The effects of physical activity on cortisol and sleep: A systematic review and meta-analysis

Len De Nys\textsuperscript{a,}\textsuperscript{*}, Kerry Anderson\textsuperscript{a}, Esther F. Ofosu\textsuperscript{a}, Gemma C. Ryde\textsuperscript{b}, Jenni Connelly\textsuperscript{a}, Anna C. Whittaker\textsuperscript{a}

\textsuperscript{a} Faculty of Health Sciences and Sport, University of Stirling, Scotland, UK
\textsuperscript{b} Institute of Cardiovascular and Medical Sciences, University of Glasgow, Scotland, UK

ARTICLE INFO

Keywords:
Physical exercise
Endocrine health
Hypothalamic-Pituitary-Adrenal axis
Sleep hygiene
Sleep quality
Cortisol

ABSTRACT

Background: Managing stress and having good quality sleep are inter-related factors that are essential for health, and both factors seem to be affected by physical activity. Although there is an established bidirectional relationship between stress and sleep, remarkably few studies have been designed to examine the effects of physical activity on cortisol, a key biomarker for stress, and sleep. Research is particularly scarce in older people despite both sleep and cortisol changing with age. This systematic literature review addresses this gap.

Methods: A systematic review was conducted following the PRISMA guidelines. Original, peer-reviewed records of intervention studies such as randomized controlled trials (RCTs) and non-RCTs with relevant control groups were eligible for inclusion. The Participant, Intervention, Comparison, Outcome (PICO) characteristics were (1) adults or older adults (2) physical activity programmes of any duration, (3) controls receiving no intervention or controls included in a different programme, (4) cortisol measurement, and subjective or objective measures of sleep.

Results: Ten original studies with low-to-moderate risk of bias were included. Findings from this review indicated with moderate- and low-certainty evidence, respectively, that physical activity was an effective strategy for lowering cortisol levels (SMD [95% CI] = 0.37 [−0.52, −0.21] \( p < 0.001 \)) and improving sleep quality (SMD [95% CI] = 0.30 [−0.56, −0.04], \( p = 0.02 \)). Caution is needed to generalize these findings to the general population, as included trials were predominantly participants with breast cancer, included few males and no older adults.

Conclusion: Cortisol regulation and sleep quality are intertwined, and physical activity programmes could improve both in several ways. Further, physical activity may benefit adults with long term conditions or current poor (mental) health states the most, although more research is needed to support this claim fully. Few intervention studies have examined the inter-relationship between cortisol and sleep outcomes in males or older adults, indicating fruitful enquiry for future research.

1. Introduction

1.1. Background and rationale

Regular physical activity, managing stress and having good quality sleep are known to be significant contributors towards healthy ageing (World Health Organization, 2015). Recent studies have shown that physical exercise can have a positive impact on physical and mental health (Penedo and Dahn, 2005). Further, it moderates stress systems (Anderson and Wideman, 2017; Duclos and Tabarin, 2016; Fragala et al., 2011) and positively affects sleep quantity and quality (Kredlow et al., 2015; Uchida et al., 2012; Vanderlinden et al., 2020; Wang and Boros, 2019). Despite this, remarkably few studies have examined the combined effect of physical exercise on both stress and sleep.

Cortisol is an adrenal hormone released in response to stress (Ockenberg et al., 1995). It displays strong circadian rhythmicity, with high levels in the morning and a steady decline towards the evening (Adam and Kumari, 2009; Stalder et al., 2016). This diurnal fluctuation is indicative of hypothalamic-pituitary-adrenal (HPA) axis reactivity (Smyth et al., 1997). A favourable cortisol profile is characterised by a
brief peek in the morning in the first 30–45 min after awakening (known as the CAR). A gradual decline follows this peak during the waking day to reach a low point around midnight (Kirschbaum and Hellhammer, 1989; Pruessner et al., 1997). Studies show that a daily cortisol curve with a typically sharp decline is associated with better physical and psychosocial health (Adam et al., 2006; Adam and Kumari, 2009). However, an accepted sampling design for cortisol involves, e.g., measurements immediately after awakening, 30-min post-awakening, at noon, in the late afternoon, and immediately prior to bed (Hellhammer et al., 2007). The known diurnal rhythmicity of cortisol (Adam and Kumari, 2009; Stalder et al., 2016), the significant intra-individual differences (Coste et al., 1994; Pruessner et al., 1997) and the significant variability between people in the shape of their diurnal cortisol rhythms (Adam et al., 2006; Smyth et al., 1997) highlight a need for measurement consistency in research in order to compare different study findings (Dickerson and Kemeny, 2004; Ryan et al., 2016). Cortisol is often used as a biomarker to indicate dysregulation of the HPA axis, triggering subsequent poor sleep and fatigue (Bower et al., 2005; Buckley and Schatzberg, 2005b), further promoting negative health states such as depression (Juster et al., 2010; Lupien et al., 2007), cancer recurrence (Sephton et al., 2013; Sephton et al., 2000) (Sapolsky, Kraemer and Spiegel, 2000), and is associated with all-cause and cardiovascular mortality (Kumari, Shipley, Stafford, and Kivimaki, 2011; Phillips et al., 2010).

Further, while sleep is an absolute necessity for human health (Grandner, 2017), sleep problems are prevalent in the global population (Léger, Poursain et al., 2008). Sleep problems are associated with poorer quality of life and several comorbidities, such as cardiovascular problems, cancer, mental and physical health issues, and contribute to all-cause mortality (Irwin, 2015; Krieger et al., 2001; Mallon et al., 2002; Simon and VonKorff, 1997). This evidence highlights the health importance of understanding how cortisol levels and sleep can be improved.

Numerous studies have investigated the reciprocal interactions between the HPA axis and sleep regulation (for a review, see, e.g. (Nollet, Wisden, and Franks, 2020; Steiger, 2002)). First, the HPA axis is implicated in sleep regulation and sleep/wake cycles (Buckley and Schatzberg, 2005b; Ono and Yamanaka, 2017; Pawlyk et al., 2008). Second, sleep hygiene and circadian rhythmicity are proven to impact cortisol profiles through a decreased efficacy of the negative-feedback regulation of the HPA axis (Spiegel et al., 1999). Sleep disruption increases stress by triggering the HPA axis and dysregulating the production of one of the main adrenal stress hormones, cortisol (Kim et al., 2015; Wright et al., 2015). In turn, stress increases sleep disruption. As both lead to downstream effects on health (Ono and Yamanaka, 2017), an integrative approach to understanding how these factors interact with each other deserves research attention.

In summary, there is established health importance of, and a bidirectional relationship between, cortisol and sleep. Both seem to be moderated by physical exercise. However, few attempts have been made to systematically summarize these data. Consequently, this systematic review aimed to assess the effects of physical exercise on the physiological stress marker cortisol and sleep outcomes.

1.2. Objectives

A systematic review and meta-analysis were carried out to investigate the impact of physical exercise interventions on stress, measured by cortisol levels, and sleep, measured either subjectively or by examination of physiological sleep architecture. It further sought to examine whether there was a difference in participant characteristics (such as in different age groups, gender or health status) or in intervention characteristics (type or duration of physical activity), and whether there was a relationship between cortisol markers and sleep outcome changes.

2. Methods

2.1. Eligibility criteria

Original, peer-reviewed articles in the English language of intervention studies, such as randomized controlled trials (RCTs), non-RCTs with relevant control groups, were eligible for inclusion. No date limits were chosen. The Population, Intervention, Comparison and Outcome (PICO) characteristics (Richardson et al., 1995) for eligibility were: (1) adults and older adults, regardless of their health condition (having a sleep complaint was not required); (2) physical exercise intervention programmes or movement-based mind-body approaches of any duration; (3) controls receiving no intervention (e.g., wait-list or usual care), or controls included in a different programme and; (4) physiological measure of stress (cortisol in saliva, blood, hair or urine samples) and measurements of sleep, either subjective (questionnaires such as Pittsburg Sleep Quality Index) or objective (architecture of sleep such as polysomnography, actigraphy, and accelerometry). No filters in search databases were used. Studies were ineligible if outcomes of interest were not measured, or because the results for the outcome of interest were not reported.

2.2. Information sources and search strategy

Searches were run in August 2021 and re-run before the final analyses in October 2021. The following electronic bibliographic databases were searched: PubMed, the Cochrane Library (the Cochrane Central Register of Controlled Trials (CENTRAL)), PsycINFO, OvidSP, CINAHL and Web of Science (no data limits were chosen). Grey literature searches were conducted searching online databases (ClinicalTrials.gov) and using the Google Scholar search engine according to the recommendations of Haddaway (Haddaway et al., 2015). Reference lists of key papers were searched and cross-references manually to supplement initial keyword searches. The exact search strategy used with suitable search terms for each database is in Appendix A.

2.3. Selection process

Search results were collected in Sciwheel software (sciwheel.com). Two reviewers independently screened each title and abstract for eligibility using Rayyan software (rayyan.qcri.org) (Ouzzani et al., 2016). Any study identified by either reviewer was included for further screening. Full-text screening of selected records was performed by the same two reviewers, and any disagreements were resolved by discussion and consensus. A third reviewer was available when no consensus was reached.

2.4. Data collection process

One reviewer (LDN) collected data from each selected record, carefully checked by the second reviewer (KA). Disagreements were resolved by discussion between the two reviewers, with a third reviewer overviewing the process. Study investigators were contacted to obtain missing information (n = 6). Data extraction from figures or graphs was conducted with WebPlotDigitizer (Rohatgi, 2021). When multiple records corresponded to a single study, the record most relevant to the review question was used to extract data of interest.

2.5. Data items

Data extraction was performed using Excel, collecting the following data items: (1) study ID, (2) design, (3) PICO characteristics, including different age groups (adults aged 26–47 years, middle-aged adults aged 48–64 years, and older adults aged 65 years or over), different exercise interventions using Frequency, Intensity, Time, Type and Duration (FITT-D) components, cortisol measurements (saliva/blood/other), and
subjective (Pittsburgh Sleep Quality Index (PSQI)/other), or objective (polysomnography/other) sleep measurement, (4) general findings and (5) statistics relevant to the research question. Where multiple subjective sleep measurements were reported, we selected one outcome (PSQI) for inclusion in the meta-analysis and for reporting main outcomes, as this questionnaire was most frequently used.

2.6. Study risk of bias assessment

To ascertain the validity of eligible studies, the two reviewers independently assessed the risk of bias to determine the adequacy of randomization, concealment of allocation, blinding of participants, personnel or data collectors, reliability and completeness of outcome data or selective reporting and an overall summary ‘Risk of Bias’ (RoB) judgement (low, some concerns, high), where the overall RoB for each study was determined by the highest RoB level in any of the domains that were assessed for both outcomes. The Cochrane Risk of Bias (RoB) 2.0 tool (Higgins et al., 2011) was used to assess RCTs. There were no non-RCTs retrieved in the review process, so no bias assessment for non-RCTs was performed.

2.7. Effect measures

Continuous measures were displayed for each outcome using mean differences (MD) and 95% confidence intervals (CI) for studies using the same scale, and standardized mean differences (SMD) and 95% CI were used to compare the same outcomes measured in different ways.

2.8. Synthesis methods

No non-RCT studies with relevant control groups were retrieved, so only RCTs were synthesized. Study characteristics were tabulated and sorted by outcome for consistency. The risk of bias in included studies was visualized with the ‘roblivis’ software (McGuinness and Higgins, 2021). Included study data were found similar enough in terms of methodological and clinical characteristics to ensure meaningful conclusions from a statistically pooled result, so meta-analyses were performed for both the cortisol and sleep outcomes. Assuming a true effect size was not the same in all studies, and that studies were performed in different populations, random-effects models to analyse data were performed. The data were based on mean, standard deviation (SD) and the number of participants assessed for both the intervention and comparison groups and used to calculate the SMDs and 95% in Review Manager 5 (The Cochrane Collaboration, 2020). These were visualised by forest plots, and any data conversions or transformations of the reported data can be found in Appendix C. The degree of heterogeneity was assessed through Chi-squared (Chi$^2$) statistics and was quantified and interpreted using the I-squared (I$^2$) statistic. To explore heterogeneity and in convergence with the research question, sub-group analyses were conducted to differentiate between age groups, gender and health status, if there were enough studies identified to make relevant comparisons. RCTs not included in the meta-analysis were synthesised narratively. If sufficient studies showed a correlation between cortisol and sleep outcomes, effect sizes based on correlation were measured for these studies (as predefined in the protocol, https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=272251).

A sensitivity analysis was conducted by making a chart with the characteristics of the retrieved trials that were relevant to the review question but did not have a control group in their design (non-RCTs), and by comparing them narratively with results of the primary analysis. Second, an effect direction plot was performed with the retrieved RCTs and non-RCTs that reported the sleep outcomes in multiple ways, based on existing guidance (Boon and Thomson, 2021). Third, intervention effects of RCTs in the meta-analysis were sub-grouped by relevant control groups (active control, usual care or waiting list) for both cortisol and sleep outcomes.

2.9. Reporting bias assessment

A funnel plot (Egger’s test) (Egger et al., 1997) was assessed to evaluate small-study effects, together with visual inspection for asymmetry (Appendix G). If asymmetry was detected, trial characteristics, protocols (if available), or methods and results sections of the trial publications would be reviewed to assess whether the asymmetry was due to publication bias or other factors, such as methodological or clinical heterogeneity of the trials. The potential impact of missing results was explored in the sensitivity analyses.

2.10. Certainty assessment

The certainty of all evidence was assessed as high, moderate, low, or very low, using the GRADE approach (Guyatt et al., 2008), for the cortisol and sleep outcomes. A starting rating of ‘High quality’ evidence was downgraded by one level if serious concerns (or by two levels for very serious concerns) became apparent in terms of risk of bias, inconsistency; indirectness, imprecision, or publication bias. GRADE wording was used to incorporate certainty of evidence (Guyatt et al., 2011). The standardized statements to describe results are based on existing consensus (Glenton et al., 2010).

3. Results

3.1. Study selection

A total of 4143 records were retrieved through database searches. After duplicate removal, 3412 records remained, from which 103 were retrieved for full-text screening. Finally, 12 reports of 10 original studies were included (Fig. 1) (Al-Sharman et al., 2019; Arikawa et al., 2013; Banasik et al., 2011; Bowden et al., 2012; Chandwani et al., 2010; Chen et al., 2013; Hilcove et al., 2021; Ho et al., 2016; 2018; Imboden et al., 2021; Payne et al., 2008; Ratcliffe et al., 2016). Reports were mainly excluded due to wrong PICO characteristics, e.g., (1) participants were children, adolescents, or young adults (18–25 years old), (2) interventions were exercises which were not movement-based (mindfulness or meditation only) (3) outcome measures were stress and sleep, but the stress component was not measured through cortisol samples. A table with characteristics of studies that were excluded after full-text screening (de Bruin, Formsma, Frijstein, and Börgels, 2017; Fouladbaksh et al., 2014; Garrido et al., 2017; Grahn Kronhed et al., 2020; Leone et al., 2018; Passos et al., 2014; (Baron et al., 2010); Vera et al., 2009; Zaccari et al., 2020), with reasons for exclusion, is in Appendix B.

3.2. Study characteristics

All included studies were RCTs (n = 10), of which one was an RCT comparing three different interventions. Key characteristics of each study are presented in Table 1 and a chart summary was made to summarize all study characteristics (Fig. 2).

3.3. Risk of bias in studies

Some concerns in terms of the overall risk of bias arose in six of the 10 studies due to domain four, ‘Bias due to outcome measurement’ (Fig. 3). Subjective questionnaires may be subject to recall or performance bias, however, all included questionnaires were valid and reliable, as standardised (a priori-defined) sleep measurement tools.

3.4. Results of individual studies

The summary statistics, effect estimate, and precision of each study are presented in a forest plot (Figs. 4 and 5). For the articles by Payne and Arikawa (Arikawa et al., 2013; Payne et al., 2008), the mean SDs had to be computed or estimated from other information (Appendix C).
3.5. Results of syntheses

The 10 RCTs included 756 participants (90% females), with mainly adults (27–64 years old), and no older adults (65+ years old). Most studies included females with breast cancer (n = 6) and used saliva samples (n = 7). The discrepancy of participants’ characteristics across studies was visualized in diagrams (Appendix D). Five studies measured the diurnal cortisol slope, one measured the cortisol awakening response (CAR) and one performed a one-point in time measurement to measure cortisol, and the Pittsburgh Sleep Questionnaire Index (PSQI) (n = 8) to measure sleep. A chart overview of study characteristics was made (Fig. 2). Four studies were of low risk of bias, whereas six arose some concerns, owing to not having an active control group. Knowledge of the assigned intervention could have influenced participant-reported outcomes in the subjective sleep questionnaires. However, all questionnaires used were validated for each of the participant groups. Further, many confounding factors could be at play in cortisol measurement. However, each of the studies used standardised (a priori-defined) protocols for the cortisol sampling, and randomisation strategies for control groups, to account for most confounding factors.

Nine RCTs directly used cortisol and PSQI outcome measures with relevant control groups (Al-Sharman et al., 2019; Arikawa et al., 2013; Banasik et al., 2011; Bowden et al., 2012; Hilcove et al., 2021; Ho et al., 2016; Imboden et al., 2021; Payne et al., 2008; Ratcliff et al., 2016), and were deemed clinically and methodologically similar enough to be included in the meta-analysis. One study was deemed not appropriate to be included in the meta-analysis, because it lacked a relevant control group, as it compared the effects of a brain wave training programme to a yoga and a mindfulness group.

Physical exercise was associated with reduced cortisol levels (SMD [95% CI] = −0.37 [−0.52, −0.21] p < .001) and improved PSQI outcomes (SMD [95% CI] = −0.30 [−0.56, −0.04], p = .02) compared to active controls or usual care among 651 and 660 (middle-aged) adults, respectively. With heterogeneity present in the PSQI outcome (Chi^2 = 18.82 with df = 8), the quantified heterogeneity was deemed moderate (I^2 = 57%). The moderate heterogeneity could be explained by excluding studies with the highest SMDs, without changing the overall effect. Further, although studies were deemed similar enough, clinical differences, e.g., differences between the health status of participants, could also explain the heterogeneity. No RCTs reported a correlation between cortisol and sleep outcomes, but there was one non-RCT that reported that decreases in cortisol were significantly correlated with
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Population</th>
<th>Intervention</th>
<th>Outcome</th>
<th>Relevant findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payne et al. (2008)</td>
<td>South-eastern United States</td>
<td>Middle-aged adults</td>
<td>Moderate-Aerobic, walking activity</td>
<td>Cortisol (times - measure) Blood sample (1x)</td>
<td>Training group — improved PSQI scores*</td>
</tr>
<tr>
<td>Banak et al., 2010</td>
<td>Washington, United States</td>
<td>Middle-aged adults</td>
<td>Mind-body, yoga</td>
<td>Fatigue Likert scale</td>
<td>Yoga group — lower cortisol at 5 pm and less fatigue</td>
</tr>
<tr>
<td>Bowden et al. (2012)</td>
<td>London, United Kingdom</td>
<td>Middle-aged adults</td>
<td>Moderate to vigorous - Aerobic, weight bearing exercise</td>
<td>Blood sample (1x - morning)</td>
<td>No differences between groups in cortisol or sleep outcomes</td>
</tr>
<tr>
<td>Arikawa et al. (2013)</td>
<td>Boston, United States</td>
<td>Adults</td>
<td>Depend</td>
<td>PSQI</td>
<td>Yoga group — less fatigue*</td>
</tr>
<tr>
<td>Chen et al. (2013)</td>
<td>Shanghai, China</td>
<td>Adults and middle-aged adults</td>
<td>Mind-body, Qigong</td>
<td>Psqi, BFI</td>
<td>Yoga group — reduced fatigue*, and steeper cortisol slope. No change in PSQI.</td>
</tr>
<tr>
<td>Ratcliff et al. (2016)</td>
<td>Texas, United States</td>
<td>Middle-aged adults</td>
<td>Mind-body, yoga</td>
<td>Psqi</td>
<td>Yoga group — improved sleep*</td>
</tr>
<tr>
<td>Ho et al. (2016)</td>
<td>Hong Kong</td>
<td>Middle-aged adults</td>
<td>Dance movement therapy</td>
<td>Psqi, Bfi</td>
<td>Yoga group — beneficial effect on cortisol slope</td>
</tr>
<tr>
<td>Al-Sharman et al. (2019)</td>
<td>Jordan</td>
<td>Multiple Sclerosis</td>
<td>Moderate - Aerobic, recumbent stepping machine</td>
<td>Psqi, ISI, actigraphy</td>
<td>Exercise group — improvements in PSQI*</td>
</tr>
<tr>
<td>Hilcove et al., 2020</td>
<td>South-western United States</td>
<td>Adults</td>
<td>Generally healthy</td>
<td>Item from PSQI</td>
<td>Exercise group — improved sleep*</td>
</tr>
<tr>
<td>Imboden et al. (2021)</td>
<td>Switzerland</td>
<td>Adults</td>
<td>Aerobic, indoor bicycle</td>
<td>Psqi, polysomnography</td>
<td>Both groups — lower cortisol and improved PSQI</td>
</tr>
</tbody>
</table>

Note. * Age groups: Adults: 26–47 years, Middle-aged adults: 48–64 years. * Marks significance in study <.05. CAR: Cortisol Awakening Response, BFI: Brief Fatigue Inventory, ISI: Insomnia Severity Index, PRFS: Piper Revised Fatigue Scale, PSQI: Pittsburgh Sleep Quality Index
increases in total sleep time and rapid eye movement (REM) sleep (Passos et al., 2014).

3.5.1. Sub-group Analyses

The magnitude of the direction of effect did not notably change among all sub-group analyses performed. Differentiating between gender was deemed not valuable as insufficient data was retrieved to assess males as a distinct subgroup. Also, none of the analyses explained the moderate statistical heterogeneity in the PSQI outcome (Appendix E). First, identified age groups were adults (26–47 years old) only (four studies (Al-Sharman et al., 2019; Arikawa et al., 2013; Hilcove et al., 2021; Imboden et al., 2021)), and adults and middle-aged adults (26–64 years old).
years old) (five studies (Banasik et al., 2011; Chen et al., 2013; Ho et al., 2018; Payne et al., 2008; Ratcliff et al., 2016)), sub-groups did not differ significantly ($p = .49$ for cortisol, and $p = .53$ for PSQI). No studies were performed on older adults (65 years and over). Second, cortisol measurement was done by measuring the diurnal cortisol slope (Banasik et al., 2011; Chen et al., 2013; Hilcove et al., 2021; Ho et al., 2018; Ratcliff et al., 2016), one point in time measurement (Al-Sharman et al., 2019; Arikawa et al., 2013; Bowden et al., 2012; Payne et al., 2008), or the CAR (Imboden et al., 2021). Sub-grouping between cortisol measurements revealed no significant differences ($p = .66$). Third, three studies used active control groups (Al-Sharman et al., 2019; Imboden et al., 2021; Ratcliff et al., 2016), while six studies used usual care or wait-list controls as comparison groups (Arikawa et al., 2013; Banasik et al., 2011; Chen et al., 2013; Hilcove et al., 2021; Ho et al., 2018; Payne et al., 2008), these two sub-groups did not differ significantly ($p = .89$ for cortisol, and $p = .73$ for PSQI). This analysis was also part of the sensitivity analysis (see below). Fourth, regarding intensity type, four studies used aerobic exercise (Al-Sharman et al., 2019; Arikawa et al., 2013; Imboden et al., 2021; J. K. Payne et al., 2008), and five studies used mind-body exercises (Banasik et al., 2011; Chen et al., 2013; Hilcove et al., 2021; Ratcliff et al., 2016). Again, these two sub-groups did not differ significantly ($p = .50$ for cortisol, and $p = .27$ for PSQI).

### 3.6. Sensitivity analyses

Sensitivity analyses were consistent with the main analysis giving a notion of the robustness of the synthesised results. Narrative analyses that included descriptive PICO characteristics of retrieved non-RCTs ($n = 5$) (Appendix F) showed similar study characteristics compared to the included RCTs. Notable differences were mostly in the ‘participant’ domain: one study included older adults (Fouladbakhsh et al., 2014), and more men were included compared to in the RCTs (30% vs. 10% respectively). Further, none of the studies looked at breast cancer, whereas six out of 10 RCTs did. A summary of other measures for sleep, including both RCTs and non-RCTs revealed seven other measurements of sleep, including four studies complementing a sleep questionnaire with measures of the architecture of sleep (polysomnography (Imboden et al., 2021; Leone et al., 2018) and actigraphy (Al-Sharman et al., 2019; Garrido et al., 2017) (Fig. 6).

#### 3.7. Reporting bias

A funnel plot by Egger revealed no visual asymmetry in the cortisol outcome, and an asymmetry in the sleep outcome due to missing studies in the area of ‘no intervention effect’ (Appendix G). Reporting bias is considered, however, this asymmetry could exist due to chance, as the analysis contained few studies with few participants (Sterne et al., 2011). Further, a sensitivity analysis that explored the impact of potential missing results in the main analysis, showed that pilot studies without control groups reported similar results compared to the included RCTs.

#### 3.8. Certainty of evidence

First, there was moderate-certainty evidence that physical exercise probably decreases cortisol levels slightly compared to active controls or usual care. Second, there was low-certainty evidence that exercise interventions may improve PSQI outcomes slightly compared to active controls or usual care. The evidence was downgraded one level for each of the outcomes due to an issue with indirectness, and another level in the PSQI outcome, due to moderate heterogeneity. The explained...
Further, patients often suffer from various side effects such as psychological distress or disruption in the HPA-axis, fatigue, and sleep disturbance (Ratcliff et al., 2016), derive greater benefit from psychological or behavioural interventions. Further, a recent meta-analysis of 61 trials evidenced that baseline distress moderates the efficacy of such interventions for cancer patients (Schneider et al., 2010). This suggests that physical exercise interventions may be most beneficial for adults with current poor (mental) health states.

As a priori described in the protocol, any statistical correlation between changes in cortisol and sleep outcomes in the identified records would be documented, but no RCTs reported this. However, it appears that women react differently to chronic stress compared to men, displaying larger increases in cortisol levels when chronically stressed (Kunz-Ebrecht et al., 2004; Schulz et al., 1998; Steptoe et al., 2000). The general picture that emerges is that there are significant gender differences for ACTH and free salivary cortisol (measured in saliva), but not for plasma cortisol (Kirschbaum et al., 1999; Kudielka et al., 2004; Kudielka, Brigitte M., & Kirschbaum, C., 2005). Further, the HPA axis is influenced by sex hormones, particularly estrogen (Gillies and McArthur, 2010), which can modify stress responsiveness via its regulation of cortisol receptors (Oldhinkel and Bouma, 2011). This evidence highlights the importance of using standardised methodological protocols to strictly distinguish between the total cortisol secretion and the bioavailable cortisol levels and to differentiate between pre- and post-menopausal states and menstrual cycle phases to draw conclusions.

A concurrent wide variability for both outcomes in the effectiveness of the physical exercise interventions is seen in most trials, as indicated by wider 95% CIs. In line with this, previous research suggests that cancer patients with higher distress or depressive symptoms (Andersen et al., 2010; Antoni et al., 2001; Danhauer et al., 2009), or elevated sleep disturbance (Ratcliff et al., 2016), derive greater benefit from psychological or behavioural interventions. Further, a recent meta-analysis of 61 trials evidenced that baseline distress moderates the efficacy of such interventions for cancer patients (Schneider et al., 2010). This suggests that physical exercise interventions may be most beneficial for adults with current poor (mental) health states.

Table 6: Summary of measures for sleep, including both RCTs and non-controlled trials (In colour please). Note. Study design: RCT: Randomised Controlled Trial; nCT: non-Controlled Trial. Outcome: PSQI: Pittsburg Sleep Quality Index; BFI: Brief Fatigue Inventory; ISI: Insomnia Severity Index; GSQ: Global sleep quality (item from PSQI), SSQ: Spiegel Sleep Questionnaire; PRFS: Piper Revised Fatigue Scale. Effect direction: upward arrow ▲ = positive health impact, downward arrow ▼ = negative health impact, sideways arrow ↔ = no change/mixed effects/conflicting findings. Sample size: final sample size (individuals) in intervention group. Preventive effects are shown by wide 95% CIs, whereas no effect is indicated by narrow CIs. Large arrow ▲ > 300; medium arrow ▲ 50-300; small arrow ▲ 50; Colours: Green: Low risk of bias, Orange: some concerns, Red: High risk of bias.

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Design</th>
<th>PSQI</th>
<th>BFI</th>
<th>ISI</th>
<th>Fatigue likert scale</th>
<th>GSQ</th>
<th>SSQ</th>
<th>PRFS</th>
<th>Polysomno-</th>
<th>Actigraphy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payne, 2008</td>
<td>RCT</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arikawa, 2010</td>
<td>RCT</td>
<td>▲</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banasik 2010</td>
<td>RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bowden, 2012</td>
<td>RCT</td>
<td></td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen, 2013</td>
<td>RCT</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chandwani, 2014</td>
<td>RCT</td>
<td></td>
<td>▲</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ho, 2016</td>
<td>RCT</td>
<td></td>
<td>▲</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al-Sharman, 2019</td>
<td>RCT</td>
<td>▲</td>
<td>▲</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hilcoe, 2020</td>
<td>RCT</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imboden, 2021</td>
<td>RCT</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fouladbakhsh, 201</td>
<td>nCT</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passos, 2014</td>
<td>nCT</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leone, 2015</td>
<td>nCT</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garrido, 2017</td>
<td>nCT</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zacari, 2020</td>
<td>nCT</td>
<td>▲</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As a priori described in the protocol, any statistical correlation between changes in cortisol and sleep outcomes in the identified records would be documented, but no RCTs reported this. However, one identified non-RCT in middle-aged adults with chronic insomnia, found that reductions in cortisol were significantly correlated with increases in total sleep time and rapid eye movement sleep (measured by polysomnography) after an aerobic exercise intervention of 12 weeks (Passos et al., 2014). In addition, previous literature also points to such an association: The conceptual model developed by Payne (2004) suggest that fatigue, sleep disturbances, and depressive symptoms may result from dysregulation of hormones such as cortisol (J. D. Payne and Nadel, 2004). Similarly, other studies in cancer research pointed out that sleep disturbance and fatigue were associated with disrupted cortisol rhythms (Schmidt et al., 2016; Tell et al., 2014). It seems that cortisol regulation and sleep quality are intertwined, and that physical exercise could improve both in several ways, but few intervention studies have
examined this, thus indicating a fruitful line of future research.

This systematic review found that all studies modulating cancer-related fatigue were on women with breast cancer (Arikawa et al., 2013; Banasik et al., 2011; Chen et al., 2013; Ho et al., 2018; J. K. Payne et al., 2008; Ratcliff et al., 2016). Limited published evidence on modulating cancer-related fatigue in males was also found in another systematic review and meta-analysis (Brown et al., 2011). Further, a review with similar findings to this review described in depth the neuroendocrine-immune responses to exercise and gender (Fragala et al., 2011), which could be of particular interest for researchers seeking to design future physical exercise interventions on cancer-related fatigue, differentiating between males and females.

Although the present research set out to differentiate between different age groups (adults aged 26–47 years, middle-aged adults aged 48–64 years, and older adults aged 65 years or over), the search retrieved no RCTs that included older adults, and only one pilot study that included nine older lung cancer survivors. This was surprising for the following reasons. First, age-related changes are associated with a dysregulation of the circadian rhythm of the HPA axis (Al-Turk and Al-Dujaili, 2016; Heaney et al., 2012), further leading to sleep problems (Van Cauter et al., 2000). Second, many of the sleep disturbances in ageing parallel the age-related changes in the HPA axis and cortisol rhythm, suggesting that the HPA axis may play a key role in age-related changes with sleep (Buckley and Schatzberg, 2005a). Third, there is an established link between sleep quality and physiological well-being in older adults (Hanson and Ruthig, 2012). Nevertheless, other trials in older adults did find improvements in quality of life, stress, and sleep outcomes after physical exercise interventions, with stress measured subjectively with questionnaires rather than with cortisol (Grahm Kronhed et al., 2020; Halpern et al., 2014; Innes and Selfe, 2012). Thus, quality research is lacking in measuring the effects of the joint effect of physical exercise interventions on cortisol and sleep in older adults. Future research should focus on this issue as well as the correlation between cortisol and sleep measures in this age group.

Another research question was to differentiate between different intervention types. The performed analysis did not find any significant sub-group differences, partially because the analysis yielded too few studies to make meaningful interpretations. However, regarding the cortisol outcome, a meta-analysis exploring the effect of different exercise interventions found that aerobic exercise in patients with major depressive disorder reduced cortisol levels the most compared to strength training (Beserra et al., 2018). Again, caution in interpretation is suggested due to the small number of studies, with substantial heterogeneity among them. Regarding sleep, reviews have shown positive effects on sleep quality across various exercise modalities (such as mind-body exercises and vigorous strength exercises) (Brupbacher et al., 2021), and levels of intensity (Lederman et al., 2019). Overall, more quality studies differentiating between different exercise types could deepen this current understanding.

4.1. Strengths and limitations

This review process carefully followed the PRISMA guidelines (Page et al., 2021). The screening and selection of articles, risk of bias assessment, and rating of the certainty of the evidence were done independently by two reviewers. Also, the data extraction performed by one reviewer was carefully screened by the second reviewer. Reporting bias was addressed by a funnel plot and sensitivity analysis and properly reported. This rigorous methodology allowed for greater confidence in the objectivity of the results. Further, a meta-analysis was performed to increase the precision of the effect estimates.

However, to interpret the findings appropriately, several limitations of the evidence included in this review need to be considered. First, this systematic review included a relatively small number of studies, only RCTs, mostly in breast cancer patients, including remarkably fewer males. In addition, it should be mentioned that positive outcomes are more likely to be published in English-language journals (Egger, Zellweger-Zähner et al., 1997), which could lead to selection bias in systematic reviews and meta-analyses. Therefore, although the findings agreed with other systematic reviews on the topic, claims made regarding the effectiveness of physical exercise programmes should be interpreted within these limits. Second, recall and/or performance bias for the studies including subjective outcome measurements (e.g., questionnaires or scales) should be considered. Objectivity would increase when these measurements are complemented with the architecture of sleep measures (e.g., polysomnography or actigraphy), however, this is often a cost/accuracy trade-off. Further, several limitations of the review process should also be reported. First, moderate heterogeneity was found when pooling the studies within the sleep outcome, complicating a meaningful summary. Therefore, the certainty was downgraded based on existing GRADE guidance (Schünemann et al., 2013), and findings were reported appropriately. Second, the pitfalls about sub-grouping were considered (Burke et al., 2015), therefore, sub-groups were mainly analysed to allow for transparency in addressing the research question and should be interpreted cautiously. Third, sensitivity analyses were mainly performed narratively. This could solely give an indication of the robustness of the synthesized results, rather than provide definitive conclusions.

4.2. Implications for practice

Findings from this review indicate that physical exercise is an evidence-based strategy within the holistic perspective to tackle stress and sleep complaints in adults, particularly in women with breast cancer. This adds to the current consensus that the benefits of physical exercise for people with long-term conditions outweigh the risks (Reid et al., 2021). Additionally, the results point out that light-to-moderate physical exercise, whether it may be aerobic or mind-body exercises, seem to be beneficial to modulate both stress and sleep. Given the noted wide variety in the effectiveness in the results and the established evidence that any physical exercise modality and intensity could improve these outcomes, it seems adults could benefit from individually tailored physical exercise interventions based on their preference and needs.

4.3. Implications for research

The generalizability of the findings of this review would be enhanced by high-quality studies including different health states, both biological genders and older adults. Previous research indicates that adults with current poor (mental) health states may benefit most from physical exercise interventions. Therefore, future research could hypothesize that baseline (mental) health states could moderate interventions effects. More specifically, the understanding of the value of physical exercise for cancer-related fatigue would benefit from the design of physical exercise interventions including males and differentiating between both biological genders. Also, given that in ageing the changes in the HPA axis and cortisol rhythm and many sleep disturbances seem to go hand in hand, quality studies on older adults would be a welcome addition to the literature. Further, more quality RCTs with a clear description of the participant’s health status and physical exercise methodology would provide bigger sample sizes and possibly increase homogeneity in future reviews exploring which exercise modality is best for different health states.

5. Conclusions

This systematic review and meta-analysis set out to investigate the effects of physical activity on cortisol and sleep and shed light on the bidirectional relationship between cortisol and sleep. Small beneficial effects of physical exercise interventions to decrease cortisol levels and improve sleep outcomes in adults were identified. To establish a greater degree of accuracy on this matter more information about (1) different
participant demographics, as most studies included women with breast cancer and no older adults, (2) the effect of different intervention types, because the analysis yielded too few studies to make meaningful interpretations, (3) the bidirectional relationship between cortisol and sleep, as there is an established link between these two, but no RCTs researching this were identified.

Other information
none.

Registration and protocol
The systematic review protocol was registered in the international prospective register of systematic reviews (PROSPERO) (https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=722251), registration number CRD42021272251.

Support
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRedIT authorship contribution statement
Len De Nys: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft. Kerry Anderson: Formal analysis. Esther Ofosu: Conceptualization, Methodology. Gemma C. Ryde: Conceptualization, Methodology, Supervision, Writing – review & editing. Jenni Connelly: Conceptualization, Methodology, Supervision, Writing – review & editing. Anna Whitaker: Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information
Supplementary data associated with this article can be found in the online version at doi:10.1016/j.psyneuen.2022.105843.

References
Adam, E.K., Hawley, L.C., Kadielka, B.M., Cacioppo, J.T., 2006. Day-to-day dynamics of sleep, as there is an established link between these two, but no RCTs researching this were identified.

Support
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRedIT authorship contribution statement
Len De Nys: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft. Kerry Anderson: Formal analysis. Esther Ofosu: Conceptualization, Methodology. Gemma C. Ryde: Conceptualization, Methodology, Supervision, Writing – review & editing. Jenni Connelly: Conceptualization, Methodology, Supervision, Writing – review & editing. Anna Whitaker: Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information
Supplementary data associated with this article can be found in the online version at doi:10.1016/j.psyneuen.2022.105843.

References
Adam, E.K., Hawley, L.C., Kadielka, B.M., Cacioppo, J.T., 2006. Day-to-day dynamics of sleep, as there is an established link between these two, but no RCTs researching this were identified.

Support
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRedIT authorship contribution statement
Len De Nys: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft. Kerry Anderson: Formal analysis. Esther Ofosu: Conceptualization, Methodology. Gemma C. Ryde: Conceptualization, Methodology, Supervision, Writing – review & editing. Jenni Connelly: Conceptualization, Methodology, Supervision, Writing – review & editing. Anna Whitaker: Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information
Supplementary data associated with this article can be found in the online version at doi:10.1016/j.psyneuen.2022.105843.

References
Adam, E.K., Hawley, L.C., Kadielka, B.M., Cacioppo, J.T., 2006. Day-to-day dynamics of sleep, as there is an established link between these two, but no RCTs researching this were identified.

Support
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRedIT authorship contribution statement
Len De Nys: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft. Kerry Anderson: Formal analysis. Esther Ofosu: Conceptualization, Methodology. Gemma C. Ryde: Conceptualization, Methodology, Supervision, Writing – review & editing. Jenni Connelly: Conceptualization, Methodology, Supervision, Writing – review & editing. Anna Whitaker: Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information
Supplementary data associated with this article can be found in the online version at doi:10.1016/j.psyneuen.2022.105843.

References
Adam, E.K., Hawley, L.C., Kadielka, B.M., Cacioppo, J.T., 2006. Day-to-day dynamics of sleep, as there is an established link between these two, but no RCTs researching this were identified.

Support
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRedIT authorship contribution statement
Len De Nys: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft. Kerry Anderson: Formal analysis. Esther Ofosu: Conceptualization, Methodology. Gemma C. Ryde: Conceptualization, Methodology, Supervision, Writing – review & editing. Jenni Connelly: Conceptualization, Methodology, Supervision, Writing – review & editing. Anna Whitaker: Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information
Supplementary data associated with this article can be found in the online version at doi:10.1016/j.psyneuen.2022.105843.


