



CEE review 08-003

THE IMPORTANCE OF NATURE FOR HEALTH: IS THERE A SPECIFIC BENEFIT OF CONTACT WITH GREEN SPACE?

Systematic Review

BOWLER, D., BUYING-ALI, L., KNIGHT, T. & PULLIN A.

Centre for Evidence-Based Conservation - School of the Environment and Natural Resources - Bangor University – Bangor - LL57 2UW

Correspondence: d.bowler@bangor.ac.uk
Telephone: +44(0) 1248 382953

Draft protocol published on website: 12 February 2008 - Final protocol published on website: 21 April 2008 - Draft review published on website: 13 October 2009 – Final review posted on website: 28 June 2010

Cite as: Bowler, D., Buyung-Ali, L., Knight, T. & Pullin, A.S. 2010. The importance of nature for health: is there a specific benefit of contact with green space? CEE review 08-003 (SR40). Environmental Evidence: www.environmentalevidence.org/SR40.html.

Summary	4
1. Background	6
2. Objectives	7
3. Methods	8
3.1 Question formulation	8
3.2 Search strategy.....	8
3.2.1 Databases.....	8
3.2.2 Web sites	8
3.2.3 Search terms	8
3.3 Study inclusion criteria	8
3.4 Study characterisation and quality assessment.....	10
3.5 Data extraction.....	11
3.6 Data analysis.....	11
4. Results.....	13
4.1 Review statistics	13
4.2 Description of studies	13
4.2.1 Sources of studies	13
4.2.2 Focus of studies	13
4.3 Study methodology.....	17
4.3.1 Methods and tools used to collect data	18
4.3.2 Quality score of studies in the meta-analyses.....	20
4.4 Quantitative synthesis	21
4.4.1 Attention.....	21
4.4.2 Emotions/moods	22
4.4.3 Physiological	26
4.4.4 Characterisation of impact on emotions/moods	28
4.4.5 Summary of results	29
4.5 Qualitative research methods	29
4.6 Narrative synthesis.....	30
4.7 Studies on exercise performance	31
5. Discussion	32
5.1 Evidence of effectiveness.....	32
5.1.1 Study findings.....	32
5.1.2 Methodology of studies	32
5.2 Reasons for variation in effectiveness	33
5.3 Review limitations.....	33
6. Review conclusions	35
6.1 Implications for policy	35
6.2 Implications for research.....	35
7. Acknowledgements	37
8. Potential conflicts of interest and Source of support.....	37
9. References	38
Appendix A: Project Stakeholders	39
Appendix B: Website search.....	40
Appendix C: Methodology assessment	42
Appendix D: Flow chart of study inclusion.....	44
Appendix E: Characteristics of studies.....	45
Appendix F: Confidence intervals of effect sizes	47
Appendix G: References and summary tables of included studies	51

Summary

1. Background

There is increasing interest in the potential role of the natural environment in human health and well-being. Natural environments could impact on health through a number of pathways, either acting directly on health and therefore having a specific benefit, or acting indirectly, for instance, by promoting health-enhancing behaviour such as participation in physical activity. Despite the discussion on this topic, the evidence for there being a specific and direct benefit from contact with the natural environment has not been systematically reviewed.

2. Objectives

To address the question: how effective is direct accessing of natural environments in the promotion of health and well-being compared with other forms of 'exposure' to the natural environment or with accessing 'synthetic' environments?

3. Methods

We searched for literature in relevant databases, websites of specialist organisations, internet search engines, and references from bibliographies. Predefined inclusion criteria were applied to each article in order to identify the subset relevant for the review. For an article to be included in the review it must have compared a health or well-being outcome following activity (passive or otherwise) in a natural environment with activity (passive or otherwise) in a synthetic environment and/or viewing a natural environment. 'Natural environment' was used in a broad sense to include any environment with green space. The methodology of studies included in the review was critically appraised and the most commonly measured health/well-being outcomes were synthesized with meta-analysis.

4. Main results

Our broad search captured over 20,000 articles, however, after applying the inclusion criteria, only 28 studies were identified as relevant for the review. These studies mainly fell into two main pools: 1) studies comparing activity in the natural environment with an indoor environment or 2) studies comparing activity in the natural environment with an outdoor, built environment. Studies were diverse in terms of the types of participants and health/well-being outcomes measured. Most studies were short-term crossover trials and took measurements before and after exposure to the different environmental settings. Meta-analyses were conducted on several physiological parameters such as blood pressure, pulse and cortisol concentrations: these showed no evidence of an effect. There was also no evidence of a consistent effect on measures of attention or concentration. The most common outcome types measured were of mood/emotions, and based on these data, there was some evidence of a positive benefit on mood after a walk or run in a natural environment when compared to a different environment.

5. Conclusions

There is some evidence that activity in a natural environment compared to a different environment can have a positive impact on mental well-being. However, this is primarily drawn from short-term tests on self-reported feelings such as 'anger/aggression', 'sadness/depression' and 'fatigue/tiredness'. The validity of these

psychological scores as measures of mental well-being is not clear. There is little evidence of an impact on physiological outcomes but this is limited by the low number of studies available which measured similar outcomes. There was insufficient data to allow comparison of differences types of exposure to nature. Clearly, a 'natural environment' has many components. It is likely that further investigation on this topic and the design of more appropriate studies would be aided by refining the hypotheses on how specifically nature might impact on health and which specific attributes are the most important. The evidence is suggestive that nature may be used within the context of public health promotion interventions but we require a more comprehensive evidence-base in order to make appropriate and effective use of natural resources.

1. Background

The presence of a link between the natural environment and human health and well-being is of current interest to a number of organisations within the public health and environmental sectors. On the basis that this link does exist, several organisations have already invested resources in initiatives which use the natural environment in some way as a means of improving public health (e.g. British Trust for Conservation Volunteer's Green Gym; Parks Victoria's Graded Walks). The aim of such initiatives is not only to promote health but also to conserve biodiversity. Natural environments can be an attractive context for public health interventions by providing a free and accessible resource that can accommodate different sorts of activities. However, implementing these schemes can be costly and can mean resources are diverted from other activities: therefore it is important to ensure that there is a strong and scientifically sound evidence-base supporting them.

The literature on the links between the environment and health covers a wide range of issues from effects of nature on psychological health and well-being (RMNO, 2004; Bird, 2007) to the encouragement of exercise (Kaczynski and Henderson, 2007). A growing number of reviews have been produced that draw together the theory and evidence and offer conclusions on the importance of the environment for health. However, it is not clear how much the conclusions from these reviews reflect an objective and unbiased synthesis of the literature. Indeed, many explicitly start out with the assumption that there is a positive link and then discuss the ways in which the environment can promote health. Few reviews specifically address the effectiveness of particular interventions, which would be of particular use to decision-makers.

Systematic review methodology is widely employed in medicine and public health as a way of synthesising the evidence for the effectiveness of a particular intervention (Sackett and Rosenberg 1995; Khan *et al.*, 2003), but is only beginning to be recognised in environmental management (Pullin & Knight 2001, Pullin & Stewart 2006). These reviews differ from narrative reviews in a number of key ways. Essentially, the systematic review process comprises the setting of a clear and focused question which the review will address, a comprehensive search of the literature (both published and unpublished), transparent criteria for including studies captured by the search, critical appraisal of studies included in the review, and extraction and synthesis of data from included studies in order to address the review question (Khan *et al.*, 2003; CRD, 2004). Each component of this methodology is important as it allows confidence to be placed in the findings of the review being both unbiased and based on the best evidence available. Systematic reviews of the effectiveness of the natural environment in promoting health and well-being could also inform policy or practice through highlighting an evidence gap.

The review question was formulated following consultation with a project stakeholder group through a workshop held in October 2007 together with email correspondence with others. A report on the outputs of this consultation process is available. Question formulation was further informed by an appraisal of literature reviews and other relevant reports on the topic of 'the natural environment and promotion of human

health'. Stakeholders collectively identified 'effectiveness of measures to improve/increase access to natural environments in order to promote health and well-being' as a core question of importance to decision-makers. This breaks down into two sub-questions:

- 1a: How effective is direct accessing of natural environments in the promotion of health and well-being compared with other forms of 'exposure' to the natural environment or with accessing 'synthetic' environments?
- 1b: What is the effectiveness of interventions aimed at increasing access by people to natural environments?

Question 1a is the focus of this review. 'Natural environment' is used broadly to refer to all types of green space, many of which would more accurately be described as 'semi-natural', particularly in the UK context. However, we have used the looser term 'natural environment' as this has been the more widely used in the literature (this is further defined in section 3.3). 'Direct access' refers to direct contact with or physically being within the natural environment while 'other forms of exposure' include viewing nature through a window, in a picture or on a video. Essentially, this question addresses whether the benefits of green space vary with the type of exposure and whether there is an added benefit from exposure to the natural environment over and above an indoor or synthetic environment. Understanding the role that nature plays in health will be important in designing interventions that promote public health in the most efficient manner.

2. Objectives

To address the question:

How effective is direct accessing of natural environments in the promotion of health and well-being compared with other forms of 'exposure' to the natural environment or with accessing 'synthetic' environments?

Subject	Intervention	Comparator	Outcome
Human health	Direct exposure to nature	Indirect exposure to any sort of natural environment or exposure to an urban or synthetic environment	Measures of human health or well-being

3. Methods

3.1 Question formulation

The review question was formed in consultation with Natural England (a UK Government organisation). Stakeholders from various organisations (Appendix A) and subject experts acting as peer-reviewers were invited to comment on the proposed question and review protocol following the normal CEE consultation process.

3.2 Search strategy

3.2.1 Databases

Searching for relevant research data was conducted using a range of databases of different disciplines (environmental, ecological, public health) and document types (peer-reviewed, theses, grey literature) to ensure a comprehensive and, as much as possible, unbiased sample of the relevant literature was obtained. In each case, no time or document type restrictions were applied. The following databases were used: PubMed; EMBASE; CINAHL; PsycINFO; Web of Science (Science and Social science); SPORT Discus; ASSIA; HMIC- Data ; LILACS; UK National Research Register archives; TRIP database; UK National Library for Health ; Index to Theses Online; Directory of Open Access Journals; Economic and Social Data Service; Database of Promoting Health Effectiveness Reviews; Trials Register of Promoting Health Interventions; Cochrane Collaboration and Campbell Collaboration.

3.2.2 Web sites

A general internet search was undertaken using the following web sites: www.google.com; www.dogpile.com and www.scirus.com, using combinations of the search terms. The first 50 hits were checked for relevance. A search of the websites of specialist organisations was also undertaken (listed in Appendix B).

3.2.3 Search terms

Combinations of the following search terms (where * denotes a wild card term) were used to search the databases and internet web sites. When appropriate, relevant Medical Subject Headings were also searched for within a database and also added to the search term list.

Health terms: Exerc*, Physical activit*, Walk*, Health, Restorati*, Recovery, Therap*, Well-being, Wellbeing, Well being, Pyscholog*, Quality of Life, Life satisfaction, Play, Social, Recreation, Self*, Personal development, Child development, Happiness, Morale, Anxiety, Depression, Stress, Pain

Environmental terms: Park, Parks, Green*, Natural environment*, Open space*, Garden*, Horticultur*, Wilderness, Countryside, Outdoors, Nature, Biodiversity, Wood*, Allotment*, Forest*

3.3 Study inclusion criteria

For an article captured by our search to be relevant for the review it was required to meet all the following criteria i.e. to include a relevant subject, intervention, outcome and comparator as specified below:

Relevant subject(s): Human population(s) in any part of the world

Types of intervention: Previous studies have used a range of words to describe the environment e.g. green space, natural environment and the outdoors. Here, we employ the term ‘natural environment’ to encompass all types of ‘green’ outdoor, urban and rural environments from gardens, allotments and parks through to woodland, forests and so-called ‘wilderness’ while recognising that many of these environments may not be truly natural. In order to be inclusive, we also included environments such as university campuses that, based on author descriptions, appeared to be reasonably ‘green’. Therefore ‘natural environment’ is used in a broad sense. The review does not include studies of the impacts of environmental hazards (such as air pollution), indoor plants or domestic pets. In addition, in this review, an intervention does not refer to policies or programmes which enable or encourage specific human behaviour. Instead, it refers broadly to a human action of choice, which may or may not be encouraged by a policy, and falls into one or more of the following categories:

- i. An activity involving direct interaction with the natural environment e.g. gardening
- ii. Physically being present in a form of natural environment e.g. hiking, sitting in a forest
- iii. Other forms of exposure to natural environments e.g. videos, pictures, views from windows
- iv. Physically being present or carrying out an activity in a non-natural, ‘synthetic’ environments e.g. any indoor environment activity such as exercising in a gym

Types of comparator: Some studies might compare health outcomes before and after activity in a natural environment; however this is not deemed sufficient to demonstrate a specific health benefit of nature as any effect could simply be due to participation in an activity. Thus, we only include studies that compare at least two types of interventions in order to ask whether there is any evidence that presence in, or activity within a natural environment, has a different effect than presence in, or activity within a different non-natural environment, or from other forms of exposure to nature. To note, based on this criteria we do not include studies which compare viewing pictures of nature with viewing pictures of urban or synthetic environments.

Types of outcome: Health and well-being has many components and so a broad range of measurable physical and mental health outcomes was included. We aimed to include relevant quantitative and qualitative data.

References captured from computerised databases were imported into an Endnote library and duplicates were removed. The above inclusion criteria were applied to each article in turn in order to identify the most relevant articles from those captured by the search. In the first instance, the inclusion criteria were applied to title only in order to efficiently remove clearly irrelevant citations. Articles remaining were then further filtered by viewing abstracts and then full text, to reach the final list of relevant articles. As web searches retrieved fewer documents and these could not be easily imported into Endnote, they were filtered initially on their title and abstract (or introduction section) and then inclusion criteria were applied again to the full text. Title and abstract inclusion criteria were applied by three reviewers (DB, LBA & TK)

with consultation in cases of uncertainty. Full text inclusion was repeated by two reviewers on all those identified as potentially relevant (DB & TK).

3.4 Study characterisation and quality assessment

On all studies that met the above inclusion criteria, basic information was extracted from articles, which included details of the intervention, participants, type of data being measured, and the basic methodology used to collect data. A detailed quality checklist appropriate for studies (experimental or observational) was devised, guided by items from an available quality assessment tool (Thomas et al. 2004). This aimed to define and therefore standardise the extraction of the key aspects of study design, which covered elements of reporting quality, external validity (generalisability), internal validity of design and any factors that may affect the appropriateness of the study as a test of the effect of nature (described in full in Appendix C).

In order to generate an overall quality rating for the articles included in our quantitative synthesis, we identified six of these features as being most important. These were used to create a quality score based on the sum of criteria being met out of the six. These six criteria were:

1) Did the study identify its target population or give sufficient information on the types of individuals participating?

This determines the extent to which the study findings can be placed in context of the type of participant.

2) How did studies recruit participants?

'Referred by a third party' was considered superior to 'self-referral' due to the possibility of self-selection bias, which affects the representativeness of the sample of the population.

3) Was randomisation used to allocate participants to groups, or in the case of a cross-over trial, to decide the order of treatments?

This affects whether there were any systematic differences between participants of intervention and comparator groups.

4) Were studies shown to be similar at base-line or were base-line differences accounted for in the analysis?

This is important to be able to attribute measurable differences to the intervention.

5) Was the method of collecting data described and likely to be reliable and valid?

This affects the confidence we can have in the results reflecting the intended measurement.

6) Were there any other differences between intervention and comparator groups that might explain differences in the data being measured?

If there were any differences between groups apart from environmental setting, this might explain any differences in the outcome rather than the effect of nature. Thus studies with evident differences between groups apart from environmental setting did not meet this criterion. However, it is important to note that failure to meet this criterion can reflect the specific hypothesis of study, for instance, if the study was interested in investigating effects of a different factor and not specifically the effect of nature.

The value of quality scores in systematic review and meta-analysis has been debated (Juni et al. 1991; Greenland 1994). The relative weighting of these criteria could be questioned and further criteria could also be considered important from alternative

perspectives. We recognise our quality score represents only a coarse attempt to appraise study methodology and investigate the implications of study methodology on study findings. A more detailed investigation of the impact of specific features of study quality rather than this ‘overall quality’ summary variable was not considered appropriate given the limited number of studies identified as relevant for the review.

3.5 Data extraction

In all cases, studies either collected quantitative data or converted essentially qualitative data (on moods/emotions) to quantitative data through the use of scales/scores. Data on the health/well-being outcomes could usually be extracted from articles in the forms of means and standard deviations (or standard errors) either from their presentation in a table or from a figure within the article. If data were not available in the article, an attempt was made to contact the author for the relevant data. In order to ensure consistency in data extraction, we adhered to the following rules:

- In cases when individuals have been measured more than once after an activity, we extracted the values taken at a time closest to the end of the activity. This is for the sake of comparison with the remaining studies as most took measurements shortly following the activity.
- In cases when individuals have been measured more than once before an activity, we extracted the values taken when individuals were still in similar environments as this was deemed to provide the best base-line data.
- In cases when a study made more than one comparison, we focused on any comparators where the activity was more similar between environments, or at least there was some form activity rather than an inactive/control group (these were additional comparisons possible from: Hartig et al. 1991b; Harte & Eifert, 1995; Plante et al. 2006). However, some studies provided more than one relevant comparative data set for different reasons: firstly, if the authors investigated the effect of an additional factor (e.g. effect of a pre-treatment task versus no task) or secondly if the authors split the data by a subgroup (e.g. for male and female participants). In these cases, we extracted the disaggregated data set (i.e. multiple points per study).

Two articles also presented additional qualitative data, which are discussed in a later section (section 4.5).

3.6 Data analysis

Our review included studies that measured a broad range of different outcomes, which could not simply be synthesised simultaneously. We chose a threshold value of four studies measuring the same outcome in order to decide whether to pursue a meta-analysis on a particular outcome. From the data available, the effect size that we could calculate was the standardised difference between the outcome after some form of activity in a natural environment versus some form of activity in a comparator environment (either ‘outdoor built’ or ‘indoor’). All effect sizes were calculated using Hedges g and were corrected with the multiplication factor $1-3/(4*(n_1+n_2)-9)$ where n_1 and n_2 is the sample size of groups 1 and 2 respectively to account for the known bias of this formula as a population estimator (Cooper and Hedges 1994).

In most cases, studies also presented data on the outcome before the test. Ideally, we could have calculated the effect size as a standardized mean gain score; however, no data were available to estimate the correlation between before and after points for individual participants, which is necessary for this calculation. Instead, we use these data to calculate a pretest effect size and tested the effect of adjusting the posttest effect size by this value accordingly to account for any differences between groups prior to exposure to the treatments. However, because participants were usually similar at base-line, similar pooled effect sizes were obtained using only the unadjusted posttest effect sizes. We present the statistics on unadjusted effect sizes and note when the adjustment affects the result (see Appendix F: Tables c and d for adjusted effect sizes and their confidence intervals).

We calculated the overall pooled effect size and its confidence interval as a weighted average based on a random effects model (weighting of each study by the inverse variance where the variance includes its effect size variance and between-study variance). Arguably fixed effects models could have been used when the heterogeneity test indicated an insignificant amount of between-study variance ('heterogeneity'), however, in these cases, similar results were obtained either way and so we proceeded with random effects models for all analyses. We identified statistically significant effects as those where the confidence interval of the effect size did not overlap zero. Heterogeneity was tested using the Q-statistic, which tests the hypothesis that all effect sizes have come from the same population i.e. that any differences between studies are only due to chance. We present the statistics and forest plots of the effect sizes based on the full data set (disaggregated data: i.e. more than one effect size per study when available) to illustrate variables that have been investigated within individual studies. However, we also repeated the analysis on the aggregated data (i.e. one effect size per study) to account for non-independence of data from the same study and also present this in the results section; similar results were obtained either way.

We were able to conduct several separate meta-analyses on different outcomes but due to the low number of studies within individual analyses, we limit our exploration of heterogeneity to specific cases as follows. When the assumption of homogeneity was rejected ($p < 0.05$), we used meta-regression to investigate any effects of the quality score or comparator type as a covariate. Both these analyses were conducted on the aggregated data set, as these variables only differed between and never within studies. These results should be interpreted cautiously given few studies could be included in each analysis. Publication bias was not formally assessed due to the low number of studies.

4. Results

4.1 Review statistics

The electronic database search yielded a large number of articles (over 20,000), which is a reflection of the widespread discussion on nature and health. Many articles were removed based on title and/or abstract as the articles could be classed as either clearly irrelevant, more concerned with a more general discussion or were promotional material on health and activity in nature (see flow diagram in Appendix D). From the database search, 19 articles were identified as relevant after examination of their full text (one of these articles contained two studies). In addition, 8 additional articles were found during the web search and in the references of articles. Thus, 28 studies from 27 articles were included in the review. All other studies were excluded from the review because they failed to measure an aspect of health or wellbeing, did not investigate this after exposure to a natural environment and/or, most commonly, did not make a comparison with a non-natural environment.

4.2 Description of studies

4.2.1 Sources of studies

All articles identified as relevant at full text examination were published in peer-reviewed journals except one report. All were written in the English except one in German and one in Polish.

4.2.2 Focus of studies

a) Comparator

Studies have been initially grouped into three ‘pools’. Two pools were identified based on the specific type of comparison that was being made, which together comprised 23 studies. Pool one included twelve studies that compared a health or well-being outcome after activity in an outdoor green environment with after activity in an *outdoor built* environment (e.g. city, urban street, urban residential area). Pool two comprised thirteen studies that compared a health or well-being outcome after activity in an outdoor green environment with after activity in an *indoor* environment (usually a gym or a laboratory). Two studies were included in both pools as an attempt was made to compare similar activities in both comparisons (note section 3.5 on data extraction: for studies making more than one comparison, we focus on comparisons when the activity was more similar between environments).

There were an additional five studies (pool three) that could potentially be classified as making an indoor versus natural environment comparison. However this was not the aim of these studies and they were usually focused on comparing “exercise performance” in a laboratory test versus an outdoor test. The green element of the outdoor environment in these cases was generally not clear.

The following sections are based on the first two of these pools of studies. Articles in pool three are not discussed until a later session (section 4.7).

b) Participants

A wide range of types of participants was used in the studies. The most common study participant was a college student (eight studies). Several studies also focused

on individuals of one sex (six used only males and three used only females). Physically active individuals such as regular runners or athletes were also reasonably common (six studies). A few studies focused on individuals with specific health conditions such as inactive adults at risk from cardiovascular disease; children with impaired vision; children with Attention Deficit Disorder /Attention Deficit Hyperactivity Disorder; adults with ‘profound mental retardation’ [author’s term] or menopausal women (see tables in Appendix E).

c) Outcomes

The most common mental health/well-being outcome was some measure of an individual’s mood or feelings (see Table 1). Fifteen of the 23 studies used at least one measure of a particular mood. Many measured more than one mood (e.g. revitalisation, anger, anxiety etc...) Investigation of effects on attention/concentration was also relatively common with five studies employing a formal test of concentration e.g. Digit Span test and two other studies using reports of children’s ADHD symptoms from their parents. Impacts on physiological variables were usually on blood pressure or pulse (seven studies), or hormone levels (six studies: mostly salivary or urinary cortisol but included salivary amylase amongst others). Less common outcomes by one or two studies were investigation of effects on immune function (e.g. immunoglobulin A concentration) and levels of physical activity, and ‘others’ included motor performance, social engagement, cerebral activity and sleeping hours.

Table 1 The number of studies for each broad type of outcome, presented for each comparison and in total.

Type of outcome	Green vs Urban	Green vs Indoor	Total no. studies
Emotions/Moods	7	8	15
Attention/Concentration	7	2	7
Cardiovascular	3	4	7
Hormonal	4	2	6
Immune function	2	0	2
Physical activity	0	1	1
Other	2	3	5

d) Intervention

To be included in the review, one environment in which the outcome was measured must have included some element of greenery/nature. This included a range of different types of environments. Six studies took measurements after an activity in a public park and three others after an activity in a nature reserve or ‘wilderness’ [author’s term]. Five studies investigated activity which took place on a university campus, which, based on the authors’ descriptions, appeared to be relatively green. One study investigated the effects of gardening in a garden and another, of ‘forest kindergartens’. Four Japanese studies were interested in the activity known as ‘forest air-bathing’ (*shrinrin-yoku*) but little detail was given on the forest characteristics. As several studies were interested in an alternative hypothesis to the effect of nature, for instance, simply the effect of an outdoor environment, information on the nature of outdoor environment was not always detailed.

In the majority of cases, at least one of the comparisons being made by a study was the same activity in both environments (outdoor built versus nature: 11 out of 12; indoor versus nature: nine out of 13). Most investigated walking (12 studies) or running (four studies). Exceptions to this were: wilderness backpacking trip versus non-wilderness vacation (various activities); outdoor walking versus leisure-centre based exercise; gardening versus yoga (with comparable amount of activity noted by the author); time spent outdoors versus within a Snoezelen room (multisensory room); attendance of forest versus regular kindergartens (similarity of normal activities not clear). However, even in these cases, there was some form of activity in each group and in any case, apart from two of these cases, these studies did not have appropriate outcomes/available data which permitted inclusion in a meta-analysis (the two cases were: pulse following yoga versus gardening in Telles & Srinivas, 1998; pulse/blood pressure following outdoor walking versus leisure-centre based activity in Isaacs et al. 2007).

In most cases, these studies investigated short-term exposure to each environment, with generally less than one hour in each environment (see Table 2).

Table 2 The number of studies investigating different lengths of exposure to the environment, presented for each comparison and in total.

Time spent within the environment	Green vs Urban	Green vs Indoor	Total no. studies
less than 1 hour	8	4	12
less than 1 day	0	3	3
less than 1 week*	2	0	2
Greater than 1 week*	0	3	3
Not clear	2	3	3

* In these cases, time in the environment was not continuous but this duration reflects the period over which the repeated exposures occurred.

<u>Participants</u>	College students	(8)	<u>Green space</u>	Park	(6)
	Adult males only	(6)		University campus	(5)
	Physically active adults	(6)		Forest*	(5)
	Children (specific groups)	(5)		Nature reserve/wilderness	(3)
	Adult females only	(3)		Outdoors*	(3)
	Adults with health conditions	(2)		Garden	(1)
	Other adult groups	(2)			
<u>Outcomes</u>	Moods	(15)	<u>Activity in green space</u>	Walking	(12)
	Concentration	(7)		Running	(4)
	Cardiovascular	(7)		Mixed activities	(4)
	Hormones	(6)		Backpacking	(1)
	Immune function	(2)		Gardening	(1)
	Physical activity	(1)		Passive activity only	(1)
<u>Methodology</u>	Randomised controlled trial	(6)	Figure 1 Summary of the different foci of the 23 studies in our review. The size of the text is a reflection of the number of studies that meet the category; the number of studies is given in brackets. Appendix E shows the number of studies in each comparator group. An asterisk is used to denote categories when either little or no description was given.		
	Observational study	(5)			
	Counterbalanced crossover trial	(5)			
	Randomised crossover trial	(4)			
	Ordered crossover trial	(3)			

4.3 Study methodology

Figure 2 presents the number of studies out of the 23 studies that met each of the six main methodological quality criteria. Many studies described the individuals participating in their study (14 out of 23), usually in terms of their age, sex, health conditions and amount of previous activity. However, most recruited participants as volunteers rather than them being referred from a third party (four out of 23). Because of this self-selection, few studies could be deemed likely to be representative of their target population.

The basic design of most studies was a crossover design, which involved each participant being exposed to both environments. In four of these cases, individuals were randomised to decide the order of treatments received. Six other studies were randomised controlled trials. However, apart from one case, the method of randomisation was not described in any of these studies. In most cases, measurements were taken before and after exposure to the environment and therefore base-line data were available to assess pretest comparability (19 out of 23).

Most studies used standard tools to collect data; we discuss the validity of these data collection tools in the next section (section 4.3.1). Over half of the studies were not deemed to control for potentially confounding variables in their comparison of different environments, which could have affected any differences between environments observed. However, this arose from a number of different factors. In several cases, this was because all participants were exposed to the environments in the same order in a cross-over trial or because the studies were observational studies. However, in many other cases, this was because of the specific hypothesis of the experiment, which meant that additional features were present in one environment but not the other (e.g. listening to various stimuli through headphones/watching a video during the indoor walk only).

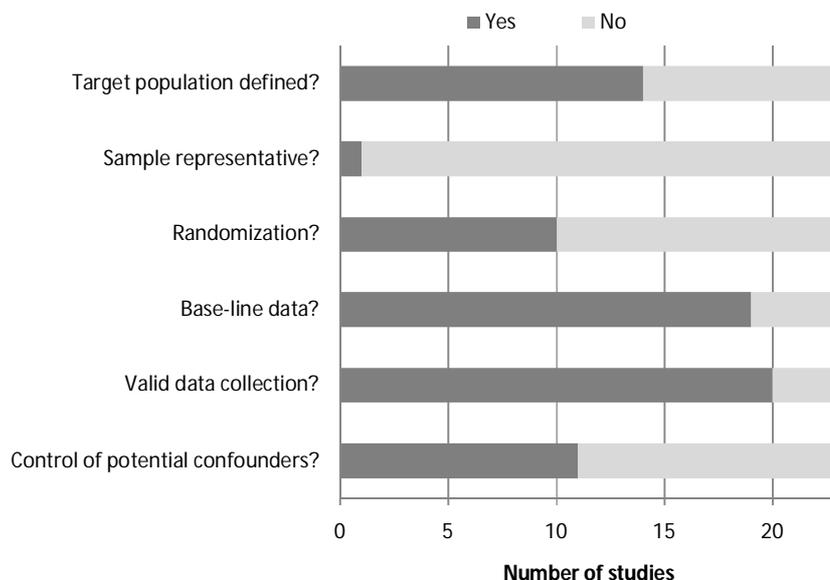


Figure 2 The number of studies that were deemed to meet each of the six main methodological criteria during assessment of study quality (total number of studies = 23).

4.3.1 Methods and tools used to collect data

We focus further on the ‘validity of data collection’ because of its importance in the strength of conclusions that can be drawn from the study findings. Ideally, assessment of data collection methods should consider both reliability (i.e. precision) and validity (i.e. accuracy) of the methods used to measure each outcome. However, different measurement tools and techniques were used for different outcomes and there was variation in the methodological information provided. We used different criteria to appraise different outcomes.

Because of the potential for ‘regression to the mean’ effect (Yudkin and Stratton 1996), we assessed the reliability of methods to collect data on blood pressure/pulse on whether there was an indication that more than one measurement had been taken at each measurement point. Assessment of methods to measure hormone concentration could not be fully appraised as most studies sent samples to a laboratory elsewhere to be analysed; this was deemed valid and reliable in the absence of further information.

Assessment of concentration/attention was usually based on standard tests (e.g. Digit Span test; Symbol Digits Modalities Test; Necker Cube Pattern Control) or another test e.g. proof reading task; these were deemed valid and reliable. Information on emotions/moods was based on self-reported data participants, obtained through use of various, reportedly ‘standard’, psychological questionnaires/scores. We present in Table 2 details on how authors justified the choice of psychological score used. In several cases, authors presented Cronbach’s alpha, which is a measure of the internal consistency in the measurement of a particular outcome (i.e. consistency in response between different items that are intended to measure that same emotion/mood construct). In other cases, authors only included a statement on validity/reliability, which usually referred to the previous use of the score, backed up with a reference. Some studies did give examples of the questions within the psychological scoring questionnaires. As an illustration of these, we present excerpts from articles below. This list is not exhaustive but illustrates the similarity of the different scores used.

- Bodin & Hartig (2003): “In light of their apparent link to restoration concepts and their previous use in research on emotional benefits of exercise, we used the revitalization and tranquility subscales from the Exercise-Induced Feeling Inventory (EFI; Gauvin & Rejeski, 1993) to measure positive emotions. Respondents indicate the extent to which given words (e.g., calm, relaxed, revived, refreshed) describe how they feel at the given moment (1 = *do not feel*; 5 = *feel very strongly*).
- Butryn & Furts (2003): “The shortened POMS is a 30 item mood-assessment rating scale which instructs subjects to indicate on a 5-point Likert scale the extent to which they feel each item, where a 0 indicates "not at all" and a 5 indicates "extremely." The instrument has six factors, which include: Tension-Anxiety, Depression-Dejection, Anger-Hostility, Fatigue-Inertia, Vigor-Activity, and Confusion-Bewilderment.”
- Hartig et al. 2003: “Pretest and postwalk measures of positive affect, attentiveness, fear arousal, sadness, and anger/aggression were obtained with Zuckerman’s (1977) Inventory of Personal Reactions (ZIPERS). Respondents indicate the extent to which statements describe how they feel that moment (e.g. I feel elated or pleased) (1 =not at all; 5 = very much).
- Teas et al. (2007): “Mood states were measured using the Positive Affect Scale (PAS) and the Negative Affect Scale (NAS), a validated standard questionnaire that includes a brief list of six mood states (Diener and Emmons, 1985). Participants were instructed to consider their current feelings and to mark how happy, joyful, delighted, pleased, depressed/sad, angry, frustrated, and worried/anxious they felt on a standard six point Likert continuum scale, with 0 representing “not at all” and 6 for “extremely.”

Table 2 The different scores used in studies and the method by which authors reported the reasons for their choice of score.

Citation	Score/index used in study	Cronbach's alpha reported	Other statement referring to its validity/reliability with reference	Reference to another paper only
Bodin & Hartig (2003)	Exercise-Induced Feeling Inventory	✓		
	Negative Mood Scale	✓		
Butryn & Furst (2003)	Profile of Mood States	✓		
	Exercise-Induced Feeling Inventory	✓		
Hartig et al. (2003)	Zuckerman Inventory of Personal Reactions		✓	
	Overall Happiness Scale		✓	
Park et al. (2007)	Calm/comfort index			
Tsunetsugu et al. (2007)	Calm/comfort index			✓
	Stress-Refresh feeling test			
Hartig et al. (1991)	Zuckerman Inventory of Personal Reactions		✓	
	Overall Happiness Scale			✓
Hull and Michael (1995)	Thayer's 4-point scale	✓		
Kerr et al. (2006)	Tension and Effort Stress Inventory	✓		
Isaacs et al. (2007)	Hospital Anxiety and Depression Scale (HADS)	✓		
Teas et al. (2007)	Positive affect scale and Negative affect scale		✓	
Peacock et al. (2007)	Profile of Mood States		✓	
Plante et al. (2006)	Activation-Deactivation Adjective Check List		✓	
Plante et al. (2007)	Activation-Deactivation Adjective Check List		✓	
Harte and Eifert (1995)	Profile of Mood States		✓	

4.3.2 Quality score of studies in the meta-analyses

We present the quality scores of the subset of studies included in the meta-analyses to illustrate the quality and number of studies in each analysis. We use the basic philosophy of a ‘harvest plot’ (Oglivie et al. 2008) to illustrate the number of studies (each represented by a bar) reporting on each specific outcome, simultaneously with the number of quality components out of the six main criteria each study met (represented by bar height; see Figure 3).

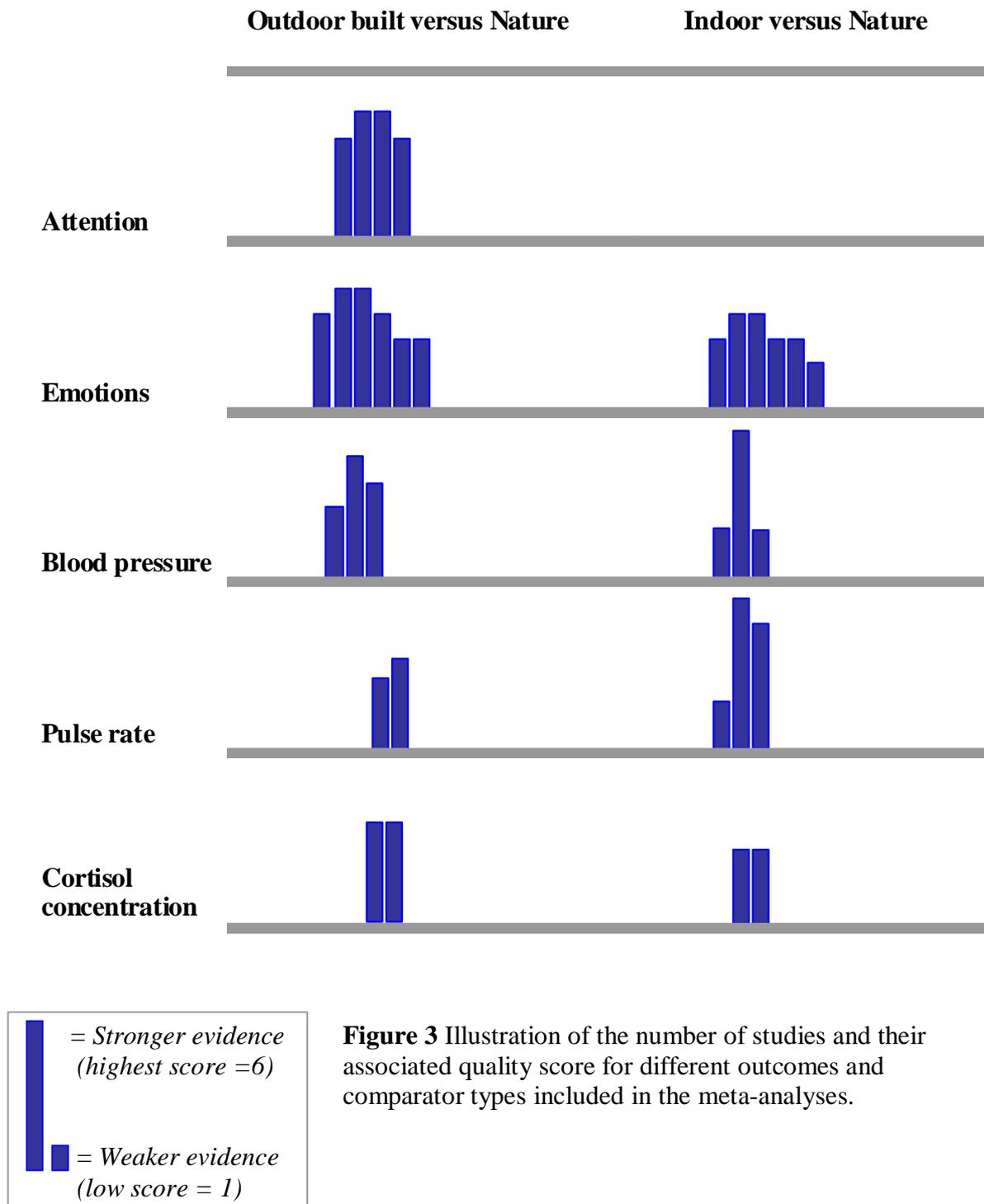


Figure 3 Illustration of the number of studies and their associated quality score for different outcomes and comparator types included in the meta-analyses.

4.4 Quantitative synthesis

We pursued a meta-analysis when there were at least four studies measuring the same variable. This is recognisably a small number but in many cases only four or five studies were available. In total, 18 of the studies contained at least one variable which could be synthesised in a meta-analysis.

Interpretation of the effect size:

The effect size (Hedges g) reflects the difference between the outcome after exposure to a natural environment compared to after a different, non-natural environment (either outdoor built or an indoor environment). As explained in section 3.6 above, **it does not reflect the difference between the pretest and posttest of activity (see Appendix F for these effect sizes).**

A positive value indicates that the effect on the outcome was *positive or at least more positive* in nature compared to non-nature. Similarly, a negative value indicates that the effect on the outcome was *negative or at least more negative* in nature compared to non-nature.

To note, we use ‘positive’ and ‘negative’ to simply reflect the quantitative value of the outcome. In some cases, a positive value does not necessarily reflect a beneficial impact on health (e.g. anxiety).

On each of the following forest plots, study descriptors denote the authors and year of the study and in cases of multiple data points per study details of the specific comparator group used to derive each effect size are also shown (see Appendix G for further details on each study).

4.4.1 Attention

For this outcome, in some cases, a study used more than one standard test to measure attention in the same group of participants in the same environment; in these cases data points were averaged over multiple tests.

Study

Bodin & Hartig (2003) male
Bodin & Hartig (2003) female
Faber-Taylor & Kuo (2008) downtown
Faber-Taylor & Kuo (2008) neighbourhood
Hartig et al. (1991)
Hartig et al. (2003) pretreatment task
Hartig et al. (2003) pretreatment no task

Summary

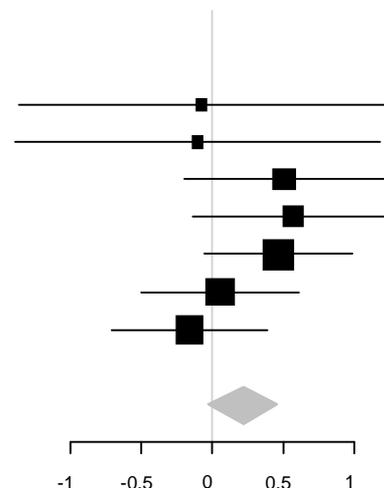


Figure 4 Forest plot showing the effect sizes (Hedges g) \pm 95 % CI for attention.

All studies measuring attention compared an outdoor built environment with a natural environment. Of the seven studies which investigated attention, three studies could not be included in the analysis; in two cases the outcomes were parental reports of

their child's ADD/ADHD symptoms and in the other, the data were not available for synthesis. The confidence interval of the effect size overlaps zero indicating that the hypothesis of no effect cannot be rejected (average $g = 0.217$, 95% CI = -0.029, 0.46457). The level of heterogeneity in effect size was not significant ($Q = 5.444$, $df = 6$, $p = 0.488$). Thus, despite some evidence of an effect in individual studies, the overall average of these four studies does not indicate a consistent difference on attention of direct exposure to a natural environment in comparison to a different non-natural environment (Fig 4).

4.4.2 Emotions/moods

The particular moods/emotions measured was dependent on the particular psychological score used (e.g. ZIPERS or POMS). This meant that different studies measured different sets of moods/emotions. Using the terminology of the authors, we initially grouped data from studies that could be considered as measuring the same mood/emotion and conducted meta-analysis on these groups.

In an attempt to increase the number of studies we could synthesize, we also considered a higher level of grouping. A simple split into 'positive' and 'negative' moods/emotions was deemed inadequate because of the variety of moods/emotions within each group e.g. the contrast between energy and calmness, both 'positive' moods. We use a framework outlined by Russell (2003) that groups moods along two dimensions: one dimension corresponds to a pleasure/displeasure axis and a second dimension corresponds to an activation/deactivation axis and represents a sense of mobilization and energy (see Russell 2003 for more discussion).

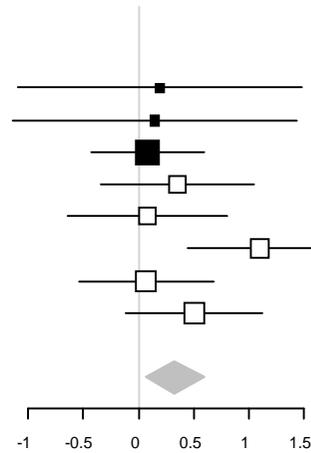
We firstly present analysis of the individual moods, denoting the specific scores used in each case in brackets, and in a subsequent section present the analysis after grouping studies in this two-dimensional framework.

Energy

This analysis included studies that measured one of the following: revitalisation (Exercise-Induced Feeling Inventory); energetic (Thayer's 4-point scale); vigour-activity (Profile of Mood States) or energy (Activation-Deactivation Adjective Check List). Two of the studies could not be included in a meta-analysis because data were not available or there was uncertainty over the data. Based on analysis of post-test differences, there is a tendency towards a slight positive effect of a natural environment on feelings of energy when compared to a different, non-natural, environment (Fig 5a; average $g = 0.328$, 95% CI = 0.068, 0.588). However, this result is less consistent after adjusting the effect size for pre-test differences (average $g = 0.147$, 95% CI = -0.089, 0.382). The amount of heterogeneity in effect size among different comparisons and studies was not significant ($Q = 8.234$, $df = 7$, $p = 0.312$).

a) Study measuring Energy

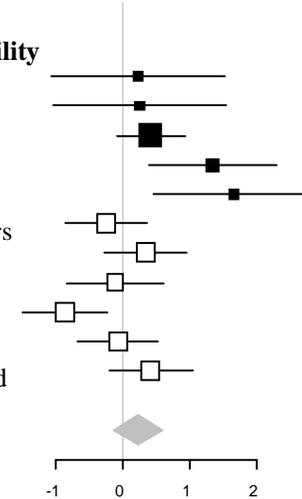
- Bodin & Hartig (2003) males
- Bodin & Hartig (2003) females
- Butryn & Furst (2003)
- Peacock et al. (2007)
- Plante et al. (2006) males, treadmill & video
- Plante et al. (2006) females, treadmill & video
- Plante et al. (2007) alone
- Plante et al. (2007) with friend



Summary

b) Studies measuring Tranquillity

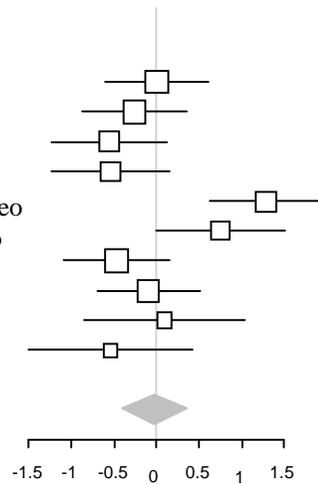
- Bodin & Hartig (2003) males
- Bodin & Hartig (2003) females
- Butryn and Furst (2003)
- Park et al. (2007)
- Tsunetsugu et al. (2007)
- Kerr et al. (2006) recreat runners
- Kerr et al. (2006) comp runners
- Plante et al. (2006) males
- Plante et al. (2006) females
- Plante et al. (2007) alone
- Plante et al. (2007) with a friend



Summary

c) Studies measuring Anxiety

- Kerr et al. (2006) comp runners
- Kerr et al. (2006) recreat runners
- Teas et al. (2007)
- Peacock et al. (2007)
- Plante et al. (2006) females, treadmill & video
- Plante et al. (2006) males, treadmill & video
- Plante et al. (2007) alone
- Plante et al. (2007) with friend
- Harte and Eifert (1995) external stimuli
- Harte and Eifert (1995) internal stimuli



Summary

Figure 5 Forest plot showing the effect sizes (Hedges $g \pm 95\%$ CI for a) energy; b) tranquillity and c) anxiety

- Nature vs Outdoor built
- Nature vs Indoor
- Pooled summary effect

Tranquillity/Calmness

This analysis included studies that measure at least one of the following: tranquillity (subscales from the Exercise-Induced feeling Inventory); 'subjective calm feeling'; placidity (TESI) or calmness (AD-ACL). One study could not be used in the analysis as the data were not available. The confidence interval of the effect overlaps zero (Fig 5b; average $g = 0.230$, 95% CI = -0.138, 0.597). However, there was support for systematic heterogeneity in effect size above that expected by sampling error ($Q = 29.271$, $df = 10$, $p < 0.001$). Studies comparing the natural environment with an urban environment tended to find more positive effects of nature than those making a comparison with an indoor environment (coef = -0.897 ± 0.381 , $p = 0.018$). There was also some indication that studies obtaining a lower quality score, at least relative to the other studies, reported an effect of larger magnitude (ignoring the sign of the effect; coef = -0.562 ± 0.268 , $p = 0.036$).

Stress/Anxiety

This analysis included studies that measured at least one of the following: worried (Negative affect Scale); anxiety (TESI); anxious (Thayers); anxiety (HADS); tension (POMS) or tension (ADCL). Two studies did not present the necessary data and therefore could not be included. All remaining studies that investigating anxiety or tension compared the natural environment with an indoor environment. For this particular emotion, a negative effect size indicates a more beneficial effect on health. The confidence interval of the effect size overlaps zero (Fig 5c; average $g = -0.032$, 95% CI = -0.412, 0.348), however, there was a significant amount of heterogeneity among studies ($Q = 28.839$, $df = 9$, $p < 0.001$). Study quality score did not explain any variation (coef = 0.083 ± 0.379 , $p = 0.83$).

Anger/Aggression

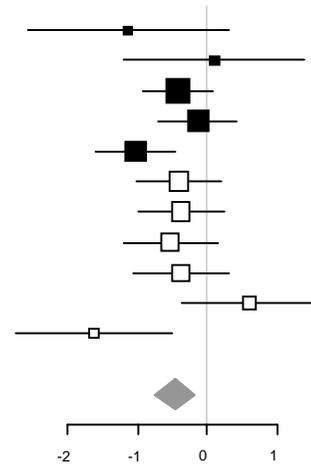
This analysis included studies that measured at least one of the following: anger (Negative Mood Scale); anger and aggression (ZIPERS); anger (TESI); angry (Negative Mood Scale) or anger (Profile of Mood States). There were two studies for which relevant data were not available. The remaining studies indicated that feelings of 'anger' were lower after activity in a natural environment compared to another environment (Fig 6a; average $g = -0.456$, 95% CI = -0.737, -0.175). There was marginally insignificant heterogeneity in effect size among studies ($Q = 17.681$, $df = 10$, $p = 0.06$).

Sadness/Depression

This analysis included studies measuring at least one of the following: sadness (ZIPERS); sad (Positive affect scale); depression (Profile of Mood states) or depression-rejection (Profile of Mood states). There were four studies for which the data were not available. Overall, the studies support lower feelings of sadness or depression after activity in a natural environment compared to a non-natural environment (Fig 6b; average $g = -0.497$, 95% CI = -0.839, -0.155). The amount of heterogeneity was not significant ($Q = 4.845$, $df = 4$, $p = 0.304$).

a) Studies measuring Anger

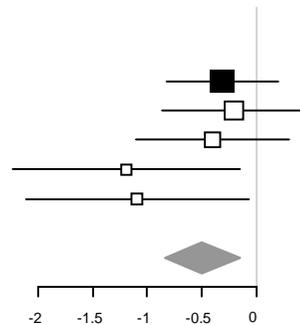
- Bodin & Hartig (2003) male
- Bodin & Hartig (2003) female
- Hartig et al. (1991b)
- Hartig et al. (2003) pretreatment task
- Hartig et al. (2003) pretreatment no task
- Kerr et al. (2006) comp runners
- Kerr et al. (2006) recreat runners
- Teas et al. (2007)
- Peacock et al. (2007)
- Harte & Eifert (1995) external stimuli
- Harte & Eifert (1995) internal stimuli



Summary

b) Studies measuring Sadness

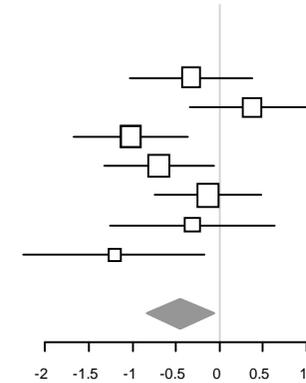
- Hartig et al. (1991b)
- Teas et al. (2007)
- Peacock et al. (2007)
- Harte & Eifert (1995) external stimuli
- Harte & Eifert (1995) internal stimuli



Summary

c) Studies measuring Fatigue

- Peacock et al. (2007)
- Plante et al. (2006) males
- Plante et al. (2006) females
- Plante et al. (2007) alone
- Plante et al. (2007) with a friend
- Harte & Eifert (1995) external stimuli
- Harte & Eifert (1995) internal stimuli



Summary

Figure 6 Forest plot showing the effect sizes (Hedges $g \pm 95\%$ CI for a) anger; b) sadness and c) fatigue

- Nature vs Outdoor built
- Nature vs Indoor
- ◆ Pooled summary effect

Tiredness/Fatigue

This analysis included studies that measured either fatigue (Profile of Mood States) or tiredness (AD-ACL). Data were not available from two studies. All studies made a comparison between the natural environment and an indoor environment. There was support for lower fatigue after activity in a natural environment compared to an indoor environment (Fig 6c; average $g = -0.453$, 95% CI = -0.841, -0.066). However, there was significant heterogeneity ($Q = 12.759$, $df = 6$, $p = 0.047$). Study quality score did not explain any variation in effect size (coef = -0.041 ± 0.229 , $p = 0.86$).

4.4.3 Physiological

Although various physiological and other health outcomes were measured by the studies, there was little overlap in outcome type between different studies. Based on previous procedures, we synthesized data on outcome for which there were at least four studies. We therefore synthesized the following: blood pressure, pulse and cortisol concentrations.

For this outcome, two of the studies that could be included did not compare the same activity in each environment (see section 4.2.2 d) comparators); exclusion of these studies makes little differences to the results.

Blood pressure

Analysis was conducted separately on data for systolic and diastolic blood pressure. There was no evidence of a difference between environments on systolic blood pressure following the activity (Fig 7a; average $g = -0.015$, 95% CI = -0.201, 0.171) nor was there any evidence of heterogeneity ($Q = 2.308$, $df = 7$, $p = 0.94$). Similarly, there was no evidence of a difference between environments on diastolic blood pressure following the activity (Fig 7b; average $g = 0.009$, 95% CI = -0.217, 0.236) nor evidence of heterogeneity ($Q = 5.982$, $df = 5$, $p = 0.308$). Overall, there was considerable uncertainty in effect sizes from individual studies due to within-study variation among participants.

Pulse or Heart rate

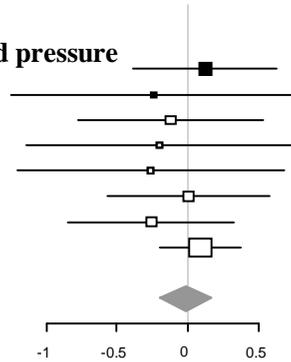
There was no evidence of a difference between environments on pulse following the activity (Fig 7c; average $g = -0.002$, 95% CI = -0.191, 0.187) nor was there any evidence of significant heterogeneity among studies ($Q = 2.967$, $df = 4$, $p = 0.564$).

Cortisol concentrations

These effect sizes were derived from either salivary (3 studies) or urinary cortisol (1 study) concentrations. There was no evidence of a difference between environments on cortisol concentrations following the activity (Fig 7d; average $g = -0.047$, 95% CI = -0.493, 0.398) nor evidence of heterogeneity among studies ($Q = 5.961$, $df = 4$, $p = 0.202$).

a) Studies measuring systolic blood pressure

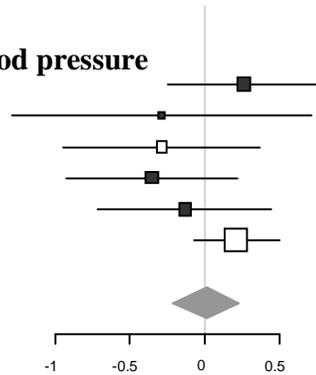
Hartig et al. (1991b)
 Tsunetsugu et al. (2007)
 Teas et al. (2007)
 Harte & Eifert (1995)
 Harte & Eifert (1995)
 Hartig et al. (2003) pretreatment task
 Hartig et al. (2003) pretreatment no task
 Isaacs et al. (2007) 10wks*



Summary

b) Studies measuring diastolic blood pressure

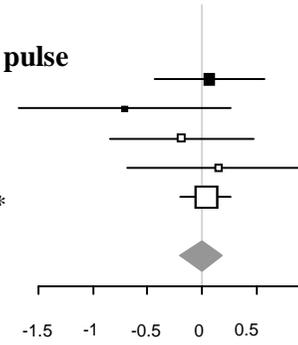
Hartig et al. (1991b)
 Tsunetsugu et al. (2007)
 Teas et al. (2007)
 Hartig et al. (2003) pretreatment task
 Hartig et al. (2003) pretreatment no task
 Isaacs et al. (2007) 10wks*



Summary

c) Studies measuring pulse

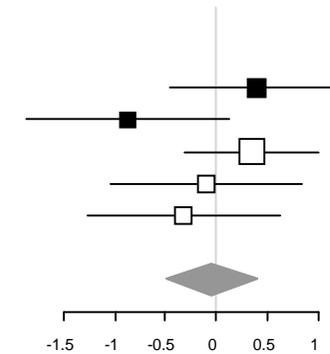
Hartig et al. (1991b)
 Tsunetsugu et al. (2007)
 Teas et al. (2007)
 Telles & Srinivas (1998)*
 Isaacs et al. (2007) 10wks*



Summary

d) Studies measuring cortisol

Park et al. (2007)
 Tsunetsugu et al. (2007)
 Teas et al. (2007)
 Harte & Eifert (1995) external stimuli
 Harte & Eifert (1995) internal stimuli



Summary

Figure 7 Effect sizes for various physiological parameters (see legend Fig 6 for further details). Studies with an asterisk (*) are those which do not compare the same form of activity between environments (see section 4.2.2 d); their exclusion from the analysis makes little difference to the results.

4.4.4 Characterisation of impact on emotions/moods

In a final analysis, moods/emotions were grouped into four categories. This allows the incorporation of moods/emotions which could not be grouped into previous analyses and therefore increases the sample size in the meta-analyses. We base this analysis on aggregated data (one average data point per study).

Each group includes any data available on the following:

- i) Pleasure-activation: all those included in ‘energy’ as above as well as positive affect; attentiveness, happiness, pleased, delighted, joy.
- ii) Pleasure-deactivation: all those included in ‘tranquillity’ as above as well as relaxation [apart from one point, this data set is the same as that in ‘tranquillity’ and therefore shows similar results]
- iii) Displeasure-activation: all those included in ‘anger’ and ‘anxiety’ as above as well as confusion.
- iv) Displeasure-deactivation: all those included in ‘fatigue’ and ‘sadness/depressions’ as above as well as boredom.

Pooling across both (indoor and outdoor built) comparators

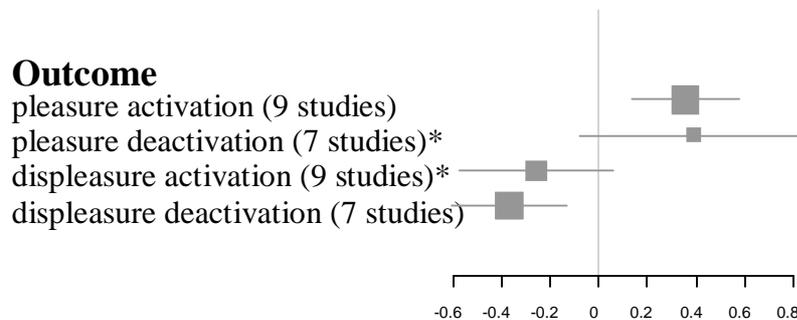


Figure 8 shows the pooled effect sizes (Hedges g) and 95% CI for each ‘mood grouping’ based on aggregated data (one average effect size per study). In brackets are shown the number of studies used to calculate the effect size and an asterisk is used to denote a significant ($p < 0.05$) heterogeneity test for a particular group.

Based on this grouping, there is consistent support for a positive influence of the natural environment compared to a different environment on pleasurable moods and a negative influence on those classified as displeasurable (i.e. less displeasure after nature compared to non-nature) (Fig 8). Due to heterogeneity within the studies, the findings are less consistent on pleasure-deactivation moods (i.e. tranquillity/calmness) and displeasure-activation (anxiety/anger). However, overall the data indicate a positive benefit on mood/emotion when comparing a natural environment with a non-natural environment, which is consistent with the previous analysis on individual moods.

4.4.5 Summary of results

We present the summary of our results for each separate outcome (Fig 9). This is based on the aggregated data set (one average data point per study) to verify the robustness of the results to non-independence of data from the same study. This also allows comparison of the effect size among different outcomes.

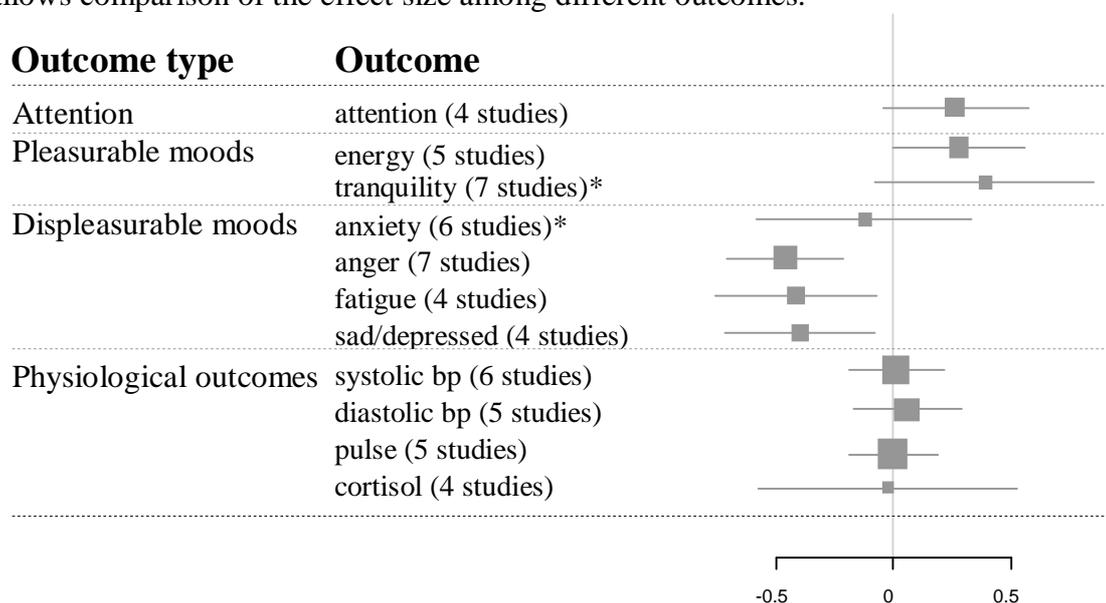


Figure 9 The pooled effect sizes (Hedges g) and 95% CI for each outcome based on aggregated data (one average effect size per study). In brackets are shown the number of studies that was used to calculate the effect size and an asterisk is used to denote a significant heterogeneity test ($p < 0.05$) for a particular group.

4.5 Qualitative research methods

Most of the previous analyses were based on self-reported assessment of the emotions/moods of participants which was converted to quantitative data through the use of scores. Our review aimed to also include any studies that collected data using qualitative research methods as long as data were collected from relevant comparator groups (and therefore satisfied the inclusion criteria for the review). There were no studies with appropriate comparators which used purely qualitative research methods however, two of the studies included in the review, which presented data using psychological scores, also employed some form of qualitative research method, although in both cases this was a small part of their investigation.

Faber Taylor et al. (2001) surveyed the parents of children with ADD regarding the effects of different activities in different environmental settings on their child's symptoms. Following analysis of the survey data, a subset of participants was invited to a focus group dinner to discuss the results. Selected comments from parents during this discussion were reported in the article. Peacock et al. (2007) compared outcomes in participants following a walk in an indoor shopping centre and a country park. As part of their questionnaire that incorporated the "Profile of Mood States" psychological score, there was also a narrative question allowing participants to provide qualitative information on their experience of the walks. These comments of the participants are presented within the report. These studies illustrate the range of

perceived benefits but they do not formally analyse their qualitative data e.g. with thematic analysis.

4.6 Narrative synthesis

An issue with synthesis of the studies included in this review is the variety of different health or well-being outcomes which have been used. A range of other outcomes were measured but insufficient independent data points (i.e. from different studies) were available to attempt a quantitative synthesis. We therefore provide a brief narrative synthesis of the other outcomes that have been measured, restricted to those which do not fall into one of the outcome types analysed above.

Two studies conducted in Japan investigated the effects of walking in a forest on measures of immune function (Tsunetsugu et al. 2007; Li et al. 2008) including variables such as secretory immunoglobulin A, NK activity, number of T cells and white blood cells, amongst others; mixed results were obtained with no clear difference between a forest environment and a city environment. Other hormones, or measures of hormone activation, apart from cortisol, have been investigated such as adrenaline (Harte and Eifert 1995; Li et al. 2008), salivary amylase (Yamaguchi et al. 2006) and noradrenaline (Harte and Eifert 1995) with mixed results on the differential effect of different environmental settings. Park et al. (2007) used a Spectroscopy system to measure haemoglobin concentrations as a measure of cerebral activity; this technique had not been used in the field before.

Cuvo et al. (2001) compared the effects on adults with ‘profound mental retardation’ (authors’ description), of activity within an indoor living and multisensory room with outdoor activities in the grounds of institution in a rural area. Three adult participants were observed, specifically for idiosyncratic behaviour such as mouthing, body rocking, body swaying and picking, as well as engagement. There was some indication of an improvement during the outdoor activity compared to the indoor environments.

Two cross-sectional studies asked parents of children with ADHD/ADD to rate their child’s symptoms after different activities, and within different environmental and social settings (Faber-Taylor et al. 2001; Kuo & Faber-Taylor 2004). Based on the parental assessment, the results support a positive impact of a natural environment compared to both an indoor and an outdoor built environment. This could have been due to the types of activities undertaken within a green environment but was still evident after controlling for the type of activity. However, the reliability of parental assessment as a measure of ADD/ADHD symptoms was not clear.

Scholz and Krombholz (2007) compared the sport-motor performance of children from 10 forest kindergartens and from 4 ‘regular’ kindergartens, measuring variables that reflect the ability to balance and jump. The authors concluded that the motor performance of the children from forest kindergartens was superior.

Most studies were short-term apart from one notable exception: Isaacs et al. 2007. In this study, the hypothesis was not the effect of nature, but it was still deemed relevant for the review as it compared 10 week programmes of leisure-centre based activities with instructor-led walking programmes through parks and open spaces (also with an

advice-only group). This study included follow up assessments at 10 weeks, 6 months and 1 year and measured a range of outcomes including self-reported exercise behaviour, blood pressure, total cholesterol and lipid subfractions, anthropometry, cardiorespiratory fitness, flexibility, strength and power, self-reported lifestyle behaviour, psychological health status and quality of life. This contrasts to most of the other studies which took measurements just before and after the activity. Isaacs et al. (2007) found that there was generally little difference in health/well-being between the two groups and even in comparison with the advice-only group.

4.7 Studies on exercise performance

Our review identified five studies which potentially could be grouped within our indoor versus natural environment comparison group. However, due to the very different hypotheses and outcome measures of these studies, we deal with these studies in a separate section here. In these cases, details on the type of outdoor environment, and particularly the level of greenery, were not clear. When an email address was available, we attempted to determine the nature of the environment from the author. We briefly outline these studies.

Adach et al. (2000) compared heart rate, lactic acid concentrations and running speed in 12 athletes at anaerobic threshold and during maximal effort during a run in a treadmill test in a laboratory and in an outdoor field environment. Bassett et al. (1985) compared treadmill running with a run along a road through farmers' fields [informed by email from the author] in order "to compare the rate of oxygen consumption for overground hill running and inclined treadmill running, at identical speeds and grades"[quote from email]. McMurray et al. (1998) compared beta-endorphins, lactate concentrations, heart rate and oxygen uptake in a 10-mile run on a laboratory treadmill and a 10-mile run outdoors; the type of outdoor environment was not clear. Hassmen (1996) compared the effects of a cycle ergometer test performed in the laboratory with an outdoor run along a park path, measuring perceived exertion and heart rate during the run in order to investigate the environmental effects on perceived exertion. Similarly Ceci and Hassmen (1991) compared treadmill running with a run along an outdoor track, measuring heart rate, lactate and perceived exertion as part of a study investigating the self-regulation of exercise intensity.

These studies may provide useful information to further understand the mechanisms of, or variation in, health benefits of getting out and about in nature. For instance, differential effects of activity in a natural environment could be explained by variation in terrain and complexity of the landscape, representing a greater physical challenge than simply running on a treadmill. However, it is likely we have only identified a small sample of these types of studies given that their hypotheses were not focused on the natural environment element of the study and thus their titles/abstracts may not have contained any of our natural environment search terms. Thus, a more focused search for these studies would need to be devised to be able to review and synthesize these studies in a comprehensive manner.

5. Discussion

5.1 Evidence of effectiveness

5.1.1 Study findings

Our review identified 23 studies which could provide tests of whether an activity in a natural environment was more or less beneficial than an activity in a more urban or synthetic environment. The studies measured a wide range of different health or well-being outcome measures, which limited the ability to synthesize data together in a formal meta-analysis. Meta-analyses could be conducted on several physiological parameters such as blood pressure and pulse, as well as hormone concentrations such as cortisol, based on the few studies available measuring each one. In none of these cases was there any indication of a differential effect of a natural environment compared to an outdoor built or indoor environment. However, analyses on self-reported moods/emotions did provide some evidence of a positive health benefit of contact with nature; there was some indication that displeasurable emotions such as anger, fatigue and sadness were lower after exposure to a natural environment compared to a different environment, and also some indication that pleasurable emotions such as tranquillity could be greater after exposure to a natural environment compared to an outdoor built environment.

5.1.2 Methodology of studies

The strength of evidence depends on the number of studies available as well as the methodological rigour of these studies. Most of these studies used a crossover design, although several were comprised of control and intervention groups, and experimentally manipulated exposure to different environments. Most studies collected data before and after exposure to the different environmental settings, and therefore allowed investigation of any base-line differences. A broad range of participants have been studied, however, college students, adult males, and physically active adults were the most common type, and participants were usually volunteers.

There were a number of drawbacks to the studies. Several of the studies were not specifically testing the hypothesis that nature has positive effects. Thus, in some cases, there were other differences between comparison groups apart from environmental settings, which may have biased estimation of the effect size. In some instances, due to different hypotheses, the natural environment being compared included environments such as green paths on a university campus and thus whether the environment was sufficiently green to provide a test of the effect of nature is not clear. Ideally, we would have compared the effects of different types of 'natural environments' covering a range of extents of green space.

The most common type of study outcome was self-reported measures of different moods or emotions. However, the validity of these scores/indices as a general measure of well-being was not clear, at least based on the information presented by authors in their articles. Given these data were self-reported, they were therefore potentially open to bias depending on prior beliefs of the participants. The blinding of participants to the research question in these studies is problematic as the hypothesis could be guessed by participants based on the study design. Thus, findings may have been affected by participants' pre-test opinions/beliefs on the likely effects of

different environments rather than any actual changes in their mental health or well-being.

A final issue with the current studies available is that most studies were focused on very short-term effects of different environments, making assessments of the participants shortly before and after the activity, and in some cases, during the activity. The studies were mostly not aiming to investigate the effects of large scale public health interventions. The longer-term implications of exposure to different environments cannot therefore be fully assessed. This clearly is a practical problem as longer-term studies may be more costly and less easy to control, in terms of exposing participants to different environments, than a study which takes place over the course of a day. However, we did identify one longer-term study, which compared 10 week programmes of leisure-centre based activities with 10 week programmes of outdoor walking, demonstrating that this sort of trial is feasible; in this case, the findings do not provide any evidence of a differential effect of each environment (Isaacs et al. 2007).

5.2 Reasons for variation in effectiveness

Because of the limited number of studies we could synthesize in any single meta-analysis, we only investigated the effect of two reasons that might explain variation in effectiveness among different studies. The first reasons we investigated was the effect of the comparator type. In only one case did we identify a difference: the effect of a natural environment on tranquillity was more positive when compared to an outdoor built environment than when compared to an indoor environment. However, investigation of comparator type was hindered by the low number of studies and for some outcomes only one comparator type was available. The second reason we explored was whether any variation in effect size could be due to study quality score. There was little indication this was important but this also was hindered by low study numbers. This only represented a coarse attempt to study the implications of study methodology on the effect size. Ideally, if more studies were available, the effect of specific aspects of study methodology would be investigated rather than an overall quality score.

Some studies tested the effects of particular factors in their investigation on the effects of different environments. These include: differences between males and females; differences between recreational and competitive runners; pre-treatment task or no task (i.e. degree of mental fatigue); different stimuli while in the indoor environment; the activity alone or with a friend and the type of built environment being compared. This list illustrates some of the variables which have been hypothesised to be important and potentially lead to variation in the effect of an environment. At present, investigation of the impact of these variables would be based on few studies and therefore we do not pursue their effects further here.

5.3 Review limitations

Search: Because of the large number of articles found during the search, the inclusion criteria had to be applied in a strict manner, which may have led to the early exclusion of relevant material.

Quality assessment and data extraction: Ideally two reviewers would have assessed the quality of all papers and extracted the necessary data for analysis as a check on repeatability. In this review, only one reviewer extracted data from papers and therefore it is susceptible to associated biases. However, the method of quality assessment was piloted on a sample of articles by two reviewers and any discrepancies discussed, in order to minimise opportunities for inconsistency in the way information was extracted. In addition, the numeric data were presented in a straightforward manner in most papers, which limits the scope for variation in extraction.

Availability of data: Some data sets were not obtainable either because we could not identify a contact email address for the author or because the author no longer had the data. In addition, we were not able to take into account the correlation between pre and post test scores and calculate the effect size as a standardised difference gain score, which would have been more appropriate. Greater sample size and this method of calculation of effect size may have increased our power to detect effects.

6 Review conclusions

6.1 Implications for policy

Can we promote activity in nature as beneficial for health?

Our review focused on the evidence for specific health benefits of nature. Thus, we did not aim to address whether activity programs in nature are effective *per se* as this could simply be an investigation of the effects of activity rather than the effects of nature. Rather, we focused only on studies which compared activity in a natural environment to a different, non-natural environment, in order to explore the specific effect of the setting.

Based on self-reported measures of moods or emotions, usually measured shortly following exposure to each environment, there was some indication that an activity in nature could have more positive effects than similar activities in a different environment; displeasurable emotions such as anger, fatigue and sadness were lower after exposure to a natural environment compared to a different environment and pleasurable emotions such as tranquillity could be greater after exposure to a natural environment compared to an outdoor built environment. However the reliability and validity of these scores or indices as a general measure of well-being and the longer-term implications on health and general well-being are not clear. In addition, for some outcomes, there was some indication of heterogeneity in the effect size. Clearly it would be important to have a greater understanding of the factors generating this variation in order to be able to target specific activities in specific settings towards attainment of these outcomes. Studies employing robust qualitative research methods may be important to this understanding. Although we did identify such studies in our initial search, the lack of a comparator in their design (which may have not have been relevant given the type of questions being investigated by these studies) meant they could not be included in this review.

Another issue is that of generalisability of the study findings to the wider general population. Policy makers should be wary of translating the findings of studies which may have questionable representativeness of their specific target populations, and which have only been studied in specific natural settings, for defined indicators, into generalised statements of universal benefits.

There was little consistent evidence of any ‘added value’ of direct exposure to a natural environment for any other type of outcome apart from self-reported moods. However, it is important to note that the lack of evidence does not necessarily mean that there is no effect; rather that a consistent effect is not apparent from the studies included in this review.

6.2 Implications for research

How should we test for direct health benefits of nature?

Testing for direct health benefits of nature is problematic given the variety of aspects of nature and the ways in which it might impact on health. The factor that limited the inclusion of many studies was the necessity of a comparator group allowing comparison of the effectiveness of a particular activity in nature with that in a different environment. However, what would constitute the most appropriate comparator is debatable and differences between a natural environment and an

alternative environment could be arguably due to factors of the alternative environment rather than those of the natural environments. Investigation of the effect of different natural environments with a range of alternative environments could aid in understanding the specific attributes of environment that are important.

Other study designs may be used to provide some supportive evidence of the effectiveness of green space, for instance, correlational studies that relate the amount of green space with individual and population parameters have inferred positive benefits of green environments (e.g. de Vries et al. 2003). However, in these cases, it can be difficult to attribute causality; even if there is an attempt to control for potentially confounding variables in analysis, it cannot be assured that all appropriate factors have been measured and included.

Different hypotheses can be proposed on how the natural environment might affect human health or well-being. Clarification of the specific hypothesis under investigation could aid in designing better studies and tease apart the mechanisms involved. One hypothesis may be that a natural environment provides a particular set of sensory stimuli, which allows mental restoration (Kaplan & Kaplan 1989). A few studies included in our review measured outcomes reflecting 'restorativeness' but this was not a commonly measured outcome and therefore was not synthesized by meta-analysis. We did not include in our review studies which compared pictures of nature versus pictures of a built environment as we were interested in the evidence for the effectiveness of getting out into a natural environment rather than simply viewing it. However, such studies have been conducted, and these and other studies which manipulate sensory stimuli, may provide some indication of the sensory impacts of nature, and any associated health and well-being benefits. A second hypothesis is that nature allows recovery from stress (Ulrich 1983) which could be investigated further with more studies measuring physiological outcomes. A third hypothesis that has been proposed to explain any association of nature with well-being is that humans might have an innate preference for natural environments, arising as a product of past evolutionary pressures (the Biophilia hypothesis; Wilson 1984). Testing such a hypothesis is clearly difficult, however, exploring links between environmental preferences and well-being may be a first step. Qualitative research methods may also prove useful in identifying potential mechanisms and the importance of the context in determining the effect of nature. However, in reality, the links between these different hypotheses mean that separating their relative importance will be difficult.

A final and alternative hypothesis is that nature simply promotes health-enhancing behaviour. For instance, the types of activities that occur in a natural environment, and not in indoor or built environments, may be particularly beneficial for health or well-being. Alternatively, a natural environment may encourage the initiation and sustainment of physical activity, for instance jogging through a park, which in turn has beneficial impacts on physical fitness and health. Under this hypothesis, nature does not necessarily have a special benefit for health itself but rather promotes health through preferences for particular environments and activities.

Overall, the studies suggest that the role of natural environments in public health deserves further investigation. There is considerable scope to investigate further the effects of nature on health and refine and test the specific hypotheses discussed in this

review. In particular, we would suggest that study of the longer-term consequences of repeated exposures to natural environments, the context-dependence of any effects and variation in effects among people warrant particular attention.

7. Acknowledgements

We thank Dave Stone (Natural England) and the stakeholder group for feedback on the review protocol. Thanks also to: Jane Teas, Jo Barton & Terry Hartig for kindly supplying data.

8. Potential conflicts of interest and Source of support

None declared. This review was funded by Natural England.

9. References

Here we list the background references, please see Appendix G for the references of articles that met the inclusion criteria of the review.

- Bird, W. (2007). *Natural Thinking*. RSPB, UK.
- Cooper, H., Hedges, L. (1994) *The handbook of research synthesis*. Russell Sage Foundation, New York.
- CRD. (2001). *Undertaking systematic reviews of research on effectiveness*, CRD Report 4 (2nd ed). National Health Service Centre for Research and Dissemination, University of York, York
- De Vries, S., Verheij R.A., Groenewegen, P.P. & Spreeuwenberg, P. (2003) Natural environments-healthy environments? An exploratory analysis of the relationship between greenspace and health. *Environ. Plann. A* 35: 1717-1713.
- Effective Public Health Practice Project. (1998). *Dictionary for the Effective Public Health Practice Project Quality Assessment Tool For Quantitative Studies*
- Greenland S (1994) Quality scores are useless and potentially misleading - Reply to Re – a Critical-Look at Some Popular Analytic Methods. *American Journal of Epidemiology* 140: 300–301
- Juni P, Witschi A, Bloch R, Egger M (1999) The hazards of scoring the quality of clinical trials for meta-analysis. *Jama-Journal of the American Medical Association* 282: 1054–1060
- Kaczynski, A.T. and Henderson, K.A. (2007). Environmental correlates of physical activity: A review of evidence about parks and recreation. *Leisure Sciences*, 29: 315-354.
- Kaplan, R. & Kaplan, S. (1989) *The experience of nature: a psychological perspective*. Cambridge University Press, Cambridge.
- Khan, K. S., Kunz, R., Kleijnen, J. & Antes, G. (2003). *Systematic reviews to support evidence-based medicine*. Royal Society of Medicine Press Ltd., London
- Ogilvie, D., Fayter, D., Petticrew, M., Sowden, A. Thomas, S., Whitehead, M., Worthy, G. (2008) The harvest plot: a method of synthesizing evidence about the differential effects of interventions. *BMC Med Res Method* 8:8.
- Pullin, A.S. and Knight, T.M. (2001). Effectiveness in Conservation Practice: Pointers from Medicine and Public Health. *Conservation Biology* 15: 50-54.
- Pullin, A. S. and Stewart G. B. (2006). Guidelines for systematic review in conservation and environmental management. *Conservation Biology*, 20: 1647-1656.
- RMNO (2004) *Nature and health. The influence of nature on social, psychological and physical well-being*. Publication no. 2004/09. The Hague: Health Council of the Netherlands and Dutch Advisory Council for Research on Spatial Planning, Environment and Nature.
- Russell, J.A. (2003) Core affect and the psychological construction of emotion. *Psychological Review* 110: 145-172
- Stevens, A., and R. Milne. (1997) The effectiveness revolution and public health. In: G. Scally (ed) *Progress in public health*. Royal Society of Medicine Press, London, pp 197-225.
- Sackett, D.L. & Rosenberg, W.M.C. (1995) The need for evidence-based medicine. *Journal of the Royal Society of Medicine* 88: 620-624.
- Thomas, B.H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing*, 1: 176-184
- Wilson, E.O. (1984) *Biophilia*. Harvard University Press, Cambridge, MA.
- Yudkin, P. and Stratton, I.M. (1996) How to deal with regression to the mean in intervention studies. *Lancet*, 347, 241-3

Appendix A: Project Stakeholders

Organisation	Project contact/s
Biological Sciences, University of Essex	Jo Barton
BRASS ESRC Centre, Cardiff University	Julie Newton
British Heart Foundation, University of Oxford	Charlie Foster
BTCV	Calum MacIntosh Yvonne Trchalik
CABE	Helen Beck
Countryside Council for Wales	James Parkin
DEFRA	John Robbins
Department of Health, England	Mala Rao
Environment Agency	Roger Milne Anthea Hawke
Forest Research	Liz O'Brien
Forestry Commission	Marcus Sangster
Glasgow Centre for Population Health	Fiona Crawford
Health Protection Agency	Richard Jarvis
Institute for Rural Health	Jane Randall-Smith
MRC Southampton University	Stephen Holgate
National Public Health Service for Wales	Andrew Jones Malcolm Ward
Natural England	Kevin Charman William Bird
NICE	Hugo Crombie
North East Community Forests	Janine Ogilvie
Offender Health, Department of Health, England	Sara Moore
RSPB	Martin Harper
Scottish Executive	George Morris
Sustainable Development Commission	Larissa Lockwood
Sustainable Development Research Network/PSI	Kate McGeevor Ben Shaw
UK Public Health Association	Angela Mawle UKPHA Health & Sustainable Environments Special Interest Group
Wales Centre for Health	Sue Mably
Welsh Assembly Government	Ronnie Alexander

Appendix B: Website search

The websites of the following organisations were searched for publications.

British Heart Foundation (incl. National Centre for Physical Activity and Health)
British Trust for Conservation Volunteers
Canadian Health Network
Commission for Architecture and the Built Environment
Centre for Child and Family Research
Countryside Recreation Network
Countryside Council for Wales
Department of Health
Environment agency
EU Cost Action E39
European Environment Agency
Forest Research
Forestry Commission
Glasgow Centre for Population Health
Greenspace (including Greenspace Scotland)
Groundwork
Health Development Agency
Health Protection Agency
Health Technology Assessment
Institute of Rural Health
Living Streets
Mental Health Foundation
Mind
Medical Research Council
National Library for Public Health
National Public Health Service for Wales
National Parks
National Trust
Natural England
The new economics foundation
NICE
OPENspace
Parks Victoria
Royal Society for the Protection of Birds
Scottish Executive
Sustainable Development Commission
Sustainable Development Research Network
Scottish Environment Protection Agency
Scottish Natural Heritage
SNIFFER
Sustrans
THRIVE
The Nature Conservancy
The Woodland Trust
UK MAB Urban Forum
UK National Research Register Archive

UK Public Health Association
World Health Organisation
Wildlife Trusts
World Wide Fund for Nature

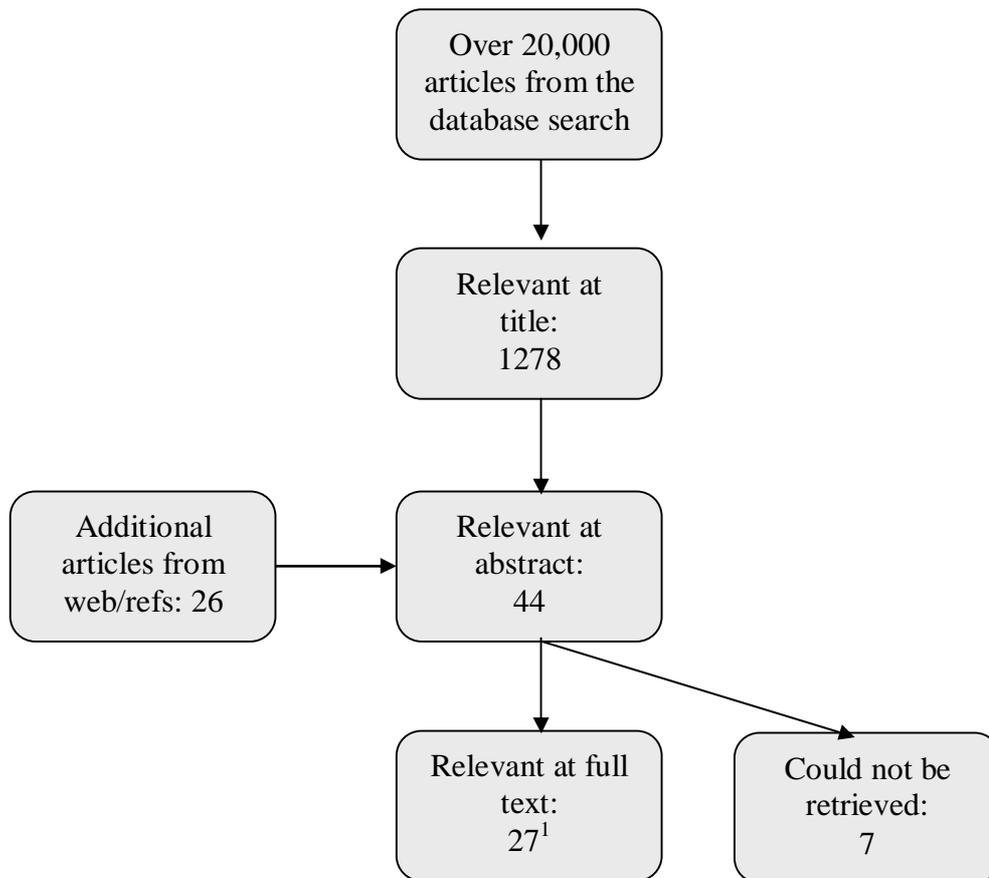
Appendix C: Methodology assessment

This section provides details on study methodology that were extracted from each article. Criteria used in the quality score are shaded. The remaining criteria are used to provide more general information of the articles in the summary tables.

Aspect of methodology	Types of answer
Was the target population of the study described?	Either inclusion criteria of participants specified or the study gives detail of the participants (e.g. age, physical activity experience) that took part in the study
Was the recruitment strategy likely to capture a comprehensive sample of the population?	Detail of recruitment method e.g. organisations contacted, advertisements.
Self-selection bias of participants	Were participants referred from an independent third party (e.g. GP) or did participants self-select?
What was the basic design of the study?	Experiment or Observational
Did the study use randomisation to allocate participants to treatments or order of treatments?	This was assessed on the basis of the use of the word 'random' or similar in the methodology. We did not restrict to only those describing the method of randomisation; only one study did this.
Were groups shown to be similar at baseline?	This was assessed on whether this data was presented/investigated or whether this was accounted for in the analysis e.g. analysis of time x environment interaction.
Was the hypothesis under investigation the effect of nature?	The hypothesis under investigation was based on either an explicit statement or from the discussion of literature in the introduction from which it could be indicated if positive impacts of nature were hypothesized.
Were the participants informed of the hypothesized positive impacts of nature?	This clearly depended on whether the authors themselves were interested in the positive impacts of nature (some were interested in the effects of simply the outdoors rather than the green outdoors). This was assessed on this basis, and whether there was a statement indicating what the participants had been informed about the study.
Was it likely that there were withdrawals?	This was based on any explicit statements of the number of individuals that started the study who also completed. Most

	studies were short-term before and after studies and therefore attrition of participants was in most cases not deemed important.
Were the tools used or method of data collection shown to be reliable measures of the outcome of interest?	For well-being outcomes using scores, this was assessed on the basis of an explicit statement that reflects on the validity/reliability of the score. For other measures e.g. blood pressure, a detailed description of the methodology was deemed sufficient.
Was the intervention fully described?	This was assessed through detail provided on the intervention exposed to participants.
Were there any other differences between nature and comparator treatments apart from environmental setting?	This depended on detail provided by the authors. Some authors were not investigating the effect of nature and therefore there were additional treatments (e.g. external stimuli when running on an indoor treadmill).

Appendix D: Flow chart of study inclusion



¹28 studies from 27 articles

Appendix E: Characteristics of studies

The following tables show the basic groupings of studies (by participants, outcomes, methodology, green space and type of activity, Tables 1 to 5 respectively), and therefore indicate the generalisability of the review to situations of interest. We focus on comparator types 1 (nature versus outdoor built) and 2 (nature versus indoor), which comprised 23 studies. We note that the number of studies reported simply reflects the number of studies identified as relevant and does not take into account the validity of data collection or whether the data could be incorporated into our meta-analysis. In addition, the total number does not necessarily equal the total number of studies, as in some cases studies could fall into more than one category and in one case the same two studies were included in both pools.

Table 1. Participants: the number of studies split by subset of the human population under study

Subgroup	Population	Green vs Urban	Green vs Indoors	No. of studies
Active individuals	Experienced backpackers	1	0	1
	Active/regular runners	2	3	5
Children	with ADD/ADHD	3	2	3
	with impaired vision	0	1	1
	at kindergarten	0	0	1
One sex only	females	1	2	3
	males	4	2	6
Adults with health problems	with CVD risk	0	1	1
	with 'mental retardation'	0	1	1
Other	College students	5	3	8
	Postmenopausal women	0	1	1
	Members of Mind groups	0	1	1

Table 2. Outcomes: the number of studies split by broad type of outcome under study

Type of outcome	Green vs Urban	Green vs Indoor	Total
1. Emotions/Moods	7	8	15
2. Concentration	7	2	7
3. Cardiovascular	3	4	7
4. Hormonal	4	2	6
5. Immune function	2	0	2
6. Physical activity	0	1	1
7. Performance	0	1	1
8. Other	2	3	5

Examples of each type of outcome:

1= revitalisation, tranquillity, anger, fear, attentiveness, happiness, fatigue, placidity, pride, delight etc....etc.....

2= Digit Span test; Proof reading task; Symbol Digits Modalities Test; ADHD symptoms.

3 = Heart rate, Blood pressure,

4 = Adrenaline (from urine), Salivary cortisol, Salivary amylase, Noradrenaline (from urine), Cortisol (from urine)

5 = NK activity, granulysin, perforin, granzymes, immunoglobulin A concentration

6 = Self reported Physical activity

7 = Oxygen consumption, Running Speed, Number of minutes cycled etc...

8 = sleeping hours, cerebral brain activity (absolute hemoglobin concentrations of the prefrontal cortex!), motor performance; engagement, various others (1 x total cholesterol, lipid subfractions, lung function, skin resistance etc...)

Table 3. Methodology: the number of studies split by the type of methodology used in the study

Methodology	Green vs Urban	Green vs Indoor	Total number
Randomised controlled trial	2	4	6
Randomised crossover trial	4	0	4
Counter-balanced crossover trial	2	3	5
Crossover trial (same order)	1	2	3
Observational	3	4	5

Table 4. Green space: The number of studies split by the type of green space under study

Type of green space	Green vs Urban	Green vs Indoor	Total number
University campus	0	5	5
Garden	0	1	1
Public Park/ Country Park	3	3	6
Nature reserve/ Wilderness	3	0	3
'Forest'	4	1	5
Outdoors – not clear how green	2	3	3

Table 5. Activity: The number of studies split by the type of activity in green space under study

Type of activity in green space	Green vs Urban	Green vs Indoor	Total number
Sitting in/passive activity	0*	1	1
Walking	7	5	12
Running	2	2	4
Backpacking	1	0	1
Gardening	0	1	1
Mixed	2	4	4

*3 studies compared the effects of nature by allowing participants to walk in a natural environment in the morning, and then sit and watch nature in the afternoon; this was repeated in a city environment. As the data from these comparisons are not independent, we have focused on the first parts of the study when participants were walking in each environment and therefore placed them in the walking group.

Appendix F: Confidence intervals of effect sizes

Our review focuses on the difference between outcomes after each environmental setting (i.e. nature versus non-nature). However, any difference does not necessarily reflect the change in outcome in each environment (e.g. a positive difference between nature and non-nature does not necessarily mean that overall there has been a positive before and after change with the activity). Therefore, we present here analysis of the before and after differences of outcomes in the natural environment. Note that we do not interpret these effects as direct impacts of nature but the combined effects of the activity and the environment.

The results show that energy increased while anxiety, anger, fatigue and sadness decreased, when comparing measurements on individuals before and after the activity. However, for a number of variables such as attention, tranquillity, blood pressure, pulse and cortisol concentrations, there were no evident changes before and after the activity in the natural environment.

Table a) Before and after differences for individuals in ‘nature’ groups (aggregated data)

Outcome	Effect size	Lower CI	Upper CI	No. studies	Summary
Attention	-0.069	-0.557	0.418	2	No effect
Energy	0.757	0.298	1.215	5	Increase [·]
Anxiety	-0.518	-0.791	-0.246	6	Decrease
Tranquillity	0.05	-0.434	0.534	7	No effect ^{**}
Anger	-0.354	-0.635	-0.073	6	Decrease
Fatigue	-0.757	-1.109	-0.406	4	Decrease
sad/depressed	-0.661	-1.158	-0.163	3	Decrease
systolic bp	-0.102	-0.459	0.254	4	No effect
diastolic bp	-0.111	-0.493	0.270	3	No effect
Pulse	0.309	-0.231	0.849	4	No effect [*]
cortisol conc.	-0.570	-1.571	0.432	4	No effect ^{**}

Superscripts on ‘Summary’ indicate the significance of heterogeneity
[·] = <0.1; ^{*} = <0.05; ^{**} = <0.01

The following table present the effect sizes and confidence intervals that are presented in the results section

Table b) Effect size and 95 % CI for unadjusted post test effect sizes

Outcome	Unadjusted means on disaggregated data ¹					Unadjusted means on aggregated data ²				
	Effect size	Lower CI	Upper CI	No. points	Summary	Effect size	Lower CI	Upper CI	No. studies	Summary
Attention	0.217	-0.029	0.464	7 (4)	No effect	0.264	-0.049	0.578	4	No effect
Energy	0.328	0.068	0.588	8 (5)	Positive	0.28	-0.007	0.567	5	No effect
Anxiety	-0.032	-0.412	0.348	10 (6)	No effect**	-0.122	-0.582	0.339	6	No effect*
Tranquillity	0.23	-0.138	0.597	11 (7)	No effect**	0.391	-0.078	0.86	7	No effect**
Anger	-0.456	-0.737	-0.18	11 (7)	Negative	-0.465	-0.712	-0.218	7	Negative
Fatigue	-0.453	-0.841	-0.07	7 (4)	Negative*	-0.418	-0.762	-0.073	4	Negative
sad/depressed	-0.497	-0.839	-0.16	5 (4)	Negative	-0.399	-0.719	-0.079	4	Negative
systolic bp	-0.015	-0.201	0.171	8 (6)	No effect	0.011	-0.192	0.213	6	No effect
diastolic bp	0.009	-0.217	0.236	6 (5)	No effect	0.058	-0.176	0.293	5	No effect
Pulse	-0.002	-0.191	0.187	5 (5)	No effect	-0.002	-0.191	0.187	5	No effect
cortisol conc.	-0.047	-0.493	0.398	5 (4)	No effect	-0.026	-0.576	0.525	4	No effect

Superscripts on 'Summary' indicate the significance of heterogeneity

· = <0.1; * = <0.05; **=<0.01

¹Disaggregated data = multiple effects size in studies with more than one comparison

²Aggregated data = one average effect size per study to account for potential non-independence

The following two tables present the results of the analysis adjusting post-test effect size by the pre-test effect size to investigate the effect of accounting for any base-line differences.

*Table c) Effect sizes and 95 % CI based on **disaggregated and adjusted** post-test data (posttest effect size adjusted by pre effect size). The p-value of the heterogeneity statistic (Q) is also presented.*

Outcome	Overall					Nature vs outdoor built					Nature vs indoor				
	Effect Size	Lower CI	Upper CI	Q (p)	No. points (studies)	Effect Size	Lower CI	Upper CI	Q (p)	No. points (studies)	Effect size	Lower CI	Upper CI	Q (p)	No. points (studies)
Attention	0.11	-0.238	0.457	0.09	7 (4)	0.11	-0.238	0.457	0.09	7 (4)	NA	NA	NA	NA	0 (0)
Energy	0.147	-0.089	0.382	0.64	8 (5)	0.043	-0.38	0.466	0.596	3 (2)	0.193	-0.091	0.477	0.434	5 (3)
Anxiety	-0.309	-0.719	0.101	<0.001	10 (6)	NA	NA	NA	NA	0 (0)	-0.309	-0.719	0.101	<0.001	10 (6)
Tranquillity	0.278	-0.177	0.734	<0.001	11 (7)	0.804	0.175	1.433	0.049	5 (4)	-0.078	-0.634	0.479	<0.001	6 (3)
Anger	-0.66	-1.053	-0.266	<0.001	11 (7)	-0.573	-1.094	-0.052	0.028	5 (3)	-0.758	-1.391	-0.125	<0.0001	6 (4)
fatigue	-0.722	-1.34	-0.104	<0.001	7 (4)	NA	NA	NA	NA	0 (0)	-0.722	-1.34	-0.104	<0.001	7 (4)
sad/depressed	-0.461	-0.765	-0.158	0.681	5 (4)	-0.313	-0.812	0.186	NA	1 (1)	-0.548	-0.931	-0.166	0.624	4 (3)
systolic bp	-0.008	-0.194	0.177	0.749	8 (6)	-0.003	-0.297	0.292	0.457	4 (3)	-0.012	-0.252	0.227	0.647	4 (3)
diastolic bp	0.07	-0.14	0.28	0.367	6 (5)	-0.007	-0.315	0.3	0.361	4 (3)	0.094	-0.312	0.5	0.203	2 (2)
pulse	0.068	-0.207	0.342	0.233	5 (5)	-0.348	-1.31	0.614	0.062	2 (2)	0.146	-0.064	0.356	0.743	3 (3)
cortisol conc.	0.311	-0.698	1.321	<0.001	5 (4)	0.758	-2.228	3.743	<0.0001	2 (2)	0.089	-0.357	0.536	0.377	3 (2)

Table d) Effect sizes and 95 % CI based on **aggregated and adjusted** post test data (posttest effect size adjusted by pre effect size). The p-value of the heterogeneity statistic (*Q*) is also presented.

Outcome	Overall					Nature vs outdoor built					Nature vs indoor				
	Effect size	Lower CI	Upper CI	Q (p)	No. studies	Effect size	Lower CI	Upper CI	Q (p)	No. studies	Effect size	Lower CI	Upper CI	Q (p)	No. studies
Attention	0.167	-0.307	0.642	0.104	4	0.167	-0.307	0.642	0.104	4	NA	NA	NA	NA	0
Energy	0.145	-0.142	0.433	0.778	5	0.023	-0.433	0.479	0.786	2	0.226	-0.144	0.596	0.538	3
Anxiety	-0.378	-0.821	0.065	0.023	6	NA	NA	NA	NA	0	-0.378	-0.821	0.065	0.023	6
Tranquillity	0.446	-0.096	0.987	0.001	7	0.927	0.175	1.68	0.031	4	-0.068	-0.68	0.543	0.051	3
Anger	-0.598	-0.851	-0.345	0.397	7	-0.548	-0.903	-0.193	0.711	3	-0.691	-1.162	-0.22	0.143	4
Fatigue	-0.684	-1.466	0.098	0.002	4	NA	NA	NA	NA	0	-0.684	-1.466	0.098	0.002	4
Sad/ Depressed	-0.434	-0.754	-0.114	0.868	4	-0.313	-0.812	0.186	NA	1	-0.519	-0.935	-0.102	0.845	3
Systolic bp	0.004	-0.198	0.206	0.845	6	0.003	-0.344	0.349	0.716	3	0.005	-0.244	0.254	0.506	3
Diastolic bp	0.086	-0.145	0.318	0.345	5	0.001	-0.4	0.401	0.284	3	0.094	-0.312	0.5	0.203	2
Pulse	0.068	-0.207	0.342	0.233	5	-0.348	-1.31	0.614	0.062	2	0.146	-0.064	0.356	0.743	3
Cortisol conc.	0.437	-0.798	1.671	<0.001	4	0.758	-2.228	3.743	<0.0001	2	0.175	-0.411	0.761	0.266	2

Appendix G: References and summary tables of included studies

Citation	Activity/Comparator	Participants	Aspects of methodology	Outcomes measures
Bodin, M. and T. Hartig (2003). "Does the outdoor environment matter for psychological restoration gained through running?" <i>Psychology of Sport and Exercise</i> 4(2): 141-153.	Running through an urban area vs large nature reserve for 60 mins	12 regular runners (6 male, 6 female)	<ul style="list-style-type: none"> - Pretest-posttest crossover trial - Participants from local running clubs and referrals - Treatment order randomised. - 7 day period between each run - Runners familiar with predefined route - Assessments at the start and end of the run 	Exercise-Induced Feeling Inventory (revitalization and tranquillity) Negative Mood Scale (anxiety/depression and anger) Attention (Digit Span test and Symbol Digits Modalities) Perceived restorativeness, Daily Hassles scale
Butryn, T. M. and D. M. Furst (2003). "The effects of park and urban settings on the moods and cognitive strategies of female runners. Effets des environnements naturels (parc) et urbains sur l' humeur et les strategies cognitives de coureuses feminines." <i>Journal of Sport Behavior</i> 4: 335-355.	Running through an urban area vs public park along a 4 mile course	31 non-elite female runners	<ul style="list-style-type: none"> - Pretest-post test crossover trial - Participants recruited from local running clubs and flyers posted in a campus fitness facility - Treatment order counterbalanced - Minimal experience with predefined routes and they ran separately - Participants not informed of research hypothesis - Assessments at the start and end of run 	Mood (Total Mood Disturbance) Exercise-Induced Feeling Inventory (Revitalization, Tranquillity and Positive Engagement) Also presents Thoughts During Running Scale (TDRS) Other outcomes measured (e.g. POMS) but a comparison between park and urban not discussed.
Cuvo, A. J., M. E. May, et al. (2001). "Effects of living room, Snoezelen room, and outdoor activities on stereotypic behavior and engagement by adults with profound mental retardation." <i>Research in Developmental Disabilities</i> 22(3): 183-204.	Activity in living and Snoezelen (multisensory) rooms compared to the outdoor area of the institution (in a rural area)	3 adults with profound mental retardation living at a state operated developmental center based on a number following criteria	<ul style="list-style-type: none"> - observations made of participants in each environment -order of environments were counter-balanced, and experienced consecutively for 15-20 mins on each research day -living room was also first and served as the base-line -there were different activities in different environments and depended on participant preference 	Idiosyncratic symptoms (e.g. body rocking, swaying, picking) and also level of engagement

			- reliability of assessment by observation checked by comparing inter-observer agreement	
Faber Taylor A, Kuo F, and Sullivan W (2001) Coping with ADD. The Surprising Connection to Green Play Setting. Environment and Behaviour. 33 Jan 2001 pp54-77	Afterschool and weekend activity in indoor, built vs green settings	96 parents of children aged 7 to 12 yrs old formally diagnosed ADD ADHD	- Observation study using a questionnaire to assess symptoms after activities - Parents recruited through flyers - Questionnaire was pretested. - Completed questionnaire which asked them to rate the effects of various activities on their child's symptoms (worse, the same, or better than usual)	ADHD symptoms as rated by their parents
Faber-Taylor, A. & F.E. Kuo. (2008) Children with attention deficits concentrate better after walk in the park. Journal of Attention Disorders.	Walk in a city park, vs through urban downtown or residential for 20 mins	17 children (15 boys, 2 girls) with ADD or ADHD	- Posttest crossover trial - Participants were children between 7 to 12 with formally diagnosed ADD or DHD - Randomised treatment order - Each walk was one week apart - Prior to walk, children completed puzzles to induce attentional fatigue - Led individually on walk - Post test following walk by assessors blind to treatment. - Extra 8 children withdrawn for various reasons	Attention (Digit Span Backwards) Other attention measures taken but not used for various reasons. Perceived experience of run
Harte, J. L. and G. H. Eifert (1995). "The effects of running, environment, and attentional focus on athletes' catecholamine and cortisol levels and mood." Psychophysiology 32(1): 49-54.	Indoor running on a treadmill in laboratory with either (1) external stimuli(e.g. listened through headphones, sound of wind, cars, people walking, birds); (2) internal stimuli (wore earphones connected to a microphone on their chest to hear breathing) compared to a 12-km outdoor run on the university campus (and also a control 'quiet inactivity' group)	10 male amateur triathletes	-pretest-posttest cross over trail -order of treatments counterbalanced -Participants obtained after sending letters out to several running clubs -1 person dropped out due to work commitments -psychological assessments and blood pressure were pretested before the activity but baseline urine samples for endocrine assessment were collected on mornings when the subjects were not attending the experiment but at the same time of day -posttest assessment taken following the activity	POMS(tension/anxiety, depression/rejection, anger/hostility, confusion/bewilderment, vigor/activity, fatigue/inertia), perceived exertion, cortisol, blood pressure, adrenaline and noradrenaline
Hartig, T., Evans, G.W., Jamner, L.D., Davis, D.S., & Gärling, T. (2003).	Room with tree view and nature walk vs viewless room/urban	112 normotensive students	- Pretest-posttest control and intervention - Block-randomisation to assign to groups	Systolic/Diastolic blood pressure

Tracking restoration in natural and urban field settings. Journal of Environmental Psychology, 23, 109-123	walk for 50 mins		<ul style="list-style-type: none"> - Led individually on walk - First pretest before arriving at the site, and posttest following walk - Half of the subjects had completed an attention-demanding task prior to the room/walk 	<p>ZIPERS (positive affect, attentiveness, fear arousal, anger/aggression)</p> <p>Overall Happiness Scale Attention (Necker Cube Pattern Control, Search and Memory test)</p>
Hartig, T., M. Mang, et al. (1991a). "Restorative effects of natural environment experiences." Environment and Behavior 23(1): 3-26.	Non-wilderness vacation (car touring, visits to family and sightseeing trips) vs extended wilderness backpacking trip (also 'no vacation' group)	Experienced, physically-fit backpackers (25 in the wilderness group and 18 in the non-wilderness group)	<ul style="list-style-type: none"> - Observation study with assessments before and after the vacation already planned by participants - Participants in the wilderness group were recruited at trailheads; other groups were recruited from local clubs and referrals - Both 4 to 7 days in duration - Some differences between groups 	<p>ZIPERS (fear, anger/aggression, sadness, positive affect, attentiveness)</p> <p>happines (OHS), life satisfaction, attention (proof-reading task)</p>
Hartig, T., M. Mang, et al. (1991b). "Restorative effects of natural environment experiences." Environment and Behavior 23(1): 3-26.	Walking in an urban area vs a park (also vs relaxation conditions) for 40 mins	34 college students (17 men, 17 women)	<ul style="list-style-type: none"> - Pretest-posttest crossover trial - Random assignment to groups - Standardized induction of mental fatigue prior to experiment - Led individually on walk - Main pretest in a different location before arrival at site but posttest at the end of walk 	<p>ZIPERS (fear, anger/aggression, sadness, positive affect, attentiveness)</p> <p>Overall Happiness Scale Attention (proof-reading task)</p> <p>Perceived restorativeness</p> <p>Blood pressure</p> <p>Pulse</p>
Hull, R. B. and S. E. Michael (1995). "Nature-Based Recreation, Mood Change, and Stress Restoration." Leisure Sciences 17(1): 1-14.	Passive leisure activities at home vs in a suburban park	20 (10 men and 10 women) drawn from first part of study on college students	<ul style="list-style-type: none"> - Observational study with assessments before, during and after leisure - Park visitors were approached to take part in the study - A subset were asked to take part in the home study when carrying out similar activity - Assessment made by questionnaires - 26 were asked to take part in the home part of the study but only 20 returned completed questionnaires 	<p>Thayer's 4-point scale (calm, energetic, tired, anxious, stress)</p>
Isaacs AJ, Critchley JA, See Tai S, Buckingham K, Westley D, Harridge SDR, et al. Exercise Evaluation Randomised Trial (EXERT): a	Leisure-centre based exercise program vs instructor-led walking programme through parks and open space (also 'advice only')	943 adults aged between 40 and 74 yrs, not currently physically active and with at least one cardiovascular risk	<ul style="list-style-type: none"> - Randomised controlled trial - Referrals from GP to invite participants - Block randomisation to assign participants to groups 	<p>Exercise performance, blood pressure, cholesterol, lipid subfractions, heart rate, lung function, anthropometry,</p>

randomised trial comparing GP referral for leisure centre-based exercise, community based walking and advice only. Health Technology Assessment 2007: 11(10).	group) on 10-wk programme	factor	<ul style="list-style-type: none"> - 10 week programmes, 2 to 3 times a week for both exercise groups - Assessments and questionnaires at 10wk, 6 months and 1 yr 	flexibility, strength, power, self-reported physical activity, HADS (anxiety, depression), SF-36 Medical Outcome Scale, Stages of Change, Barriers to Exercise Scale, EuroQol, Decision Balance Scale, health service usage
Kerr, J. H., H. Fujiyama, et al. (2006). "Psychological responses to exercising in laboratory and natural environments." Psychology of Sport and Exercise 7(4): 345-359.	Running on a laboratory treadmill vs along a tree-lined university footpath for 5km	22 male recreational runners and 22 male competitive runners	<ul style="list-style-type: none"> - pre and posttest crossover design - Order of activity counter-balanced - ran individually, maintaining similar intensity by pacemaker - each run separated by 1 week -Assessment made before and after activity 	TESI (relaxation, anxiety, excitement, boredom, placidity, anger, provocativeness, sullenness, pride, humiliation, modesty, shame, gratitude, resentment, virtue, guilt, external and somatic tension and effort
Kuo and Taylor. (2004). A potential natural treatment for attention-deficit/hyperactivity disorder: evidence from a national study. American Journal of Public Health, 94: 1580-1586.	Afterschool and weekend activity in indoor, built vs green settings	452 parents of children aged 5 to 18 yrs old who had been formally diagnosed with ADHD	<ul style="list-style-type: none"> - Observation study using a questionnaire to assess symptoms after activities - Parents recruited through adverts - Completed questionnaire which asked them to rate the effects of various activities on their child's symptoms (worse, the same, or better than usual) 	ADHD symptoms as rated by their parents
Li, Q., K. Morimoto, et al. (2008). "Visiting a forest, but not a city, increases human natural killer activity and expression of anti-cancer proteins." International Journal of Immunopathology and Pharmacology 21(1): 117-127.	Walking during a 3 day trip to city vs 3 day trip to forest	12 health males (11 in city group)	<ul style="list-style-type: none"> - Pretest-post test crossover trial - Participants selected from four large companies - All received same order of treatment (forest and city trip at different times of year) - Not clear if individuals walked together as a group - Pretest was 3 days before trip while post test was taken on day 2 and day 3 (and day 7 and 30 for forest groups only) 	Immune function (NK activity, number of NK cells, white blood cells, number of T cells, granulysin, perforin, granzymes A/N) Adrenaline Sleeping hours
Park, B. J., Y. Tsunetsugu, et al. (2007). "Physiological effects of Shinrin-yoku (taking in the atmosphere of the forest)--using salivary cortisol and cerebral	walking (morning) and watching (afternoon) in a day in a city vs a day in the forest for 20 mins each time	12 healthy males	<ul style="list-style-type: none"> - Pretest-posttest crossover trial - Divided into two groups randomly to decide treatment order -Experienced each environment on 	Cerebral activity (absolute haemoglobin concentrations of the prefrontal cortex) Salivary cortisol

activity as indicators." J Physiol Anthropol 26(2): 123-8.			consecutive days - Activity in the environment alone. - Pretest in the morning at accommodation before arriving at site, and also just before and after each activity at site, and final posttest on returning to the place of accommodation	Perceived comfort/calm
Peacock, J., Hine, R. & Pretty, J. (2007). The mental health benefits of green exercise activities and green care. Mind week report	Walk around a shopping centre vs in a country park for 1h	20 members (7 male, 13 female) of local Mind associations aged between 31 and 70	- pretest - posttest crossover study - Same order of activity for all participants (all walked together for both groups) - walks led by a MIND organizer - each walk was separated by 1 week - Continuous walking was preferred and a certain level of social interaction encouraged	Rosenberg-Self esteem scale, POM (anger, confusion, depression, fatigue, tension, vigour), TMD (mood)
Plante, T. G., C. Cage, et al. (2006). "Psychological Benefits of Exercise Paired With Virtual Reality: Outdoor Exercise Energizes Whereas Indoor Virtual Exercise Relaxes." International Journal of Stress Management 13(1): 108-117.	Walk outdoors within a green university campus vs on a treadmill watching a virtual reality video of the same walk (and 'virtual reality video only' group) for 20 mins	112 (47 male and 65 female) psychology students	- pretest-posttest comparison groups - Participants volunteered for the study to fulfil a research participation requirement for their course - Randomisation to assign participants to groups - walking at similar speeds, not clear if they were alone -Assessment before and after the activity	AD-ACL(energy, calmness, tension, tiredness), PACES (physical activity enjoyment) and Marlowe Crowne Social Desirability Scale
Plante, T. G., C. Gores, et al. (2007). "Does Exercise Environment Enhance the Psychological Benefits of Exercise for Women?" International Journal of Stress Management 14(1): 88-98.	Walk on a treadmill in the university fitness facility vs in a green university campus (and also with a friend or alone) for 20 mins	88 female undergraduate students aged 18 to 22 yrs old	- pretest-posttest comparison groups - Randomisation to assign students to groups - walking at a moderate speed - in both cases, they were tested alone and with a friend -Assessments before and after the activity	AD-ACL(energy, calmness, tension, tiredness) and PACES (physical activity enjoyment)
Scholz, U. and H. Krombholz (2007). "Untersuchung zur koerperlichen Leistungsfahigkeit von Kindern aus Waldkindergaerten und Regelkindergaerten. A study of the physical performance ability of children from wood kindergartens and from regular kindergartens. Enquete sur la	Attendance of a forest kindergarten compared to a regular kindergarten	129 children in total - 45 children from 10 wood kindergartens, 42 children from 2 urban kindergartens and 42 children from 2 rural kindergartens aged at least 60 months with no physical problems	- observational study investigating on the implications of movements in a natural environments for motor ability -tests developed by the Institute of Early Education in Munich -no base-line data before attending wood kindergarten Measurements taken on motor ability as	Sport motor performance ability (balancing forwards, backwards, jumps, swinging etc..)

condition physique d'enfants frequentant des jardins d'enfants forestiers et de ceux qui frequentent des jardins d'enfants reguliers." Motorik Mar 1: 17-22.			well as the familial environment (parents occupation/family activities) and other activities outside of kindergarten for each child	
Teas, J., T. Hurley, S. Ghumare ; K. Ogoussan. (2007) Walking Outside Improves Mood for Healthy Postmenopausal Women. Clinical Medicine : Oncology. 1: 35-43	Walking indoors on a treadmill in a gym vs along a university path (grassy area with trees, flower beds) for 1 hr	19 self-reported postmenopausal status (no menstruation for at least 12 months), regularly engaged in recreational exercise at 3 hrs/week as well as answering 'no' to all questions on the Physical Activity Readiness Questionnaire (PAR-Q)	<ul style="list-style-type: none"> - pretest - posttest crossover study - Recruited from adverts and word of mouth - Same order of activity for all participants (all walked outside on the same day) - walks separated by 1 week - Instructed to walk at a similar pace - walked together - the gym environment also had music 	Positive and Negative affect scale (pleased, sad, frustrated, worried, angry, delighted, joy) Salivary cortisol and alpha amylase
Telles, S. and R. B. Srinivas (1998). "Autonomic and respiratory measures in children with impaired vision following yoga and physical activity programs." International Journal of Rehabilitation & Health 4(2): 117-22.	Yoga vs physical activity in the garden for 3 wks	24 children with impaired vision aged between 12 to 17 yrs	<ul style="list-style-type: none"> - pretest-posttest comparison groups - Participants selected from a special school for persons with visual impairment - Randomisation to assign participants to activities - 3 weeks of the activities, for 1 hr, 5 days a week - Assessment before and after this 3-week period 	Heart rate, breath rate and skin resistance
Tsunetsugu, Y., B. J. Park, et al. (2007). "Physiological effects of Shinrin-yoku (taking in the atmosphere of the forest) in an old-growth broadleaf forest in Yamagata Prefecture, Japan." J Physiol Anthropol 26(2): 135-42.	walking (morning) and watching (afternoon) in a day in a city vs a day in the forest for 15 mins each time	12 male university students	<ul style="list-style-type: none"> - Pretest-posttest crossover trial - Divided into two groups randomly to decide treatment order - Experienced each environment on consecutive days - Carried out activity alone in the environment - Pretest in the morning at accommodation before arriving at site, and also just before and after each activity at site, and final posttest on returning to the place of accommodation 	Systolic/diastolic blood pressure Pulse Heart rate Salivary cortisol Immunoglobulin A Perceived comfort/calm Stress-Refresh feeling test
Yamaguchi, M. Deguchi, M. & Y Miyazaki (2006). The effects of exercise in forest and urban environments on sympathetic nervous activity of normal	walking (morning) and watching (afternoon) in a day in a city vs a day in the forest for 20 mins each time	10 male university students	<ul style="list-style-type: none"> - Pretestpost test crossover trial - Divided into two groups to determine treatment order - Not clear if the walking alone. 	Salivary amylase

young adults. Journal of International
Medical Research 34: 152-159.

- Pretest in the morning at accommodation
before arriving at site, and also just before
and after each activity at site, and final
posttest on returning to the place of
accommodation

Other studies on exercise performance

Adach, Z., M. Zakrzewski, et al. (2000). "Zachowanie sie wybranych wskaźników fizjologicznych podczas testu o wzrastającej intensywności wykonywanego w laboratorium i w warunkach terenowych. Changes of chosen physiological parameters during increasing intensity test in laboratory and outdoor conditions." *Wychowanie Fizyczne i Sport* 1: 3-11.

Bassett Jr, D. R., M. D. Giese, F.J.Nagle, A.Ward, D.M. Raab, B. Balke. (1985) "Aerobic requirements of overground versus treadmill running." *Med Sci Sports Exerc* 17: 477-81

Ceci and Hassmen. 1991. Self-monitored exercise at three different RPE-intensities in treadmill versus field running. *Medicine and Science in Sports and Exercise*, 6: 732-738

Hassmen, P. (1996). "Environmental effects on ratings of perceived exertion in males and females. Effets experimentaux sur les indices de perception de l'effort chez des hommes et des femmes." *Journal of Sport Behavior* Aug 3: 235-245.

McMurray, R. G., M. J. Berry, et al. (1988). "The effect of running in an outdoor environment on plasma beta endorphins." *Annals of Sports Medicine* 4: 230-233.