inferential challenge for research based on psychophysiological measures. For example, several studies reported that the ambient temperature, which are known to relate to HRV (e.g., Liu, Lian, & Liu, 2008; Matsumoto et al., 1999; Wu et al., 2013), was higher in the control environments (e.g., Lee et al., 2009; Tsunetsugu et al., 2007). As such, the higher levels of HRV may be confounded by physiological processes related to the adaptation to ambient temperatures. To increase the inferential usefulness of the research, the effect of ambient temperature and humidity should be considered. Impacts of confounding factors that are not easily controlled per protocol, e.g., ambient temperature and humidity, should be considered in the statistical modeling, for example by mixed effects modeling (for example, see Calogiuri et al. (2015) or Dettweiler et al. (2017)).

Transient factors. In many of the studies, some or extensive action was taken to control for transient, confounding factors, e.g., physical activity and intake of food and beverages. In some studies (Park, 2009; Park et al., 2007; Tsunetsugu et al., 2007), vagally mediated HRV was measured dynamically while participants were walking. Although a few of these studies reported that attempts were made to equalize, e.g., walking speed and exhaustion, these were not quantitatively measured and controlled and could potentially introduce disturbances in the data. Given the importance of these factors for the reliability of psychophysiological measures, such as HF HRV (Quintana & Heathers, 2014), this should remain a high priority.

Sample sizes. While the studies included in the meta-analyses were homogenous in terms of outcome measures, type of activities and participant factors, issues such as small sample sizes and inconsistent baseline levels were likely factors that introduced heterogeneity within- and between-studies. In an effect size distribution analysis of almost 300 HRV effect studies, Quintana (2017) reported that, for case-control studies, an appropriate sample size would be 233 participants to identify a small effect size. Although prepost designs were not included in this analysis and the results were not necessarily transferable to designs included in this review (Quintana, 2017), the research was generally underpowered. Consequently, there is a need for larger, well-controlled studies to qualify the estimated effect sizes and to perform subgroup and moderator analyses. This could contribute to an understanding of the effects as well the heterogeneity that was addressed earlier. Subgroup and moderator analyses of special importance could relate to, for example, characteristics of natural environments (e.g., forest versus park, urban versus rural), control conditions (e.g., pleasant/not stressful versus unpleasant/stressful), participant characteristics (e.g., stressed versus not stressed or adults versus children), and timing of the cortisol measurements.

Study designs. Randomized crossover designs were frequently used and are indeed appropriate for filtering out individual-specific influences on acute effects. For an increased focus on accumulated effects of repeated exposures, which seems a worthwhile direction for future research, the performance of traditional between-subject randomized controlled trials (RCT) would be preferred. In RCTs, participants are randomly assigned to either an intervention or control group, which prevents selection bias and balances intervention and control groups with respect to many known and unknown confounding variables. The RCT is widely seen as the foremost approach to insuring the internal validity of a study and, ultimately, the best basis for causal inference. While RCTs are not always feasible and ethical, adaptive designs such as dynamic waitlisted designs (C. H. Brown et al., 2009), time series, or regression discontinuity designs (Gottfredson et al., 2015) may be useful.

Quality of reporting or quality of conduct. We received additional information about outcome measures from several studies (51%, *n* = 14) but less successfully gathered information about quality indicators that were unclear (*n* = 5). Consequently, a few studies could have been rated higher if all information had been available. As highlighted by Ohly et al. (2016) in the context of assessing studies that were potentially developed for other purposes than what was the focus of their systematic review, some quality indicators may have been rated poorly due to lack of reporting for the research question at hand. However, future studies of psychophysiological indicators would benefit from taking the quality items into account during study planning and conduct. This would improve the basis for accumulating high-quality evidence from which directions for practice could be formulated.

Strengths and Limitations of this Review

A notable weakness in these meta-analyses is related to small sample sizes in the included studies and that at least two of the meta-analyses were at risk of being influenced by small-study effects, indicated by asymmetric funnel plots. Small-study effects, which can encompass publication bias where small studies are more frequently published when they report treatment effects, distort the estimated pooled effect sizes, rendering them a less accurate representation of the “true” effect sizes. Although the number of studies included in the funnel plots was small and indications of publication bias were tentative, we recommend that researchers as well as journals publish null findings and results that counter hypotheses.

As described in the methods section, the review was nested within a larger review with a broader set of inclusion criteria. While the comprehensiveness of the literature search might have elicited a high retrieval rate by covering diverse fields of research and search terms, the approach would be immensely time-consuming to reproduce. Furthermore, the search strategy was based on the assumption that a broad, thematic search relating to health, well-being, and psychological stress would include studies utilizing psychophysiological outcomes. As such, the dependent variables were not predefined or included in the search strategy. While this could be speculated to result in the omission of relevant papers, the retrieval rate within the specific field of this review was higher than previously seen (Bowler et al., 2010; Haluza et al., 2014; Twohig-Bennett & Jones, 2018). In comparison to the most recent review by Twohig-Bennett and Jones (2018), we retrieved nine studies not included in their review (Aspinall et al., 2015; Brown et al., 2014; Dettweiler et al., 2017; Gidlow, Randall, Gillman, Smith, & Jones, 2016; Hohashi & Kobayashi, 2013;W. Kim et al., 2009; Lee et al., 2009; Matsuura et al., 2011; Park et al., 2008), while we missed five studies included in theirs (Grazuleviciene et al., 2016; Jia et al., 2016; Song, Ikei, Igarashi, et al., 2015; Song et al., 2013; Tsunetsugu et al., 2013). To acknowledge the contributions made by these five additional studies, we performed post hoc quality assessments and included study characteristics and quality in supplementary material C. The studies generally reported positive findings but did not alter the overall conclusions of our review. The studies were rated as being of low-to-moderate quality and shared the limitations observed in the body of evidence discussed above.

No one in the author group was proficient in Asian languages, and some studies that could potentially have been relevant were excluded from this review (e.g., Joung et al., 2015; Park et al., 2014; Song, Lee, Ikei, et al., 2015). Additionally, studies exploring the effects of contact with nature through, for example, gardening, views through windows or virtual nature were not included in this review.

Conclusions

While intuitively and theoretically sound, the empirical support for a stress-reducing impact of natural environments is tentative. The majority of the studies reported positive effects, but small-study effects might bias the body of evidence. Where possible, random-effects meta-analyses were performed to calculate pooled effect sizes. Meta-analyses indicated that seated relaxation ($g = .5$, $p =.06$) and walking ($g = .3$, $p =.02$) in natural environments enhanced vagally-mediated HRV more than the same activities in control conditions. Cortisol concentration measures were inconsistent. Future research would benefit from including larger sample sizes, increased population diversity (in terms of sociodemographic factors, medical conditions and diagnoses, age, and sex), blinding of outcome assessors (for group or condition assignment) and participants (for research question and aims), and thorough descriptions of natural and control environments and conditions, as well as participant recruitment and inclusion criteria. Further attention to quantitative assessment and control for potential confounding factors, such as temperature and physical activity, as well as inconsistent baseline levels, is warranted. Last, we recommend that researchers preregister trials to enhance transparency and accountability in the research field.

References

References marked with an asterisk indicate studies included in the meta-analyses.

Akselrod, S., Gordon, D., Ubel, F. A., Shannon, D. C., Berger, A. C., & Cohen, R. J. (1981). Power spectrum analysis of heart rate fluctuation: A quantitative probe of beat-to-beat cardiovascular control. *Science (New York, N.Y.)*, *213*(4504), 220–222. doi: 10.1126/science.6166045

Aspinall, P., Mavros, P., Coyne, R., & Roe, J. (2015). The urban brain: Analysing outdoor physical activity with mobile EEG. *British Journal of Sports Medicine*, *49*(4), 272–276. doi: 10.1136/bjsports-2012-091877

*Beil, K., & Hanes, D. (2013). The influence of urban natural and built environments on physiological and psychological measures of stress— A pilot study. *International Journal of Environmental Research and Public Health*, *10*(4), 1250–1267. doi: 10.3390/ijerph10041250

Bertone, Z. J. (2015). *Escaping the matrix: A quantitative study of the integration of yogic practice in nature*. Institute of Transpersonal Psychology. Retrieved from http://search.proquest.com/openview/2b79b71772d3f28eafccd9c71932e674/1?pq-origsite=gscholar&cbl=18750&diss=y

Bowler, D. E., Buyung-Ali, L. M., Knight, T. M., & Pullin, A. S. (2010). A systematic review of evidence for the added benefits to health of exposure to natural environments. *BMC Public Health*, *10*(1), 456. doi: 10.1186/1471-2458-10-456

Brown, C. H., Ten Have, T. R., Jo, B., Dagne, G., Wyman, P. A., Muthén, B., & Gibbons, R. D. (2009). Adaptive Designs for Randomized Trials in Public Health. *Annual Review of Public Health*, 30(1), 1–25. doi: 10.1146/annurev.publhealth.031308.100223

Brown, D. K., Barton, J. L., Pretty, J., & Gladwell, V. F. (2014). Walks4Work: Assessing the role of the natural environment in a workplace physical activity intervention.

Scandinavian Journal of Work, Environment & Health, *40*(4), 390–399. doi: 10.5271/sjweh.3421

Calogiuri, G., Evensen, K., Weydahl, A., Andersson, K., Patil, G., Ihlebæk, C., & Raanaas, R. K. (2015). Green exercise as a workplace intervention to reduce job stress. Results from a pilot study. *Work*, *53*(1), 99–111. doi: 10.3233/WOR-152219

Centre for Reviews and Dissemination. (2009). *Systematic reviews: CRD's guidance for undertaking reviews in health care*. York, UK. Retrieved from https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf

Chalmers, J. A., Quintana, D. S., Abbott, M. J.-A., & Kemp, A. H. (2014). Anxiety disorders are associated with reduced heart rate variability: A meta-analysis. *Frontiers in Psychiatry*, *5*(80). doi: 10.3389/fpsyt.2014.00080

Clay-Warner, J., & Robinson, D. T. (2015). Infrared thermography as a measure of emotion response. *Emotion Review*, *7*(2), 157–162. doi: 10.1177/1754073914554783

Critical Appraisal Skills Programme. (2013). *CASP UK critical appraisal checklists*. UK. Retrieved from https://casp-uk.net/

Dettweiler, U., Becker, C., Auestad, B. H., Simon, P., & Kirsch, P. (2017). Stress in school. Some empirical hints on the circadian cortisol rhythm of children in outdoor and indoor classes. *International Journal of Environmental Research and Public Health*, *14*(5), 475. doi: 10.3390/ijerph14050475

Effective Public Health Practice Project. (2013). *Quality assessment tool for quantitative studies*. Canada: McMaster University. Retrieved from https://merst.ca/ephpp/

Gatersleben, B., & Andrews, M. (2013). When walking in nature is not restorative? The role of prospect and refuge. *Health & Place*, *20*, 91–101. doi: 10.1016/j.healthplace.2013.01.001

*Gidlow, C. J., Jones, M. V., Hurst, G., Masterson, D., Clark-Carter, D., Tarvainen, M. P., … Nieuwenhuijsen, M. (2016). Where to put your best foot forward: Psycho-physiological responses to walking in natural and urban environments. *Journal of Environmental Psychology*, *45*, 22–29. doi: 10.1016/j.jenvp.2015.11.003

Gidlow, C. J., Randall, J., Gillman, J., Smith, G. R., & Jones, M. V. (2016). Natural environments and chronic stress measured by hair cortisol. *Landscape and Urban Planning*, *148*, 61–67. doi: 10.1016/j.landurbplan.2015.12.009

Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of Evidence for Efficacy, Effectiveness, and Scale-up Research in Prevention Science: Next Generation. *Prevention Science: The Official Journal of the Society for Prevention Research*, *16(7)*, 893–926. doi: 10.1007/s11121-015-0555-x

Grazuleviciene, R., Vencloviene, J., Kubilius, R., Grizas, V., Danileviciute, A., Dedele, A., … Nieuwenhuijsen, M. J. (2016). Tracking restoration of park and urban street settings in coronary artery disease patients. *International Journal of Environmental Research and Public Health*, *13*(6), 550. doi: 10.3390/ijerph13060550

Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *BMJ*, *331*(7524), 1064–1065. doi: 10.1136/bmj.38636.593461.68

Haluza, D., Schönbauer, R., & Cervinka, R. (2014). Green perspectives for public health: A narrative review on the physiological effects of experiencing outdoor nature. *International Journal of Environmental Research and Public Health*, *11*(5), 5445–5461. doi: 10.3390/ijerph110505445

Han, J.-W., Choi, H., Jeon, Y.-H., Yoon, C.-H., Woo, J.-M., & Kim, W. (2016). The effects of forest therapy on coping with chronic widespread pain: Physiological and psychological differences between participants in a forest therapy program and a control group. *International Journal of Environmental Research and Public Health*, *13*(3), 255. doi: 10.3390/ijerph13030255

Hartig, T. (2008). Green space, psychological restoration, and health inequality. *Lancet (London, England)*, *372*(9650), 1614–1615. doi: 10.1016/S0140-6736(08)61669-4

Hartig, T., Mitchell, R., Vries, S. de, & Frumkin, H. (2014). Nature and health. *Annual Review of Public Health*, *35*(1), 207–228. doi: 10.1146/annurev-publhealth-032013-182443

Hohashi, N., & Kobayashi, K. (2013). The effectiveness of a forest therapy (shinrin-yoku) program for girls aged 12 to 14 years: A crossover study. *Stress Science Research*, *28*, 82–89. doi: 10.5058/stresskagakukenkyu.28.82

Jia, B. B., Yang, Z. X., Mao, G. X., Lyu, Y. D., Wen, X. L., Xu, W. H., … Wang, G. F. (2016). Health effect of forest bathing trip on elderly patients with chronic obstructive pulmonary disease. *Biomedical and Environmental Sciences*, *29*(3), 212–218. doi: 10.3967/bes2016.026

Joung, D., Kim, G., Choi, Y., Lim, H., Park, S., Woo, J.-M., & Park, B.-J. (2015). The prefrontal cortex activity and psychological effects of viewing forest landscapes in autumn season. *International Journal of Environmental Research and Public Health*, *12*(7), 7235–7243. doi: 10.3390/ijerph120707235.

Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., & Koo, B.-H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation*, *15*(3), 235–245. doi: 10.30773/pi.2017.08.17

Kim, W., Lim, S.-K., Chung, E.-J., & Woo, J.-M. (2009). The effect of cognitive behavior therapy-based psychotherapy applied in a forest environment on physiological changes and remission of major depressive disorder. *Psychiatry Investigation*, *6*(4), 245. doi: 10.4306/pi.2009.6.4.245

Kjellgren, A., & Buhrkall, H. (2010). A comparison of the restorative effect of a natural environment with that of a simulated natural environment. *Journal of Environmental Psychology*, *30*(4), 464–472. doi: 10.1016/j.jenvp.2010.01.011

Kühn, S., Düzel, S., Eibich, P., Krekel, C., Wüstemann, H., Kolbe, J., … Lindenberger, U. (2017). In search of features that constitute an “enriched environment” in humans: Associations between geographical properties and brain structure. *Scientific Reports*, *7*(1), 11920. doi: 10.1038/s41598-017-12046-7

Lakens, D. (2017). Equivalence tests. *Social Psychological and Personality Science*, *8*(4), 355–362. doi: 10.1177/1948550617697177

*Lee, J. (2014). Influence of forest therapy on cardiovascular relaxation in young adults. *Evidence-Based Complementary and Alternative Medicine*, *(2)*. doi: 10.1155/2014/834360

*Lee, J., Park, B.-J., Ohira, T., Kagawa, T., & Miyazaki, Y. (2015). Acute effects of exposure to a traditional rural environment on urban dwellers: A crossover field study in terraced farmland. *International Journal of Environmental Research and Public Health*, *12*(2), 1874–1893. doi: 10.3390/ijerph120201874

*Lee, J., Park, B.-J., Tsunetsugu, Y., Kagawa, T., & Miyazaki, Y. (2009). Restorative effects of viewing real forest landscapes, based on a comparison with urban landscapes. *Scandinavian Journal of Forest Research*, *24*(3), 227–234. doi: 10.1080/02827580902903341

*Lee, J., Park, B.-J., Tsunetsugu, Y., Ohira, T., Kagawa, T., & Miyazaki, Y. (2011). Effect of forest bathing on physiological and psychological responses in young Japanese male subjects. *Public Health*, *125*(2), 93–100. doi: 10.1016/j.puhe.2010.09.005

Li, Q., Kobayashi, M., Kumeda, S., Ochiai, T., Miura, T., Kagawa, T., … Kawada, T. (2016). Effects of forest bathing on cardiovascular and metabolic parameters in middle-aged males. *Evidence-Based Complementary and Alternative Medicine*, *2016*, e2587381. doi: 10.1155/2016/2587381

Li, Q., Otsuka, T., Kobayashi, M., Wakayama, Y., Inagaki, H., Katsumata, M., … Kagawa, T. (2011). Acute effects of walking in forest environments on cardiovascular and metabolic parameters. *European Journal of Applied Physiology; Heidelberg*, *111*(11), 2845–2853. doi: 10.1007/s00421-011-1918-z

Liu, W., Lian, Z., & Liu, Y. (2008). Heart rate variability at different thermal comfort levels. *European Journal of Applied Physiology*, *103*(3), 361–366. doi: 10.1007/s00421-008-0718-6

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews. Neuroscience*, *10*(6), 434–445. doi: 10.1038/nrn2639

Maller, C., Townsend, M., Pryor, A., Brown, P., & St Leger, L. (2006). Healthy nature healthy people: ‘Contact with nature’ as an upstream health promotion intervention for populations. *Health Promotion International*, *21*(1), 45–54. doi: 10.1093/heapro/dai032

Mao, G.-X., Cao, Y.-B., Lan, X.-G., He, Z.-H., Chen, Z.-M., Wang, Y.-Z., … Yan, J. (2012). Therapeutic effect of forest bathing on human hypertension in the elderly. *Journal of Cardiology*, *60*(5–6), 495–502. doi: 10.1016/j.jjcc.2012.08.003

*Mao, G.-X., Lan, X.-G., Cao, Y.-B., Chen, Z. M., He, Z.-H., Lv, Y.-D., … Yan, J. (2012). Effects of short-term forest bathing on human health in a broad-leaved evergreen forest in Zhejiang Province, China. *Biomedical Environmental Sciences*, *25*(3), 317–324. doi: 10.3967/0895-3988.2012.03.010

Marques, A. H., Silverman, M. N., & Sternberg, E. M. (2010). Evaluation of stress systems by applying noninvasive methodologies: Measurements of neuroimmune biomarkers in the sweat, heart rate variability and salivary cortisol. *Neuroimmunomodulation*, *17*(3), 205–208. doi: 10.1159/000258725

Matsumoto, T., Miyawaki, T., Ue, H., Kanda, T., Zenji, C., & Moritani, T. (1999). Autonomic responsiveness to acute cold exposure in obese and non-obese young women. *International Journal of Obesity*, *23*(8), 793–800. doi: 10.1038/sj.ijo.0800928

Matsuura, A., Nagai, N., Funatsu, A., Irimajiri, M., Yamazaki, A., & Hodate, K. (2011). Comparison of the short-term effects of horse trekking and exercising with a riding simulator on autonomic nervous activity. *Anthrozoös*, *24*(1), 65–77. doi: 10.2752/175303711X12923300467401

McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *New England Journal of Medicine*, *338*(3), 171–179. doi: 10.1056/NEJM199801153380307

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Medicine*, *6*(7), e1000097. doi: 10.1371/journal.pmed.1000097

Morita, E., Fukuda, S., Nagano, J., Hamajima, N., Yamamoto, H., Iwai, Y., … Shirakawa, T. (2007). Psychological effects of forest environments on healthy adults: Shinrin-yoku (forest-air bathing, walking) as a possible method of stress reduction. *Public Health*, *121*(1), 54–63. doi: 10.1016/j.puhe.2006.05.024

Mygind, L., Kjeldsted, E., Hartmeyer, R., Mygind, E., and Bentsen, P. (2018). Forskningsoversigt Over Effekter af Friluftsliv på Mental, Fysisk og Social Sundhed, 330. Copenhagen. Retrieved from The Danish Outdoor Council website: https://www.friluftsraadet.dk/publikationer

Mygind, L., Kjeldsted, E., Hartmeyer, R., Mygind, E., Bølling, M., & Bentsen, B. (2019). Immersive Nature-Experiences as Health Promotion Interventions for Healthy, Vulnerable, and Sick Populations? A Systematic Review and Appraisal of Controlled Studies. *Frontiers in Psychology* (10): 943. doi: 10.3389/fpsyg.2019.00943

Ochiai, H., Ikei, H., Song, C., Kobayashi, M., Miura, T., Kagawa, T., … Miyazaki, Y. (2015). Physiological and psychological effects of a forest therapy program on middle-aged females. *International Journal of Environmental Research and Public Health*, *12*(12), 15222–15232. doi: 10.3390/ijerph121214984

Ohly, H., White, M. P., Wheeler, B. W., Bethel, A., Ukoumunne, O. C., Nikolaou, V., & Garside, R. (2016). Attention Restoration Theory: A systematic review of the attention restoration potential of exposure to natural environments. *Journal of Toxicology and Environmental Health. Part B, Critical Reviews*, *19*(7), 305–343. doi: 10.1080/10937404.2016.1196155

*Park, B.-J. (2009). Physiological effects of forest recreation in a young conifer forest in Hinokage Town, Japan. *Silva Fennica*, *43*(2009).

Park, B.-J., Ka, J.-N., Lee, M. S., Kim, S.-A., Park, M.-W., Choi, Y., … Kim, G. (2014). Physiological effects of walking and viewing on human at a urban arboretums. *Journal of Korean Society of Forest Science*, *103*(4), 664–669. doi: 10.14578/jkfs.2014.103.4.664

Park, B.-J., Tsunetsugu, Y., Ishii, H., Furuhashi, S., Hirano, H., Kagawa, T., & Miyazaki, Y. (2008). Physiological effects of shinrin-yoku (taking in the atmosphere of the forest) in a mixed forest in Shinano Town, Japan. *Scandinavian Journal of Forest Research*, *23*(3), 278–283. doi: 10.1080/02827580802055978

*Park, B.-J., Tsunetsugu, Y., Kasetani, T., Hirano, H., Kagawa, T., Sato, M., & Miyazaki, Y. (2007). Physiological effects of shinrin-yoku (taking in the atmosphere of the forest)—using salivary cortisol and cerebral activity as indicators. *Journal of Physiological Anthropology*, *26*(2), 123–128. doi: 10.2114/jpa2.26.123

Park, B.-J., Tsunetsugu, Y., Kasetani, T., Kagawa, T., & Miyazaki, Y. (2010). The physiological effects of shinrin-yoku (taking in the forest atmosphere or forest bathing): evidence from field experiments in 24 forests across Japan. *Environmental Health and Preventive Medicine*, *15*(1), 18–26. doi: 10.1007/s12199-009-0086-9

Quintana, D. S. (2017). Statistical considerations for reporting and planning heart rate variability case-control studies. *Psychophysiology*, *54*(3), 344–349. doi:10.1111/psyp.12798

Quintana, D. S., & Heathers, J. A. J. (2014). Considerations in the assessment of heart rate variability in biobehavioral research. *Frontiers in Psychology*, *5,* 805. doi: 10.3389/fpsyg.2014.00805

Rose, G. (2001). Sick individuals and sick populations. *International Journal of Epidemiology*, *30*(3), 427–432. doi: 10.1093/ije/30.3.427

Ryan, R., Hill, S., Broclain, D., Oliver, S., Prictor, M., & Cochrane Consumers and Communication Review Group. (2013, June). *Study design guide*. Retrieved from http://cccrg.cochrane.org/author-resources

Sahlin, E., Lindegård, A., Hadzibajramovic, E., Grahn, P., Vega Matuszczyk, J., & Ahlborg, G. (2016). The influence of the environment on directed attention, blood pressure and heart rate—An experimental study using a relaxation intervention. *Landscape Research*, *41*(1), 7–25. doi: 10.1080/01426397.2014.982079

Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). Small-Study Effects in Meta-Analysis. In G. Schwarzer, J. R. Carpenter, & G. Rücker (Eds.), *Meta-Analysis with R* (pp. 107–141). Cham: Springer International Publishing. doi: 10.1007/978-3-319-21416-0_5

Shonkoff, J. P., Boyce, W. T., & McEwen, B. S. (2009). Neuroscience, molecular biology, and the childhood roots of health disparities: Building a new framework for health promotion and disease prevention. *Journal of American Medical Association*, *301*(21), 2252–2259. doi: 10.1001/jama.2009.754

Song, C., Ikei, H., Igarashi, M., Takagaki, M., & Miyazaki, Y. (2015). Physiological and psychological effects of a walk in urban parks in fall. *International Journal of Environmental Research and Public Health*, *12*(11), 14216–14228. doi: 10.3390/ijerph121114216

*Song, C., Ikei, H., Kobayashi, M., Miura, T., Taue, M., Kagawa, T., … Miyazaki, Y. (2015). Effect of forest walking on autonomic nervous system activity in middle-aged hypertensive individuals: A pilot study. *International Journal of Environmental Research and Public Health*, *12*(3), 2687–2699. doi: 10.3390/ijerph120302687

Song, C., Joung, D., Ikei, H., Igarashi, M., Aga, M., Park, B.-J., … Miyazaki, Y. (2013). Physiological and psychological effects of walking on young males in urban parks in winter. *Journal of Physiological Anthropology*, *32*(1), 18. doi: 10.1186/1880-6805-32-18

Song, C., Lee, J., Ikei, H., Kagawa, T., Miyazaki, Y., & Park, B.-J. (2015). Physiological and psychological effects of walking around and viewing a lake in a forest environment. *Journal of Korean Forest Society*, *104*(1), 140–149.

Sung, J., Woo, J.-M., Kim, W., Lim, S.-K., & Chung, E.-J. (2012). The effect of cognitive behavior therapy-based “forest therapy” program on blood pressure, salivary cortisol level, and quality of life in elderly hypertensive patients. *Clinical and Experimental Hypertension*, *34*(1), 1–7. doi: 10.3109/10641963.2011.618195

Takai, N., Yamaguchi, M., Aragaki, T., Eto, K., Uchihashi, K., & Nishikawa, Y. (2004). Effect of psychological stress on the salivary cortisol and amylase levels in healthy young adults. *Archives of Oral Biology*, *49*(12), 963–968. doi: 10.1016/j.archoralbio.2004.06.007

Takayama, N., Korpela, K., Lee, J., Morikawa, T., Tsunetsugu, Y., Park, B.-J., … Kagawa, T. (2014). Emotional, restorative and vitalizing effects of forest and urban environments at four sites in japan. *International Journal of Environmental Research and Public Health*, *11*(7), 7207–7230. doi: 10.3390/ijerph110707207

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996). Heart rate variability: Standards of measurement, physiological interpretation and clinical use. *Circulation*, *93*(5), 1043–1065.

Thayer, J. F., Åhs, F., Fredrikson, M., Sollers, J. J., & Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neuroscience & Biobehavioral Reviews*, *36*(2), 747–756. doi: 10.1016/j.neubiorev.2011.11.009

Thompson, J. E. S. (2014). *The impact of an 8-week green-exercise programme on systemic health, and on markers associated with cardiovascular disease risk.* Cardiff Metropolitan University. Retrieved from https://repository.cardiffmet.ac.uk/handle/10369/7526

Toda, M., Den, R., Hasegawa-Ohira, M., & Morimoto, K. (2013). Effects of woodland walking on salivary stress markers cortisol and chromogranin A. *Complementary Therapies in Medicine; Kidlington*, *21*(1), 29–34. doi: 10.1016/j.ctim.2012.11.004

Tost, H., Champagne, F. A., & Meyer-Lindenberg, A. (2015). Environmental influence in the brain, human welfare and mental health. *Nature Neuroscience*, *18*(10), 1421–1431. doi: 10.1038/nn.4108

Tsunetsugu, Y., Lee, J., Park, B.-J., Tyrväinen, L., Kagawa, T., & Miyazaki, Y. (2013). Physiological and psychological effects of viewing urban forest landscapes assessed by multiple measurements. *Landscape and Urban Planning*, *113*, 90–93. doi: 10.1016/j.landurbplan.2013.01.014

*Tsunetsugu, Y., Park, B.-J., Ishii, H., Hirano, H., Kagawa, T., & Miyazaki, Y. (2007). Physiological effects of Shinrin-yoku (taking in the atmosphere of the forest) in an old-growth broadleaf forest in Yamagata Prefecture, Japan. *Journal of Physiological Anthropology*, *26*(2), 135–142. doi: 10.2114/jpa2.26.135

Tsunetsugu, Y., Park, B.-J., & Miyazaki, Y. (2010). Trends in research related to "Shinrin-yoku" (taking in the forest atmosphere or forest bathing) in Japan. *Environmental Health and Preventive Medicine*, *15*(1), 27–37. doi: 10.1007/s12199-009-0091-z

Twohig-Bennett, C., & Jones, A. (2018). The health benefits of the great outdoors: A systematic review and meta-analysis of greenspace exposure and health outcomes. *Environmental Research*, *166*, 628–637. doi: 10.1016/j.envres.2018.06.030

Tyrväinen, L., Ojala, A., Korpela, K., Lanki, T., Tsunetsugu, Y., & Kagawa, T. (2014). The influence of urban green environments on stress relief measures: A field experiment. *Journal of Environmental Psychology*, *38*(Supplement C), 1–9. doi: 10.1016/j.jenvp.2013.12.005

van den Berg, M., Wendel-Vos, W., van Poppel, M., Kemper, H., van Mechelen, W., & Maas, J. (2015). Health benefits of green spaces in the living environment: A systematic review of epidemiological studies. *Urban Forestry & Urban Greening*, *14*(4), 806–816. doi: 10.1016/j.ufug.2015.07.008

Vigo, D., Thornicroft, G., & Atun, R. (2016). Estimating the true global burden of mental illness. *The Lancet. Psychiatry*, *3*(2), 171–178. doi: 10.1016/S2215-0366(15)00505-2

Wu, S., Deng, F., Liu, Y., Shima, M., Niu, J., Huang, Q., & Guo, X. (2013). Temperature, traffic-related air pollution, and heart rate variability in a panel of healthy adults. *Environmental Research*, *120*, 82–89. doi: 10.1016/j.envres.2012.08.008

Yamaguchi, M., Deguchi, M., & Miyazaki, Y. (2006). The effects of exercise in forest and urban environments on sympathetic nervous activity of normal young adults. *Journal of International Medical Research*, *34*(2), 152–159. doi: 10.1177/147323000603400204

Yu, Y.-M., Lee, Y.-J., Kim, J.-Y., Yoon, S.-B., & Shin, C.-S. (2016). Effects of forest therapy camp on quality of life and stress in postmenopausal women. *Forest Science and Technology*, *12*(3), 125–129. doi: 10.1080/21580103.2015.1108248

Table 1

Eligibility criteria

Item	Criteria
Intervention/exposure	Direct experience of public, predominantly green space, excluding private gardens, indoor nature, views to nature and virtual nature
Comparison	For within-subjects studies: Similar exposure or activity in a predominantly synthetic environment, e.g., indoor or urban For between-subjects studies: a no treatment or treatment as usual condition
Outcome	Psychophysiological stress response markers, not predefined
Methods and design	Quantitative research using within- and between-subjects, randomized and non-randomized controlled designs
Language	Danish, Swedish, Norwegian, German or English
Time	Published after 2003, search finalized in May 2017
Type publication	Peer-reviewed publications
Target group	No restrictions

Table 2

Study characteristics

Study information			Study sample		Study design					
Ref.	Ctry.	Size	Participants	Age	Activity	Exposure	Natural	Control	Outc.	Rep. results
Randomized trial										
D.K. Brown et al. (2014)	UK	73 i: 27 c1: 27 c2: 19	Office workers from international firm, ♀♂	40 (*SD*: 10.6)	Walking (measured during 1) relaxation 2) stressful task 3) after stressful task)	8 weeks (2*20 min weekly)	Green space (trees, maintained grass, paved footpaths, country lanes)	c1: Same activity in urban (industrial) environment c2: Waitlist	HF	%
Quasi-randomized trial										
Calogiuri et al. (2016)	NO	11 i: 6 c: 5	Healthy sedentary or moderately active individuals, ♀♂	49 (*SD*: 8)	Exercise (biking and ~~rubberband~~rubber band exercises)	2 days (2*45 min a day)	Forest	Indoor (gym-hall)	C, Ca	- ¦
Mao et al. (2012)	CHN	20 i: 10 c: 10	University students, ♂	20.79 (*SD*: .54)	Walking	2 days (2*1.5 hours on the same day (morning and afternoon))	Forest (broad-leafed, evergreen)	Urban (downtown area)	C	+
Cross-over trial										
Beil and Hanes (2013)	USA	15	Healthy "Non-Hispanic White" adults, ♀♂	42.3 (*SD*: N.R.)	Seated relaxation	20 min	1) ”Very natural” 2) ”Natural”	1) ”Very built” 2) ”Built”	C, SA	%
Gidlow, Jones et al. (2016)	UK	38	Healthy, unstressed individuals, ♀♂	40.9 (*SD*: 17.6)	Walking	30 min	1) Park 2) Park with lake	Calm urban	C, HF	%

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Hohashi and Kobayashi (2013)	JP	11	Junior high school students, ♀	13.1 (*SD*: N.R.)	Seated relaxation (following a 30 min walk in same environment)	15 min	Forest	Urban	SA	%
Lee et al. (2009)	JP	12	University students, ♂	21.3 (*SD*: 1.1)	Seated relaxation	15 min	Forest (pine and beech)	Urban (commercial area)	C	+
Lee et al. (2011)	JP	12	University students, ♂	21.2 (*SD*: .9)	Seated relaxation	15 min	Forest (broad-leafed deciduous trees)	Urban (commercial area)	C, HF	+
Lee et al. (2014)	JP	48	Healthy individuals, ♂	21.1 (*SD*: 1.2)	Walking	12-15 min	Forest (4 areas, flat)	Urban (4 areas, flat)	ln(HF)	+
Lee et al. (2015)	JP	12	University students, ♂	22.3 (*SD*: 1.3)	Seated relaxation	15 min	Rural landscape (terraced paddy field)	Urban (area near train station)	C, ln(HF)	- [II]
Matsuura et al. (2011)	JP	23 (HF) 26 (SA)	University students, ♀♂	19-25, (mean & *SD* N.R.)	Horseback riding	30 min	Natural trail (on horse)	Indoor (using riding ~~S~~simulator)	HF, SA	%
Park et al. (2007)	JP	12	University students, ♂	22.8 (*SD*: 1.4)	1) Walking 2) Seated relaxation	20 min	Forest (oak)	Urban	C, H	- [III]
Park et al. (2008)	JP	12	University students, ♂	21.3 (*SD*: 1.1)	Seated relaxation	15 min	Forest	Urban (parking lot)	HF	- [IV]
Park et al. (2009)	JP	12	University students, ♂	21.8 (*SD*: .8)	1) Walking 2) Seated relaxation	15 min	Forest	Urban	HF	- [V]
Park et al. (2010)	JP	1) 260 (C), 264 (HF) 2) 74 (C), 72 (HF)	University students, ♂	21.7 (*SD*: 1.5)	1) Seated relaxation 2) Walking	1) 14 min, 2) 16 min	Forest	Urban	C, HF	+

Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted

Song et al. (2015)	JP	20	Middle-aged individuals (prehypertensive or stage 1 hypertension), not taking medication, ♂	58 (*SD*: 10.6)	Walking	17 min	Forest	Urban	ln(HF)	+
Tÿrvainen et al. (2014)	FI	77	Healthy individuals, ♀♂	47.6 (*SD*: 8.68)	1) Seated relaxation 2) Walking	1) 15 min 2) 30 min	i1: Urban forest i2: Urban park	Urban built environment	C	%
Tsunetsugu et al. (2007)	JP	12	University students, ♂	22 (*SD*: 1)	1) Walking 2) Seated relaxation	15 min	Forest (birch)	Urban (area near train station)	C, HF	- [VI]
Yamaguchi et al. (2006)	JP	10	University students, ♂	23.2 (*SD*: 1.1)	1) Seated relaxation 2) Walking	20 min	Forest	Urban (area near train station)	SA	%
<u>Controlled before-and-after study</u>										
Dettweiler et al. (2017)	DE	48 i: 37 c: 11	Primary school students, ♀♂	11.23 (*SD*: .46)	School	1 day (and one school term)	Forest (measured in 1) fall 2) spring 3) summer)	Classroom (measured in 1) fall 2) spring 3) summer)	C	+
Han et al. (2016)	KR	61 i: 33 c: 28	Individuals with chronic pain, ♀♂	i: 41.6 (*SD*: 6.5) c: 37.5 (*SD*: 8.4)	Forest therapy (based on Cognitive Behavior Therapy, for individuals with chronic pain, also includes mindfulness, physical activity, music therapy etc.)	2 days	Forest	No treatment	SDNN	+

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Kim et al. (2009)	KR	63 i: 23 c1: 19 c2: 21	Individuals with major depressive disorder, ♀♂	i: 38.6 (*SD*: 11) c: 43.6 (*SD*: 13.6) c1: 43.3 (8.35)	Forest therapy (using Cognitive Behavior Therapy-Based Psychotherapy)	4 weeks (4 sessions (3 hours each))	Experimental forest (Hong-Reung ~~aroretum~~arboretum)	c1) Same treatment in hospital c2) Treatment as usual/waitlist	C, HF	+
Sung et al. (2012)	KR	56 i: 28 c: 28	Individuals with hypertension, ♀♂	i: 63 (*SD*: 11) c: 66 (*SD*: 7)	Forest therapy (using Cognitive Behavior Therapy)	8 weeks	Forest	Treatment as usual	C	+
Within-subjects (% counterbalancing)										
Aspinall et al. (2015)	UK	12	University students, ♀♂	30.08 (*SD*: N.R.)	Walking	approx. 10 min	Park	Urban shopping stress and commercial area	FC	+
Li et al. (2011)	JP	16	Healthy individuals, ♂	57.4 (*SD*: 11.6)	Walking	1 day (2*2 hours on the same day (morning and evening))	Forest	Urban	A, NA, D	+
Li et al. (2016)	JP	19	Middle-aged individuals (predominantly with stage one hypertension), ♂	51.2 (*SD*: 8.8)	Walking (2.6 km)	1 day (2*80 min (morning and evening))	Forest	Urban	A, NA, D	+

Note. C = control group; c1 = control group number 1; c2 = control group number 2; i = intervention group; N.R. = not reported. A = adrenaline; C = (salivary and serum) cortisol; Ca = cortisol awakening response; D = dopamine; FC = activity in frontal cortex of the brain; H = hemoglobin concentration in the prefrontal area of the brain; HF = high frequency heart rate variability; ln(HF) = natural logarithm of high frequency heart rate variability; NA = noradrenaline; SA = salivary amylase. "+" = Significant difference between natural and control condition were reported; "-" = Mixed results were reported (roman numerals refer to elaborative text in notes); "%" = No significant differences were reported.

[I] Authors reported no difference between natural and control condition in cortisol levels in serum. Two measures for cortisol awakening response were utilized of which one indicated an improvement related to the green-exercise program and the other no significant effect compared to the control condition. [II] Cortisol levels were improved more

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

during activity in natural than control environment. ln(HF) was measured dynamically during both activities. ln(HF) levels tended to be higher during activities in the natural environments, but at individual time points, no significant difference could be observed. However, when values for the individual time points were summed, the level of ln(HF) was significantly higher in the natural environment. [III] No difference in cortisol levels between natural and control environment after walk. After seated relaxation, cortisol level was lower in the natural environment. The absolute concentration of total hemoglobin in the left prefrontal area was reduced more in the forest area pre to post both walking and seated relaxation. [IV] HF was measured dynamically during both activities. HF levels tended to be higher during activities in the natural environments, but at some time points, no significant difference could be observed. [V] HF was measured dynamically during both activities. HF levels tended to be higher during activities in the natural environments, but at several time points, no significant difference could be observed. [VI] HF was measured dynamically during both activities. HF levels tended to be higher during seated relaxation in the natural environments, but at some time points, no significant difference could be observed. During the walk, no significant differences were reported. Cortisol levels were lower post natural environments for both activities, but baselines levels of cortisol were lower before walking and tending to be lower pre seated relaxation before exposure to the natural environments.

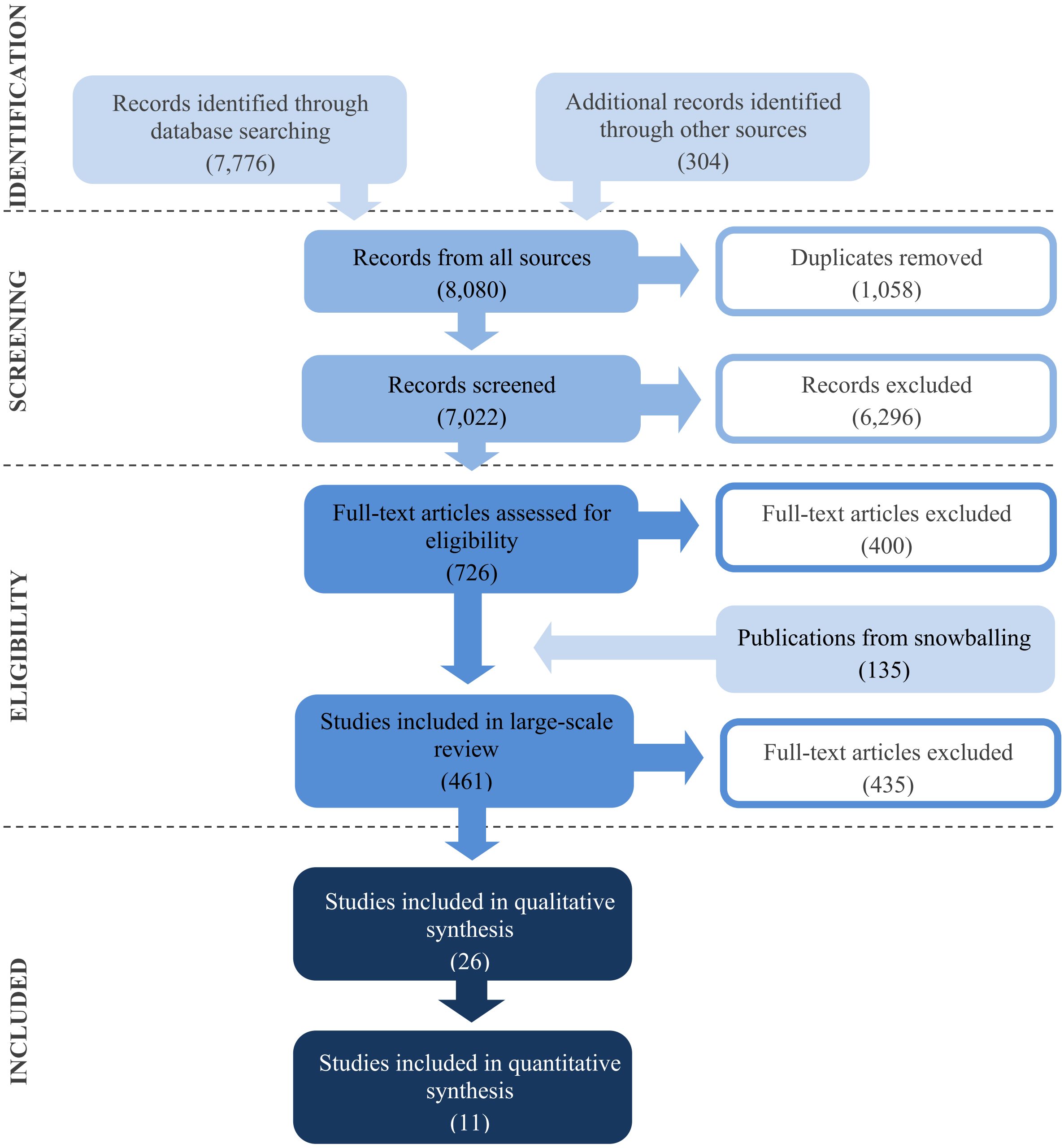


Figure 1. Flow chart.

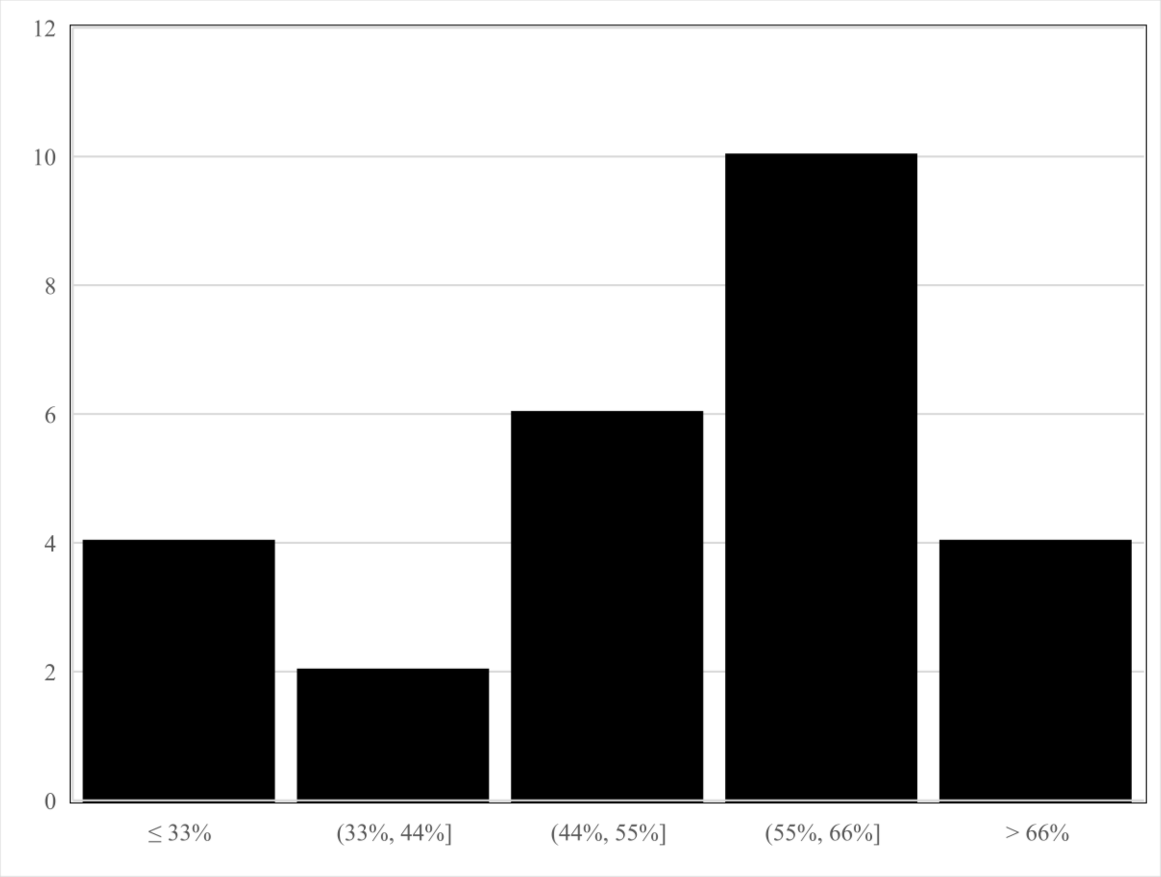
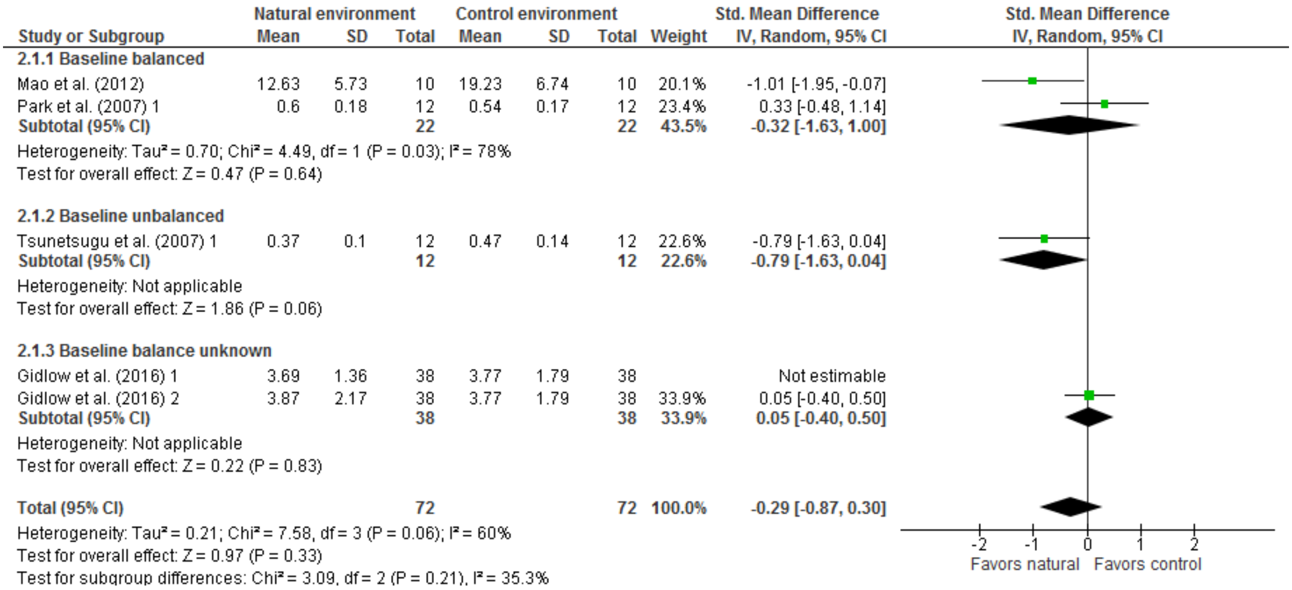


Figure 2. Distribution of quality assessments of individual studies (*n* = 26). Scores below and equal to 33% were considered low quality and scores above 66% high quality. We have added three more groupings to nuance the scores within the moderate category.



	Natural environment			Control environment				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
2.1.1 Baseline balanced								
Mao et al. (2012)	12.63	5.73	10	19.23	6.74	10	20.1%	-1.01 [-1.95, -0.07]
Park et al. (2007) 1	0.6	0.18	12	0.54	0.17	12	23.4%	0.33 [-0.48, 1.14]
Subtotal (95% CI)			**22**			**22**	**43.5%**	**-0.32 [-1.63, 1.00]**
Heterogeneity: Tau² = 0.70; Chi² = 4.49, df = 1 (P = 0.03); I² = 78%								
Test for overall effect: Z = 0.47 (P = 0.64)								
2.1.2 Baseline unbalanced								
Tsunetsugu et al. (2007) 1	0.37	0.1	12	0.47	0.14	12	22.6%	-0.79 [-1.63, 0.04]
Subtotal (95% CI)			**12**			**12**	**22.6%**	**-0.79 [-1.63, 0.04]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 1.86 (P = 0.06)								
2.1.3 Baseline balance unknown								
Gidlow et al. (2016) 1	3.69	1.36	38	3.77	1.79	38		Not estimable
Gidlow et al. (2016) 2	3.87	2.17	38	3.77	1.79	38	33.9%	0.05 [-0.40, 0.50]
Subtotal (95% CI)			**38**			**38**	**33.9%**	**0.05 [-0.40, 0.50]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.22 (P = 0.83)								
Total (95% CI)			**72**			**72**	**100.0%**	**-0.29 [-0.87, 0.30]**
Heterogeneity: Tau² = 0.21; Chi² = 7.58, df = 3 (P = 0.06); I² = 60%								
Test for overall effect: Z = 0.97 (P = 0.33)								
Test for subgroup differences: Chi² = 3.09, df = 2 (P = 0.21), I² = 35.3%								

Figure 3. Forest plot of pooled effect sizes (Hedges' *g*) for serum and salivary cortisol related to of walking in natural versus control environment. Lower levels of cortisol indicate findings in favor of the natural environment. df = degrees of freedom; CI = confidence intervals; Chi^2 = test of independence; I^2 = extent of heterogeneity in effect size across studies; P = statistical significance (more commonly symbolized as *p*); SD = Standard Deviation; Std. Mean Difference = Standardized mean difference (measure of effect size, also referred to as Hedges' *g*); Tau^2 = variance of the true effect sizes, Z = standard deviations away from the mean.

Gidlow, Jones et al. (2016) (in the forest plot referred to as Gidlow et al. (2016)) reported on the effects of two types of natural environments, i.e., 1) a park with a lake and 2) a park, on the same 38 participants. Means and SD's from both conditions are presented in the forest plots, but only one of the conditions was included in the pooled analysis. We chose to include the park condition as the baseline values were numerically closer to each other in this condition than those for the condition that included a park with a lake.

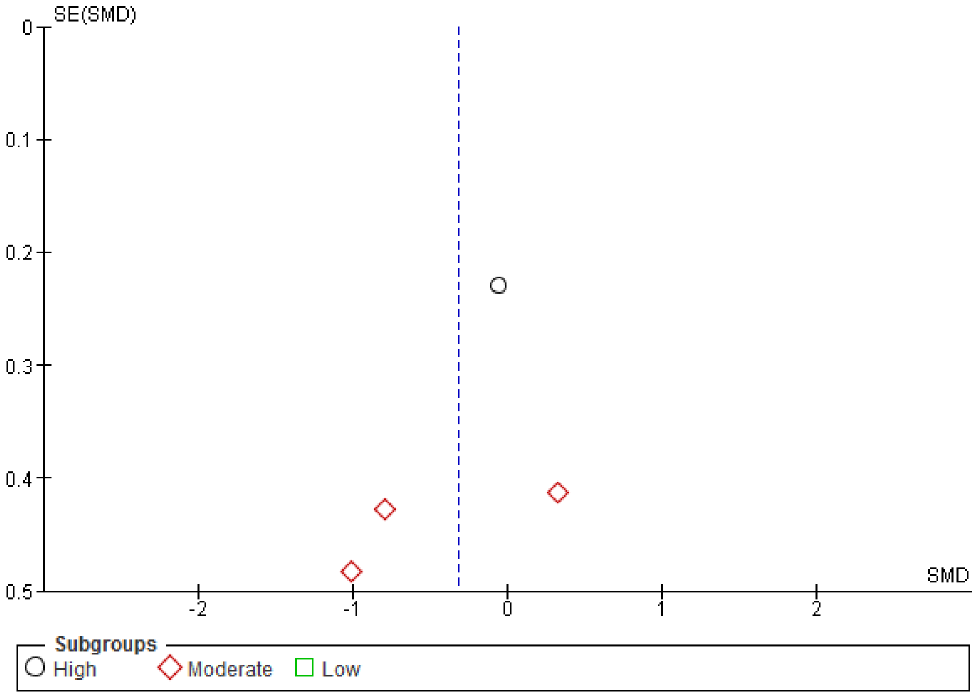
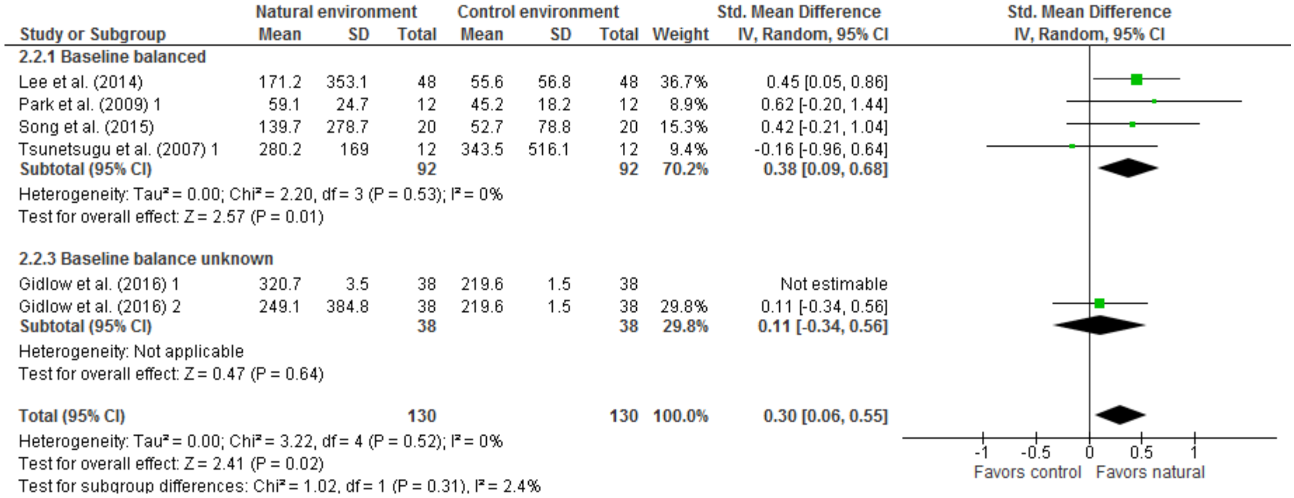


Figure 4. Funnel plot of studies exploring effects of walking in natural and control conditions on cortisol concentration, divided by study quality. Higher values of SMD indicates higher effect sizes and higher values of SE(SMD) indicate larger within-study variability in effects. The dotted line indicates the pooled effect size. SE = Standard Error; SMD = Standardized Mean Difference.



Study or Subgroup	Natural environment Mean	SD	Total	Control environment Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.2.1 Baseline balanced								
Lee et al. (2014)	171.2	353.1	48	55.6	56.8	48	36.7%	0.45 [0.05, 0.86]
Park et al. (2009) 1	59.1	24.7	12	45.2	18.2	12	8.9%	0.62 [-0.20, 1.44]
Song et al. (2015)	139.7	278.7	20	52.7	78.8	20	15.3%	0.42 [-0.21, 1.04]
Tsunetsugu et al. (2007) 1	280.2	169	12	343.5	516.1	12	9.4%	-0.16 [-0.96, 0.64]
Subtotal (95% CI)			**92**			**92**	**70.2%**	**0.38 [0.09, 0.68]**
Heterogeneity: Tau^2 = 0.00; Chi^2 = 2.20, df = 3 (P = 0.53); I^2 = 0%								
Test for overall effect: Z = 2.57 (P = 0.01)								
2.2.3 Baseline balance unknown								
Gidlow et al. (2016) 1	320.7	3.5	38	219.6	1.5	38		Not estimable
Gidlow et al. (2016) 2	249.1	384.8	38	219.6	1.5	38	29.8%	0.11 [-0.34, 0.56]
Subtotal (95% CI)			**38**			**38**	**29.8%**	**0.11 [-0.34, 0.56]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 0.47 (P = 0.64)								
Total (95% CI)			**130**			**130**	**100.0%**	**0.30 [0.06, 0.55]**
Heterogeneity: Tau^2 = 0.00; Chi^2 = 3.22, df = 4 (P = 0.52); I^2 = 0%								
Test for overall effect: Z = 2.41 (P = 0.02)								
Test for subgroup differences: Chi^2 = 1.02, df = 1 (P = 0.31), I^2 = 2.4%								

Figure 5. Forest plot of pooled effect sizes (Hedges' *g*) for HF HRV related to walking in natural versus control environment. Higher levels of HF HRV indicate findings in favor of the natural environment. df = degrees of freedom; CI = confidence intervals; Chi^2 = test of independence; I^2 = extent of heterogeneity in effect size across studies; P = statistical significance (more commonly symbolized as *p*); SD = Standard Deviation; Std. Mean Difference = Standardized mean difference (measure of effect size, also referred to as Hedges' *g*); Tau^2 = variance of the true effect sizes, Z = standard deviations away from the mean.

Gidlow, Jones et al. (2016) (in the forest plot referred to as Gidlow et al. (2016)) reported on the effects of two types of natural environments, i.e., 1) a park with a lake and 2) a park, on the same 38 participants. Means and SD's from both conditions are presented in the forest plots, but only one of the conditions was included in the pooled analysis. We chose to include the park condition as the baseline values were numerically closer to each other in this condition than those for the condition that included a park with a lake.

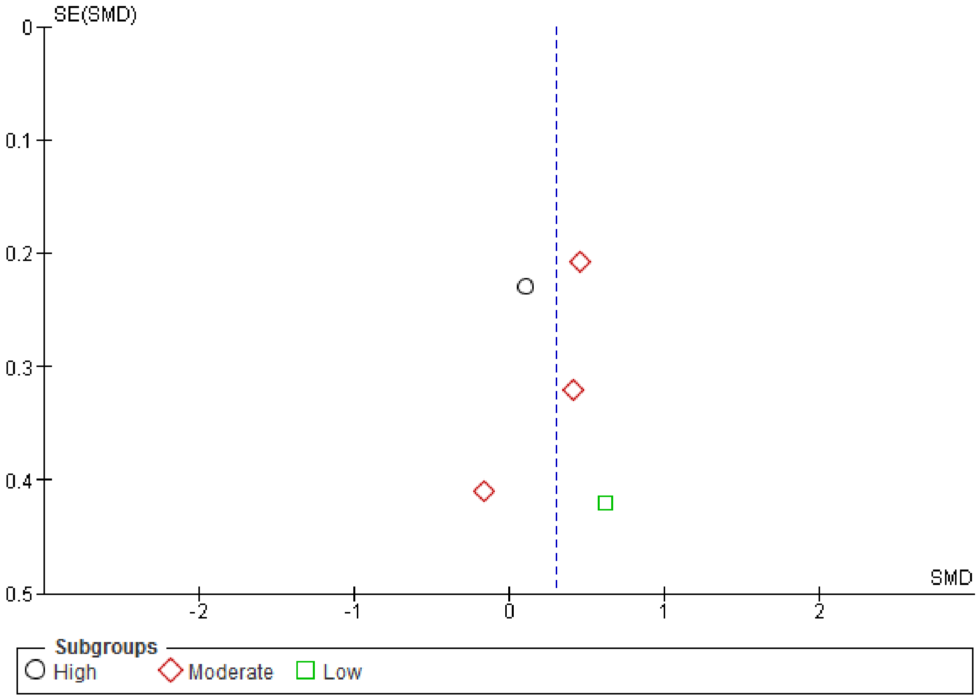
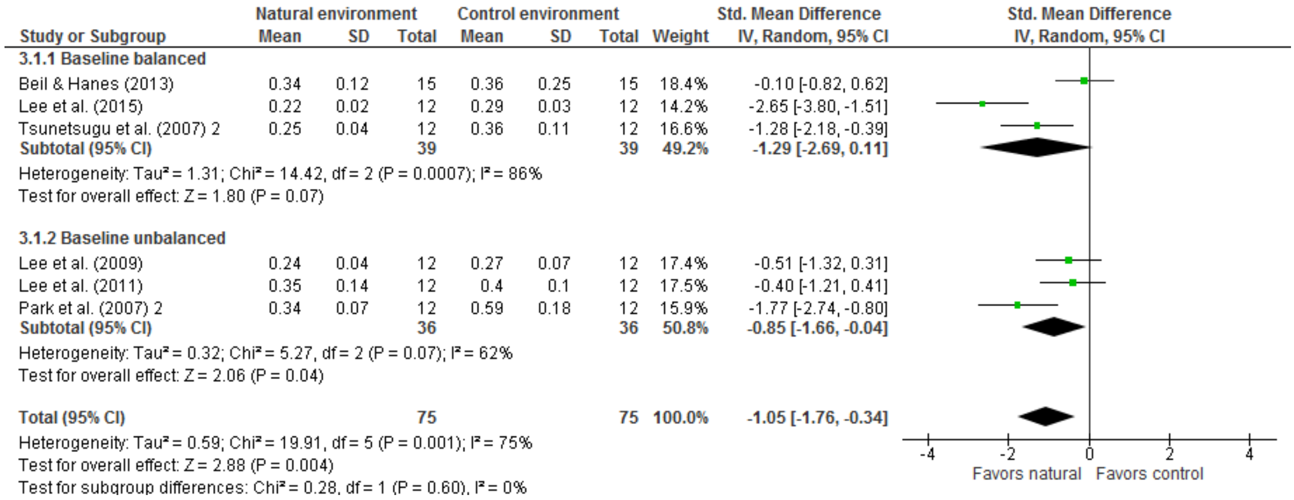


Figure 6. Funnel plot of studies exploring effects of walking in natural and control conditions on HF HRV, divided by study quality. Higher values of SMD indicates higher effect sizes and higher values of SE(SMD) indicate larger within-study variability in effects. The dotted line indicates the pooled effect size. SE = Standard Error; SMD = Standardized Mean Difference.



Study or Subgroup	Natural environment Mean	SD	Total	Control environment Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
3.1.1 Baseline balanced								
Beil & Hanes (2013)	0.34	0.12	15	0.36	0.25	15	18.4%	-0.10 [-0.82, 0.62]
Lee et al. (2015)	0.22	0.02	12	0.29	0.03	12	14.2%	-2.65 [-3.80, -1.51]
Tsunetsugu et al. (2007) 2	0.25	0.04	12	0.36	0.11	12	16.6%	-1.28 [-2.18, -0.39]
Subtotal (95% CI)			**39**			**39**	**49.2%**	**-1.29 [-2.69, 0.11]**
Heterogeneity: Tau² = 1.31; Chi² = 14.42, df = 2 (P = 0.0007); I² = 86%								
Test for overall effect: Z = 1.80 (P = 0.07)								
3.1.2 Baseline unbalanced								
Lee et al. (2009)	0.24	0.04	12	0.27	0.07	12	17.4%	-0.51 [-1.32, 0.31]
Lee et al. (2011)	0.35	0.14	12	0.4	0.1	12	17.5%	-0.40 [-1.21, 0.41]
Park et al. (2007) 2	0.34	0.07	12	0.59	0.18	12	15.9%	-1.77 [-2.74, -0.80]
Subtotal (95% CI)			**36**			**36**	**50.8%**	**-0.85 [-1.66, -0.04]**
Heterogeneity: Tau² = 0.32; Chi² = 5.27, df = 2 (P = 0.07); I² = 62%								
Test for overall effect: Z = 2.06 (P = 0.04)								
Total (95% CI)			**75**			**75**	**100.0%**	**-1.05 [-1.76, -0.34]**
Heterogeneity: Tau² = 0.59; Chi² = 19.91, df = 5 (P = 0.001); I² = 75%								
Test for overall effect: Z = 2.88 (P = 0.004)								
Test for subgroup differences: Chi² = 0.28, df = 1 (P = 0.60), I² = 0%								

Figure 7. Forest plot of pooled effect sizes (Hedges' *g*) for salivary cortisol related to seated relaxation in natural versus control environment. Lower levels of salivary cortisol indicate findings in favor of the natural environment. df = degrees of freedom; CI = confidence intervals; Chi^2 = test of independence; I^2 = extent of heterogeneity in effect size across studies; P = statistical significance (more commonly symbolized as *p*); SD = Standard Deviation; Std. Mean Difference = Standardized mean difference (measure of effect size, also referred to as Hedges' *g*); Tau^2 = variance of the true effect sizes, Z = standard deviations away from the mean.

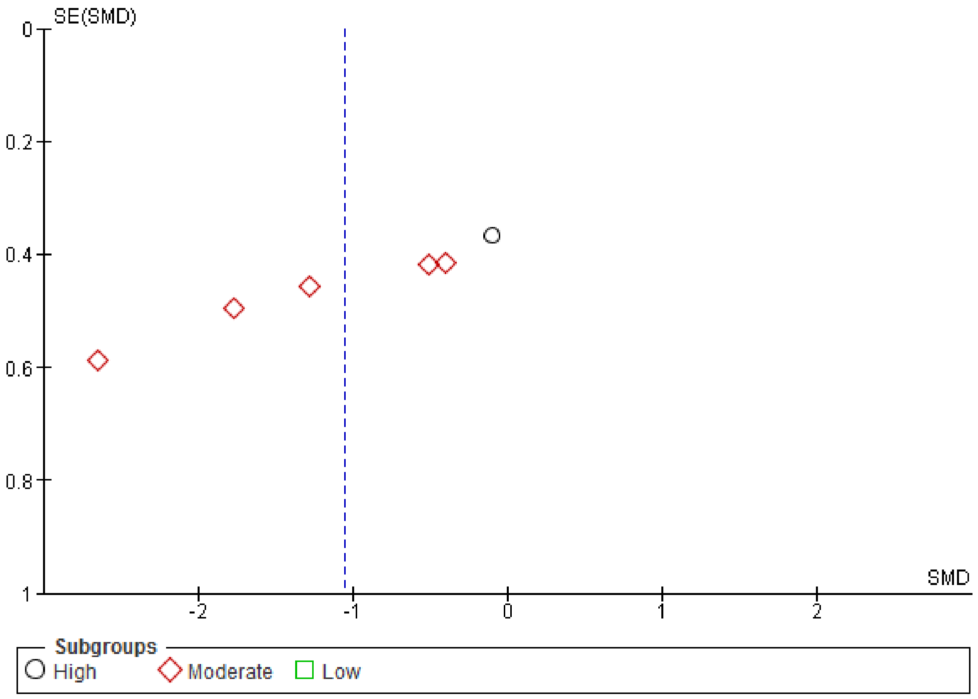
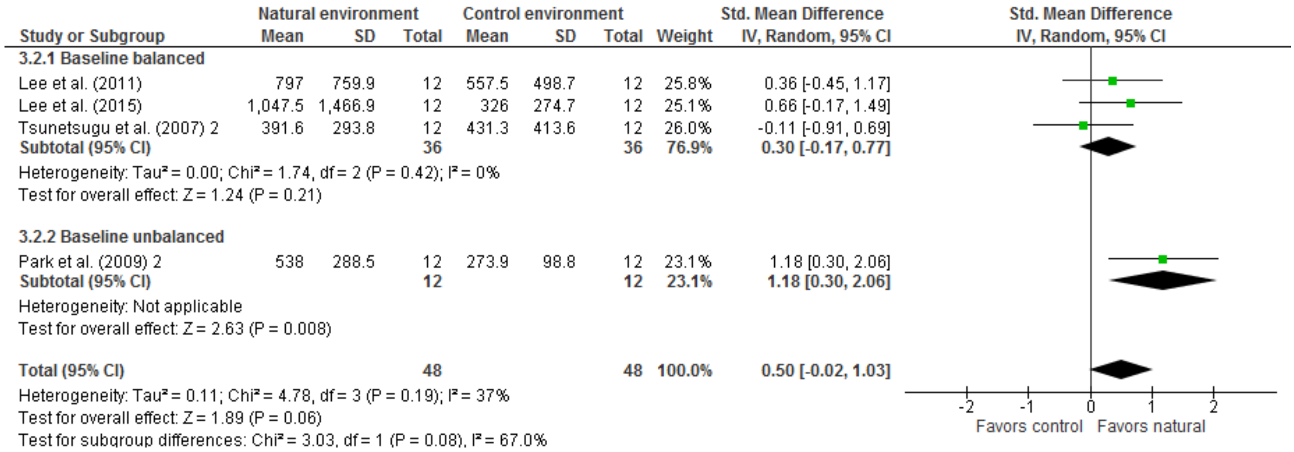


Figure 8. Funnel plot of studies exploring effects of seated relaxation in natural and control conditions on cortisol concentration, divided by study quality. Higher values of SMD indicates higher effect sizes and higher values of SE(SMD) indicate larger within-study variability in effects. The dotted line indicates the pooled effect size. SE = Standard Error; SMD = Standardized Mean Difference.



Study or Subgroup	Natural environment Mean	Natural environment SD	Natural environment Total	Control environment Mean	Control environment SD	Control environment Total	Weight	Std. Mean Difference IV, Random, 95% CI
3.2.1 Baseline balanced								
Lee et al. (2011)	797	759.9	12	557.5	498.7	12	25.8%	0.36 [-0.45, 1.17]
Lee et al. (2015)	1,047.5	1,466.9	12	326	274.7	12	25.1%	0.66 [-0.17, 1.49]
Tsunetsugu et al. (2007) 2	391.6	293.8	12	431.3	413.6	12	26.0%	-0.11 [-0.91, 0.69]
Subtotal (95% CI)			**36**			**36**	**76.9%**	**0.30 [-0.17, 0.77]**
Heterogeneity: Tau² = 0.00; Chi² = 1.74, df = 2 (P = 0.42); I² = 0%								
Test for overall effect: Z = 1.24 (P = 0.21)								
3.2.2 Baseline unbalanced								
Park et al. (2009) 2	538	288.5	12	273.9	98.8	12	23.1%	1.18 [0.30, 2.06]
Subtotal (95% CI)			**12**			**12**	**23.1%**	**1.18 [0.30, 2.06]**
Heterogeneity: Not applicable								
Test for overall effect: Z = 2.63 (P = 0.008)								
Total (95% CI)			**48**			**48**	**100.0%**	**0.50 [-0.02, 1.03]**
Heterogeneity: Tau² = 0.11; Chi² = 4.78, df = 3 (P = 0.19); I² = 37%								
Test for overall effect: Z = 1.89 (P = 0.06)								
Test for subgroup differences: Chi² = 3.03, df = 1 (P = 0.08), I² = 67.0%								

Figure 9. Forest plot of pooled effect sizes (Hedges' *g*) for HF HRV related to seated relaxation in natural versus control environment. Higher levels of HF HRV indicate findings in favor of the natural environment. df = degrees of freedom; CI = confidence intervals; Chi^2 = test of independence; I^2 = extent of heterogeneity in effect size across studies; P = statistical significance (more commonly symbolized as *p*); SD = Standard Deviation; Std. Mean Difference = Standardized mean difference (measure of effect size, also referred to as Hedges' *g*); Tau^2 = variance of the true effect sizes, Z = standard deviations away from the mean.