Exercise and sleep: a systematic review of previous meta-analyses

George A Kelley and **Kristi Sharpe Kelley**

Department of Biostatistics, Robert C. Byrd Health Sciences Center, West Virginia University, Morgantown, WV, USA

Abstract

Objective—Conduct a systematic review of previous meta-analyses on exercise and sleep outcomes in adults and a meta-analysis of studies nested within these meta-analyses.

Methods—Meta-analyses of randomized controlled exercise interventions were included by searching nine electronic databases and cross-referencing. Dual-selection and data abstraction were conducted. Methodological quality of meta-analyses was assessed using AMSTAR and quality of evidence using GRADE. Random-effects models were used to pool results from the individual studies included in each meta-analysis.

Results—Three meta-analyses representing 950 adults were included. Methodological quality ranged from 36% to 64% while quality of evidence was very low to low. Statistically significant improvements ($P \leq 0.05$) were observed for the apnea-hypopnea index (AHI), overall sleep quality, global score, subjective sleep, and sleep latency. The number-needed-to-treat (NNT) and percentile improvements ranged from 4 to 7 and from 18.1 to 26.5, respectively. When overall sleep quality results from individual studies nested within different meta-analyses were pooled, statistically significant standardized mean difference (SMD) improvements were observed (−0.50, 95% CI −0.72 to −0.28). The NNT and percentile improvement were 7 and 19, respectively.

Conclusions—Exercise improves selected sleep outcomes in adults. To increase public health reach, a large, well-designed, and more inclusive meta-analysis is needed.

Keywords

apnea; exercise; meta-analysis; sleep; systematic review

1 INTRODUCTION

Sleep disorders are considered to be a major public health epidemic in the United States, affecting an estimated 50 to 70 million US adults.[1] These disorders have been associated

with motor vehicle crashes and industrial disasters as well as medical and other occupational errors.[1] In addition, sleep disorders in adults have been associated with an increased risk for chronic diseases that include hypertension, type 2 diabetes, depression, obesity and cancer.[1] Furthermore, adults with sleep disorders report a lower quality of life and are less productive than those without a sleep disorder.[1] Most notably, sleep disordered adults are at an increased risk for all-cause mortality.[1] In terms of economics, the total costs associated with sleep disorders have been estimated to be as high as $166 billion per year in the United States.[2]

While pharmacologic interventions are a common treatment for sleep disorders,[3,4] statistically significant adverse events have been reported, including an increased risk for falls and cognitive impairment among older adults.[3] In addition, the magnitude of benefit has not been firmly established.[3,4] Exercise, a low-cost, nonpharmacologic intervention that is readily available to the vast majority of adults, offers a potential complementary or alternative approach for improving sleep and is particularly appealing in a public health setting.[5]

Currently, systematic reviews with meta-analysis are considered by many to be the gold standard for determining the effects of an intervention on an outcome.[6,7] However, given that multiple systematic reviews with meta-analysis now exist on the same topic,[8] it becomes difficult to make evidence-based decisions regarding the intervention effects on the outcome of interest in the population of interest. Therefore, it is now necessary to systematically study previous systematic reviews of meta-analyses in order to provide healthcare personnel and policy-makers with the information they need to make better decisions regarding the effectiveness of an intervention on the outcome of interest, provide researchers with information to inform future original studies, and provide meta-analysts with information to inform future meta-analyses, including whether an updated meta-analysis should be conducted on the topic of interest.

Previous meta-analyses have reached conflicting conclusions regarding the effects of exercise on sleep.[9–17] In addition, to the best of the authors' knowledge, no previous systematic review of systematic reviews with meta-analysis examining the effects of exercise on sleep in adults has been performed. Therefore, given multiple systematic reviews with meta-analysis on exercise and sleep as well as the conflicting findings of such,[9–17] the need to systematically review multiple meta-analyses for both applied and research purposes,[8] and the absence of any previous systematic review of systematic reviews with meta-analysis of randomized controlled trials on this topic, the objectives of this study were to conduct a systematic review of previous meta-analyses on exercise and sleep outcomes in adult humans, and conduct a meta-analysis on exercise and sleep outcomes based on the individual studies nested within these meta-analyses.

2 METHODS

2.1 Study eligibility

This systematic review of previous systematic reviews with meta-analysis is registered in the International Prospective Register of Systematic Reviews (PROSPERO) trial registry

(CRD42015023449). In addition and where applicable, the general guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Statement were followed.[18]

The inclusion criteria for this study, established a priori, were as follows: (1) previous systematic reviews with meta-analysis of randomized controlled trials or data reported separately for randomized controlled trials if the meta-analysis included other study designs, (2) adult humans ≥ 18 years of age, (3) exercise (aerobic, strength or both) as the intervention, (4) published and unpublished (dissertations and master's theses) studies in any language up through June of 2015, and (5) exercise minus control group differences in sleep as an outcome in the original meta-analysis and reported as the standardized mean difference (SMD) effect size or calculable using the SMD if at least two studies were pooled. The focus on limiting meta-analyses to randomized controlled trials is based on previous research suggesting that they are the only way to control for unknown confounders and that nonrandomized controlled trials tend to overestimate the effects of treatment(s) in healthcare interventions.[19,20] The SMD was selected as the metric of choice given the different instruments used to assess selected sleep outcomes as well as the desire to compare results using the same metric. Broadly, studies were excluded based on at least one of the following: (a) inappropriate population (e.g., children and/or adolescents), (b) inappropriate intervention (e.g., pharmacological trial), (c) inappropriate comparison (e.g., aerobic exercise versus drug, analyzing the difference in the exercise group while not accounting for the control group, etc.), (d) inappropriate outcome (e.g., anxiety), and (e) inappropriate study type (e.g., systematic review with no meta-analysis included, meta-analysis that did not report data separately for randomized controlled trials only, etc.).

2.2 Data sources

Potentially eligible studies were derived by electronic database searches, cross-referencing from retrieved articles and inspection of the first author's files. For electronic searching, nine databases were searched from their inception up to June 14, 2015 using the graphical user interface for each database. Databases searched included PubMed, Sport Discus, Web of Science, Scopus, PsychInfo, Cochrane Database of Systematic Reviews, Physiotherapy Evidence Database (PEDro), Database of Abstract of Reviews of Effects (DARE), and Proquest. Scopus was included because it has been reported to provide coverage of EMBASE, a database that was not accessible to the investigators.[21] Keywords or forms of keywords used in the database searches included exercise, physical fitness, randomized, systematic review, meta-analysis, sleep, apnea, and insomnia. A copy of the search strategies used for each database is shown in Supplementary File 1. Following duplicate removal both electronically and manually, the overall precision of the searches was calculated by dividing the number of studies that met the eligibility criteria by the total number of studies screened after removing duplicates.[22] The number needed to read (NNR) was then calculated as the reciprocal of the precision.[22] All studies were stored in Reference Manager, version 12.0.[23]

2.3 Study selection

Both authors selected all studies independent of each other. They then convened and reviewed their selections for concurrence. Any incongruities were resolved by consensus.

2.4 Data abstraction

Prior to coding studies, coding sheets were developed using Microsoft Excel (2010).[24] The coding sheets could hold up to 284 items from each included meta-analysis. The major categories of variables coded included (a) study characteristics (source, year, impact factor of journal, etc.), (b) participant characteristics (age, gender, condition(s), etc.), (c) intervention characteristics (length, frequency, intensity, duration, type of exercise, compliance, etc.), and (d) data for sleep outcomes (sample sizes, means, variances, etc.) at both the pooled meta-analytic level as well as for each study included in each meta-analysis. All data were coded by both authors, independent of each other. After coding was completed, all items were reviewed by both authors for correctness. Any inconsistencies were resolved by consensus. Using Cohen's kappa statistic,[25] the overall agreement rate prior to correcting any differences was kappa = 0.96, considered to be "excellent".[26]

2.5 Methodological quality

Methodological quality of each meta-analysis was assessed using The Assessment of Multiple Systematic Reviews (AMSTAR) Instrument.[27–30] This instrument was chosen because of (a) its construct validity (intra-class correlation coefficient = 0.84), (b) inter-rater reliability (kappa = 0.70), and (c) feasibility (average of 15 minutes per study to complete).[29] Responses of "Yes," "No," "Can't Answer," or "Not Applicable" are possible for this 11-item questionnaire. The "Can't Answer" response is chosen when an item is not described but relevant while the "Not Applicable" response is chosen when an item is not relevant (e.g., assessment of publication bias not possible because of a lack of studies).[27–30] For consistency with the other questions, the question "Was the status of publication (i.e., grey literature) used as an inclusion criterion?" was changed to "Was the status of publication (i.e., grey literature) as an inclusion criterion avoided?" Both authors evaluated the methodological quality of each study independent of each other. They then met and reviewed every rating for agreement. Disagreements were resolved by consensus. The overall agreement rate prior to correcting discrepancies was kappa = 0.58, considered to be "good".[26]

In addition to AMSTAR, the overall quality of the evidence was assessed using the Grades of Recommendations Assessment, Development and Evaluation (GRADE) instrument.[31] Overall quality was classified as either very low, low, moderate, or high.[31] To assess impact, the total number of times that each included meta-analysis was cited as well as the average number of citations per year was calculated. This was accomplished using version 4.17 of Publish or Perish (Google Scholar Citation mechanism)[32] on August 5, 2015.

2.6 Data synthesis

2.6.1 Summary findings for sleep outcomes from each meta-analysis—The summary findings from each meta-analysis were extracted with a focus on random versus fixed-effect models because the former incorporates between-study heterogeneity into the analysis when pooling results.[33,34] The SMD was the primary metric of interest along with its 95% confidence intervals (CI), z statistic and alpha value. If sufficient data were available or it was feasible, these data were calculated if not reported in the original study. The magnitude of effect for each SMD from each meta-analysis was categorized as either trivial

(<0.20), small (0.20 to 0.49), medium (0.50 to 0.79) or large (0.80).[35] Two-tailed alpha levels 0.05 for z were considered statistically significant. In addition, Q, a measure of heterogeneity, was also extracted or calculated for each outcome if data were available to do so.[36] An alpha value 0.10 was considered to represent statistically significant heterogeneity.[37] The I^2 statistic, a gauge of inconsistency, was also extracted or calculated if appropriate data were provided.[37] Values of I^2 were categorized as either low (0% to <25%), moderate (25% to <50%), large (50% to <75%) or very large (75%).[37]

It was assumed, a priori, that none of the included meta-analyses would include prediction intervals (PI).[38–40] Consequently, 95% PI were computed if the findings were statistically significant and the requisite data from each study included in each meta-analysis were provided.[38–40] Prediction intervals are used to approximate the treatment effect in a new study[38–40] and may be more applicable for decision analysis.[41]

To enhance practical application, the number-needed-to-treat (NNT) was estimated for any summary findings that were reported as statistically significant. This was accomplished using a control group risk of 30%.[6] Additionally, Cohen's U_3 index was calculated to estimate the percentile gain in the intervention group.[42] Results for small-study effects (publication bias, etc.) were also abstracted or calculated using the regression-intercept approach of Egger et al.,[43] assuming that adequate information were available and the number of SMDs was 10.[44] One-tailed alpha values 0.05 for the intercept were considered to represent statistically significant small-study effects.

2.6.2 Meta-analysis based on studies nested within included meta-analyses—

To increase generalizability, the investigators also conducted their own meta-analysis based on available sleep outcome results from the individual studies nested within each included meta-analysis and while avoiding duplication, that is, results of the same study reported in two or more different meta-analyses. Data synthesis included the abstraction and pooling of results (sample sizes, SMD, variance statistics, etc.) from each study included in each meta-analysis into one overall finding for similar outcomes (e.g., overall sleep quality). All analyses were limited to data reported in the retrieved meta-analyses because the focus of the current investigation was on each meta-analysis and not the original studies. Pooling of studies was accomplished using a random-effects, method-of-moments model.[45] Heterogeneity was assessed using the Q statistic[36] and inconsistency using I^2.[37] A two-tailed, z-based alpha value 0.05 for the SMD was considered statistically significant. In addition, 95% CI were calculated. Based on recent recommendations, small-study effects were examined using funnel plots and Egger's regression intercept test.[44] A one-tailed alpha value 0.05 for the intercept was considered to represent statistically significant small-study effects. Outliers were considered to be those in which the alpha values for the standardized residuals were 0.05. In addition, influence analysis was conducted with each study deleted from the model once as well as cumulative meta-analysis, ranked by year. Furthermore, 95% PI, NNT based on a control group risk of 30% and percentile improvement using Cohen's U_3 index were calculated. Finally, simple random-effects meta-regression (method-of-moments approach) was used to examine the association between changes in sleep and the meta-analysis from which the results were derived.[45] A two-tailed, z-based alpha value 0.05 for the slope (β_1) was considered statistically significant. Negative SMD's were

considered to represent improvements in sleep. All analyses were carried out using Comprehensive Meta-Analysis (version 3.3)[46] and Microsoft Excel 2010.[24]

3 RESULTS

3.1 Characteristics of included meta-analyses

A total of 392 references were initially identified. After removing duplicates both electronically and manually, 283 (72.2%) remained. Of the 283 independent citations screened, three aggregate data meta-analyses met all eligibility criteria.[9,14,15] Search precision after removing duplicates was 0.01, while the NNR was 94. A flow diagram that describes the search process is shown in Figure 1 while a list of excluded studies, including the reasons for exclusion, can be found in Supplementary File 2. Studies were excluded based on an inappropriate study design (41.4%) as well as an inappropriate intervention (36.8%), outcome (17.9%), population (3.6%), and comparison (0.4%). Table 1 describes the general characteristics of each meta-analysis. As can be seen, the three included studies were published between 2013 and 2015, included 2 to 9 studies and between 63 and 599 men and women (total N = 950).[9,14,15] One study reported receiving government funding for their work.[9] All three meta-analyses included different types of populations[9,14,15] in adults up to 72 years of age.[9,14,15] One meta-analysis was limited to participants with obstructive sleep apnea.[15] Length of training for the studies included in each meta-analysis ranged from 5 to 52 weeks, frequency from 3 to 10 times per week, and duration from 20 to 90 minutes per session.[9,14,15] The one meta-analysis that provided information on adverse events reported that six studies did not provide any information while one study reported that adverse events were minimal.[9] Both supervised and unsupervised exercise were performed and included both aerobic and/or strength training.[9,14,15] For the two meta-analyses that reported data,[9,14] intensity of training was classified as moderate to vigorous. For the one meta-analysis that reported information, compliance, defined as the percentage of exercise sessions attended, ranged from 32.4% to 93.3% for the studies included in their review.[9] One meta-analysis only included studies in which walking was part of the exercise intervention.[9] Sleep outcomes were assessed using the apnea-hypopnea index (AHI)[15] and Pittsburgh Sleep Quality Index (PSQI).[14] Another meta-analysis reported the assessment of overall sleep quality but specific measures for the assessment of such from each of the included studies were not provided.[9]

3.2 Methodological quality and impact

The results for each meta-analysis using the AMSTAR instrument are shown in Supplementary File 3. As can be seen, the overall quality of the meta-analyses using the AMSTAR instrument ranged from 36% to 64%.[9,14,15] All of the studies were considered to have provided adequate information regarding (a) an a priori design, (b) description of study characteristics, (c) assessment of study quality, and (d) appropriate methods for combining studies.[9,14,15] In contrast, none of the studies provided a reference list of excluded studies, including the reasons for exclusion, as well as appropriate information regarding conflicts of interest, especially with respect to conflicts of interest from each of the studies included in the meta-analyses.[9,14,15] The results for the other five criteria were mixed.[9,14,15]

In relation to impact, the total number of times that each meta-analysis was cited were 0,[9] 24,[15] and 56.[14] When adjusted for the number of years that each meta-analysis was available, citation rates were 0,[9] 12,[15] and 18.7.[14]

3.3 Data synthesis

3.3.1 Results from each meta-analysis—Overall results for the three included meta-analyses are shown in Table 2.[9,14,15] As can be seen, the number of ES's in each meta-analysis were small, ranging from 2 to 9, while the number of participants ranged from 63 to 599.[9,14,15] Statistically significant improvements were observed for all three meta-analyses.[9,14,15] These included the AHI,[15] overall sleep quality,[9,14] subjective sleep[14] and sleep latency.[14] No statistically significant differences were observed for sleep duration, efficiency, disturbance, or daytime functioning.[14]

For those results that were statistically significant, no statistically significant heterogeneity or inconsistencies were observed for the AHI.[15] However, statistically significant heterogeneity and a large amount of inconsistency were observed for overall sleep quality in both meta-analyses that assessed such[9,14] as well as subjective sleep, sleep latency, sleep duration, sleep efficiency and sleep disturbance in the one study that reported this information.[14] Nonoverlapping prediction intervals were observed for overall sleep quality in the study by Chiu et al.[9] but not Yang et al.[14] Overlapping prediction intervals were also observed for subjective sleep as well as sleep latency.[14] None of the studies reported results for potential small-study effects (publication bias, etc.).[9,14,15]

The NNT and percentile improvement estimates for statistically significant findings are shown in Table 3. Assuming a control group risk of 30%, the NNT for sleep outcomes ranged from a low of 4 for sleep latency to a high of 7 for overall sleep quality.[14] Overall percentile improvements were similar, ranging from 18.1 to 26.5. Corresponding 95% CI were best for sleep quality in the study by Chiu et al.,[9] and worst for sleep latency in the study by Yang et al.[14] Using the GRADE instrument, the overall quality of evidence ranged from very low to low (Supplementary File 4).

3.3.2 Results of pooling different studies from different meta-analyses for the same outcomes—Based on the availability of evidence, the pooling of studies from each meta-analysis was limited to overall sleep quality from two of the included meta-analyses.[9,14] This included 14 different studies representing 887 participants.[9,14] As shown in Figure 2, a moderate and statistically significant SMD improvement in overall sleep quality was observed ($z = 4.6$, P < 0.001) along with statistically significant heterogeneity ($Q = 30.7$, P = 0.004) and a large amount of inconsistency ($I^2 = 57.7\%$, 95% CI 23.4% to 76.6%).

No statistically significant outliers were observed (P for all standardized residuals > 0.05). Visual inspection of the funnel plot (Fig. 3) as well as Egger's regression intercept test suggests that statistically significant small-study effects were present ($\beta_0 = -3.24$, P = 0.04). As can be seen in Figure 4, influence analysis revealed that SMD results were statistically significant with each study deleted from the model once, with overall changes differing by 0.11 (20.8%). Cumulative meta-analysis demonstrated that improvements in sleep quality

have remained statistically significant since the first study was published in 1997 (Fig. 5). No association was observed for changes in overall sleep quality and the meta-analysis from which results were derived (β_1 = 0.05, 95% CI −0.41 to 0.52, P = 0.82). Overlapping prediction intervals were observed(95% PI, −1.19 to 0.20). The NNT was 7(95% CI 5 to 10), while the percentile improvement was 19 (95% CI 11.2 to 26.1).

4 DISCUSSION

4.1 Findings

In the ideal scenario supporting the effects of any intervention on an outcome, results will be (a) statistically significant with nonoverlapping and narrow CI, (b) homogeneous with no inconsistency, (c) have narrow and nonoverlapping PI, (d) be free of all bias, including small-study effects, and (e) be stable with a magnitude of change that is practically important. However, satisfying all these criteria is probably highly unlikely.

Rather, one must draw inferences based on imperfect findings. Such is the case with the current investigation. From the investigative team's perspective, the overall findings of the current study suggest that exercise improves selected sleep outcomes in the sample of adults included. This interpretation is reinforced by (a) statistically significant improvements in the AHI, overall sleep quality, subjective sleep, and sleep latency, (b) low NNT (4 to 7) for statistically significant outcomes, (c) percentile improvements (18.1 to 26.5) for statistically significant outcomes, and (d) nonoverlapping PI for overall sleep quality in the meta-analysis by Chiu et al.[9] In contrast, the potential benefits of exercise on sleep in adults may be questioned given (a) statistically significant heterogeneity for three of the five statistically significant outcomes, (b) a large amount of inconsistency for four of the five outcomes in which statistically significant improvements were observed, (c) overlapping PI for global sleep score, subjective sleep and sleep latency, (d) lack of statistically significant improvements for sleep duration, sleep efficiency, sleep disturbance and daytime functioning, (e) based on AMSTAR assessment, a quality rating of less than 50% for two of the three meta-analyses conducted,[9,14] and (f) based on GRADE assessment, low- to very low-quality levels of evidence for all outcomes.

The potential benefits of exercise on overall sleep quality may be especially promising given that the investigative team, based on the results reported in the original meta-analyses, were able to successfully combine the results for overall sleep quality from two of the included meta-analyses.[9,14] These findings included (a) statistically significant improvements in overall sleep quality, (b) a low NNT,[7] (c) a large percentile improvement,[19] (d) stability of findings when each study was deleted from the model once, (e) statistical significance of findings since the conduct of the first study in 1997, and (f) no statistically significant association between the meta-analysis from which the results were derived. However, these findings may have been weakened by (a) statistically significant heterogeneity, (b) a large amount of inconsistency, (c) overlapping PI, (d) potential small-study effects, and (e) based on GRADE, the very low to low quality of evidence.

The statistically significant findings observed in this study for selected outcomes are somewhat less than the use of pharmacological interventions to improve sleep. For example,

the approximate 19.7% improvement in AHI in the meta-analysis by Araghi et al.,[15] compares to improvements ranging from 25% to 45% with the use of various pharmacologic agents in participants with obstructive sleep apnea.[47] However, the authors of this prior study concluded that there was insufficient evidence to recommend drug therapy in the treatment of obstructive sleep apnea.[47] In contrast, the current findings for overall sleep quality compare favorably to a recent meta-analysis of mind-body interventions in cancer patients in which statistically significant SMD improvements of −0.43 (95% CI −0.24 to −0.62) as well as long-term improvements up to three months (SMD, −0.29 95% CI −0.52 to −0.06) were reported.[48] The present findings also compare favorably to a recent meta-analysis on meditative movement therapies in adults >60 years of age and in which SMD improvements of −0.70 (95%CI −0.96 to −0.43) were reported.[49] Thus, it appears that exercise may serve as a complementary or alternative approach for improving sleep in adults.

4.2 Implications for research

There are at least six recommendations for future research using the meta-analytic approach to examine the effects of exercise on sleep outcomes in adults. First, the methodological quality of the meta-analyses themselves could be improved. This includes (a) the inclusion of studies regardless of publication status or providing a strong rationale for not doing so, (b) providing a bibliography of all excluded studies, including the reasons for exclusion, and (c) providing a description of potential conflicts of interest, including potential sources of support, for each of the studies included in each meta-analysis.

Second, based on citation rates, the impact of the included metaanalyses appears to be small. One possible explanation may be that this work is published in journals that do not have a large readership, thereby compromising the reach of this potentially beneficial non-pharmacological intervention. A second possible reason may be the fact that guidelines for the treatment of sleep problems such as insomnia are heavily focused on pharmacological versus non-pharmacological therapies.[50] Another reason may be that the number of researchers and practitioners interested in sleep problems may be less than those interested in other conditions such as cancer and cardiovascular disease.

Third, future meta-analyses should provide practical information so that practitioners and policy-makers can make better evidence-based decisions with respect to the effects of exercise on sleep in adults. These include, but are not necessarily limited to, statistics such as relative changes, NNT and/or percentile improvements.

Fourth, future meta-analyses should include PIs as well as CIs. The use of PIs can help to establish expected outcome effects in a new study and may also be more valid for decision-making.[40] However, it's important to understand that PIs are based on random mean effects while confidence intervals are not.[40]

Fifth, the three meta-analyses included in the current study were limited to less than 10 effect sizes for each outcome as well as participants with certain characteristics: obstructive sleep apnea,[15] cancer,[9] and older adults with sleep problems.[14] Given the former, it would appear plausible to suggest that a larger more inclusive meta-analysis would be a more

powerful research design and have greater applicability across a wider range of participants, one of the very reasons for conducting a meta-analysis. This approach may be particularly important if viewed from the perspective of increasing public health reach. In addition, a larger and more inclusive meta-analysis would provide one with a greater opportunity to examine for potential predictors with respect to changes in selected sleep outcomes.

Sixth, future meta-analyses should report information on adverse events for all included studies, including whether the original studies provided such information. This is important for balancing the benefits and harms of any intervention, including exercise.

Based on the current findings, three recommendations for future randomized controlled trials appear to be warranted. First, a need exists for the inclusion of data on the cost-effectiveness of exercise interventions on sleep outcomes in adults given that none of the meta-analyses reported such data.[9,14,15] This is of course assuming that the original studies included in each of the metaanalyses did not provide this information. Second, a need exists for multi-arm randomized controlled trials that directly compare the dose-response effects of exercise on selected sleep outcomes in adults, including what type of exercise, aerobic, strength training, or both, may be best for improving selected sleep outcomes. More accurate data on this topic should lead to better treatment in the population of interest. Third, future randomized controlled trials need to report complete information on any adverse events experienced by the participants during the intervention.

4.3 Implications for practice

The findings of the current review provide important information for practice. First, despite the low quality of evidence as well as lack of statistically significant results for several sleep outcomes, exercise appears to improve selective sleep outcomes, including more global measures of sleep. While no specific recommendations directed solely at sleep outcomes can be made and further research is needed, it would appear pragmatic to suggest that adherence to current and broad guidelines for exercise be recommended. These include at least 150 minutes per week of moderate-intensity activity such as brisk walking or 75 minutes or more each week of vigorous-intensity activity such as jogging.[51] Some combination of the two is also acceptable.[51] Additionally, at least two days per week of muscle strengthening activities that exercise the major muscle groups of the body (legs, hips, back, abdomen, chest, shoulders, and arms) should be performed.[51] However, it is important to note that these are general recommendations.[51]

4.4 Strengths and potential limitations of this study

There are at least three *strengths* of the current study. First, to the best of the authors' knowledge, this is the *first* systematic review of previous systematic reviews with meta-analysis directed at determining the effects of exercise on selected sleep outcomes in adults. This is important for (a) determining the effects of exercise on sleep outcomes, (b) providing recommendations on the reporting and conduct of future research, and (c) providing evidence regarding the prioritization of exercise over alternative treatments such as pharmacologic interventions.[8] As a result, a summary of previous meta-analyses addressing the effects of exercise on sleep outcomes is now available, thereby contributing important

evidence for advancing future research, practice and policy-making. Second, the additional analyses that were conducted but not available in the original meta-analyses (NNT, percentile improvement, PIs)[9,14,15] aided in strengthening the evidence from which conclusions could be drawn from the included studies. Importantly, the calculation of PIs provides future researchers with valuable information in the planning and conduct of randomized controlled studies aimed at determining the effects of exercise on sleep in adults. Third, to the best of the investigative team's knowledge, this is the first systematic review of previous systematic reviews with meta-analysis in which a meta-analysis was conducted based on a similar outcome from different studies included in different meta-analyses.[9,14] Such an approach is a cost and time efficient way to increase statistical power for primary endpoints and enhance generalizability.

The current study may be subject to at least three *possible* limitations. First, the number of studies and subsequent effect sizes included in each meta-analysis was small and limited to very narrowly defined populations. As a result, the findings may not be generalizable to other populations. Despite this, it's important to realize that two is the minimum number of studies necessary for conducting a meta-analysis.[6] Second, the results of the included meta-analyses may have suffered from small-study effects (publication bias, etc.). However, the assessment of such in the original meta-analyses was not justifiable since all three meta-analyses included less than 10 effect sizes and a minimum of 10 effect sizes is recommended before any such analyses is performed.[44] Third, the current study inherited the potential biases included in not only the original studies but also the meta-analyses themselves.[9,14,15] This includes, but is not necessarily limited to, ecological fallacy, specifically Simpson's paradox. Finally, while the inclusion of the study by Araghi et al.[15] may be questioned given that it was conducted in sleep apnea participants using the more objective AHI, the significance of the change compared to the other included studies that used more general and subjective sleep quality measures in participants that did not appear to have sleep apnea appeared to be similar.

5 CONCLUSIONS

The results of the current study suggest that exercise is associated with improvements in selected sleep outcomes in the sample of adults included in the meta-analyses. To increase public health reach, a need exists for a large, well-executed and more inclusive systematic review with meta-analysis on this topic.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Institute of Medicine. Sleep disorders and sleep deprivation: an unmet public health problem. Washington, DC: The National Academies Press; 2006.

2. University of Maryland. Sleep disorders. 1-29-2015.

3. Glass J, Lanctot KL, Herrmann N, Sproule BA, Busto UE. Sedative hypnotics in older people with insomnia: meta-analysis of risks and benefits. BMJ. 2005; 331:1169. [PubMed: 16284208]

4. Smith MT, Perlis ML, Park A, Smith MS, Pennington J, Giles DE, et al. Comparative meta-analysis of pharmacotherapy and behavior therapy for persistent insomnia. American Journal of Psychiatry. 2002; 159:5–11. [PubMed: 11772681]

5. Brady TJ, Jernick SL, Hootman JM, Sniezek JE. Public health interventions for arthritis: expanding the toolbox of evidence-based interventions. Journal of Women's Health (Larchmt). 2009; 18:1905–1917.

6. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. 2011 The Cochrane Collaboration.

7. Shorten A, Shorten B. What is meta-analysis? Evidence-Based Nursing. 2013; 16:3–4. [PubMed: 23178310]

8. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Medical Research Methodology. 2011; 11:15. [PubMed: 21291558]

9. Chiu HY, Huang HC, Chen PY, Hou WH, Tsai PS. Walking improves sleep in individuals with cancer: a meta-analysis of randomized, controlled trials. Oncology Nursing Forum. 2015; 42:E54–E62. [PubMed: 25806892]

10. Mishra SI, Scherer RW, Geigle PM, Berlanstein DR, Topaloglu O, Gotay CC, et al. Exercise interventions on health-related quality of life for cancer survivors. Cochrane Database of Systematic Reviews. 2012; 8:CD007566.

11. Montgomery P, Dennis J. Physical exercise for sleep problems in adults aged 60+ Cochrane Database of Systematic Reviews. 2002:CD003404. [PubMed: 12519595]

12. Tomlinson D, Diorio C, Beyene J, Sung L. Effect of exercise on cancer-related fatigue: a meta-analysis. American Journal of Physical Medicine & Rehabilitation. 2014; 93:675–686. [PubMed: 24743466]

13. Yang B, Xu J, Xue Q, Wei T, Xu J, Ye C, et al. Non-pharmacological interventions for improving sleep quality in patients on dialysis: systematic review and meta-analysis. Sleep Medicine Reviews. 2014; 23C:68–82.

14. Yang PY, Ho KH, Chen HC, Chien MY. Exercise training improves sleep quality in middle-aged and older adults with sleep problems: a systematic review. Journal of Physiotherapy. 2012; 58:157–163. [PubMed: 22884182]

15. Araghi MH, Chen YF, Jagielski A, Choudhury S, Banerjee D, Hus-sain S, et al. Effectiveness of lifestyle interventions on obstructive sleep apnea (OSA): systematic review and meta-analysis. Sleep. 2013; 36:1553–1562E. [PubMed: 24082315]

16. Kubitz KA, Landers DM, Petruzzello SJ, Han M. The effects of acute, chronic exercise on sleep. A meta-analytic review. Sports Medicine. 1996; 21:277–291. [PubMed: 8726346]

17. Iftikhar IH, Kline CE, Youngstedt SD. Effects of exercise training on sleep apnea: a meta-analysis. Lung. 2014; 192:175–184. [PubMed: 24077936]

18. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioan-nidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Annals of Internal Medicine. 2009; 151:W65–W94. [PubMed: 19622512]

19. Sacks HS, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. American Journal of Medicine. 1982; 72:233–240. [PubMed: 7058834]

20. Schulz KF, Chalmers I, Hayes R, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995; 273:408–412. [PubMed: 7823387]

21. Burnham JF. Scopus database: a review. Biomedical Digital Libraries. 2006; 3:1–8. [PubMed: 16522216]

22. Lee E, Dobbins M, DeCorby K, Mcrae L, Tirilis D, Husson H. An optimal search filter for retrieving systematic reviews and meta-analyses. BMC Medical Research Methodology. 2012; 12:51. [PubMed: 22512835]

23. Reference Manager [computer program]. Version 12.0.3. Philadelphia, PA: Thompson ResearchSoft; 2009.

24. Microsoft Excel [computer program]. Version 2007. Redmond, WA: Microsoft Corporation; 2010.

25. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin. 1968; 70:213–220. [PubMed: 19673146]

26. Cicchetti DV, Sparrow SA. Developing criteria for establishing inter-rater reliability of specific items: applications to assessment of adaptive behavior. American Journal of Mental Deficiency. 1981; 86:127–137. [PubMed: 7315877]

27. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Medical Research Methodology. 2007; 7:10. [PubMed: 17302989]

28. Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS ONE. 2007; 2:e1350. [PubMed: 18159233]

29. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. Journal of Clinical Epidemiology. 2009; 62:1013–1020. [PubMed: 19230606]

30. Kang DY, Wu YX, Hu D, Hong Q, Wang JL, Zhang X. Reliability and external validity of AMSTAR in assessing quality of TCM systematic reviews. Evidence-Based Complementary and Alternative Medicine. 2012; 2012:1–7.

31. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008; 336:924–926. [PubMed: 18436948]

32. Harzing's Publish or Perish [computer program]. Version 4.27. 2007

33. Mosteller F, Colditz GA. Understanding research synthesis (meta-analysis). Annual Review of Public Health. 1996:1–23.

34. Borenstein, M., Hedges, L., Higgins, J., Rothstein, H. Introduction to meta-analysis. West Sussex: John Wiley & Sons: 2009.

35. Cohen J. A power primer. Psychological Bulletin. 1992; 112:155–159. [PubMed: 19565683]

36. Cochran WG. The combination of estimates from different experiments. Biometrics. 1954; 10:101–129.

37. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003; 327:557–560. [PubMed: 12958120]

38. Graham PL, Moran JL. Robust meta-analytic conclusions mandate the provision of prediction intervals in meta-analysis summaries. Journal of Clinical Epidemiology. 2012; 65:503–510. [PubMed: 22265586]

39. Kelley GA, Kelley KS. Impact of progressive resistance training on lipids and lipoproteins in adults: another look at a meta-analysis using prediction intervals. Preventive Medicine. 2009; 49:473–475. [PubMed: 19804794]

40. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. Journal of the Royal Statistical Society: Series A. 2009; 172:137–159.

41. Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewers checklist. Medical Decision Making. 2013; 33:679–691. [PubMed: 23804511]

42. Cohen, J. Statistical power analysis for the behavioral sciences. New York: Academic Press; 1988.

43. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. BMJ. 1997; 315:629–634. [PubMed: 9310563]

44. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011; 343:d4002. [PubMed: 21784880]

45. Dersimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clinical Trials. 1986; 7:177–188. [PubMed: 3802833]

46. Comprehensive meta-analysis [computer program]. Version 3.3. New Jersey: Englewood; 2015.

47. Mason M, Welsh EJ, Smith I. Drug therapy for obstructive sleepapnoea in adults. Cochrane Database of Systematic Reviews. 2013; 5:CD003002.

48. Chiu HY, Chiang PC, Miao NF, Lin EY, Tsai PS. The effects of mind-body interventions on sleep in cancer patients: a meta-analysis of randomized controlled trials. Journal of Clinical Psychiatry. 2014; 75:1215–1223. [PubMed: 25188331]

49. Wu WW, Kwong E, Lan XY, Jiang XY. The effect of a meditative movement intervention on quality of sleep in the elderly: a systematic review and meta-analysis. Journal of Alternative & Complementary Medicine. 2015; 21:509–519. [PubMed: 26120865]

50. Schutte-Rodin S, Broch L, Buysse D, Dorsey C, Sateia M. Clinical guideline for the evaluation and management of chronic insomnia in adults. Journal of Clinical Sleep Medicine. 2008; 4:487–504. [PubMed: 18853708]

51. Physical Activity Guidelines Advisory Committee. Physical Activity Guidelines Advisory Report. Washington, DC: U.S. Department of Health and Human Services; 2008.
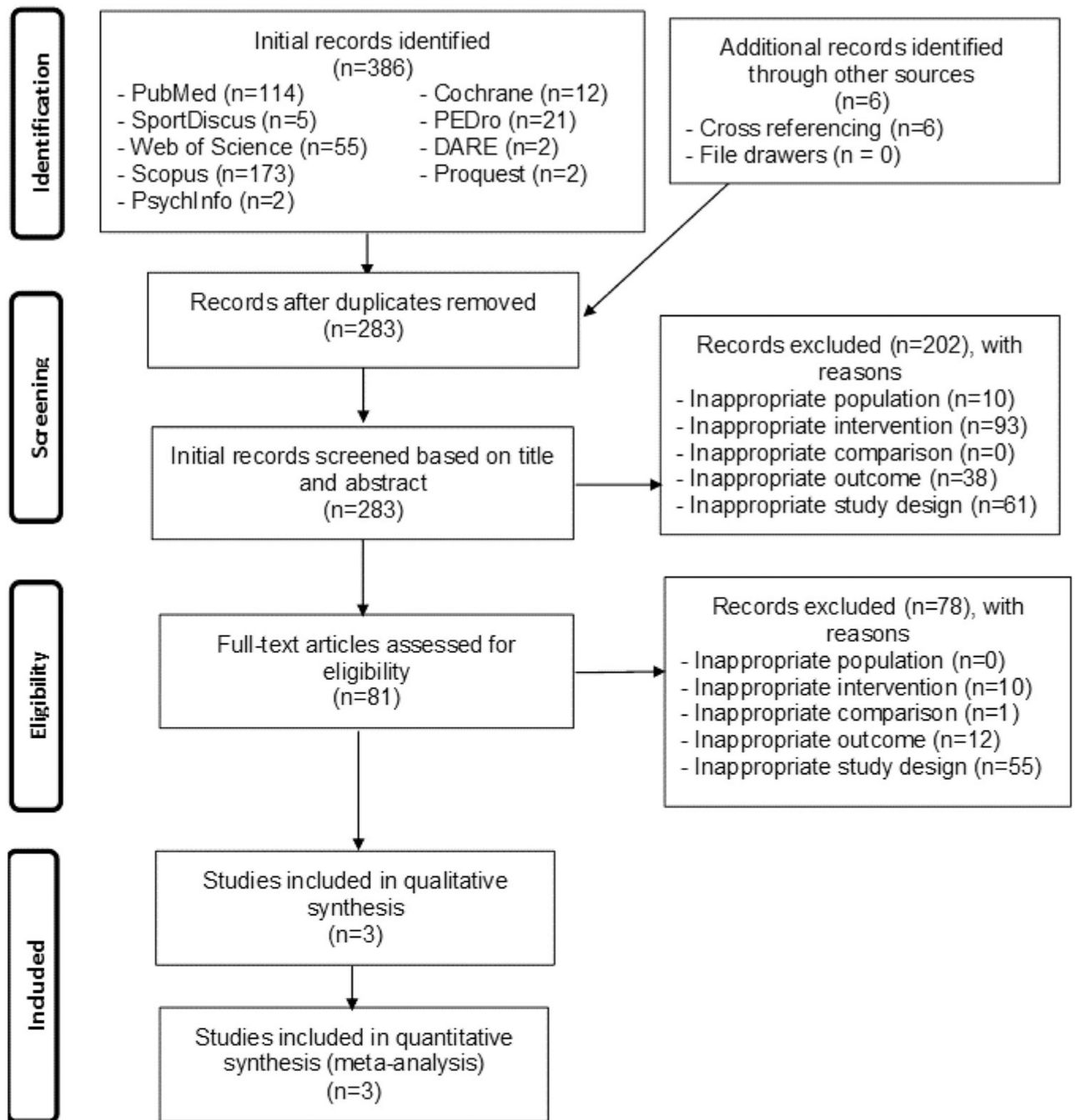
FIGURE 1.
Flow diagram for selection of articles

Group by Meta-analysis	Original Study	Statistics for each study			Point estimate and 95% CI
		Point estimate	Lower limit	Upper limit	
Chiu et al., 2015	Cheville et al., 2013	-0.86	-1.40	-0.31	
Chiu et al., 2015	Donnelly et al., 2011	-0.55	-1.22	0.13	
Chiu et al., 2015	Mock et al., 1997	-0.67	-1.26	-0.09	
Chiu et al., 2015	Rogers et al., 2014	-0.15	-0.75	0.45	
Chiu et al., 2015	Sprod et al., 2010	-0.16	-0.79	0.46	
Chiu et al., 2015	Tang et al., 2010	-1.23	-1.73	-0.73	
Chiu et al., 2015	Wang et al., 2011	-0.79	-1.27	-0.32	
Chiu et al., 2015	Wenzel et al., 2013	-0.21	-0.56	0.14	
Chiu et al., 2015	Wiskemann et al., 2011	-0.13	-0.51	0.26	
Chiu et al., 2015		-0.52	-0.79	-0.25	
Yang et al., 2012	Elavsky & Mcauley, 2007	-0.15	-0.55	0.25	
Yang et al., 2012	Irwin et al., 2008	-0.35	-0.91	0.21	
Yang et al., 2012	King et al., 1997	-1.23	-1.89	-0.57	
Yang et al., 2012	King et al., 2008	-0.12	-0.61	0.37	
Yang et al., 2012	Singh et al., 1997	-0.81	-1.59	-0.04	
Yang et al., 2012		-0.47	-0.86	-0.08	
Overall		-0.50	-0.72	-0.28	

-2.00   -1.00   0.00   1.00   2.00
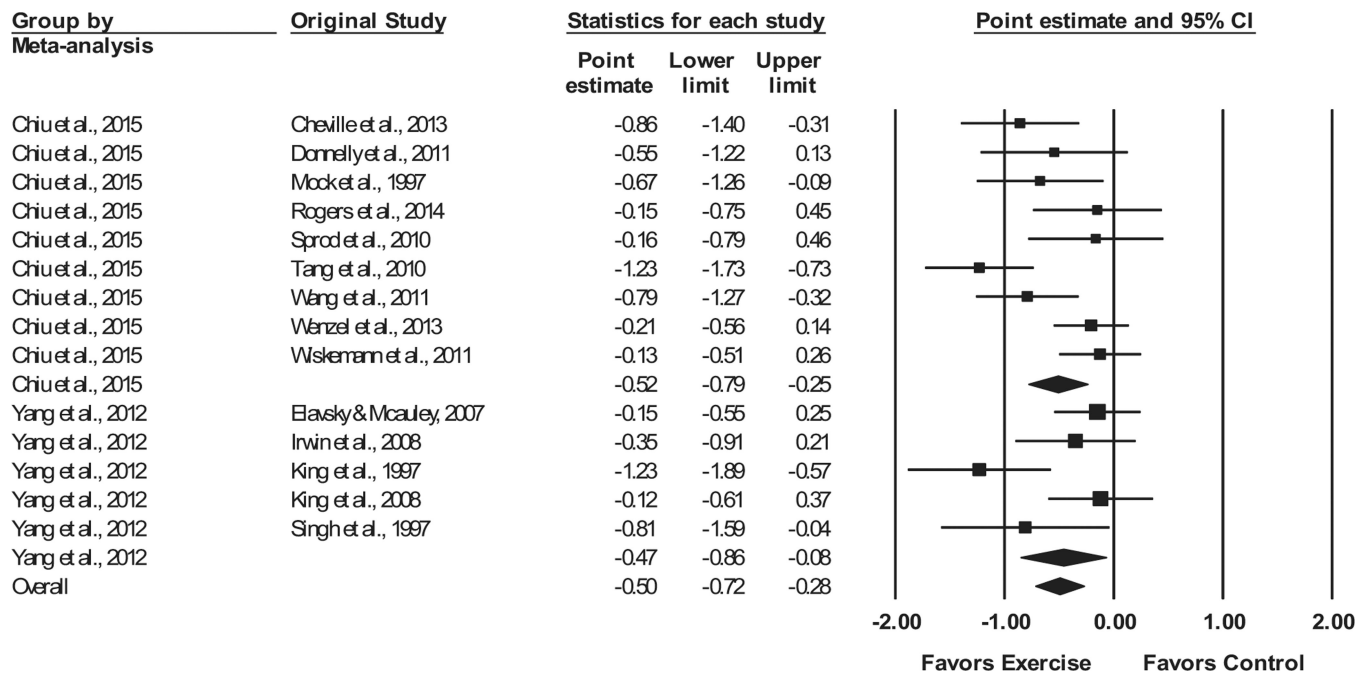
Favors Exercise   Favors Control

FIGURE 2.

Forest plot for changes in overall sleep quality. The black horizontal lines represent the 95% confidence intervals while the squares represent the point estimate. The first two black diamonds represent the overall point estimate and 95% confidence intervals from each meta-analysis, while the third black diamond represents the overall pooled point estimate and 95% confidence intervals from all individual studies included in each meta-analysis. All analyses are based on the random-effects model
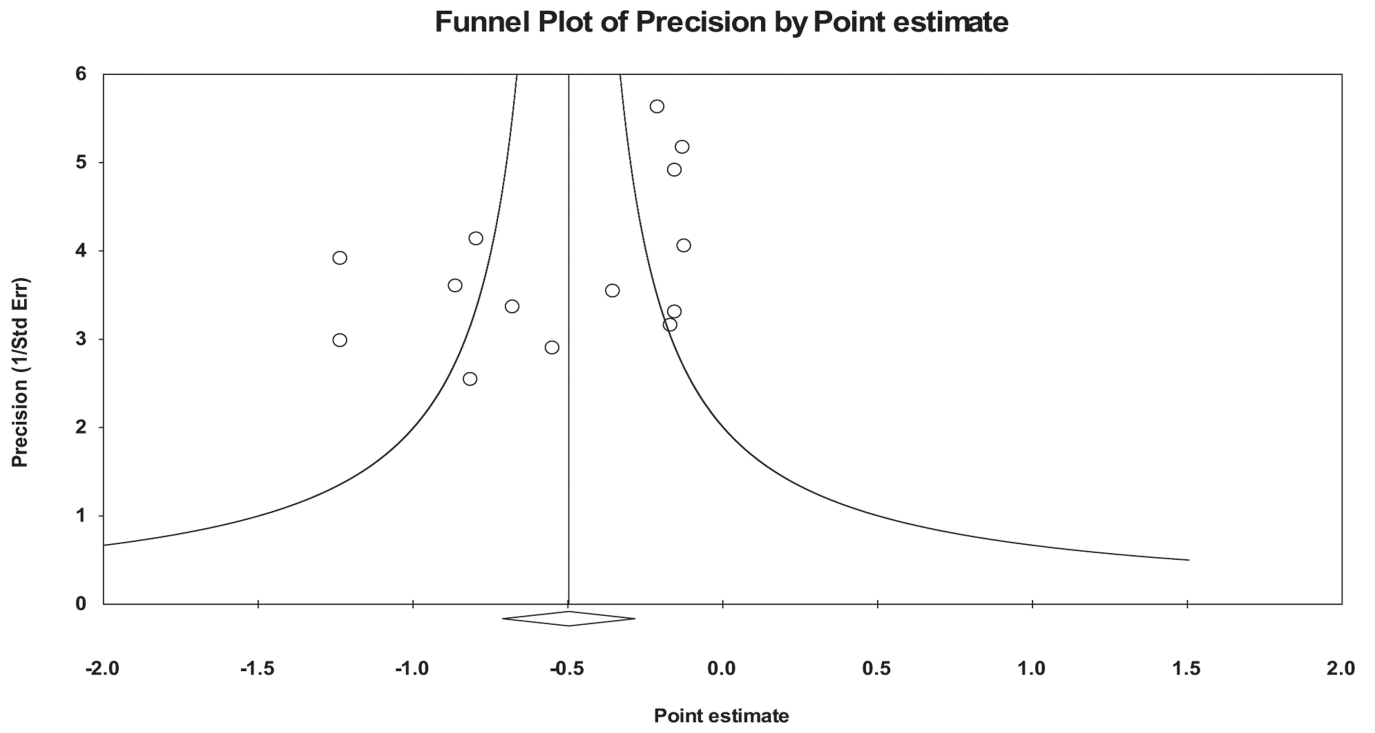
Funnel Plot of Precision by Point estimate

FIGURE 3.
Funnel plot for overall changes in sleep quality

Original Study

Point estimate (95% CI) with study removed

Original Study	Point	Lower limit	Upper limit
Tang et al., 2010	-0.42	-0.61	-0.23
King et al., 1997	-0.45	-0.65	-0.24
Cheville et al., 2013	-0.47	-0.69	-0.25
Wang et al., 2011	-0.47	-0.70	-0.25
Singh et al., 1997	-0.48	-0.70	-0.26
Mock et al., 1997	-0.49	-0.71	-0.26
Donnelly et al., 2011	-0.50	-0.72	-0.27
Irwin et al., 2008	-0.51	-0.74	-0.28
Sprod et al., 2010	-0.52	-0.75	-0.30
Rogers et al., 2014	-0.52	-0.75	-0.30
Wenzel et al., 2013	-0.53	-0.76	-0.30
King et al., 2008	-0.53	-0.75	-0.30
Elavsky & Mcauley, 2007	-0.53	-0.76	-0.31
Wiskemann et al., 2011	-0.53	-0.76	-0.31
	-0.50	-0.71	-0.28

-1.00   -0.50   0.00   0.50   1.00
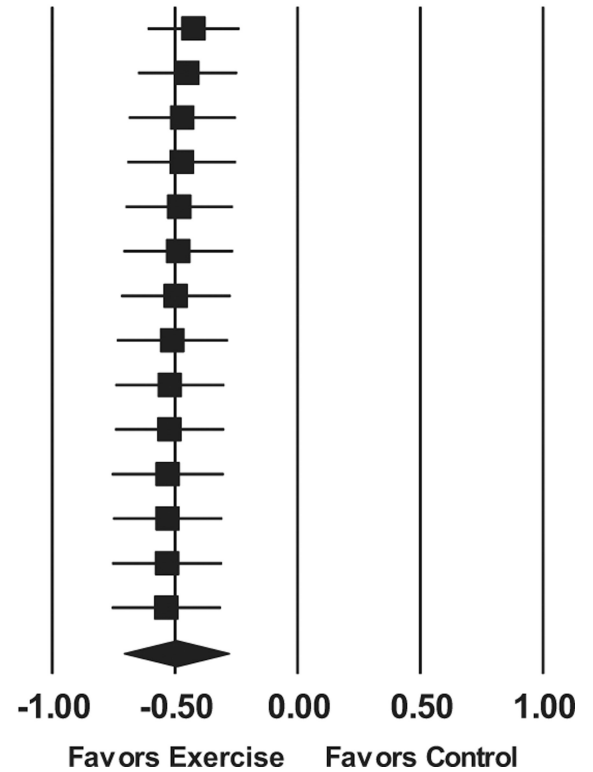
Favors Exercise   Favors Control

FIGURE 4.

Influence analysis for changes in overall sleep quality with each study deleted from the model once. The black horizontal lines represent the 95% confidence intervals, while the squares represent the point estimate. The black diamond represents the overall point estimate and 95% confidence intervals

Original Study	Cumulative statistics			Cumulative point estimate (95% CI)
	Point	**Lower limit**	**Upper limit**	
Mock et al., 1997	-0.67	-1.26	-0.09	
King et al., 1997	-0.93	-1.47	-0.39	
Singh et al., 1997	-0.89	-1.27	-0.51	
Elavsky & Mcauley, 2007	-0.67	-1.16	-0.17	
Irwin et al., 2008	-0.59	-0.97	-0.20	
King et al., 2008	-0.49	-0.83	-0.16	
Sprod et al., 2010	-0.44	-0.74	-0.15	
Tang et al., 2010	-0.57	-0.90	-0.23	
Donnelly et al., 2011	-0.56	-0.86	-0.26	
Wang et al., 2011	-0.59	-0.86	-0.31	
Wiskemann et al., 2011	-0.53	-0.79	-0.28	
Cheville et al., 2013	-0.56	-0.80	-0.32	
Wenzel et al., 2013	-0.52	-0.75	-0.30	
Rogers et al., 2014	-0.50	-0.71	-0.28	
	-0.50	-0.71	-0.28	

-2.00   -1.00   0.00   1.00   2.00
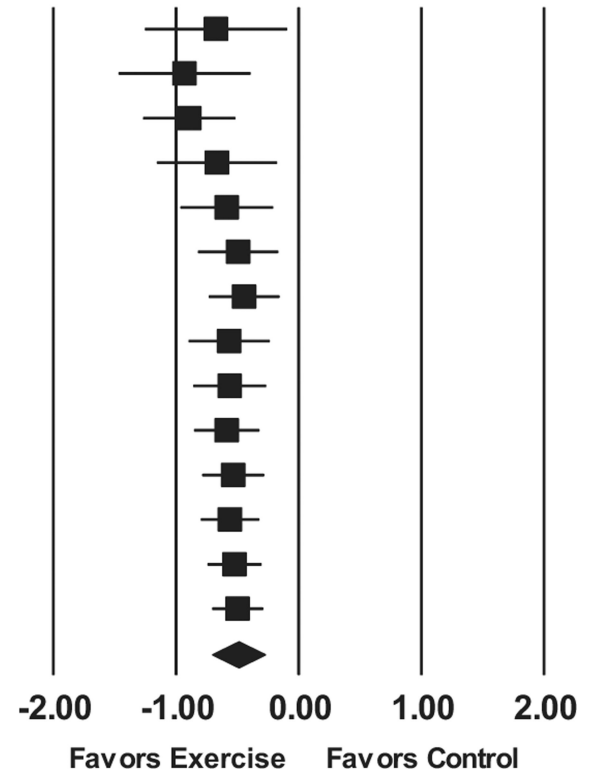
Favors Exercise   Favors Control

FIGURE 5.
Cumulative meta-analysis, ranked by year, for point estimate changes in overall sleep quality. The black diamond represents the overall point estimate and 95% confidence intervals

TABLE 1

General characteristics of included meta-analyses the included meta-analyses

Study	Year	Country	Studies	Participants	Interventions	Sleep assessment
Araghi et al.[15]	2013	UK	2	63[a] men and women with obstructive sleep apnea, 46–54 years of age ($\bar{X} = 50$)	Exercise lasting 12 weeks	AHI
Chiu et al.[9]	2015	China	9	599[a] men and women with cancer (42% breast), 18 years of age ($\bar{X} \pm SD$, 54.4 $\pm$ 5.7)	Supervised and unsupervised walking included in interventions lasting 5–35 weeks ($\bar{X} = 10$), frequency 3–10× week ($\bar{X} = 4.5$), duration 20–90 minutes per session ($\bar{X} = 37.5$), intensity classified as moderate in 7 studies, moderate to vigorous in 1, and not reported in another, compliance 32.4–93.3% ($\bar{X} = 77.0\%$)	–
Yang et al.[14]	2012	China	6	305 older adults (171 exercise, 134 control) with sleep problems, 48.6–72 years of age ($\bar{X} \pm SD$, 62.7 $\pm$ 7.4)	Aerobic and/or strength training 10–52 weeks ($\bar{X} \pm SD$, 21 $\pm$ 15), frequency 3–5× week ($\bar{X} \pm SD$, 4 $\pm$ 1), duration 10–60 minutes per session, intensity 60–75% HRR (1 study), 60–85% HRR (1 study), 60–85% MHR (1 study), 55–75% MHR (1 study)	PSQI (global score, subscales of subjective sleep, sleep latency, sleep duration, sleep efficiency, sleep disturbance, daytime functioning)

Note: $\bar{X} \pm SD$, mean $\pm$ standard deviation. Description of meta-analyses limited to those studies nested within each meta-analysis that met all eligibility criteria for this study. Data presented limited to what was reported or could be calculated from reported data. Number of participants limited to those in which results were calculated. –, data not provided or insufficient data to calculate; HRR, heart rate reserve; MHR, maximum heart rate; PSQI, Pittsburgh Sleep Quality Index; AHI, Apnea-hypopnea Index.

[a] Separate sample sizes not available for exercise and control groups.

TABLE 2

Overall posttreatment changes in sleep from included meta-analyses

Study	ES/participants (No.)	$\bar{X}$(95% CI)	Z(p)	Q(p)	I^2(%)	Tau^2	PI (95%)
Araghi et al. (2013)[15]							
AHI	2/63	−0.72 (−1.23, −0.21)	**2.79 (0.005)**[a]	0.005 (0.94)	0	<0.001	-
Chiu et al. (2015)[9]							
Overall sleep quality	9/599	−0.52 (−0.79, −0.25)	**3.77 (0.0002)**[a]	20.3 (0.009)[a]	61	0.019	−0.98, −0.06
Yang et al. (2012)[14]							
Global score	5/288	−0.47 (−0.86, −0.08)	**2.35 (0.02)**[a]	9.99 (0.04)[a]	60	0.12	−1.74, 0.80
Subjective sleep	5/239	−0.47 (−0.73, −0.20)	**3.47 (0.0005)**[a]	9.06 (0.06)	56	0.13	−1.70, 0.76
Sleep latency	5/239	−0.58 (−1.08, −0.08)	**2.26 (0.02)**[a]	12.2 (0.02)[a]	67	0.21	−2.25, 1.09
Sleep duration	6/305	−0.10 (−0.53, 0.33)	0.47 (0.64)	15.6 (0.008)[a]	68	0.19	-
Sleep efficiency	5/239	−0.35 (−0.84, 0.15)	1.38 (0.17)	12.0 (0.02)[a]	67	0.20	-
Sleep disturbance	4/245	−0.25 (−0.96, 0.47)	0.68 (0.50)	20.4 (0.0001)[a]	85	0.44	-
Daytime functioning	5/262	−0.23 (−0.48, 0.02)	1.81 (0.07)	6.9 (0.14)	42	-	-

Note: No., number; ES, effect size; $\bar{X}$ (95% CI), mean difference and 95% confidence intervals; Z (p), Z-value and probability value for Z; Q (p), Cochran's Q statistic and associated alpha (p) value for Q; I^2, I^2 statistic for inconsistency; PI, prediction intervals; negative values represent improvements in sleep; −, data not calculated or calculable; boldfaced values represent statistically significant changes.

[a] Statistically significant (P 0.05).

TABLE 3

NNT and percentile improvement for statistically significant sleep outcomes

Study	NNT (95% CI)	U$_3$ index (95% CI)[a]
Araghi et al. (2013)[15]	5 (4, 13)	26.5 (8.5, 39)
AHI		
Chiu et al. (2015)[9]		
Overall sleep quality	6 (5, 12)	19.8 (9.9, 28.5)
Yang et al. (2012)[14]		
Global score (PSQI)	7 (5, 34)	18.1 (3.2, 30.5)
Subjective sleep (PSQI)	5 (7, 14)	18.1 (7.9, 26.7)
Sleep latency (PSQI)	4 (6, 34)	21.9 (2.8, 36.0)

Note: NNT, number needed to-treat, calculated from SMD and 95% confidence intervals and assuming a control group risk of 30%; 95% CI, 95% confidence intervals.

[a] Cohen's U$_3$ index for percentile improvement; AHI, apnea-hypopnea index; PSQI, Pittsburgh Sleep Quality Index.